



HAL
open science

Knowledge Modeling and Multilingual Information Extraction for the Understanding of the Cultural Heritage of Silk

Thomas Schleider

► **To cite this version:**

Thomas Schleider. Knowledge Modeling and Multilingual Information Extraction for the Understanding of the Cultural Heritage of Silk. Computer Aided Engineering. Sorbonne Université, 2022. English. NNT: 2022SORUS280 . tel-03971231

HAL Id: tel-03971231

<https://theses.hal.science/tel-03971231>

Submitted on 3 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PhD Thesis

In Partial Fulfilment of the Requirements for the
Degree of Doctor of Philosophy from Sorbonne University
Specialization: Data Science

Knowledge Modeling and Multilingual Information Extraction for the Understanding of the Cultural Heritage of Silk

Thomas Schleider

Defended on 30/09/2022 before a committee composed of:

Reviewer	Béatrice MARKHOFF , University of Tours, France
Reviewer	Stefanos KOLLIAS , National Technical University of Athens, Greece
Examiner	Marieke VAN ERP , KNAW Humanities Cluster, Amsterdam, The Netherlands
Examiner	Paolo PAPOTTI , EURECOM, Sophia Antipolis, France
Guest	Jean-Claude MOISSINAC , Telecom Paris, France
Thesis Director	Nicholas EVANS , EURECOM, Sophia Antipolis, France
Thesis Co-Director	Raphäel TRONCY , EURECOM, Sophia Antipolis, France

Dedicated...



Abstract

Modeling any type of human knowledge is a complex effort and needs to consider all specificities of its domain including niche vocabulary. This thesis focuses on such an endeavour for the knowledge about the European silk object production, which can be considered obscure and therefore endangered. However, the fact that such Cultural Heritage data is heterogenous, spread across many museums worldwide, sparse and multilingual poses particular challenges for which knowledge graphs have become more and more popular in recent years. Our main goal is not only into investigating knowledge representations, but also in which ways such an integration process can be accompanied through enrichments, such as information reconciliation through ontologies and vocabularies, as well as metadata predictions to fill gaps in the data. We will first propose a workflow for the management for the integration of data about silk artifacts and afterwards present different classification approaches, with a special focus on unsupervised and zero-shot methods. Finally, we study ways of making exploration of such metadata and images afterwards as easy as possible.



Abrégé

La modélisation de tout type de connaissance humaine est un effort complexe qui doit prendre en compte toutes les spécificités de son domaine, y compris le vocabulaire de niche. Cette thèse se concentre sur un tel effort pour la connaissance de la production européenne d'objets en soie, qui peut être considérée comme obscure et donc en danger. Cependant, le fait que ces données du patrimoine culturel soient hétérogènes, réparties dans de nombreux musées à travers le monde, éparses et multilingues, pose des défis particuliers pour lesquels les graphes de connaissances sont devenus de plus en plus populaires ces dernières années. Notre objectif principal n'est pas seulement d'étudier les représentations des connaissances, mais aussi de voir comment un tel processus d'intégration peut être accompagné d'enrichissements, tels que la réconciliation des informations par le biais d'ontologies et de vocabulaires, ainsi que la prédiction de métadonnées pour combler les lacunes des données. Nous proposerons d'abord un flux de travail pour la gestion de l'intégration des données sur les artefacts de la soie, puis nous présenterons différentes approches de classification, en mettant l'accent sur les méthodes non supervisées et les méthodes de type "zero-shot". Enfin, nous étudions les moyens de rendre l'exploration de ces métadonnées et des images par la suite aussi facile que possible.

Contents

Abstract	i
List of Figures	ix
List of Tables	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Research context: the SILKNOW project	2
1.3 Research Questions	3
1.3.1 How can we represent domain-specific knowledge coming from museum records?	3
1.3.2 How can we most effectively extract structured information across different languages from textual descriptions without the need of extensive amounts of annotated training data?	4
1.3.3 How can we make the exploration of a knowledge graph about cultural heritage easier?	4
1.4 Summary of contributions	4
1.5 Thesis outline	5
2 Related Work	7
2.1 The Semantic Web	7
2.1.1 Ontology modelling	10
2.1.2 Controlled vocabularies	12
2.1.3 Knowledge Graphs with Cultural Heritage data	14
2.1.4 ETL pipeline	15
2.2 Information Extraction	16
2.2.1 Word Embeddings	17
2.2.2 Transformer Architecture and Attention	18
2.2.3 Language models	18
2.2.4 Named entity recognition	19
2.2.5 Text classification	20

Contents

2.3	Summary	22
3	Developing a Knowledge Graph about the production of silk artifacts	23
3.1	Data Model	24
3.1.1	Requirements	24
3.1.2	The SILKNOW Ontology	24
3.1.3	Modeling Metadata Predictions	26
3.1.4	Evaluation with Competency Questions	27
3.2	Controlled vocabularies	33
3.2.1	The SILKNOW Thesaurus	33
3.2.2	Applying tools to our Knowledge Graph	36
3.2.3	Evaluation	37
3.3	Data Harvesting and Conversion	38
3.3.1	Developing a web crawler and scraper for public museums	38
3.3.2	Converter software	40
3.3.3	Evaluation	43
3.4	Data Access	53
3.4.1	Graphical interface for the Thesaurus - SKOMOS	53
3.4.2	Access through semantic queries - SPARQL Endpoint	56
3.4.3	Access for web developers - SPARQL Transformer	57
3.4.4	Access for web development - RESTful API	57
3.4.5	Evaluation	59
3.5	Representing human knowledge based on multilingual museum records	72
4	Predicting metadata gaps	73
4.1	Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts	74
4.1.1	Datasets	78
4.1.2	Methods	83
4.1.3	Experiments and Results	91
4.1.4	Discussion	95
4.1.5	Conclusion	102
4.2	Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles	104
4.2.1	Approach	105
4.2.2	Evaluation	107
4.2.3	Conclusion and Future Work	109
4.3	Prompt-guided Zero-Shot Information Extraction (ProZe)	110
4.3.1	Methods	111
4.3.2	Datasets	114

4.3.3	Evaluation	116
4.3.4	Discussion	118
4.4	Using transformer-based QA and CQ systems for metadata predictions	119
4.4.1	Question generation for key information extraction from texts about silk fabrics	120
4.4.2	Conclusion and Future Work	123
4.5	Predicting museum metadata gaps through classification	123
5	Exploring the European Silk Heritage	135
5.1	ADASilk	135
5.1.1	A user-friendly interface for non-experts	136
5.1.2	Integration of research and engineering work of SILKNOW project partners	139
5.1.3	Evaluation	140
5.2	Retrieval of images of silk textiles through domain expert rules and our knowledge graph	146
5.2.1	Approach	146
5.2.2	Evaluation	154
5.2.3	An Exploratory Search Engine for Finding Similar Objects	156
5.2.4	Conclusion	157
5.3	Enabling the exploration of the cultural heritage of silk artifacts	158
6	Conclusion	159
6.1	Summary of the Research	160
6.2	Limitations and Further Perspectives	161
	Publications list	165
	Résumé en français	167
6.1	Introduction	167
6.1.1	Motivation	167
6.1.2	Contexte de la recherche : le projet SILKNOW	168
6.1.3	Les questions de recherche	168
6.1.4	Résumé des contributions	169
6.1.5	Plan de la thèse	170
6.2	Développement d'un graphe de connaissances sur la production d'objets en soie	170
6.2.1	Le modèle de données	171
6.2.2	Modélisation des prédictions de métadonnées	172
6.2.3	Évaluation avec des questions sur les compétences	173
6.2.4	Controlled vocabularies	173
6.2.5	Collecte et conversion des données	175
6.2.6	Accès aux données	177

Contents

6.2.7	Représentation de la connaissance humaine à partir de documents muséographiques multilingues	178
6.3	Prévision des lacunes en matière de métadonnées	179
6.3.1	Prédire les lacunes des métadonnées des musées par la classification . .	179
6.4	Explorer le patrimoine européen de la soie	180
6.4.1	Permettre l'exploration du patrimoine culturel des objets en soie	181
	Appendix: Full list of competency questions	183
	Bibliography	202

List of Figures

2.1	The Semantic Web Stack	8
2.2	Example conceptual diagram for a graph-based ontology	11
2.3	Conventional ETL Diagram	15
2.4	The Transformer - model architecture [127]	18
2.5	Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating ques-tions/answers) [36]	19
3.1	Illustration of the representation of the CDMT Terassa / IMATEX record 4537 inside the SILKNOW knowledge graph	25
3.2	Graph showing the prediction of the production technique (damask) with a high confidence score (0.9173) using the textual analysis software.	28
3.3	Illustration of the linking through String2Vocabulary	36
3.4	Results of questions 1-3	37
3.5	Examples from museum websites of how metadata originally looks like before we apply any software tool	41
3.6	Part of the mapping table for the exemplary record "08.48.46" of MET	42
3.7	Illustration of how mapping rules are implemented with the converter software	42
3.8	The table reporting an example of the UNIPA record named Caccamo 7	45
3.9	Screenshot of the Faceted Browser view of the same Caccamo 7 record as in figure 3.8	45
3.10	Results of the first validation	47
3.11	First example - records from GARIN	48
3.12	Second example - records from RISD	49
3.13	An example of record where location has not been reported	50
3.14	An example of a too short description	51
3.15	An example of a record from which the right location may be inferred	52
3.16	Homepage of Skosmos configured to browse the SILKNOW Thesaurus	53

List of Figures

3.17	General page showing metadata of the SILKNOW Thesaurus	54
3.18	Detailed view of the Acanalado (ligamento) concept in Spanish ("Extended tabby" in English)	55
3.19	Example of the concept "Moiré" in the SILKNOW Thesaurus	56
3.20	General architecture of ADASilk exploratory search User Interface	58
3.21	A request execution sample for users I, II and III in a thread group of 5 users . .	61
3.22	Data gathered for the different access methods of the Knowledge Graph given the defined tests	65
3.23	The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the internal ADASilk API	66
3.24	The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the internal ADASilk API	67
3.25	The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the public SILKNOW API	68
3.26	The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the SILKNOW Public API	69
3.27	The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the SPARQL API	70
3.28	The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the SPARQL API	71
4.1	Examples from three different museums with missing categorical properties: a) no subject depiction for the record 37.80.1 from the Metropolitan Museum of Art; b) no material for the record Cl. XXIV n. 1748 from the Musei di Venezia; c) no technique for the record GMMP-733-002 from the French Mobilier National	73
4.2	A record from the MET museum with a missing property represented in the knowledge graph using our ontology and controlled vocabularies	80
4.3	Network architecture of the CNN for multitask image classification. The input image scaled to 224 x 224 pixels is presented to a pre-trained ResNet-152 (grey) to extract generic features. The resulting 2048-dimensional feature vector is mapped to a domain-specific joint representation of 128 dimensions by two fully connected layers (blue). The task-specific classification branches consist of one softmax layer each (orange) that delivers the class scores for the corresponding variable. K_{mat} , K_{ts} , K_p , and K_{te} denote the number of class labels for the tasks <i>material</i> , <i>timespan</i> , <i>place</i> , and <i>technique</i> , respectively.	86
4.4	Multitask architecture: a shared XLM-R based encoder followed by task specific classification heads. The input to each classification head is the output of the transformer "C" corresponding to the input token "[CLS]".	88

4.5	Task specific classification head: a fully connected (FC) layer followed by a tanh activation, followed by the output projection FC layer. Dropout is applied before both FC layers.	89
4.6	Architecture of the multimodal classifier. Each classifier based on a single modality takes its own independent decision, D_c , which serves as input to the multimodal classifier. The final decision D is taken by the multimodal classifier,, predicting a task-specific label and assigning it to the record.	90
4.7	Association between features of the tabular classifier, measured using Cramer's V.	97
4.8	Multimodal classifier confusion matrices: predicted vs true labels.	99
4.9	Agreement between modality predictions (Cohen's kappa).	100
4.10	Example of disagreement between classifiers in the material task: the object includes cotton, hard to see in the image but clear in the description which includes the passage "filled with cotton".	101
4.11	Confusion matrices for the different modalities, task: timespan.	102
4.12	Confusion matrices for the different modalities, task: place.	102
4.13	Confusion matrix for the property technique on the English subset for the ZSC method. The Y-axis represents the true labels and the X-axis the predicted ones.	108
4.14	"Embroidery" was correctly predicted by our ZSC approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.	109
4.15	"Velvet" was predicted instead of "Embroidery" by our ZSC approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.	110
4.16	ProZe neighborhoods demo. (1) The user is asked to select a label (2) The user can input a text to prompt and guide the language model. (3) The user can visualize the label neighborhood, with added and removed nodes highlighted, and is shown a detailed list of all the changes resulting from the prompt.	115
5.1	The ADASilk home page (https://ada.silkknow.org/): the user is invited to either enter a textual search term or to browse the collections of objects using shortcuts for the most common weaving techniques, materials and depicted subjects. The image-based search can be used through the camera icon	137
5.2	An example of a search using textual query terms with auto-completion.	138
5.3	The advanced search page enables the user to refine a search using facets. Multiple values can be used for each facet.	138
5.4	Respondents who agreed that the objects were optimally shown in STMaps. Percentages shown per TA; this graph allows us to compare them	145
5.5	Comparison of STMaps concerning positive opinions among TAs. We show the positive answers to the questions. Percentages shown per TA; this graph allows us to compare them.	145

List of Figures

5.6	Excerpt of the knowledge graph: a textile object coming from the CDMT Terasse museum which has been produced in Italy in the 16th century, with the Brocatelle technique, using silk bombyx mori as material and showing the motif of a crown.	147
5.7	Top-k-scores as a function of k for all evaluated scenarios. The score gives the percentage of query images for which there was at least one meaningful result among the k most similar images delivered by the image retrieval module. . . .	156
5.8	Percentage P_m [%] of query images for which the image retrieval module delivered at least m meaningful images among the k=10 nearest neighbours for Scenarios A-E.	157
5.9	Objects that are visually similar with respect to an object produced in 1725-1735 in France using the embroidery technique and coming from the Art Institute of Chicago (ARTIC) museum.	157
6.1	Illustration de la représentation du CDMT Terassa / IMATEX record 4537 dans le SILKNOW knowledge graph	172
6.2	Partie de la table de correspondance pour l'enregistrement exemplaire "08.48.46" du MET	176
6.3	Illustration de la manière dont les règles de cartographie sont mises en œuvre avec le logiciel du convertisseur.	177
6.4	Exemples de trois musées différents avec des propriétés catégorielles manquantes : a) pas de description du sujet pour l'enregistrement 37.80.1 du Metropolitan Museum of Art ; b) pas de matériel pour l'enregistrement Cl. XXIV n. 1748 du Musei di Venezia ; c) pas de technique pour l'enregistrement GMMP-733-002 du Mobilier National français.	180

List of Tables

1.1	Overview of all GitHub and GitLab repositories that contains code that has been at least partial relevant for our work in this thesis and related publications. . . .	6
3.1	Summary of the data model evaluation through competency questions (excluding the Spanish ones). Coverage is given both for questions for which any sort of useful query was possible and for questions that could be answered with at least one result	29
3.2	Coverage of the thesaurus concepts in the museums. Showing results for thesaurus in each language separately over the museums for that language.	35
3.3	Project objectives related to the exploitation outcomes, target audiences and the Thesaurus.	38
3.4	Complete table of museum and collection sources. Number of records reflects the number of actually records successfully converted and represented inside the SILKNOW Knowledge Graph. *Garin 1820 had been successfully integrated, but is temporarily deactivated due to an ongoing rights discussion.	39
3.5	Structure of the thread group of users and the number of requests performed for each stress test on one of the three access methods offered by SILKNOW	60
3.6	Test stress execution timetable	62
4.1	Class structure and class distribution of the records.	83
4.2	Names of the museums contributing to the dataset with their identifiers (ID) used in this section, and distribution of the 28,077 records over the museums for the training (train.), validation (val.) and test sets.	84
4.3	Modality statistics of all records in the dataset that provide a class label for at least one of the variables. The values are given for the training (train.), validation (val.) and test sets as well as for the total dataset.	85
4.4	Examples of text descriptions present in our dataset.	85
4.5	Text length in characters and space delimited tokens.	86
4.6	Language distribution of text descriptions based on language of the museum. .	86
4.7	Tabular Classification, one example input row per task. Note: time label format changed to roman numbers for ease of readability.	89

List of Tables

4.8	Hyperparameters tuned (image classification). An optimal variant is obtained with $\eta=1e-4$, $\omega_R=1e-3$, $NL_{RT}=30$ (i.e., 10 residual blocks), with the focal loss $E_F(\mathbf{w})$.	91
4.9	F1 scores (F1) and overall accuracies (OA) of the image classifier obtained by minimizing the Softmax loss (eq. 4.1) and the focal loss (eq. 4.2) both for the validation and the test sets (evaluated per record). Δ gives the difference between the quality metrics achieved using the focal loss and the softmax loss.	125
4.10	Hyperparameter tuning. Hyperparameters, the investigated range range, and the value chosen to be the best in 50 random trials according to macro-F1 evaluated on the validation set.	125
4.11	F1 scores (F1) and overall accuracies (OA) obtained in the multitask experiment both for the validation set and the test set (text classification)	125
4.12	Hyperparameter tuning for the multimodal classifier: hyperparameters, the investigated range of values (Range) and interval of the search, and best values for each task, chosen by grid search according to macro-F1 evaluated on the validation set.	126
4.13	F1 (F1) and overall accuracies (OA) obtained in the experiment both for the validation set and the test set (tabular classification).	126
4.14	Tabular classifier: feature importance per task (information gain).	126
4.15	Hyperparameter tuning for the multimodal classifier: hyperparameters, the investigated range of values (Range) and interval of the search, and best values chosen by grid search according to macro-F1 evaluated on the validation set. These hyperparameters apply to the multimodal classifier using the complete set of input modalities as shown in Figure 4.6.	127
4.16	F1 scores (F1) and overall accuracies (OA) on the test set of the multimodal classifier with and without the raw tabular data as additional input. The last two columns give the differences between OA and F1 scores of the two variants (ΔOA [%] and $\Delta F1$, respectively).	127
4.17	Feature importance (gain) for the multimodal classifier per modality for all tasks, both with and without raw tabular data.	127
4.18	Multimodal classifier feature importance (gain) per task per tabular data feature.	127
4.19	Mean F1 scores (F1) and overall accuracies (OA) of the different classifiers evaluated on the entire test set. Samples for which a modality was missing are considered as errors for the corresponding modality-specific classifier. In case of the multimodal classifier, the numbers are identical to those for the variant considering raw tabular data in Table 4.16.	128
4.20	Average F1 scores of the multimodal classifier using input modalities, with and without the raw tabular data (average over all tasks).	128
4.21	Examples of misleading text descriptions. Emphasis added to highlight the misleading text snippets.	128

4.22	Number of objects exported for each property (material, technique, depiction)	129
4.23	Results for the material property across approaches	129
4.24	Mapping between the concepts used in our knowledge graph and ConceptNet	130
4.25	Results for the technique property across approaches	131
4.26	Results for subject depiction property across approaches	131
4.27	Mapping between the concepts used in the SILKNOW knowledge graph and ConceptNet (ProZe and ZeSTE)	132
4.28	Prediction scores for the news datasets (the top score in each metric is emboldened).	132
4.29	Prediction scores for the domain-specific datasets (the top score in each metric is emboldened).	132
4.30	Auto-evaluation scores based on matches between the target label and the generated question(-answers). Comparison with the label prediction accuracy of two Zero-Shot classification methods that have been performed on the same dataset (*For ZeSTE the results of its application on a minimally different, but comparable dataset are stated here). The baseline is representing the class distribution.	133
4.31	Two generated output texts per T5-based model. All examples represent cases in which the target label could be matched with the output and the Prompt-guided ZS classification method used on the same dataset predicted a wrong label.	133
5.1	Set of general questions and results to find out the feeling of users as regards efficiency and reliability of the promised functionalities and how comfortable they feel.	142
5.2	Set of general questions added to the SUS questionnaire to find out opinions about the integration of ADASilk and STMaps in specific domains	143
5.3	Rules defined by the cultural heritage experts to define pairs of similar images. Nr: number of the rules.	148
5.4	Overall accuracies [%] per variable for the different scenarios of similarity as well as the best performing experiment of test step 1 (SIR_LR_4). The highest score per variable is highlighted in bold font. The second column contains the weight α_s of the loss function term related to semantic similarity and, thus, indicates whether semantic similarity is considered ($\alpha_s > 0$) or not ($\alpha_s = 0$); the last column gives average values over all variables. In case of the variable Production Material, the first value refers to the classification results based on a binary classification procedure; the second value refers to the results including the most probable class of samples assigned to the background for all classes.	155
5.5	Average F1-Scores [%] per variable for different scenarios of similarity as well as the best performing experiment of test step 1 (SIR_LR_4). For more details, see the caption of Table 5.4	156

List of Tables

- 6.1 Overview of all GitHub and GitLab repositories that contains code that has been at least partial relevant for our work in this thesis and related publications. . . . 163
- 6.2 Résumé de l'évaluation du modèle de données à travers les questions de compétences (à l'exception des questions espagnoles). La couverture est donnée à la fois pour les questions pour lesquelles toute sorte d'interrogation utile était possible et pour les questions auxquelles on pouvait répondre par au moins un résultat. 174

Chapter 1

Introduction

1.1 Motivation

From all the cultural heritages of Europe, the historical knowledge about how to produce silk fabrics is probably one of its most obscure. In the past, silk fabrics and objects made from it used to be some of the most expensive and most desired trade goods in the world for many centuries. The historical existence of the so-called “Silk Road” hints at the original far-away origin in China and Far East Asia. The knowledge about how to weave silk items, and also the use of the necessary silkworms spread to Europe at least centuries if not millennia later after its original discovery.

Silk textiles have always been associated with luxury in Europe, specifically to the clothes, furniture and decorations of aristocrats, and many silk items can also be considered to be pieces of art. Behind every silk object stands also the craftsmanship and skills of artisans or, for more recent examples, the technological and scientific progress benefiting the construction of better looms. Lastly, the history of European silk production is connected to a very important early milestone in the history of computer hardware: the French master weaver and silk merchant Joseph Marie Jacquard invented the earliest programmable loom. This so-called Jacquard Loom was controlled by punched cards, pieces of paper holding digital data through presence or absence of holes in predefined positions. It was used for the production of silk objects and was able to use the common silk weaving techniques brocade and damask a.o.

Many museums throughout the world still own historical European silk items, from flags, canopies, tapestries and costumes to fans and sword sheaths, specifically objects from the 15th century and later. Fortunately, public access to their metadata and photos is often possible. The knowledge about their full historical domain and the specifics of particularly European-made silk objects, is however, nowadays spread-out, unknown to many and can therefore be considered as endangered.

The identification and conservation of cultural heritage artifacts requires consistent inventory and archival of their metadata and other surrounding information, such as images and their own metadata. Many museums and libraries, including ones that own or exhibit historical European silk objects, have already digitized most parts of their collections and made them publicly available - either through a web interface or even providing specific APIs. What we can, however, observe, is that next to such basic digital cataloging endeavors many possible digital tools are not applied, especially when it comes to congruent integration of all such relevant metadata of existing objects. Furthermore, when it concerns all expert knowledge inside the domain of European silk fabrics: There has simply been no single place, in the physical world or online, where an audience can access information about all such items and all relevant background information, for example about how they were weaved or what motifs they display.

1.2 Research context: the SILKNOW project

SILKNOW is a H2020-funded research project (2018-2021) aiming at understanding, conserving and disseminating the European Silk Heritage from the 15th to the 19th century. It is a multi-disciplinary project and of the goals is to apply computing research methods to the needs of museums, education, tourism, media and creative industries. Its original and full list main goals are as follows:

- Semantically relating digitized European silk heritage, enabling data interoperability across different collections, for advanced searching abilities.
- Building a “Virtual Loom” to clone weaving techniques. This will allow users to discover the complexity, artistic and artisanal values of ancient silk textiles, while preserving them for future generations.
- Improve the understanding of the European silk heritage, thanks to visual tools that show the spatio-temporal relationships of data, including an open-access, multilingual thesaurus.

The work done in this thesis is relevant to all three of these goals. We have been addressing the semantic data integration, which is at the core of both the "Virtual Loom" and the spatio-temporal map and were also working on the technical implementation of the open-access and multilingual thesaurus.

The following results were targeted at the time of the start of SILKNOW:

- To provide many institutions, custodians of an immense textile heritage, with ICT re-

sources that allow them to open their hidden wealth of European heritage to worldwide audiences.

- To contribute with strategies and best practices for the better curation of digital data in textile heritage institutions, particularly among those of small-to-medium size.
- To facilitate better strategies and the design of innovative tourism services about silk heritage, enriched through digital contents.
- To create enhanced didactic tools, scale models (in computer graphics and 3D printouts) of historical textiles that visualize their internal structure.
- To spark creative efforts by modern designers, putting silk heritage within the reach of today's consumers by well-informed reuse of its motifs.
- To pave the way for further R&D+I in 3D printing for the textile industries.
- To produce teaching units in digital format, according to different levels of the Common European Framework of Reference for Languages.
- To become a resource for project-based learning assignments, through a specific tutorial developed collaboratively with participating schools' staff.
- To connect content providers with fresh and interesting content, improving public knowledge of the Western Silk Roads, as well as their tremendous impact on our international relations, industry, technology and culture.
- To support regional policy makers in the implementation of their smart specialization strategies, with a focus on digital Cultural Heritage.

1.3 Research Questions

1.3.1 How can we represent domain-specific knowledge coming from museum records?

Creating any type of knowledge base or specifically a knowledge graph that represents expert knowledge, requires a specific workflow. We have to first develop or decide on a specific ontology. Domain experts need to be able to map semantically heterogeneous data from original museum records fields onto classes and properties part of the target ontology model. Such mapping rules need to be both applied through software and if necessary re-adjusted based on how well they are actually applicable. At this stage the use or design of controlled vocabularies should be considered as well. Implementation of such semantic mapping rules plus string matching with concepts of a controlled vocabulary is not trivial and success not

guaranteed. The quality of a developed and enriched knowledge graph is hard to evaluate and using Competency Questions has become a standard, but is still mostly a very manual and subjective process for which more automation could be considered. Finally, giving proper access to a knowledge graph to different types of end-users is another challenge that needs to be overcome.

1.3.2 How can we most effectively extract structured information across different languages from textual descriptions without the need of extensive amounts of annotated training data?

The recent progress in natural language processing and more specifically in information extraction can help us to address problems like missing metadata in knowledge based systems that are based on heterogeneous and multilingual source data. In particular, it is common to train text classification models from complete metadata records in order to predict missing categorical values in other records. However, such models do require a significant amount of annotated data for training, which is expensive to get for such specific expert domains. One alternative for such cases is the use of unsupervised approaches for metadata prediction leveraging on zero-shot learning, transfer learning and language models.

1.3.3 How can we make the exploration of a knowledge graph about cultural heritage easier?

Nodes or objects inside a cultural heritage knowledge graph can be considered similar in different ways, either based on their metadata or (if applicable) available images of an object. Even within this division different ways of measurement can be established, e.g. through the selection or weighting of different textual metadata properties or visual properties. One way of deciding on a similarity measurement is through domain expert rules or another form of human evaluation.

1.4 Summary of contributions

This thesis contributed to research with the following outcomes:

- A data model and a thesaurus for and about historical silk objects from Europe that are stored and exhibited in museums. These contributions have been implemented strongly based on the input, knowledge and design of domain experts and historians.
- The development SILKNOW Knowledge Graph with which the museum metadata and images got finally integrated. It has been implemented and uploaded with further

Semantic Web technologies and tools, like SPARQL and a triplestore, and is published in the web of data to make all our work accessible for everyone.

- A set of tools for web crawling, harvesting, downloading and converting museum data with our data model towards our graph format, replacing strings into concept URIs and therefore linking entities with our thesaurus. Furthermore, we can contribute tools that offer API access for web developers to our SPARQL endpoint.
- Exploration of several approaches, most of them perform zero-shot classification, to fill metadata gaps by predicting missing values.
- An exploratory search engine called ADASilk, to offer a simple graphical web interface for non-experts, that offers advanced search based on many of our data enrichments in the KG and also integrations of several other software tools of our SILKNOW project partners. Finally, we integrated an image retrieval model into ADASilk that can be used to find similar images of silk objects. We trained this model by leveraging the knowledge of domain experts by formulating similarity rules a.o.

1.5 Thesis outline

The remainder of this thesis is organized as follows:

- Chapter 2 is dedicated to the exploration of the state of the art on several concepts crucial to the research work of this thesis, which can be mainly grouped into two topics: The Semantic Web and Information Extraction.
- In chapter 3, we will describe the development process of the multilingual SILKNOW knowledge graph. This effort consists of designing a data model, creating a controlled vocabulary for concepts of silk weaving, a data harvesting and conversion process and constructing data access through an API.
- Chapter 4 is about the detailed methods and approaches for predicting metadata gaps in the our graph. We mainly compare various zero-shot methods to more classical supervised approaches.
- In chapter 5, we describe different ways of exploring the European Silk Heritage based on our developed and enriched knowledge graph.
- Finally, we present a summary in chapter 6 as well as highlighting both limitations that are known to us and outline further perspectives.

Chapter 1. Introduction

GitHub / GitLab repositories	Chapters	Publications
https://github.com/silknow/crawler	3	The SILKNOW Knowledge Graph
https://github.com/silknow/converter	3	The SILKNOW Knowledge Graph
https://github.com/silknow/thesaurus	3	The SILKNOW Knowledge Graph
https://github.com/silknow/knowledge-base	3	The SILKNOW Knowledge Graph
https://github.com/silknow/skosmos	3	The SILKNOW Knowledge Graph
https://github.com/silknow/api	3	
https://github.com/silknow/adasilk	3, 5	
https://github.com/silknow/image-classification	4	Multimodal Metadata Assignment for Cultural Heritage Artifacts
https://github.com/silknow/text-classification	4	Multimodal Metadata Assignment for Cultural Heritage Artifacts
https://github.com/silknow/ZSL-KG-silk	4	Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles
https://gitlab.eurecom.fr/schleide/proze	4	ProZe: Explainable and Prompt-guided Zero-Shot Text Classification
https://gitlab.eurecom.fr/schleide/qg4textunderstanding	4	
https://github.com/silknow/image-retrieval	5	Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules
https://github.com/silknow/image-retrieval-server	5	Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules

Table 1.1: Overview of all GitHub and GitLab repositories that contains code that has been at least partial relevant for our work in this thesis and related publications.

Related Work

Following up on the introduction, we can group the research work of this thesis into two main categories: the Semantic Web (Section 2.1) and all its related tools and Information Extraction (Section 2.2), a sub-field of Natural Language Processing (NLP), especially in the form of classification problems.

In this chapter, we will introduce and describe the key concepts of these fields and all other topics related to them and the thesis. We want to give special attention to the current state of the art and especially highlight recent practices.

2.1 The Semantic Web

The Semantic Web¹ and its related technologies are an extension of the World Wide Web with the ultimate goal of making Internet data machine-readable (see Figure 2.1,² for an illustration of the full stack.). The concept of it has been proposed by Tim Berners-Lee, best known as the inventor of the original World Wide Web. Human knowledge is supposed to be represented with it in a way that is more open and accessible for the public and enabling interconnections between dataset, which makes a great case for integration work of museum metadata. The Semantic Web emerged in the field of online data management due to a growing number of interconnected datasets representing various subjects of human knowledge. Important examples of such datasets are GeoNames³, which covers all geographical metadata for all countries on Earth including eleven million place names, and DBpedia⁴, a project with the aim of containing all information of Wikipedia⁵ extracted as structured content. Both datasets allow semantic queries of relationships and properties of all their linked data.

¹<https://www.w3.org/standards/semanticweb/>

²https://en.wikipedia.org/wiki/Semantic_Web#/media/File:Semantic_web_stack.svg

³<https://www.geonames.org/>

⁴<https://www.dbpedia.org/>

⁵<https://www.wikipedia.org/>

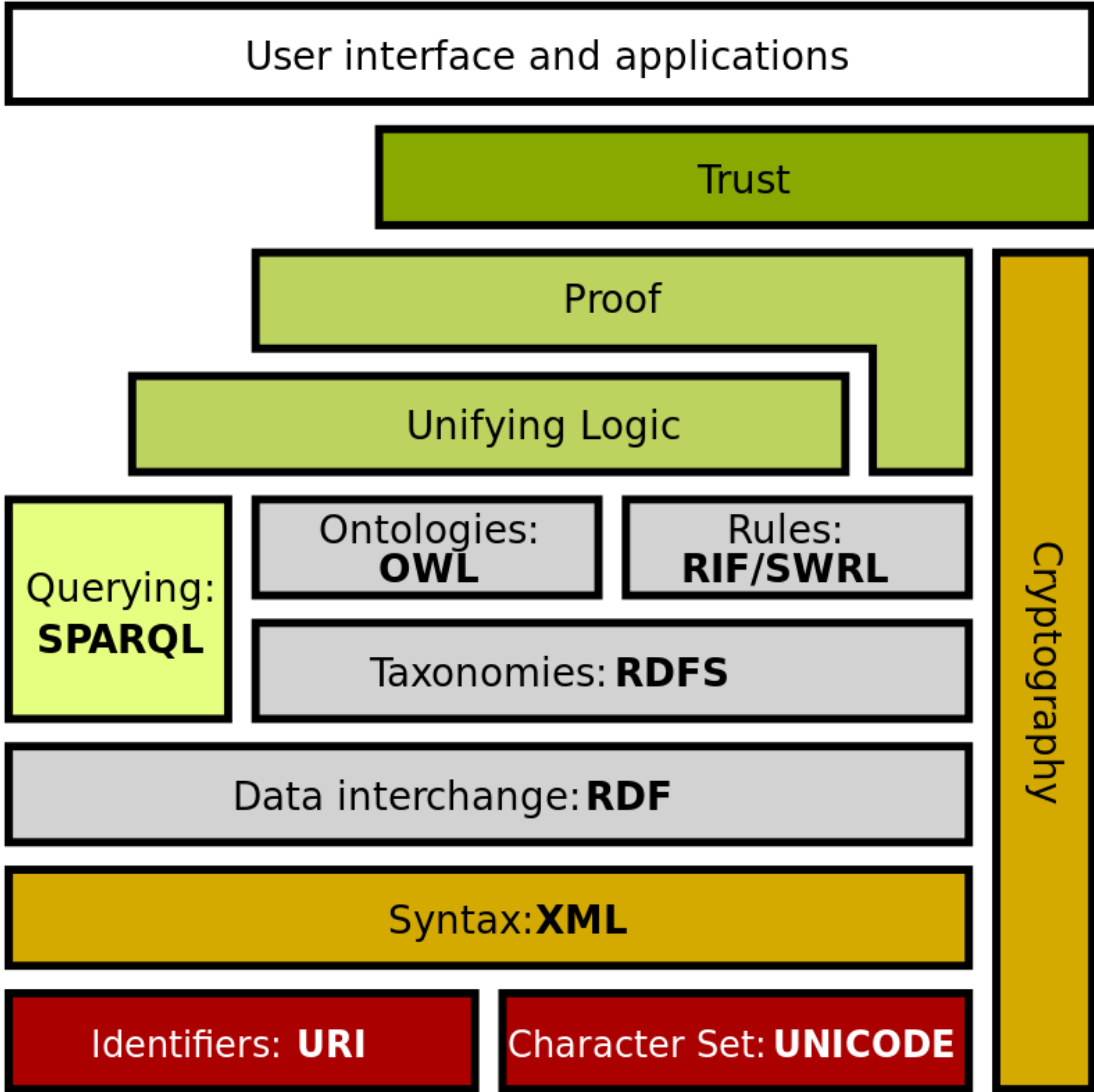


Figure 2.1: The Semantic Web Stack

At the core of unique Semantic Web technologies stands the Resource Description Framework (RDF) ⁶, originally designed by the World Wide Web Consortium (W3C) ⁷. RDF represents a directed graph composed of triple statements consisting of: a subject node, a predicate arc and an object node. Each of these parts can be identified by a Uniform Resource Identifier (URI). The object node can hereby also have a literal value, for example a string or a number. RDF allows a variety of syntax notations and file formats, such as JSON-LD or RDF/XML, but a very common one is Terse RDF Triple Language or Turtle ⁸, which has been specifically created to express such semantic triplets. Its syntax is close to the SPARQL Protocol and RDF Query Language (SPARQL) ⁹, which is another important Semantic Web technology.

To represent human knowledge with RDF it is necessary to pre-define the type of nodes that can be part of a triplet, or with other words: to define a fixed set of possible entities and relations between nodes. In Computer and Information Science, such a data model or formalized set of categories, properties, and relations is called an ontology. This concept of ontology is not entirely the same as the philosophical branch with the same name, but in both cases an ontology attempts to model all interconnected properties and relations between objects, entities and events in accordance to a system of categories.

One set of classes with certain properties to provide basic elements for the description of such ontologies is RDF Schema (Resource Description Framework Schema), most often abbreviated as RDFS. As the name implies, it provides a data-modelling vocabulary for RDF data and is therefore an extension of the basic RDF vocabulary.

Another Semantic Web technology that supports us with tools to implement such full data model is, e.g. the Web Ontology Language (OWL) ¹⁰, characterized by formal semantics and already including many RDFS components, in comparison with which it is more expressive. The Simple Knowledge Organization System (SKOS) ¹¹ is on the other hand often used to represent controlled vocabularies and taxonomies, and allows for instance a specification of preferred versus alternate labels for an object. Another part of the infrastructure for the semantic web is the Rule Interchange Format (RIF), a framework of web rule language dialects supporting rule interchange on the Web. A rule can be imagined as simple as an IF - THEN construct, a very common notion in computer science. However, the work in this thesis is not relying on RIF at any point.

⁶<https://www.w3.org/RDF/>

⁷<https://www.w3.org/>

⁸<https://www.w3.org/TR/turtle/>

⁹<https://www.w3.org/TR/rdf-sparql-query/>

¹⁰<https://www.w3.org/OWL/>

¹¹<https://www.w3.org/TR/skos-reference/>

2.1.1 Ontology modelling

Expressing Cultural Heritage data through an ontology (see also Figure 2.2, ¹²) can be difficult and time consuming, as data has to be usually collected from many different sources that generally do not use standard formats. For museum data, there are existing models that make such a process less challenging. We consider the CIDOC Conceptual Reference Model (CRM) ¹³ to be one of them. CIDOC-CRM is an event-centric ontology through which everything can be represented as an event. A man-made object has, for example, been produced at some point in time. Certain materials might have been used during this production that has taken place at a specific location in a specific century. Taking CIDOC-CRM as the starting point for building a Cultural Heritage ontology is following existing examples and practices, e.g. the PARCOURS project [93], which aims at supporting semantic interoperability in the conservation-restoration domain.

CIDOC-CRM offers not only classes and properties for such an event-based representation, but is also easily expandable with classes from other ontologies. Hence, CIDOC-CRM comes also with useful extensions such as CRMSci (Scientific Observation Model) and CRMdig (model for provenance metadata). CIDOC-CRM is the outcome of more than 20 years of development by ICOM's International Committee for Documentation (CIDOC) [40], has been an official ISO standard since 2006 and this status has been renewed in 2014. CIDOC-CRM has been implemented in OWL DL (a sublanguage of OWL with slightly less expressiveness than OWL Full, but less computation demands, and named in correspondence with Description Logics) as Erlangen CRM/OWL (ECRM) ¹⁴. ECRM is also the implementation that has been used as part of the research work in context of this thesis.

As important as creating or using an existing ontology for representing human knowledge is to define a way to evaluate such a knowledge base. Competency questions are a common method for assessing ontology-based research work [15]. They need to get defined rather early to define the scope of the knowledge model. Throughout the whole length of development it can also help to evaluate the ontology itself. But they are not only an evaluation method: they allow us to constrain the scope of the ontology. If domain experts are not asking about a certain aspect of a domain through competency questions, certain classes and properties of a data model might be unnecessary or are simply too broad for a target domain of human knowledge.

A set of competency questions needs to be written in natural language first. Therefore there is a challenge not only in trying to get answers to them through an ontology, but also in formalizing or translating them into a query language like SPARQL in order to use them in

¹²https://en.wikipedia.org/wiki/Knowledge_graph#/media/File:Conceptual_Diagram_-_Example.svg

¹³<https://www.cidoc-crm.org/>

¹⁴<http://erlangen-crm.org/>

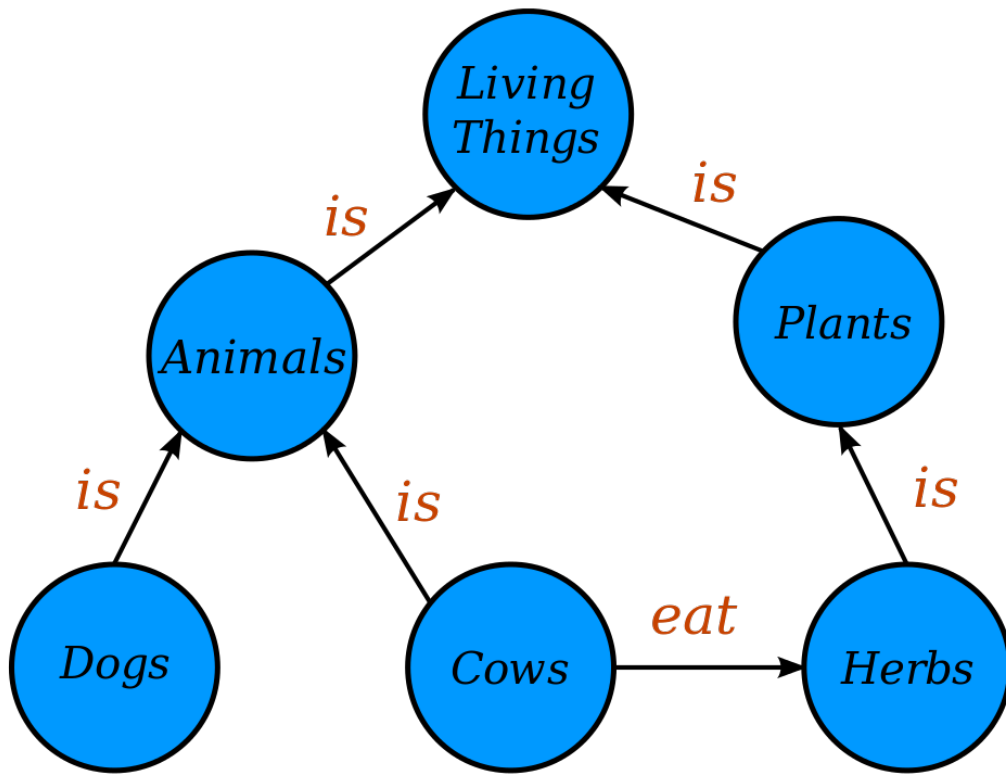


Figure 2.2: Example conceptual diagram for a graph-based ontology

both directions.

2.1.2 Controlled vocabularies

Another way of organizing knowledge and data is to create controlled vocabularies. Such a vocabulary can be defined as an organized arrangement of terms used to index content and/or retrieve content through browsing or searching [105]. In the beginning of section 2.1, we already introduced two parts of the Semantic Web stack which are crucial for the implementation: OWL, which can be used to build vocabularies, and SKOS to design knowledge organization systems.

A museum can be understood as a huge data base where cultural objects are stored. In order to properly identify these objects, the documentation area emerges as a specific and important area in the museum. Documenting a cultural object means to register and to catalog it. Doing it properly is the precondition to ensure the physical persistence of objects as the registration of a cultural asset assumes its importance as cultural heritage that requires conservation and protection. Indeed, the basic element for conservation is to classify objects, understanding it as symbolic organization of meanings: “cultural artifacts constitute the network that sustains their institutions, they are symbols that are defined as the locally objectified sites of meaning.” [74]. In other words, the conservation of cultural heritage begins with its registration and identification, tasks that are carried out through inventories and catalogs, which are the traditional tools for the study, analysis and especially protection of heritage [4].

In order to describe a cultural asset, proper terminology stands out as one fundamental pillar [10]. Information professionals, curators, conservators and general audience will be the end-users of these tools. Indeed, controlled vocabularies are essentials to provide access to museum collections not only to inside users (registrars, curatorial departments, conservators, education department), but also to external users who wish to know more about a subject without knowing the specific term of its search [8].

A thesaurus is defined in general, as a controlled vocabulary that has a semantic network of unique concepts [56] that enhances information retrieval, as it is based in queries based in categorized deductions [49]. It also links the object with the user as it allows to use a language that facilitates the research of a cultural asset and its related information. Moreover, the vast amount of metadata associated to it allows not only to document and describe the object, but also to find likenesses or differences between similar cultural assets, and to associate them, allowing users to find new connections [104].

Although some institutions and public administrations are striving to use standard vocabularies, most museums have generated their own methods of classification. The terminology used in the description varies widely according to different cataloging schools, fashions and

curators in charge of this task. At the same time, museums around the world develop their own controlled vocabularies, that they see more fitting in order to describe their collections [52]. It is the case of The Textile Museum Thesaurus from the Textile Museum in Washington, or the Museon Arlaten. We can also mention the Domus system of Spain, hosted by the Documentary Standardization of Museums [21], or French databases such as Joconde and Gallica.

On the other hand, some standardization efforts have been carried out, such as the UNESCO thesaurus or the Getty Art & Architecture Thesaurus (ATT). Also, we can cite other generic thesauri, applicable to all types of cultural, movable or immovable property: CDWA, Object ID, ULAN, TGN, Iconclass, etc. Although they are useful for their own institutions, the result is a multitude of vocabularies that are not shared, complicating interoperability.

However, the cultural heritage domain and the silk heritage in particular are characterized by large, rich and heterogeneous datasets [8]. In this sense, the silk heritage vocabulary can change according to who (careers: weavers vs historians / disciplines: art historians vs. anthropologists) and where (Europe or North America) the term is being used [54]. This has resulted in the use of different terminologies in specialized organizations when describing their collections which makes comparisons among the same type of objects, techniques, designs quite complicated, not only in different languages but also in the same language. Furthermore, cultural heritage data is being transformed into public Linked Data, especially in large-scale aggregators such as Europeana [52]. Plus, the Semantic Web technologies lead to a new approach in managing Cultural Heritage data interoperability [60]. Responding to these challenges, the SILKNOW thesaurus emerges as a thesaurus that aims to improve silk heritage knowledge by building an open-access thesaurus based on SKOS model. This thesaurus is multilingual and standardizes terminology providing conservators, researchers and other users an important tool, that allows systematic and coherent cataloging of museum collections, in order to avoid the lack of common criteria when dealing with these kinds of records.

When using controlled vocabularies, it is possible to interpret literal values and to replace them by concepts defined in thesaurus and identified by URIs following the Linked Data principles [18]. Such concepts can be pre-defined by domain experts and also provide different languages and therefore translated labels and definitions of such concepts. A fully linked controlled vocabulary, which can be either a taxonomy or a full thesaurus, can provide explicit and authored reference points to a knowledge base.

Record linking

Recognizing and matching equivalent text data is an important step for the integration and harmonization of heterogeneous data. This task can be called record linkage, data matching

or entity resolution. It constitutes also an important foundation for information extraction (see below) as it is a way to preserve existing structure in available text data before the problem of unstructured data can be solved. One way of matching such data is through the use of a controlled vocabulary. GeoNames makes for example matching and linking place names easier, as it already contains millions of locations with all their different writings across many languages.

2.1.3 Knowledge Graphs with Cultural Heritage data

Many mentioned technologies are related to the topic of general knowledge bases, but if data is expressed through RDF it is explicitly given the shape of a graph. Although the term knowledge graph is at least as old as 1972¹⁵, and Dbpedia, which has been launched in 2007, has always been graph-based, it took until Google's introduction of their own Knowledge Graph¹⁶ in 2012 until this term has been fully established in its current form. Since then many other companies, such as Facebook, LinkedIn, Airbnb, Microsoft, Amazon, Uber and eBay¹⁷ have highlighted to maintain a knowledge graph as part of their data infrastructure.

There are also more and more successful examples of KG development in the Cultural Heritage realm. For example, ARCO [22] is a knowledge graph about the Italian Cultural Heritage that at least indirectly reuses some CIDOC-CRM classes and properties. The Dutch Rijksmuseum collection is available as linked open data [38].

Kerameikos¹⁸ describes itself as "A Linked Open Greek Potter Project". It uses the CIDOC-CRM ontology. The researchers are working on a user-interface to showcase images of vases and 3D content related to typologies, geographic visualizations and amongst others distribution analyses for particular painters etc. Kerameikos has a reconciliation API for OpenRefine to help researchers normalize vase data. The project describes the largest outstanding task as the creation of automated harvesting of Linked Art-compliant JSON-LD data and incorporating that into a knowledge graph.

Cultura Italia¹⁹ integrates cultural metadata from several Italian institution into a KG that can be accessed by text search, SPARQL queries and iSPARQL queries. The KG is based on CIDOC-CRM as well.

DOREMUS is a project in which EURECOM was a partner which has integrated musical metadata from France's most important institutions regarding musical libraries. It shares many

¹⁵https://en.wikipedia.org/wiki/Knowledge_graph

¹⁶<https://developers.google.com/knowledge-graph>

¹⁷<https://kgkg.factnexus.com/@3782-167.html>

¹⁸<http://www.kerameikos.org>

¹⁹<http://www.culturaitalia.it/>

similarities with our research objectives. An Exploratory Search Engine named OVERTURE²⁰, which is built on top of its KG, has been developed and it also contains a recommendation system: for every work or artist explored, the search engine will show similar works.

2.1.4 ETL pipeline

Extract, transform, load (ETL) describes the procedure of copying data from one or more sources into a target system, which ultimately represents the data in a different way than the sources (see also Figure 2.3, [66], ²¹). Existing already since the 1970s [35], ETL is a popular concepts until today for any problem concerning integration of data from heterogeneous sources.

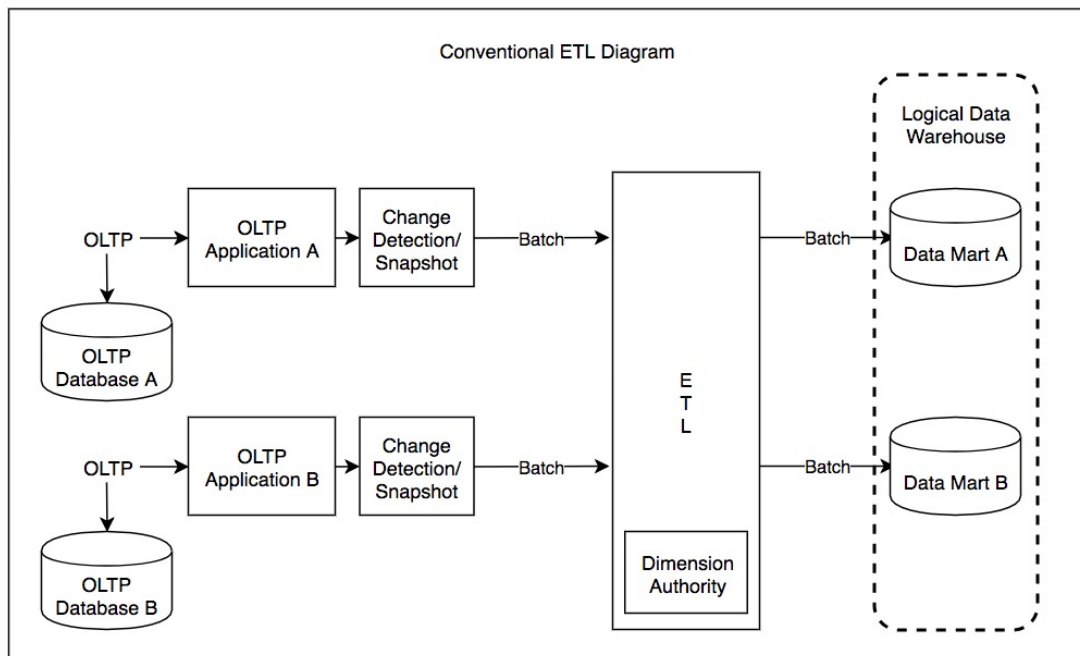


Figure 2.3: Conventional ETL Diagram

A common data format is crucial for any extraction. If it is not possible to choose a format that is shared by all sources, one has to be chosen and all data needs to be converted into it already at this stages. Possible formats are relational databases, XML, JSON and flat files. Web crawling and scraping are also important methods to ensure extraction into one common format.

In the transformation stage, a set of rules or functions are applied to this extracted or downloaded data. Most often this includes at least data cleansing. In order to integrate data into a knowledge graph, which would conclude the final load stage of an ETL process, a mapping

²⁰<https://overture.doremus.org/>

²¹https://en.wikipedia.org/wiki/Extract,_transform,_load#/media/File:Conventional_ETL_Diagram.jpg

needs to be created and implemented at the transformation part. The following subsections describe two paradigms for a mapping of relational data to RDF.

Direct Mapping

The Direct Mapping of relational data to RDF ²² is defined as a simple transformation of relational data and can be used to materialize RDF graphs. Relational datasets are still heavily used due to a.o. their efficiency and precise definitions, making it possible for tools like SQL to modify and retrieve its contents. Compared to R2RML (see below) neither structure nor target vocabulary can be changed. It can be considered the default and automatic way to translate relational databases into RDF.

R2RML

The RDB to RDF Mapping language (R2RML)²³ has been created to express customized mappings from relational databases to RDF datasets. These mappings are themselves RDF graphs, written in Turtle syntax. Every R2RML mapping is specifically designed for certain database schemas and target vocabularies. The output is an RDF dataset as defined in SPARQL.

Standardized mappings allow the migration of data views as RDF across databases. Compared to a direct mapping from relational databases to RDF (see above) a mapping author can define highly customized views over the relational data and R2RML defines itself also as a relaxed variant of Direct Mapping as a default mapping. Data and schema of it is taken as an input and an RDF graph is generated that is called the direct graph.

2.2 Information Extraction

Since the early 2010s many Natural Language Processing (NLP) techniques have been mostly relying on supervised machine learning and deep learning classification techniques that require labelled training data. Such is the case with information extraction as well: the task of automatically extracting structured information from both unstructured and semi-structured documents. Information extraction is one of the oldest NLP techniques and exists since the late 1970s [6]. Only more recently, especially with since the existence of the World Wide Web, are we however both facing a sheer huge and even more rapidly growing amount of (digitalized) documents with many unstructured texts, but also the computation power to apply advanced NLP tools on an equally massive scale.

While general purpose transformer-based language models (see also subsections 2.2.2 and 2.2.3

²²<https://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927>

²³<https://www.w3.org/TR/r2rml/>

further below) are more specifically the most used since a few years, fine-tuning them often require numerous examples [39]. One possible solution to address this lack of labelled data is to use active learning [137] in conjunction with pre-trained models such as BERT [36] that comes with possible bias problems [97]. In Cultural Heritage domains, several works are trying to compensate for the lack of labelled data. One method is to leverage human annotations through crowd-sourcing together with extracted visual and textual features and automatic annotation through transfer learning [116]. In general, Convolutional Neural Networks (CNN) are often used for image, audio and video data [27], whereas Recurrent Neural Networks (RNN) is also used for textual Cultural Heritage data [64]. Regular neural network [11] for textual data are also used together with word embeddings.

CNNs and RNNs are both part of a broader family of machine learning based on artificial neural networks with representation learning, called Deep Learning. Word embeddings can a.o. be generated by such neural networks, but not exclusively. See subsection 2.2.1 for more information about this type of numerical word representation.

More recently, zero-shot learning approaches have attracted a lot of attention for their ability to offer text classification without relying on training data. It is a form of automatic classification for which textual documents unseen by the model can be analyzed and classified. Several frameworks have been proposed over the years, based on BERT [138], [133] or other large pre-trained models [129]. There are approaches like ZeSTE²⁴ (Zero Shot Topic Extraction) [57] that provide a framework for extracting topics from textual documents using the ConceptNet common-sense knowledge graph. In addition, the framework provides explainability of its classification results using the ConceptNet KG neighborhood.

2.2.1 Word Embeddings

Creating word embeddings is a Natural Language Processing (NLP) method with which words or phrases can be mapped to vectors of real numbers. Starting in the early 2000s several authors have been working on such neural-network based approaches [13], [115], [30]. Vectors that are close to each other in such vector space represent hereby words or phrases that are semantically related. Word2vec [89], developed in 2013, was one of the first widely used word embedding algorithms. Older count-based traditional models could still compete in some cases, but only with certain parameter modifications [76]. But other models like Fasttext [19] have fully surpassed the performance of word2vec and count-based models. The researchers were inspired by an idea from 1993 [114] and incorporate character n-grams into a skip-gram model with Fasttext. Therefore it takes subword information like morphology into account whereas word2vec, Paragraph Vectors [73] and other contemporary models did not. Word embeddings can be used for text classification although a way has to be chosen

²⁴<https://github.com/D2KLab/ZeSTE>

how entire sentences or even documents are finally represented. Word embeddings can for example be averaged per sentence before being fed to a classifier. A hidden and reusable variable [63] can then be constructed.

2.2.2 Transformer Architecture and Attention

Transformer networks [127] are based on an attention mechanism: Mapping a query and a set of key-value pairs to an output, where query, keys, values and outputs are all different vectorial representations of the input (see also Figure 2.4). A weighted sum of the values (the attention distribution) is then computed as an output. This attention mechanism allows every piece (word) of the input, almost regardless of its length, to continuously draw information from the whole, thus foregoing the need for recurrence or convolution to capture such internal relations between the input elements that are so important in all language-related tasks.

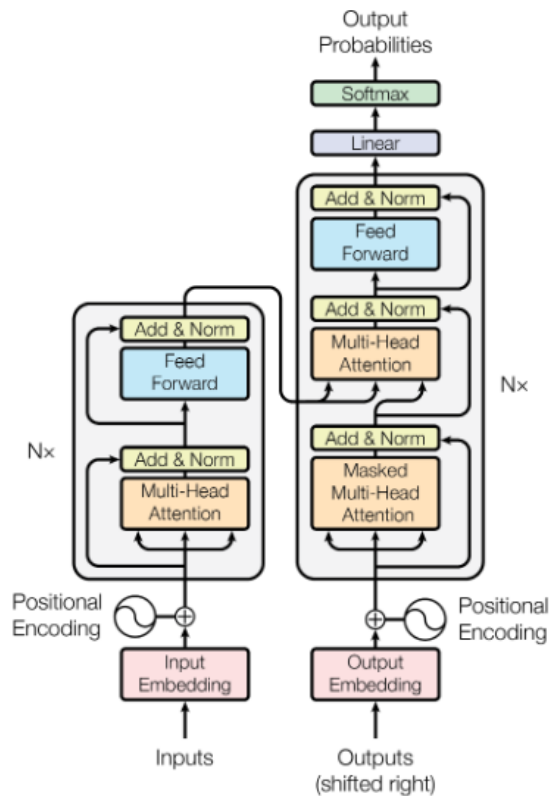


Figure 2.4: The Transformer - model architecture [127]

2.2.3 Language models

Since the introduction of transfer learning [96], [96, 124], previously learned knowledge can be effectively used to handle or improve upon the performance of other problems. Together

with breakthroughs in the application of neural networks in the form of convolutional neural networks (CNNs) for image-based tasks in the 2010s [26,59], pre-trained models became a new standard for many machine learning tasks. In the sub-field of NLP, more shallow pre-trained word-embeddings [90] used to be more commonly used than pre-trained models because the features learned for specific tasks were not easy to transfer to another. With the introduction of the Transformers architecture [118], however, it was shown how generic such models can be, and it has become the standard to use such pre-trained deep models for many NLP tasks.

Many models, training schemes and architectures, have since been based on Transformers, and the most influential of them is BERT [36]. Its defining feature is its ability to pre-train deep bidirectional representations (see also Figure 2.5. Many variants of BERT have been created since then. Such pre-trained language models remain part of the most successful approaches for a wide range of NLP tasks, such as text classification. Despite the wide availability of these language models, many classification experiments require also annotated and balanced training data to make a model properly associate text segments with labels, which is often either expensive or not available at all, especially when the domain is niche.

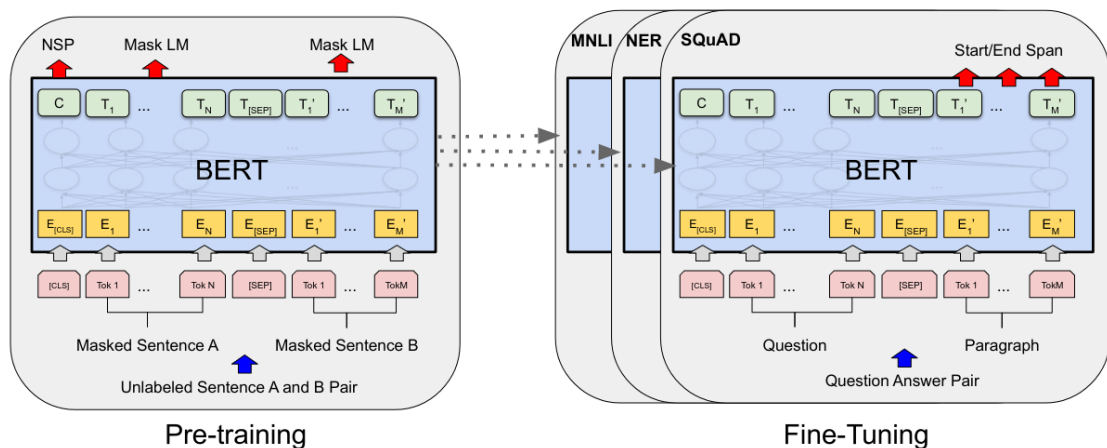


Figure 2.5: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers) [36]

2.2.4 Named entity recognition

Named entity recognition (NER), belonging to the domain of Information Extraction (IE), is the task of identifying predefined semantic types like for example location, material or color inside textual descriptions. As the aforementioned progress in the field of NLP has consequently also relevance for NER, it is not surprising that NER tools based on pre-trained Transformer-models,

specifically BERT, have proven successful and are becoming a new paradigm.

The current state-of-the-art model for the task OntoNotes 5.0 (92.07 per) is based on BERT [79]. Officially, the model that is state-of-the-art for the task CoNLL03 (93.5 per) right now is not BERT-based [9], but uses a very similar pre-trained bi-directional transformer model.

2.2.5 Text classification

In recent years, methods relying on large and generic pre-trained language models have shown to be very successful for many NLP downstream tasks. Meanwhile, because of the exponential growth of complex documents in the world, a need for machine learning methods with the capacity to accurately classify text has arisen. The definition of text classification is to be the task of assigning a fitting category to a text, where the categories depend on the domain and topic, and profited from those recent milestones in NLP. Such systems can be split into four parts: Feature extraction, dimension reductions, classifier selection and evaluations [69].

Data-less or zero-shot classification methods are able to address the specific disadvantage of the need of a lot and very balanced data and are in recent years often based on aforementioned Transformer- and BERT-based models [134, 139]. With its rising popularity, there are now more attempts to benchmark and evaluate zero-shot text classification approaches. [135] provides a survey of the recent advances in the field, while proposing *Entail*, a zero-shot classification model based on using language models fine-tuned on the task of Natural Language Inference to classify documents. Some zero-shot classification models also takes advantage of “prompt-based learning” [83], a new paradigm used for many NLP tasks that allows to extract information out of Language Models.

Compared to traditional supervised learning often used for text classification, the original input x is modified using a template into a text string prompt x' with unfilled slots. The language model has then to fill this gap to gain a string from which the final output y can be derived. Just recently, many of such prompt-based approaches have been created [109, 119, 141]. Some works also started to use prompting for domain adaptation [12]. Tuning pre-trained language models with task-specific prompts has been a promising approach for text classification. Previous studies suggest, in particular, that prompt-tuning has remarkable superiority in low-data scenarios over the generic fine-tuning methods with extra classifiers.

There is a growing amount of work interested in explainable methods for text classification [7]. Notably, one direction is to generate explanations and to develop evaluations that measure the extent and likelihood that an explanation and its label are associated with each other in the model that generated them [61, 98, 130]. However, none of these techniques totally compensates for the obscurity associated to language models. There are models like ZesTE (Zero Shot Topic Extraction) [57], however, which are not based on a pre-trained language model, but

provide explainability of their classification results using ConceptNet as a prediction support.

Question-Answer Generation and Text Classification

Leveraging question answering or question generation for information extraction is not new, as it has been studied even before the emergence of deep learning or transformer-based models [131], but it is still rather rarely studied. Despite this, recently several promising models have been recently proposed such as e.g. QuAChIE for the Chinese language [106].

A unified multi-task learning framework for joint extraction of entities and relation that consisted importantly on a sub-task including question generation based QA with a transformer-based Seq2Seq model is another new example [140]. An important feature was the detection of subjects and objects without relying on NER models in this pipeline. In this paper, we only consider a pipeline for text classification and could therefore not use this framework, but consider it relevant that an information extraction task has been pre-processed through question generation.

Finally, we would like to present one more recent approach which leverages question generation for entity and relation extraction [55]. In this case, the question answer model was created by training BERT on the SQuAD dataset. The input texts are pre-processed with a NER model and then uses a phrase generation method to frame the questions. In general, these few recent examples show promising results, but we are not aware of any recent work about pipelines that consist of question generation and text classification.

While the task of Question Generation (QG) has not received as much attention as its sibling task of Question Answering (QA), it is a relevant task to text understanding. In particular, domain adaptation in QA often involves using the task of QG, in order to create domain specific datasets on which language models can be fine-tuned [37]. Most recent approaches rely on pre-trained transformers and often consider question generation and answer generation as dual tasks that can be combined in different ways during training [5, 23]. Another approach was to simplify QG by using a single transformer-based model for answer agnostic end-to-end question generation [85].

There have recently been remarkable efforts to make transformer-based question generation easily usable, such as the three models available at https://github.com/patil-suraj/question_generation which obtained competitive results on the SQuAD benchmark and represent different ways of treating the QG-QA paradigm. All these models are T5 based and fine-tuned on the Stanford Question Answering Dataset (SQuADv1) dataset [103]. SQuAD contains context paragraphs, each associated to sets of questions (100 000 in total) and the corresponding answer spans in these paragraphs. Three of these models can be described as follows:

- Single-task QA-QG model: Hereby [23], the text is first split into different sentences. Then, the T5 model extracts elements that could qualify for answer like span (often NER for SQuAD) for each sentences and generates question-answer pairs. It therefore produces at least one question per sentence.
- Multi-task QA-QG: This approach [5] fine-tunes T5 in a multi-task way: it uses the task prefixes from T5 to extract an answer, generate a question, find the answer to the question and finally compare it to the results with the initial extracted answer.
- End-to-end QG: The T5 model is trained to generate multiple questions simultaneously by providing the context paragraph [85]. This model is answer-agnostic and generates up to three questions per paragraph.

2.3 Summary

This chapter represents a summary of literature and related to the state-of-the-art of two main research areas relevant to this thesis: Firstly, the semantic web and secondly, information extraction. The next chapters will further explore the application of these fields to the domain of European silk fabric production.

Integrating heterogeneous data about a human domain into a knowledge representation system is a prime use case for the semantic web and RDF. With CIDOC-CRM a state-of-the-art data model for representing museum data is available as a foundation for creating an ontology that is appropriate as a starting point to solve the problem that such an effort constitutes.

Beyond the integration of structured data, many advanced NLP techniques are nowadays at our disposal to tackle the automatic extraction from at least not fully structured textual data. Especially transformer-based methods can nowadays help to fill dataset gaps through predictions, even across languages. Both supervised, unsupervised and zero-shot methods are able to address such gaps in the form of text classification.

Finally, there has been a recent progress in question-answer generation models, also mostly transformer-based, which opens up a new perspective on completing knowledge representations of a specific human domain such as ours.

The following chapters investigate solutions for the problems that we encounter with our data, and constitute contributions to the here mentioned research fields.

Chapter 3

Developing a Knowledge Graph about the production of silk artifacts

The SILKNOW knowledge graph (KG) ¹ lies at the center of all efforts to create a unified representation of the metadata of European silk textiles, particularly from the 15th to the 19th century. All the data used in the experiments in this paper have been extracted or downloaded from 20 sources, most of them are public online museum records, for which we built a crawling and harvesting software. In addition to that, we have data from SILKNOW project partners Garin 1820 and the University of Palermo (Sicily Cultural Heritage). For the dataset used in the experiments of this paper a full export of all objects of the knowledge graph has been performed, which consists of the metadata of 40,873 unique silk objects before any preprocessing steps. This export includes in total 74,527 unique image files.

Modeling data that is as heterogeneous as ours requires a particular workflow and despite many tested methods being available to us for such an endeavor, creating a multilingual domain-specific knowledge graph is not a straightforward process. The structure of this chapter is as follows: Section 3.1 describes the data model from collecting the requirements to designing the final ontology. Section 3.2 continues by introducing how controlled vocabularies, especially our own thesaurus, have been designed and evaluated. In section 3.3 we present the process of collecting the data and converting into its final format based on our data model. Finally, section 3.4 shows the full stack of tools through which we give access to our knowledge graph, including a SPARQL Triplestore, a RESTful API and ultimately an exploratory search engine.

This chapter covers the following publications and submissions:

- Thomas Schleider, Raphaël Troncy, Mar Gaitán, Ester Alba, Jorge Sebastián, Dunja Mladenić, Avguštin Kastelic, M. Beshar Massri, Arabella León, Marie Puren, Pierre Vernus, Dominic Clermont, Franz Rottensteiner, Maurizio Vitella, Georgia Lo Cicero.

¹<https://zenodo.org/record/5743090>

The SILKNOW Knowledge Graph. Submitted to the Special Issue on Cultural Heritage and Semantic Web, Semantic Web Journal.

- Thomas Schleider; Raphaël Troncy. **Exploring the European Silk Cultural Heritage through the SILKNOW Knowledge Graph.** In International Conference on Silk heritage and Digital Technologies (Weaving Europe), 30 November 2020-4 December 2020. Online.

3.1 Data Model

A data model is generally an abstract model for the organization of elements of data and a standardization of their relations both to each other and the properties of real-world entities. The one that we finally rely on for our knowledge graph about silk objects consists of several parts that we will describe in this section.

3.1.1 Requirements

Each data model needs to fulfill a set of requirements to be useful in its application. Given our domain and data origin we needed to make sure that we can support not only multilingual texts, but specifically the languages used in the original records, which are English, Spanish, French and Italian. As our data is only about a very specific human topic, such as history or chemistry, in our cases historical silk fabrics, we also require mostly a domain ontology. With such specificity comes also the need to be able to hand-tune and extend such a data model.

The metadata about European silk fabrics comes always from either collections or museums, which makes an ontology necessary which offers classes and properties to represent an object and eventually photos or images of it appropriately. An important aspect of museum objects, especially when historic, is that they are often both unique products that have once been created by hand and in virtually all cases have been found somewhere else than the museum location. Therefore, they at least once changed location and owner.

3.1.2 The SILKNOW Ontology

The SILKNOW ontology on which our knowledge graph is built is strongly based on the CIDOC Conceptual Reference Model (CIDOC-CRM), which we first introduced in section 2.1.1.

Small parts of the total ontology for the SILKNOW Knowledge Graph are based on several properties of schema.org² and the W3 time ontology³. The majority of the classes and prop-

²<https://schema.org/>

³<https://www.w3.org/TR/owl-time/>

erties used in SILKNOW come from the current published version of CIDOC-CRM (6.2) and its extensions, the Scientific Observation Model (CRMsci) [41] and CRM Digital (CRMdig) [42]. The former is a formal ontology for integrating metadata about scientific observation, and the latter is to encode metadata about the steps and methods of production of digitization products. The complete usage and implementation of these ontologies and data models can be retrieved from GitHub where it is part of the converter software ⁴.

In order to aggregate numerous data sets collected from various sources, it is necessary to harmonize them by designing and implementing a unique and complete data model. To define the SILKNOW data model we first analyse the structure of records from several institutions especially the Victoria and Albert Museum, the British Museum, the Musée des Tissus in Lyon, the Garín collection at the Museu de la Seda in Moncada, the Musée des Arts Décoratifs in Paris, the Museum Baselland, and the French Joconde Database. We also used the ICOM guidelines for Museum Object, the Europeana data model, the norms and methods relative to the inventory keeping in French museums (*arrêté du 25 mai 2004*), and the French Harmonized Model for the production of cultural data. From this analysis we elaborated the data dictionary, i.e., a list of information groups or metadata interesting for the SILKNOW project. Then we selected in CIDOC-CRM the classes and properties useful to express these metadata.

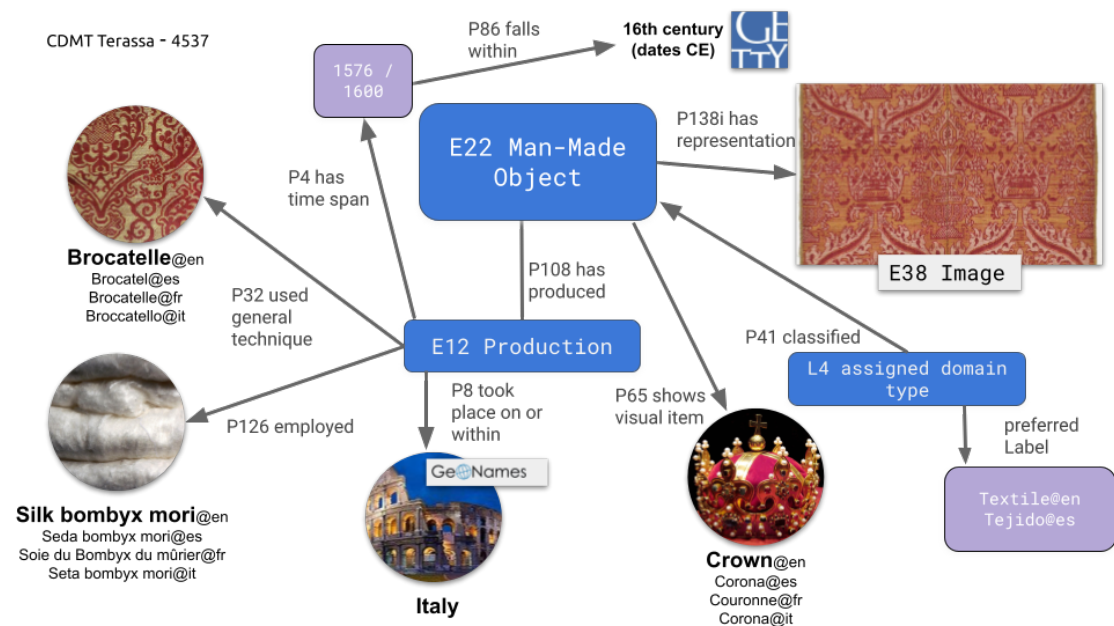


Figure 3.1: Illustration of the representation of the CDMT Terassa / IMATEX record 4537 inside the SILKNOW knowledge graph

We have chosen the CIDOC-CRM because it has been designed to express the underlying semantics of documentation on Cultural Heritage [95]. Moreover, it is an international standard,

⁴<https://github.com/silknow/converter/tree/master/src/main/java/org/silknow/converter/ontologies>

recognized as an ISO norm. It has already been used in several research projects, including EU-funded projects, such as Ariadne which developed an extension of CIDOC-CRM suitable for archeological documentation [43]. SILKNOW is using version 6.2. It is an event-centric data model, very flexible and extensible by nature: while it consists of a limited set of classes and properties, it is mainly a core ontology allowing the development of more specialised extensions. In other words, it is possible to add new sub-classes and sub-properties to express more specific relationships and properties, without modifying the basic structure of the model. The classes and properties selected for the SILKNOW ontology are publicly accessible and documented via OntoMe, an ontology management system, developed by the LARHRA research center [72].

On the one hand, the bottom up approach adopted by SILKNOW spurred us to use CRMsci as a global schema for integrating metadata about scientific observations, performed by domain experts on silk-related artefacts. On the other hand, CRMdig was used to express the relationships between data sets and metadata records describing them.

After evaluating the pertinence of the ontology by providing mapping rules between metadata examples and the SILKNOW ontology, it was observed that, so far, all fields can be represented by using existing classes and properties from the ontology.

Generally, scientific observations are expressed with free-text fields analysing the structure and the decoration of fabrics, and/or presenting the historical context of their production or use. This first mapping aimed at storing these metadata “as they are”; but the complex semantics included in data about the creative and productive process of silk textiles cannot accurately be represented with the basic CRM entities and its existing extensions. In order to address the complexity of textile data integration, it requires elaborating new CRM classes and properties.

3.1.3 Modeling Metadata Predictions

An aspect of our research contribution was to not only integrate data into a knowledge graph, but to experiment with data enrichments through various methods. Keeping this in mind, our ontology needed to be able to represent such enrichments accordingly. A major part of these enrichments consists of predictions of various metadata gaps in the data: for example missing production dates or weaving techniques. Chapter 4 will properly explain this part of the research work, whereas we will here focus on additions to the data model that are crucial for this integration.

To model the prediction as part of the SILKNOW Knowledge Graph ontology we added classes and properties of the Provenance Data Model (Prov-DM), more specifically the PROV ontology

(PROV-O)⁵, an OWL2 ontology. It makes it possible to map PROV-DM to RDF. Being a W3C recommendation, it allows expression of important elements of the predictions, both for the ones based on images, from text descriptions and based on categorical values. These different predictions can be represented using different `prov:activity` classes each. The image, text description or categories this prediction is based on is represented with the property `prov:used`. The exact date of the prediction is represented with `prov:atTime` and `prov:wasAssociatedWith` connects the activity class to the `prov:SoftwareAgent` class, which is used to describe the particular algorithm and experiment used. The actually predicted metadata value is represented with `rdf:Statement`, connected to `prov:activity` via a `prov:wasGeneratedBy` property. The confidence score of the prediction is expressed through the property L18 ("has confidence score") from our own SILKNOW Ontology. The predicted value is expressed in form of an URI with `rdf:object`, the type of the predicted property through `rdf:predicate` and its fitting CIDOC-CRM property type. The property `rdf:subject` is, at last, connecting the statement to the production class (E12) of the object in the Knowledge Graph. Every prediction is inserted in the appropriate part of the existing KG. For example, if a material value gets predicted, it gets inserted with the CIDOC-CRM property `P126_employed` at the production class of the object. See figure 3.2 for an illustration of the data model. As the prediction models were only trained on group labels, they can solely predict such. Therefore it sometimes needs to get mapped back to a more concrete concept of the SILKNOW Thesaurus. If for example "Damask" gets predicted in form of its facet link `http://data.silknow.org/vocabulary/facet/damask` it will get converted into `http://data.silknow.org/vocabulary/168`, as facet links are too general for concrete category values. All predictions are converted one after another using the described data model and saved in form of the Turtle file format and uploaded and stored as its own graph identified by `http://data.silknow.org/predictions`. This makes it possible to always identify and eventually separate predictions from original values from the museums. All in all, 98,379 predictions exist for 19,248 distinct objects and are uploaded into the SILKNOW Knowledge Graph.

3.1.4 Evaluation with Competency Questions

In accordance with section 2.1.1, formulating competency questions was an important part of correctly constraining our ontology and preparing an evaluation of our data model. We published all competency questions, which have been formulated by domain experts at the beginning of the SILKNOW project, on GitHub⁶, where a corresponding query and a list of results can be directly retrieved. A full list of them can also be found in the Appendix of this thesis. Although we did not only have English, but also Spanish questions, we did not come to

⁵<https://www.w3.org/TR/prov-dm/>

⁶https://github.com/silknow/converter/tree/master/competency_questions

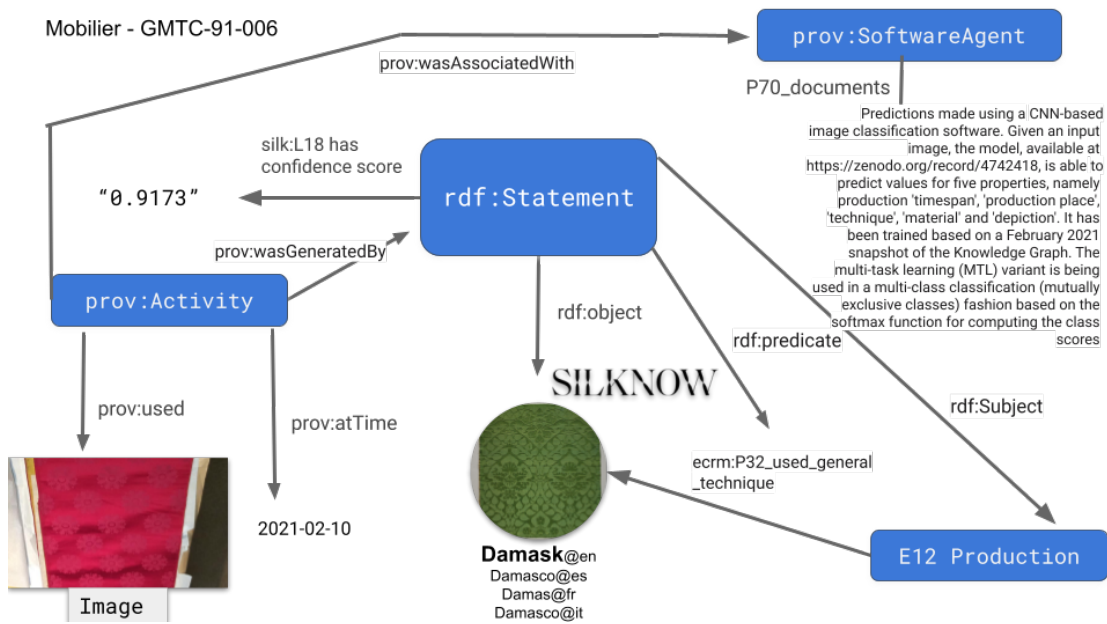


Figure 3.2: Graph showing the prediction of the production technique (damask) with a high confidence score (0.9173) using the textual analysis software.

using the latter for our data model evaluation, yet.

Qualitative Analysis

Despite some questions guiding our ontology in specific directions, many questions were still not possible to answer at the end of the project. The questions were posed mostly without regards to feasibility and purely from the point of view of a human expert and the requirements of the field. They were also designed without insight into the final selection of data and sources. On the other hand, we tried to be as strict as possible when translating them to queries. This means that if we were not sure we can properly address a question with the help of explicitly available - e.g. enriched and linked entities in the knowledge graph - information, we would still try to avoid refraining to string searches or other ways that would bypass the actual data model. To illustrate this, we want to go through most questions we have not answered in this subsection. Many of these could still be answered as part of future work.

A. Location 2. Where were Mudejar-style fabrics produced? The Mudejar style is in the Thesaurus <http://data.silknow.org/vocabulary/672>, but we do not categorically link any "styles" as of now. A string search results in only one object⁷ where the style is described in a text description (S04_Observation) and could be derived from the very specific production place ("Hotel Spa La Casa Mudejar Hospederia").

⁷<https://data.silknow.org/object/40115d26-e537-3bee-a097-700d6bece810>

Topic	Nb of Questions	Query Possible	Actual Results
Location	9	33,3%	33,3%
Time	6	50%	40%
Time and Location	4	25%	25%
Material	8	12,5%	37,5%
Artists	8	25%	37,5%
Artists and Time	3	66,7%	33,3%
Artists and Location	3	0%	0%
Style	7	28,6%	28,6%
Type of Items	4	75%	25%
Type of Items and Materials	2	100%	100%
Type of Items, Materials and Style	2	0%	0%
Type of Items and Location	2	50%	50%
Type of Items and Time	1	100%	100%
Type of Items, Time and Location	2	50%	50%
Type of Items, Time, Location and Material	2	0%	0%
TOTAL	64	39,1%	28,1%

Table 3.1: Summary of the data model evaluation through competency questions (excluding the Spanish ones). Coverage is given both for questions for which any sort of useful query was possible and for questions that could be answered with at least one result

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

3. Where was the production center called Tiraz? We have a technique in the Thesaurus called Tiraz⁸, but it is no specific location / production place. The technique is not linked. With a string search you can find the word in the description of some objects (e.g.⁹).

4. What was la Fabrique Lyonnaise? We do not have it as any production place / there is no Geonames location called like this. It appears in some text descriptions (e.g.¹⁰)

5. Which items have been produced in Italy and are now preserved in France? We do not use "E9_Move", because the movement is usually described as a transfer from a person to a museum. Therefore we only have "E10_Transfer_of_custody".

7. What Valencian fabrics are located in the Spanish royal collections? We can make a query to look for fabrics produced in Valencia, but there is no production place or entity called "Spanish royal collections" right now.

9. Give me a list of textile factories in a Florence We are not sure yet how to answer this as we have Florence only as a possible production place, but not as a location for (textile) factories.

B. Time

3. Which fabric became popular in Italy in the fifteenth century? There is probably a way to query this effectively, as we could use the hierarchy of the Thesaurus concept Fabric¹¹. But right now we cannot come up with how to query it as we did not link a lot of such "sub-groups".

4. What kinds of fabrics / weaving techniques / designs were most frequent in 18th-century France? Please give me a list of the top 5 (or 10, 15...) occurrences in a particular field. Similar to the question above.

7. What are the most common decorative motifs in the Hispanic Middle Ages? Should be an easy query, but in the end we did not come to the point where we integrated / linked time periods like this.

C. Time and location

3. Give me all the items that are preserved in the Musée des Tissus de Lyon, and that have been produced between 1650 and 1750. Similar problem as A6.

4. Who (person, institution ...) was the main textile French producer during the XVII? Asked like this we cannot retrieve a direct answer. In the future, we might be able to produce a list with the most common designers of the given time, but this is not perfectly precise.

⁸<https://skosmos.silknow.org/thesaurus/en/page/380>

⁹<https://data.silknow.org/object/cd30428c-8554-3476-84bb-851aed29e604>

¹⁰<https://data.silknow.org/object/551784ae-f8fc-329d-9daf-7633ec32a443>

¹¹<https://skosmos.silknow.org/thesaurus/en/page/649>

D. Materials

2. When does the "a pizzo" design become popular? We don't even have this design in the Thesaurus.

3. When does the "bizarre" design become popular? We do have the "bizarre motif" design in the Thesaurus .

4. What is the Blonda?, 5. What is the Buratto? These questions are purely about concepts in the Thesaurus questions, but not perfectly "English", Blonda is a Spanish name of a concept, Buratto an Italian one. Best answer to these questions would be a direct link to the Thesaurus, which includes definitions of concepts.

6. Where does the name of the Batista fabric come from? Batista (fabric) does not exist in the Thesaurus in any language.

E. Artists

2. Give me all the information you have on Philippe de la Salle! Right now we cannot query background information of persons.

3. Give me all the items inspired by a work of Giambologna

We can not really retrieve anything "inspired" by another artist.

5. Give me all the items designed by Italian artists, 6. Are there items designed by French artists in the 17th century? Similar to question E2.

8. Who were the printers or engravers that produced graph paper for making mise-en-cartes? Right now, no object is linked with the concept "mise-en-cartes".

G. Artists and location

1 .Give me all the designers who were born in England, 2. Give me all the designers who were trained in Italy, 3. Give me all the designers who were trained in Italy and in France We cannot yet retrieve further metadata of linked persons (yet) to fetch their birthplaces.

H. Style

1. Who is the Revel style name after? Same problem as above that this question requests background information on a linked person. We have a query for this, but it only shows all items made by Reel.

2. Give me all the items that have been influenced by oriental fashion. We do not have anything in the KG right now that gives us a way to query objects influenced by a certain style

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

or fashion (or visual item / depiction).

4. Give me all the items with hearts and flowers on them We don't have "hearts" in the Thesaurus as a depiction right now.

5. Give me all the items with purple There is no way to query for colours right now, except string matching.

6. Who was the introducer of the realistic style in textiles? Again, we would need access to metadata of (linked) persons. We do not have an entry for "realistic style" in the Thesaurus either. For an optimal query both entities should be interlinked as well.

7. Give me examples of textile designs that appear in paintings. We don't have "textile designs" as a concept or depiction group.

I. Type of items

2. Give me all the dresses that have been worn with a petticoat We do have "women's clothes" in our category vocabulary, but not dresses. For "petticoat" it's the same.

3. Give examples of textiles that conserve both the fabric and the mise-en-carte There is no (linking of any) object with the concept mise-en-carte.

4. When do the first mise-en-carte appeared? Same as directly above.

K. Type of items, materials and style

1. Give me all the scarves with cotton and with hearts on them It is quite a redundant (and theoretically easy) query, but as in H4 we do not have "heart" as a concept.

2. Give me examples of imitations or revivals of textiles during the 18th century This is a very imprecise question. 18th century textiles are no problem right now. But we cannot query "imitations" or "revivals", this would add much more complexity.

L. Types of items and location

2. What textiles belonged to the collector Mariano Fortuny? There is no trace of Mariano Fortuny in our knowledge graph right now (not even text descriptions). Thanks to the "acquisition" class we could however in theory identify persons that were former owners.

O. Type of items, time and location

2. Give me those ornamental motifs from classical antiquity that appear in fabrics, mises-en-carte and designs ... Organized by chronology, location, place of origin ... "Ornamental motif" does not exist as an entity. We could query time-spans from "classical antiquity" but

we would manually select the years. The term “designs” is not clear according to our current list of concepts / the thesaurus. Sorting / grouping the output would of course be no problem.

3.2 Controlled vocabularies

3.2.1 The SILKNOW Thesaurus

Development method involving experts

Silk heritage experts were involved in order to develop the SILKNOW thesaurus. These experts included art historians, historians, weavers, engineers and philologists. Multidisciplinarity was essential in order to select terms, trace their evolution, historical and current use, and how some terms evolved in time and space (e.g. local variations). As the SILKNOW thesaurus is symmetrical, all terms needed to be translated, textile specialists used specialized sources, which in some cases provided translations in other languages (such as the Castany Saladrigas dictionary, 1949). In other cases, direct translations were needed, a scope note was added when necessary or the source language was used as loan. Nevertheless, every translation was made following ISO directions for a thesaurus [33].

In order to compile the thesaurus¹², inductive and deductive methods were undertaken [94]. Around 80% of terms originated from inductive work; i.e., they were included in the thesaurus as soon as they were found in the literature. Specialized sources were used, such as specialized textile dictionaries, historical sources, glossaries, and other thesauri. The other 20% was deductive due to museum records and previous knowledge from the researchers. An extensive research was undertaken, not only taking into account specialized vocabularies, but also using historical sources and selecting the most representative and accurate ones.

Next, terms and concepts were controlled and described by adding scope notes, qualifiers and synonyms. A Preferred Term (PT) was used to refer to a unique concept, whenever polysemy arose, qualifiers were added. In order to make clearer what those concepts meant, scope notes were added following specialized literature. Finally, these definitions were reviewed by international experts.

The next logical step was to categorize those terms. The SILKNOW thesaurus is based on the Getty AAT (Art and Architecture Thesaurus)¹³ structure, as it is one of the most well-known thesauri in the cultural heritage field. Three relationships were established:

- Hierarchical: when the relationship between terms is broader and narrower. Parents were also placed according the AAT structure when possible. As the silk heritage termi-

¹²<https://github.com/silkknow/thesaurus>

¹³<https://www.getty.edu/research/tools/vocabularies/aat/>

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

nology is extensive and not easy to classify, compilers had to add new guide terms and subfacets in order to make it as accurate as possible.

- **Equivalence:** This relationship concerns when different names refer to the same concept as they are synonyms or quasi-synonyms. E.g. bobillo → bocillo. Either noun is accepted to designate this type of lace, however bobillo acts as the Preferred Term.
- **Associative relationships:** when different terms are conceptually closely related, but not hierarchically. E.g. acanalado → otoman. Both terms refer to a type of tabby, however they are not the exact same concept.

Finally, as the SILKNOW thesaurus was initially thought to standardize museums records, experts tried to make it as wide as possible in order to expand silk heritage knowledge. Looms, equipment, iconography, colours, botanical elements were added. This will help researchers to connect these data not only in museum's collections, but also in other research areas. In living heritage, for example, it is possible to see how some of these motifs are used in other contexts. By using this thesaurus, researchers, museum professionals, students and cultural heritage specialists will improve museum information and international research thanks to a free and easily accessible tool.

Thesaurus coverage

The SILKNOW thesaurus was validated on textual data of the selected museums in several natural languages. The frequency of individual thesaurus concepts that are present in the specific museum was calculated. Spanish, English and French translations of the thesaurus were each compared to resources in the corresponding language. The program for the calculation of coverage was written in Python. Pre-processing was done using the Natural Language Toolkit library (NLTK) [16] which contains the Snowball Stemmer. It was used on all the terms and their synonyms from the thesaurus, as well as all the words from online resources.

Table 3.2 gives the results showing that 76% of the terms from the Spanish thesaurus are present in the Spanish museums, followed by 87% for the English thesaurus and 90% for the French thesaurus. In more detail, the two Spanish datasets CERES and IMATEX contain 361 and 326 terms from the Spanish thesaurus respectively, 308 of them occur in both museums. Both museums contain 379 terms from the Spanish SILKNOW thesaurus.

For each online resource (a dataset from a database or museum information system) a feature vector representing all its phrases was computed using QMiner platform [1]. The result was a set of n-grams with the maximum size of three words and a corresponding number of occurrences. From here a subset was generated where all the concepts that can be found in the thesaurus were removed from the feature vector.

Museum	Thesaurus Concepts	Coverage
CERES	361	72 %
IMATEX	326	65 %
Spanish Total	379	76 %
VAM	262	82 %
RISD	210	66 %
MET	205	64 %
IMATEX	182	57 %
English Total	279	87 %
MTMAD	255	89 %
MAD	201	70 %
Joconde	158	55 %
French Total	259	90 %

Table 3.2: Coverage of the thesaurus concepts in the museums. Showing results for thesaurus in each language separately over the museums for that language.

The upcoming validation of the thesaurus (see section 3.2.3) has shown that it includes most of the silk related vocabulary that is used in the considered resources. The phrases which occur in the resources and are not included in the thesaurus are mostly common ordinary words or words not related to silk terminology.

One way to enrich data materialized in the KG is to turn strings (literal expressions) into things (objects identified by URIs in the Linked Data paradigm). For this, we use the tool `string2vocabulary`¹⁴ and either take existing controlled vocabularies that already provide identities to online things or to manually create such a vocabulary.

The `String2Vocabulary` module is a generic component that complements the `SILKNOW` converter. Its purpose is to substitute specific literals with URIs from controlled vocabularies. For example, the period in which a silk artefact has been produced, or the weaving techniques which has been used, or the place where the artefact has been distributed and sold can all be terms belonging to controlled vocabularies and reference systems, such as Geonames, Wikidata or the Getty AAT thesaurus. The `String2Vocabulary` module is built with Gradle and Apache Jena in open source. Figure 3.3 shows an illustration of the result of its linking.

The matching itself works as follows: We map all values from museum sources that contain information about specific properties, like e.g. material, as materials in the knowledge graph (we use the CIDOC-CRM property "P126 employed" in this case) and then we end up with

¹⁴<https://github.com/DOREMUS-ANR/string2vocabulary>

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

all these values correctly semantically annotated, but as pure strings. String2vocabulary tries to match every string in such a category with all labels that exist in a controlled vocabulary. If the string is "Silk" it should match and get replaced with the URI <http://data.silknow.org/vocabulary/368> as it represents the concept of silk and has the property `skos:prefLabel` with the values: "Seda"@es, "Seta"@it, "Soie"@fr, "Silk"@en. We have a few pre-processing mechanism in place to increase to matching even if a string is not written in the exactly same way, like matching independent of capital letters or plural forms. On top of that, we developed also a way to check if a string can get matched with an index of the right category, e.g. a material with a material, or if the string is actually rather a technique and then we change the property inside the KG from `P126_employed` to `P32_used_general_technique`. This is used for mixed fields or fields which semantic category is unclear.

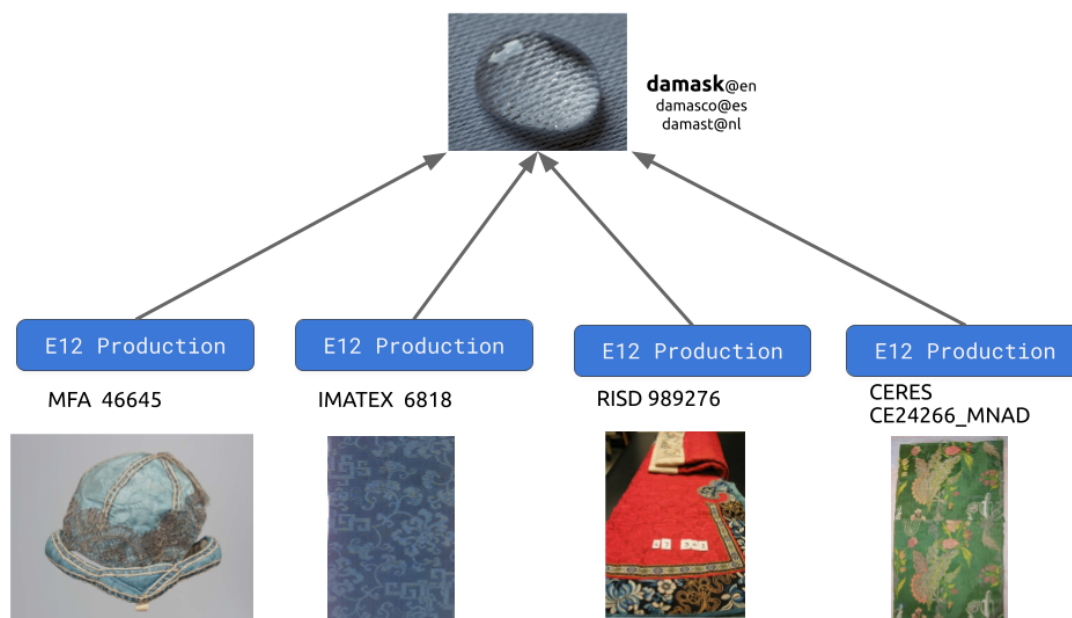


Figure 3.3: Illustration of the linking through String2Vocabulary

3.2.2 Applying tools to our Knowledge Graph

Using one or several controlled vocabularies is an essential part of SILKNOW as it allows us to efficiently use the explicit knowledge of domain experts and native speakers of several languages. In addition to using the widely adopted Getty Art & Architecture Thesaurus (AAT)¹⁵ as well as Geonames, we are working on several manually created ones: for some properties, like the categorisation of objects, the existing vocabularies are either not fine-grained enough for silk items or do not support all languages we use.

¹⁵<https://www.getty.edu/research/tools/vocabularies/aat/>

3.2.3 Evaluation

The intellectual aspects of the SILKNOW Thesaurus (concepts, hierarchies, associated terms, references, etc.) were evaluated in two ways [48]. Firstly, we conducted an internal evaluation with domain experts and another with online questionnaires. English, French, Italian and Spanish translations of the Thesaurus were each compared to resources in the corresponding language. In English, 87.92% concepts were covered, in French 86.09%, in Italian 54.13% and in Spanish 77.67%. In all cases these results correspond to the version of the thesaurus available at that moment, less complete than the current one (now including 660 terms). Then, while we performed the online evaluation of ADASilk during the months of December 2020 to March 2021, we also created an online questionnaire to evaluate the Thesaurus. It must be noted that we had few respondents due to its high degree of specialization and the Corona virus crisis. From the 17 respondents, the most used language was Italian, followed by Spanish, French and English. Figure 3.4 shows the results of the following questions: Regarding the current hierarchy, 82.35% of the respondents found it useful (Q1). We wanted to know if the respondents would apply it in their museums, as one of the main objectives of the Thesaurus was to improve the documentation of these records (Q2). It is worth mentioning that 70.58% would recommend it to other professionals (Q3). Finally, of the respondents, 47.05% would use it for research; 29.41% for cataloguing and inventory purposes, and 23.52% to standardize and review their museum catalogues.

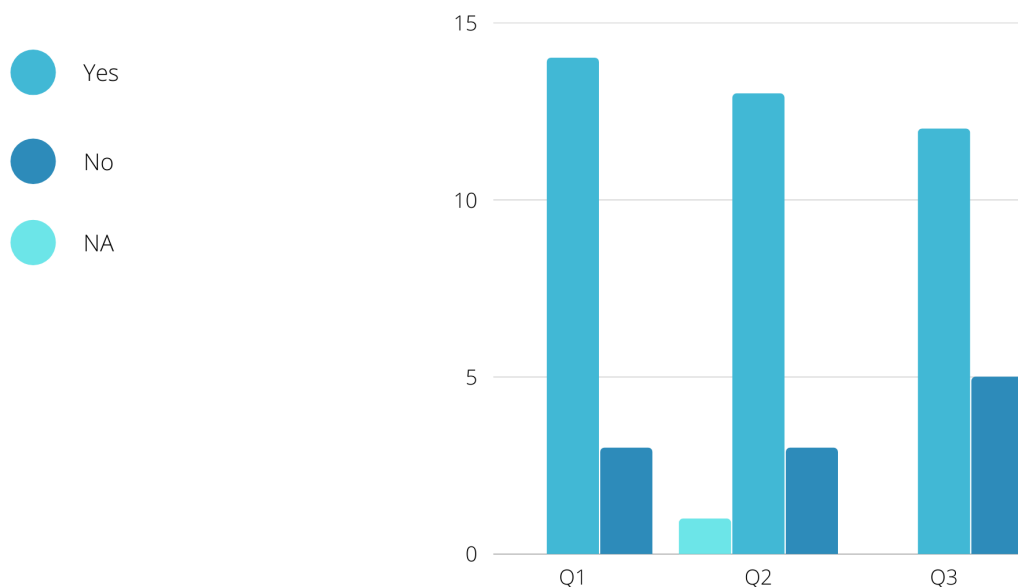


Figure 3.4: Results of questions 1-3

Moreover, the Thesaurus can be used by several audiences, but it was built mainly for cultural heritage audiences. This tool fulfils the project's objectives of increasing data interoperability

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

Project Objectives	Exploitation Outcomes	Target audiences	SILKNOW tools	Results
Advanced searching and semantically relating digitized European silk textile heritage, based on data interoperability across different collections. Moreover, we will focus on small to medium sized heritage institutions, whose digital data tends to be obsolescent, insufficiently curated and not standardized.	Multilingual, Linked Open Data (LOD), Thesaurus concerning silk heritage	Cultural Heritage	SILKNOW Thesaurus	The SILKNOW Thesaurus proved to be useful for domain experts, especially for research purposes. We disseminated the Thesaurus to over more than 50 museums. If they apply some of the concepts used here, we will fulfil the project's objective of improving interoperability across different collections.

Table 3.3: Project objectives related to the exploitation outcomes, target audiences and the Thesaurus.

across different collections as shown in Table3.3.

Finally, the Thesaurus has been analyzed by the External Advisory Board of SILKNOW, who have made some suggestions which are currently being implemented, such as improving visualization of the UI, especially the one related to hierarchy. They also suggested visualizing the entire facets and hierarchies from the AAT, and not just those proper to the SILKNOW Thesaurus.

3.3 Data Harvesting and Conversion

3.3.1 Developing a web crawler and scraper for public museums

With our crawler ¹⁶ we are able to download datasets from 18 sources either via API or manual website crawling. The crawler is made in Node.js. It uses Axios for HTTP requests, and Cheerio for DOM parsing when necessary. All of the data is made publicly available by the respective museums or collections. We receive two more datasets directly from the Garin 1820 and the University of Palermo (UNIPA) collections as they are part of SILKNOW.

When available, we use the REST API of the website / source. The response is converted in JSON if needed, and then stored in the data folder. Images are downloaded and stored locally separately. When there is no API available, we are forced to scrape the HTML pages and to collect the information we need, including, but not limited to: title, description, date, place, material type, and images. The scraping is done using Cheerio, a library that parses HTML markup and provides an API for traversing/manipulating the resulting data structure.

The final output of the crawler is a unified JSON format: each JSON file contains two properties

¹⁶<https://github.com/silknow/crawler>

3.3 Data Harvesting and Conversion

Name of Museum or Collection	Internal Abbreviation	Country	Number of Records
Metropolitan Museum of Art	MET	USA	8317
Victoria and Albert Museum	VAM	UK	7991
CDMT Terrassa	IMATEX	Spain	6127
Rhode Island School of Design	RISD	USA	3338
Boston Museum of Fine Arts	MFA	USA	3297
Garin 1820	garin	Spain	3101*
Red Digital de Colecciones de Museos de España	CER	Spain	1296
Collection du Mobilier National	mobilier	France	1295
Musée d'Arts et d'Industrie de Saint-Etienne	musee-st-etienne	France	1195
Musei di Venezia	venezia	Italy	1163
Musée des Arts Décoratifs	MAD	France	763
Musée des Tissus	MTMAD	France	663
Sicily Cultural Heritage	UNIPA	Italy	438
Art Institute of Chicago	ARTIC	USA	431
Musée du Louvre	louvre	France	399
Joconde Database of French Museum Collections	joconde	France	375
Museo de Arte Sacro El Tesoro de la Concepción	el-tesoro	Spain	277
Paris Musées	paris-musees	France	247
Smithsonian Institution	smithsonian	USA	147
Versaille	versaille	France	73
TOTAL			40873

Table 3.4: Complete table of museum and collection sources. Number of records reflects the number of actually records successfully converted and represented inside the SILKNOW Knowledge Graph. *Garin 1820 had been successfully integrated, but is temporarily deactivated due to an ongoing rights discussion.

with single values, the ID of the record and the source URL. The latter can either be a link to the crawled website or directly to a machine-readable format like JSON via API. After that each crawled JSON file contains two arrays, one called “fields” with sets of different properties which depend on the original data. The other one is an array with all the images together with their respective URLs. Inside the "fields" array the substructure is as follows: every field has exactly one label and then either one value or an array of values. In case of UNIPA the original format from the collection, which was Excel sheets, is converted to this common JSON format with the crawler. In the case of Garin, all the integration takes place in the converter and not in the crawler.

Selecting the sources Investigating for the selection of the right sources for this project had been an ongoing process throughout the whole duration of the SILKNOW project. A Harvesting Logbook has been created for this purpose with which domain experts, historians and computer scientists worked together on selecting museums and collections that fulfilled certain requirements. Sources needed to have metadata and images about historical silk objects from Europe, but also give us an option to access their images and metadata either through an API or a public website. Often this was correlated: only museums or collections with APIs or search functions could properly get evaluated with regards to their content. In total 22 external sources had been listed by the end of this project. Of these, we integrated the data of 18 to at least a certain degree. Some sources had ultimately rejected, because their content was considered non-relevant, some because data access was difficult or not possible to the degree we needed it to be. Table 3.4 shows a complete list of all sources and how many records of them we finally integrated into our knowledge graph. It also includes the internal abbreviation that we used for them, which appear also in this thesis. Figure 3.5 gives an impression of how museum data looks like before we download and convert it at this stage.

3.3.2 Converter software

In all but the aforementioned case of Garin 1820 and the University of Palermo / Sicily Cultural Heritage this common JSON format is then taken as the foundation for the converter software¹⁷ in order to output Terse RDF Triple Language (Turtle) / TTL files that can finally be uploaded to a Triplestore based on the Virtuoso Universal Server.

The challenges faced by the data conversion can be divided into two categories, dealing with ambiguity of data due to the nature of natural language in general, but also due to human errors, like unclean digitization including spelling or other mistakes. Some ways to deal with these problems are to use regular expressions (Regex) to pre-process data or also a constant re-evaluation of the semantic mapping.

¹⁷<https://github.com/silknow/converter>

How does the “raw” museum data look like?

Object Details		Période de création	
Title:	Bed curtain border	1er quart 20e siècle	
Date:	late 19th century	Millésime de création	
Culture:	Russian	1912	
Medium:	Silk, metal	Matériaux - techniques	
		soie, tissé, satin, broché	
		Mesures	
		Hauteur en cm 38 ; largeur en cm 14.5	
Metropolitan Museum of Art (MET)		Joconde	

Figure 3.5: Examples from museum websites of how metadata originally looks like before we apply any software tool

The other main problem is that we do not only have already categorized data, like specific fields that give us short and clearly defined information about e.g. the material or the production year of an object. We also need to make sense and integrate a lot of information that is only available as shorter or longer textual descriptions. For this, we needed more advanced methods from the field of Information Extraction, which will be presented in chapter 4.

The data conversion mostly consists of a translation of the mapping rules into algorithms that assign the classes and properties we defined based on the SILKNOW ontology according to the original fields in the museum metadata. For example, a mapping rule can state that the values of the museum field "Place" have to be mapped as a `E53_Place` class, which is also the value of the property `P8_took_place_on_or_within` attached to the class `E12_Production`. Figure 3.6 shows a screenshot of a mapping table containing several rules that have been implemented with the converter and Figure 3.7.

The RDF conversion is based on a manual mapping for each dataset where fields with labels like "Técnica" and their values are mapped to properties like `P32_used_general_technique`. Two of the most central classes in our knowledge graphs are `E22`, which is used to represent "Man-made objects", and `E12`, which is a class for the "Production" of an object and properties like the production date and the material used.

Some strings need to be parsed with Regular Expressions (regex), for example the Dimensions field, to extract the exact width and height with its respective unit correctly. For instance, the

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

Field	Example	Main Class	Path	Comments
Title	Passementerie		E22_Man-Made-Object P102 has title E35_Title (Passementerie) E17_Type Assignment P41 classified E22_Man-Made-Object E17_Type Assignment P42 assigned E55_Type (Passementerie) E17_Type Assignment P14 carried out by E40_Legal-Body (Victoria and Albert Museum) E17_Type Assignment P2 has type E55_Type (Title)	
Description	'Empty'	S4_Observation	S4_Observation 08 observed E22_Man-Made-Object S4_Observation P3 has note E62_String S4_Observation P2 has type E55_Type (Description)	
Date	18th century	E12_Production	E12_Production P4 has time-span E52_Time-Span P78 is identified by E49_Time-Appellation	
Culture	Italian	E12_Production	E12_Production P7 took place at E53_Place	
teaserText	Date: 18th century Accession Number: 08.48.46	WON'T BE MAPPED		
url	https://www.metmuseum.org/art/collection/search/213382?sortBy=AccessionNumber&department=12&what=Silk%7cTextil	WON'T BE MAPPED		

Figure 3.6: Part of the mapping table for the exemplary record "08.48.46" of MET

Mapping rules - created by domain experts

Field label	Value	Class / Property	Full path	Annotation
Medium	Ribbed silk and wool ground embellished with metallic and silk yarn embroidery	E12_Production	E12_Production P3 has note E62_String	This field gives information on the material and the technique. If they are extracted: E12_Production P126 employed E57_Material (silk, wool, metal) E12_Production P32 use general technique E55_Type (ribbed ground, embroidery)
Materials	silk, metallic yarn	E12_Production	E12_Production P126 employed E57_Material	
Techniques	plain weave, embroidery, embroidering, embroidered, appliqué (preferred spelling), appliqué, ribbed	E12_Production	E12_Production P32 use general technique E55_Type E55_Type P3 has note E62_String	

Mapping excerpt for a record from the Rhode Island School of Design (RISD) museum

3

Figure 3.7: Illustration of how mapping rules are implemented with the converter software

regex pattern

```
(\\d+(?:\\.\\d+)?) x (\\d+(?:\\.\\d+)?) cm
```

in the MET converter makes sure to extract numbers before and after an x if the value ends with “cm”. Furthermore, it makes sure to detect both integers as well as decimals. In some other cases we have one field in the JSON called e.g. "Auteur/exécutant" (Joconde) and it includes two different types of information: an actor and the role of the actor. In case of Joconde we can split it relatively easily as the order of them is always the same. The role becomes the property `ecrm:P2_has_type` of the class `ecrm:E7_Activity`, whereas for the actor its own class `ecrm:E39_Actor` gets created, which is also connected to the former by the property `ecrm:P14_carried_out_by`.

For production dates we also developed a complex parsing and interpretation system to properly represent all dates and to make it possible to search for objects by their date. Originally many string literals were in different formats or some time periods were named differently in the different museums and languages. We can now interpret both single years, year ranges, centuries and most periods in all languages of our datasets. Every unique year or year range gets a unique URI, e.g. <http://data.silknow.org/timespan/1843> for the year 1843 that is linked with every occurrence of that year all across the data. In addition to that, we use the property `P86_falls_within` to link every year with its corresponding century on Getty AAT.

Before some fields get mapped to classes and properties in RDF, their string values are getting checked if they are matching with some values in controlled vocabularies: places with Geonames, materials, techniques and motifs with the Silknow Thesaurus and the Getty Art & Architecture Thesaurus (AAT). In case of a match, the original string of the field of the dataset gets replaced with a URI of the concept in one of these vocabularies. For example: the technique "Embroidery" has the link <http://data.silknow.org/vocabulary/87> and a string that can be identified as either "embroidery", a synonym or translation like "Bordado" (Spanish) would trigger this linking.

3.3.3 Evaluation

The quality of the converter software has once been validated during a late stage of its development [128].

This subsection describes this validation in all details and is split into different parts for Objectives, the method, the results and finally which improvements have been added to improve the quality with regards to the identified problems.

Objectives

This validation focused on the faceted browser (available at <http://data.silkknow.org/fct/>) and followed two different rationales. The first validation activity aimed at verifying the coherence of what the knowledge graph presents, in comparison with the information stored in the records provided by the members of the consortium. The second one aimed at validating the correct interpretation of the Location attribute.

During the first validation activity, we involved domain experts from the UNIPA. Data providers in connection with this partner had provided several records with information about objects and textiles:

1. Museo Diocesano di Caccamo
2. Museo Diocesano di Palermo
3. Museo di Termini Imerese
4. Cattedrale di Palermo
5. Duomo di Monreale

the records from UNIPA were imported into the knowledge graph based on the Data Model described in section 3.1.

This validation activity objective is to verify whether all the attributes present in the UNIPA records have been correctly reported in the Knowledge Graph and are available through the web faceted browser.

During the second validation activity, we involved domain experts from UNIPA to validate data related to the production location, i.e., the place of origin where the textile was made. Experts had to verify if all data inserted in this field were correct, whether they effectively represent the place of origin and not the place where they are located now, or any previous location.

Method

The method employed for the first validation activity consists in comparing the original database records to its correspondence in the web faceted browser. Figure 3.8 shows an example of an original record and Figure 3.9 the corresponding record in the web faceted browser.

3.3 Data Harvesting and Conversion

Time chronology	Prima metà del XVIII secolo (1745 – 1750)
Geography	Italia
Region production	Sicilia o Campania
Description	Pianeta, stola e manipolo
Technique	Taffetas broccato à liage rêpris
Museum	Chiesa Madre di Caccamo: Chiesa di San Giorgio Martire, Piazza Duomo n.2, 900120
Language	Italiano
Dimensions	cm 108 x 68
State of preservation	discreto; locali slegature delle trame supplementari
Width	non rilevabile
Pattern unit	cm 48,5 x n.r.; numero dei campi: 1(?); tipo di campo: a ritorno.
Warp	1 ordito, di fondo, organzino di seta, 2 capi, S, colore avorio. Scalatura: 4 fili; Riduzione:
Weft	5 trame; I di fondo, seta, 4 capi, STA, colore avorio; II broccata, oro filato ritorto avvolto su
Construction	fondo in taffetas prodotto da tutti i fili e da tutte le trame di fondo. Opera creata da
Embroidery	assente
Description of the pattern	motivo a impostazione simmetrica e speculare. Lungo un asse verticale si dispongono
Galloon	a fuselli, in argento filato e filo di seta, cm 4, con motivo a nastro sinuoso con valva, nelle
Lining	in tela di lino di colore azzurro
Historical Critical Information	Il parato è composto da una pianeta, una stola, un manipolo. Dimensioni: cm 108 x 68 ;
Images (names of the images in the document)	SCHEDA88.jpg; SCHEDA88.1.jpg
Author of the technical analysis	R. Civiletto
Author of the Historical Critical Information	M. Vitella

Figure 3.8: The table reporting an example of the UNIPA record named Caccamo 7

About: <http://data.silknow.org/object/a7737895-1023-39b2-98ec-a23e5a2983c1> [Goto Source](#) [Not Distinct](#) [Permalink](#)

An Entity of Type: [scdm:E22_Man-Made_Object](#), within Data Space: [data.silknow.org](#) associated with source document(s)

Type: [E22 Man-Made Object](#) Command: [Start New Facet](#) Go

Attributes	Values
rdfs:type	E22 Man-Made Object
rdfs:label	Pianeta, stola e manipolo, Prima metà del XVIII secolo (1745 – 1750), Sicilia o Campania ⁽ⁱ⁾
rdfs:comment	non rilevabile ⁽ⁱ⁾ cm 48,5 x n.r.; numero dei campi: 1(?); tipo di campo: a ritorno. ⁽ⁱ⁾ 5 trame; I di fondo, seta, 4 capi, STA, colore avorio; II broccata, oro filato ritorto avvolto su anima in seta ritorta, 2 capi, S, colore giallo; III oro filato avvolto su anima in seta, 2 capi, S, colore giallo; IV di accompagnamento, oro lamellare; V broccata, seta, 3 capi, S, nei colori: rosa salmone, marrone, violetto, aragosta. Scalatura: 1 passata; Riduzione: 16 passate/cm; Proporzioni: 2 di fondo/1 broccata. ⁽ⁱ⁾ motivo a impostazione simmetrica e speculare. Lungo un asse verticale si dispongono grandi composizioni vegetali caratterizzate da cespi di piccole foglie verdi che si affiancano ad altre dorate, carnose e sfrangiate con punta arricciata. A esse si intercalano infiorescenze di rosa, peonia, mazzolini di fiori di campo, mentre in basso la composizione è definita da un festone di minuti fiori. ⁽ⁱ⁾ Il parato è composto da una pianeta, una stola, un manipolo. Dimensioni: cm 108 x 68 ; cm 216 x 22 ; cm 100 x 20. Lornato floreale, disposto con speculare simmetria e caratterizzato da grande rapporto disegnativo, ripropone una tipologia una tipologia decorativa diffusa verso la metà del XVIII secolo. Nella composizione floreale si riconoscono citazioni desunte dai tessuti bizzarre e Revel: si scorgono larghe foglie frangiate e una grande infiorescenza centrale che insieme convivono in un'originale composizione la cui disposizioni ricorda le più serrate soluzioni adottate quasi un cinquantennio prima per le stoffe stilisticamente definite "a pizzo". Si tratta di un modulo decorativo che con simili espressioni è stato già riscontrato in Sicilia, come attestano la pianeta della maggior chiesa di Termini Imerese (cfr. R. Civiletto – M. Vitella, scheda n. 15 in M. C. Di Natale – M. Vitella, <i>Ori e stoffe</i> , ..., 1997, pp. 82 – 83) e il parato di provenienza ignota custodito presso la Galleria Regionale della Sicilia di Palazzo Abatellis (cfr. E. D'Amico Del Rosso, scheda n. 79 in I paramenti, ..., 1997, p. 125). La bassa risoluzione di orditi al centimetro quadrato potrebbe far ritenere il tessuto in esame di manifattura siciliana (cfr. R. Orsi Landini, <i>Tessuti bizzarre di produzione siciliana in Splendori</i> , ..., 2001, p. 678.). (R. Civiletto; M. Vitella) ⁽ⁱ⁾ -more-
sameAs	
dc:identifier	Caccamo7
P3 has note	non rilevabile ⁽ⁱ⁾ cm 48,5 x n.r.; numero dei campi: 1(?); tipo di campo: a ritorno. ⁽ⁱ⁾ 5 trame; I di fondo, seta, 4 capi, STA, colore avorio; II broccata, oro filato ritorto avvolto su anima in seta ritorta, 2 capi, S, colore giallo; III oro filato avvolto su anima in seta, 2 capi, S, colore giallo; IV di accompagnamento, oro lamellare; V broccata, seta, 3 capi, S, nei colori: rosa salmone, marrone, violetto, aragosta. Scalatura: 1 passata; Riduzione: 16 passate/cm; Proporzioni: 2 di fondo/1 broccata. ⁽ⁱ⁾ motivo a impostazione simmetrica e speculare. Lungo un asse verticale si dispongono grandi composizioni vegetali caratterizzate da cespi di piccole foglie verdi che si affiancano ad altre dorate, carnose e sfrangiate con punta arricciata. A esse si intercalano infiorescenze di rosa, peonia, mazzolini di fiori di campo, mentre in basso la composizione è definita da un festone di minuti fiori. ⁽ⁱ⁾ Il parato è composto da una pianeta, una stola, un manipolo. Dimensioni: cm 108 x 68 ; cm 216 x 22 ; cm 100 x 20. Lornato floreale, disposto con speculare simmetria e caratterizzato da grande rapporto disegnativo, ripropone una tipologia una tipologia decorativa diffusa verso la metà del XVIII secolo. Nella composizione floreale si riconoscono citazioni desunte dai tessuti bizzarre e Revel: si scorgono larghe foglie frangiate e una grande infiorescenza centrale che insieme convivono in un'originale composizione la cui disposizioni ricorda le più serrate soluzioni adottate quasi un cinquantennio prima per le stoffe stilisticamente definite "a pizzo". Si tratta di un modulo decorativo che con simili espressioni è stato già riscontrato in Sicilia, come attestano la pianeta della maggior chiesa di Termini Imerese (cfr. R. Civiletto – M. Vitella, scheda n. 15 in M. C. Di Natale – M. Vitella, <i>Ori e stoffe</i> , ..., 1997, pp. 82 – 83) e il parato di provenienza ignota custodito presso la Galleria Regionale della Sicilia di Palazzo Abatellis (cfr. E. D'Amico Del Rosso, scheda n. 79 in I paramenti, ..., 1997, p. 125). La bassa risoluzione di orditi al centimetro quadrato potrebbe far ritenere il tessuto in esame di manifattura siciliana (cfr. R. Orsi Landini, <i>Tessuti bizzarre di produzione siciliana in Splendori</i> , ..., 2001, p. 678.). (R. Civiletto; M. Vitella) ⁽ⁱ⁾ -more-
P43 has dimension	68 108
P65 shows visual item	Floral motif
P138 has representation	http://data.silknow.org/image/6613b6ea-700f-3fa8-a40c-5239ef5e466c http://data.silknow.org/image/71be2262-1353-3c40-9ead-e711e2d7738a
P58 has section definition	http://data.silknow.org/object/a7737895-1023-39b2-98ec-a23e5a2983c1/dimension/1/pattern/1
P9 consists of	http://data.silknow.org/object/a7737895-1023-39b2-98ec-a23e5a2983c1/activities/1

Figure 3.9: Screenshot of the Faceted Browser view of the same Caccamo 7 record as in figure 3.8

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

We validated 29 records about objects and information from the above said museums. From the analysis of how the record in the faceted browser is structured against how the data providers have created the record, we generalized the following object identifiers to check:

1. Construction and Technique from the original record are combined in the `P32_used_general_technique` entity inside the `P108_has_produced` entity. Since information about construction and technique are two different rows in the original records, how many of them have been combined?
2. Some fields in the original records (Historical Critical Information, Warp, Weft, Width and Description of pattern) are combined in `P3_has_note` in the SILKNOW data model. How many of them have been combined in the faceted browser?
3. `crmsci:08_observed` put together several rows of the original records. Does this class contain all the rows, all previous information, for each record?
4. Availability of information about the dimensions of the fabric (or the object at large). Is that information available, in the original record and in the faceted browser, through the class `P34_has_dimension` (Y/N)?

Regarding the second validation, we navigated the Knowledge graph starting from “Man-Made Object” (E22 with reference to the ontology) and the related `P108_has_produced` entity. We considered objects mainly coming from two museums namely the one labelled with the identifier 95.71.XXX and GP00XXX but also from a few others. For each object, we checked the production page (by clicking on `P108_has_produced_entity` and the related `P8_took_place_on_or_within` entity).

Results

As regards the first validation, results are reported in Figure 3.10 and can be summarized as follows:

1. More than 50% of records in the faceted browser have information about an objects' dimensions in the field `P34_has_dimension`. For instance, some records are related not to individual pieces but to sets of textiles (for instance Caccamo7 presents 3 pieces in a single record)
2. In 100% of records, construction and techniques have been adequately combined in `P32_used_general_technique`
3. 100% of historical critical information, warp, weft, width and description of a pattern has been adequately combined

3.3 Data Harvesting and Conversion

Object identifier	Validation#1				P34 has dimension (Y/N)	some typos		
	Construction and Technique are combined in the P32 used general technique	Historical critical information, Warp, Weft, Width, Description of pattern	In 08 observed each observation refers to one record og	PatternUnit, Lining, Galloon, Embroidery, State of preservation are lacking				
DiocesanoPA6	y	y	y	y	y			
Caccamo7	y	y	y	y	N	3 manufatti in uno stesso record		
CattedralePa2	y	y	y	y		Dimensions only refer to Planeta, dimensions of stola, manipolo, borsa		
DiocesanoPA2	y	y	y	y	y			
Monreale1	y	y	y	y	y			
Caccamo6	y	y	y	y	N			
Caccamo1	y	y	y	y	y			
DiocesanoPa3	y	y	y	y	y			
DiocesanoPa8	y	y	y	y	n			
Terminimerese4	y	y	y	y		Dimensions only refer to one Planeta, dimensions of the other three e Planetas,manipolo, velo di calice and		
Caccamo8	y	y	y	y	n			
DiocesanoPA4	y	y	y	y	y			
Terminimerese2	y	y	y	y		Dimensions only refer to one Planeta, dimensions of the other Planetas, 3 manipolo, 1 velo di calice, 2 stolas, 1		
Caccamo4	y	y	y	y	n			
Terminimerese6	y	y	y	y	y			
Caccamo2	y	y	y	y	y			
Caccamo3	y	y	y	y	n			
Caccamo5	y	y	y	y	n			
CattedralePa1	y	y	y	y		Dimensions only refer to the Piaviale, dimensions of the Planetas, manipolo, stola, borsa e palla lack	In P32used_genera_Technique the	
DiocesanoPA1	y	y	y	y	y			
DiocesanoPa7	y	y	y	y	n			
Monreale2	y	y	y	y	y			
CattedralePA3	y	y	y	y		The dimension of one stola lacks		
DiocesanoPA5	y	y	y	y	n			
Monreale3	y	y	y	y	y			: Isolati observatio 4 refuso >>
Terminimerese1	y	y	y	y		only one dimension (wxh) instead of 4		
Terminimerese5	y	y	y	y		only one dimension (wxh) instead of 3		
Terminimerese7	y	y	y	y				
Terminimerese5	y	y	y	y	y			un e senza accento

Figure 3.10: Results of the first validation

4. 100% of the previous information are in the crmsci : 08_observed entity

Regarding the second validation, results are summarized as follows. Figure 3.11 shows that all records coming from Garin and owning the identifier T000877 present in the class E2_production, a value of P8_took_place_on_or_within equals to “chalet Garin”. This value cannot be associated with the production place, but with the current storage location.

For records mapped from RISD, Figure 3.12 shows that the description of the object does not allow the expert to understand whether the value of “P8_took_place_on” is the right place of origin. Several records, like the one shown in Figure 3.13 do not present the value of “P8_took_place_on”, so they cannot be evaluated.

Some records (see Figure 3.14) present a description of the objects that do not allow the expert to establish if the value in P8_took_place_on is the place of origin. More details in the description are needed. Most records from Italian museums (see Figure 3.15) present the right location with the exception of the ones from Monreale ones.

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

GARIN

For all the objects with this kind of identifier this might be the location in which they are now

Record 1: <http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa>

Attributes	Values
rdf:type	E22_Man-Made_Object
rdfs:comment	
P138 has representation	http://data.silkknow.org/image/86d0f363-3c09-3c43-b01f-b54407656446 http://data.silkknow.org/image/614728a2-52a7-3724-903b-17b31d838862
P1 is identified by	T000877
P3 has note	
P43 has dimension	http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/dimension/h http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/dimension/n
dc:identifier	T000877
P108 has produced of	http://data.silkknow.org/production/ab560250-a79f-3c25-8090-f6e137782216
P129 is about of	T000877
P24 transferred title of	http://data.silkknow.org/event/88640a4-5ed6-3709-9a30-80325a7e43d4
P25 moved of	http://data.silkknow.org/event/505cbe49-0e83-3961-9900-720fe60e12f
P34 concerned of	http://data.silkknow.org/assessment/2a7e1694-4f42-3b3a-9242-e3772d62fbc9
P39 measured of	http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/dimension/measurement
P41 classified of	http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/type_assignment/1 http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/type_assignment/2
crmsci:OB observed of	http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/observation/1 http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa/observation/2

Record 2: <http://data.silkknow.org/production/ab560250-a79f-3c25-8090-f6e137782216>

Attributes	Values
rdf:type	E12_Production
P108 has produced	http://data.silkknow.org/object/fda80292-91af-380c-a09b-5864e232dbaa
P126 employed	Silk Cotton
P32 used general technique	Hand loom
P4 has time-span	Primer to the siglo XX
P8 took place on or within	chalet garin
P9 consists of	http://data.silkknow.org/production/ab560250-a79f-3c25-8090-f6e137782216/activity/1
P129 is about of	T000877

Figure 3.11: First example - records from GARIN

3.3 Data Harvesting and Conversion

OPENLINK SOFTWARE

About: Textile [Goto Source](#) [Not Distinct](#) [Permalink](#)
An Entity of Type : [scrm:E22_Man-Made_Object](#), within Data Space : [data:silknow.org](#) associated with source [document\(s\)](#)
Type: [E22 Man-Made Object](#) Command: [Start New Facet](#) Go

Attributes	Values
rdfs:type	E22_Man-Made_Object
rdfs:label	Textile
P102_has_title	Textile
P138_has_representation	http://data.silknow.org/image/0c44e48c-640b-357b-920b-06efef3dfba3 http://data.silknow.org/image/b8272e3-4f2a-33d5-9264-62ed4f564ce5 http://data.silknow.org/image/71b9a0aa-3b1e-3a3f-acea-ef5f1b966f1e
P1_is_identified_by	1988.082.9
P43_has_dimension	http://data.silknow.org/object/63093f99-f80-3062-a525-5a33315a92e1/dimension/h http://data.silknow.org/object/63093f99-f80-3062-a525-5a33315a92e1/dimension/e
dc:identifier	1988.082.9
is P108_has_produced_of	http://data.silknow.org/production/9181de91-3d71-3a63-bcfe-9fe6bb2bc108
is P129_is_about_of	ID: 1988.082.9 , Filename: 1019271.json
is P24_transferred_title_of	http://data.silknow.org/event/0a277704-4b21-3c1b-b276-7b878b9ea04a
is P30_transferred_custody_of	http://data.silknow.org/event/85fa7397-9cfe-3d34-8ce3-8235c75d4dca
is P39_measured_of	http://data.silknow.org/object/63093f99-f80-3062-a525-5a33315a92e1/dimension/measurement
is P41_classified_of	http://data.silknow.org/object/63093f99-f80-3062-a525-5a33315a92e1/type_assignment/1

RISD
No description, it is not possible to say that P8 took place is the right place of origin, maybe expert form the original museum should be involved

OPENLINK SOFTWARE

About: [http://data.silknow.org/production/9181](#)
An Entity of Type : [scrm:E12_Production](#), within Data Space : [data:silknow.org](#) associated with source [document\(s\)](#)
Type: [E12 Production](#) Command: [Start New Facet](#) Go

Attributes	Values
rdfs:type	E12_Production
P108_has_produced	Textile
P126_employed	Silk
P4_has_time-span	late 1600s - early 1700s
P8_took_place_on_or_within	Netherlands
is P129_is_about_of	ID: 1988.082.9 , Filename: 1019271.json

Alternative Linked Data Documents: [ODE](#) Content

Figure 3.12: Second example - records from RISD

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

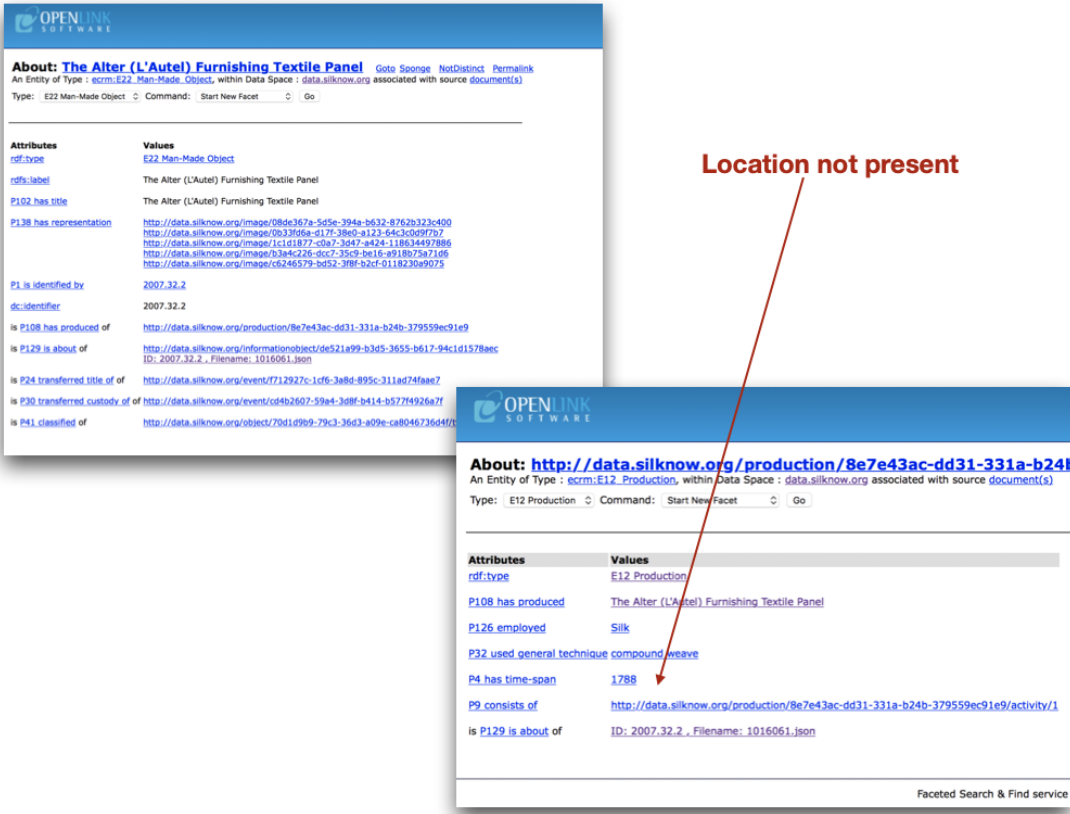


Figure 3.13: An example of record where location has not been reported

3.3 Data Harvesting and Conversion

1. Description here
2. from the above description it is not possible for UNIPA expert to establish the production origin

About: Sampler
An Entity of Type : `ecrm:E22_Man-Made_Object`, within Data Space : `data.silknow.org` associated with source `document(s)`
Type: `E22 Man-Made Object` Command: `Start New Facet` Go

British Galleries: This is the earliest surviving piece of Martha Edlin's needlework, completed when she was eight years old young girl's needleworking skills, through a range of stitches and techniques. [2/03/2003]

Attributes	Values
<code>rdf:type</code>	<code>E22 Man-Made Object</code>
<code>rdfs:label</code>	Sampler
<code>rdfs:comment</code>	Sampler
<code>E138 has representation</code>	http://data.silknow.org/instance/13108467-8415-3d96-ba09-e064e046614 http://data.silknow.org/instance/74947474-36ca-3d99-8228-cb2e6d4409c http://data.silknow.org/instance/0648c29-0932-3773-494f-a0091195c16e
<code>E139 identified by</code>	T-433-1990
<code>E139 has note</code>	
<code>P43 has dimension</code>	http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/dimension/h http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/dimension/v
<code>dc:identifier</code>	T-433-1990
<code>is P108 is composed of of</code>	Textiles and Fashion Collection
<code>is P108 has produced of</code>	http://data.silknow.org/production/d4ec41ba-a4d3-3ebb-ba07-85671ad99cb
<code>is P129 is about of</code>	O10005
<code>is P30 transferred custody of of</code>	http://data.silknow.org/information/object/3bed21c2-197c-323f-9040-634887b80449
<code>is P39 transferred custody of of</code>	http://data.silknow.org/object/20fa0a75-7869-3767-831e-b738e1ad8e7
<code>is P39 measured of</code>	http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/dimension/measurement
<code>is P41 classified of</code>	http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/type_assignment/1 http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/type_assignment/2 http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/type_assignment/3
<code>is <code>ecrm:Q8</code> observed of</code>	http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/observation/1 http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/observation/2 http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/observation/3 http://data.silknow.org/object/0399045f-21e1-3488-80fc-5620eb7e6d7/observation/4

About: <http://data.silknow.org/production/d4ec41ba-a4d3-3ebb-ba07-85671ad99cb/activity/1>
An Entity of Type : `ecrm:E12_Production`, within Data Space : `data.silknow.org` associated with source `document(s)`
Type: `E12 Production` Command: `Start New Facet` Go

Attributes	Values
<code>rdf:type</code>	<code>E12 Production</code>
<code>P108 has produced</code>	Sampler
<code>P126 employed</code>	linen (material)
<code>P32 used general technique</code>	embroidering
<code>P4 has time-span</code>	1668 (dated)
<code>P8 took place on or within</code>	England
<code>P9 consists of</code>	http://data.silknow.org/production/d4ec41ba-a4d3-3ebb-ba07-85671ad99cb/activity/1
<code>is P129 is about of</code>	O10005

Faceted Search & Find service v1.16.10

Figure 3.14: An example of a too short description

3.4 Data Access

To integrate SPARQL queries and their output into web development can be a challenge, even when the output format is JSON: It contains unnecessary metadata, each value has a datatype and is part of a bigger array with its own name and the attributes "type" and "value" or identical bindings that for example only differ in the language tag are not automatically merged and displayed multiple times. Mapping the results to another structure can be difficult, especially if avoiding to hard-code queries into the application's code.

3.4.1 Graphical interface for the Thesaurus - SKOMOS

We deployed the Skosmos open source tool [125] developed at <https://github.com/NatLibFi/Skosmos> to visualize the SILKNOW thesaurus¹⁸. The user interface¹⁹ is localized in English, Spanish, French and Italian and adapts according to the preferred language of the user's web browser. Skosmos is configured to load the data from the SILKNOW RDF endpoint and generates its view doing SPARQL queries directed to <http://data.silknow.org/sparql>.

Figure 3.16 depicts the welcome page and enables to select the SILKNOW thesaurus, which was originally based on the Getty Art & Architecture Thesaurus²⁰, but is now heavily extended and much more specialized.

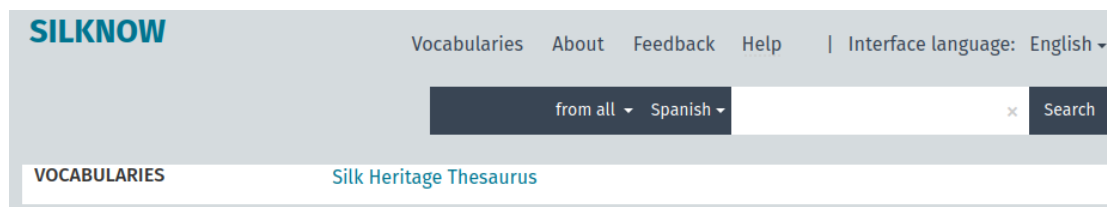


Figure 3.16: Homepage of Skosmos configured to browse the SILKNOW Thesaurus

The Figure 3.17 depicts the general metadata of the SILKNOW thesaurus. The metadata includes the creation date and the last modification date of the thesaurus. The current version 2.93 of the SILKNOW thesaurus contains 666 concepts. On the left side, the user can click on any concept to have more detailed information. Two views are offered: the alphabetical order of the concepts or their organization in a hierarchy following the broader / narrower relationships.

The Figure 3.18 depicts the detailed view of the Acanalado concept, which definition, in Spanish, reads:

¹⁸<https://github.com/silknow/skosmos>

¹⁹<https://skosmos.silknow.org/thesaurus/en/>

²⁰<https://www.getty.edu/research/tools/vocabularies/aat/>

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

The screenshot displays the SILKNOW website's 'Silk Heritage Thesaurus' page. The interface includes a navigation bar with 'Vocabularies', 'About', 'Feedback', and 'Help', along with an 'Interface language: English' dropdown. A search bar at the top right contains 'Content language English' and a search button. On the left, there are tabs for 'Alphabetical' and 'Hierarchy', with an alphabetical index (A-Z) below. A list of terms is shown, with 'Anacosta' highlighted. The main content area, titled 'Vocabulary information', provides metadata for the selected term:

LABEL	Silk Heritage Thesaurus
CREATOR	http://data.silknow.org/actor/SILKNOW
CONTRIBUTOR	Georgia Lo Cicero. Università degli Studi di Palermo Maurizio Vitella. Università degli Studi di Palermo Eliseo Martínez. Universitat de València Arabella León. Garín 1820 Pasquale Lisena. EURECOM María Roca. Universitat de València Isabel Insa. Universitat de València Florence Charpigny. CNRS-LARHRA Raphaël Troncy. EURECOM Catherine Vermorel. CNRS-LARHRA Pierre Vernus. CNRS-LARHRA Thibault Ehrhart. EURECOM Marie Puren. CNRS-LARHRA Mar Gaitán. Universitat de València Thomas Schleider. EURECOM
VERSION	2.93
CREATED	Friday, November 9, 2018 00:00:00
LAST MODIFIED	Monday, September 13, 2021 00:00:00
CREATED ON	Monday, September 13, 2021 00:00:00
TYPE	http://www.w3.org/2004/02/skos/core#ConceptScheme

Figure 3.17: General page showing metadata of the SILKNOW Thesaurus

SILKNOW Vocabularies About Feedback Help | Interface language: English ▾


Silk Heritage Thesaurus Content language: Spanish ▾ Search

Alphabetical Hierarchy

- Proceso y producción de textiles
- Sericicultura
- Teñido
- Tisaje
- Bordado
- Elementos interfuncionales
- Basta
- Cordelina
- Cruce
- Flotante
- Hilo de trama
- Hilo de urdimbre
- Hilo de vuelta
- Intervalo
- Ligadura
- Ligamento
 - Curso de ligamento
 - Escalonado
 - Fondo (ligamento)
 - Ligamento compuesto
 - Ligamento ligero
 - Ligamento pesado
 - Ligamento por trama
 - Ligamento por urdimbre
 - Ligamento simple
 - Ligamento derivado
 - Acanalado (ligamento)**
 - Acanalado alterno
 - Acanalado con dos urdimbres
 - Acanalado contrastrado
 - Acanalado de las Indias
 - Acanalado longitudinal
 - Acanalado oblicuo
 - Acanalado transversal
 - Gro
 - Radiado
 - Ligamento fundamental
 - Sombreados

... > Ligadura > Ligamento > Ligamento simple >
Ligamento derivado > Acanalado (ligamento)

PREFERRED TERM **Acanalado (ligamento)**



DEFINITION Ligamento derivado del tafetán en el que los extremos de la urdimbre o los picos de la trama, o ambos, se mueven en grupos de dos o más. Pueden de base regular o irregular

BROADER CONCEPT [Ligamento derivado](#)

NARROWER CONCEPTS [Acanalado alterno](#)
[Acanalado con dos urdimbres](#)
[Acanalado contrastrado](#)
[Acanalado de las Indias](#)
[Acanalado longitudinal](#)
[Acanalado oblicuo](#)
[Acanalado transversal](#)
[Gro](#)

RELATED CONCEPTS [Acanalado \(atributo\)](#)
[Tafetán \(ligamento\)](#)

BIBLIOGRAPHIC CITATION Castany Saládrigas, Francisco. Diccionario de tejidos. Etimología, origen, arte, historia y fabricación de los más importantes tejidos clásicos y modernos. Gustavo Gil S.A., 1949
CIETA. vocabulario Técnico Tejidos Español, francés, inglés, italiano. Lyon: Centre International d'Etude des Textiles Anciens, 1963
Dávila Corona, Rosa, Duran i Pujol, Montserrat, y García Fernández, Máximo. Diccionario histórico de telas y tejidos castellano-catalán. Salamanca:

Figure 3.18: Detailed view of the Acanalado (ligamento) concept in Spanish ("Extended tabby" in English)

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

"Ligamento derivado del tafetán en el que los extremos de la urdimbre o los picos de la trama, o ambos, se mueven en grupos de dos o más. Pueden de base regular o irregular"

Furthermore, this concept is defined as narrower than the concept <https://skosmos.silknow.org/thesaurus/es/page/640> (Spanish - Ligamento derivado / English - Derived weave). It is itself the broader concept of 8 other concepts and it is a related concept to both "Acanalado (atributo)" and "Tafetán (ligamento)". Figure 3.19 gives another impression of a concept in the Thesaurus and highlights some of its main features.

Moiré as a skos:Concept

The screenshot shows the SILKNOW Thesaurus entry for 'Moiré (technique)'. The entry includes a definition, broader concept, related concepts, entry terms, bibliographic citation, belongs to group, and in other languages. Annotations with arrows point to specific parts of the entry:

- Broader: Techniques of weaving** points to the 'BROADER CONCEPT' field.
- Related: Cannele** points to the 'RELATED CONCEPTS' field.
- Alternative labels and translations** points to the 'ENTRY TERMS' field.
- Alternative labels and translations** also points to the 'IN OTHER LANGUAGES' field.

The URL <https://skosmos.silknow.org/thesaurus/> is displayed at the bottom right of the screenshot.

Figure 3.19: Example of the concept "Moiré" in the SILKNOW Thesaurus

3.4.2 Access through semantic queries - SPARQL Endpoint

Once expressed with RDF, we upload all data to a SPARQL endpoint²¹ from which it can be queried. The Knowledge Graph constitutes the foundation for all further data-driven work and tools that are part of SILKNOW in general. Additionally, we offer a Faceted Browser, a RESTful API as well as an exploratory search engine to make the data more easily available which are further detailed in the following subsections.

The SPARQL Query Language is a declarative query language (like SQL) for performing data manipulation and data definition operations on data represented as a collection of RDF statements [125].

A SPARQL query has a solution modifier (or head) and a query body. The solution modifier

²¹<https://github.com/silknow/knowledge-base>

provides the basis for categorizing different types of SPARQL query solutions. The query body comprises a collection of RDF statement patterns that represent the entity relationships to which a query is scoped. The solution modifier includes read-oriented data access (SELECT, ASK, DESCRIBE, CONSTRUCT) and write-oriented data access (CREATE, INSERT, UPDATE, DELETE, CLEAR, DROP).

A SPARQL Query Service is an HTTP Service (also known as a Web Service) that offers an API (Application Programming Interface) for performing declarative data definition and data manipulation operations on data represented as RDF sentence collections, via GET, POST, and PATCH operations that support query solution (result set) delivery using a variety of negotiable document types. SPARQL Queries are executable directly from any computer using cURL ²², a command line tool and library for transferring data with URLs. The endpoint can be queried in a way to get the results in the JSON format, which is the favorite format of web developers.

3.4.3 Access for web developers - SPARQL Transformer

With a combination of grlc ²³ and SPARQL Transformer [81] we were able to create an easy API access for the SILKNOW knowledge graph, which makes it possible for web developers to directly work with a more suitable format ²⁴. SPARQL Transformer relies on a single JSON object for defining which data should be extracted from the endpoint and in which shape. SPARQL bindings are merged on the base of the identifiers and the grlc API framework. With its graphical interface the knowledge graph can also be searched with SPARQL Transformer for any strings of the type time, location, material or technique and the output is displayed in a simpler JSON format.

3.4.4 Access for web development - RESTful API

Representational state transfer (REST) is an architectural style for distributed hypermedia systems. It is nowadays a widely accepted guidance and style for web APIs. Such APIs that stick to REST constraints are called RESTful APIs. Part of their definition when HTTP-based are the following methods to perform actions on resources: GET, POST, PUT, DELETE.

Based on SPARQL transformer and grlc (see subsection above) our exploratory search engine ADASilk (see also section 5.1) aims to provide access to the data stored in the SILKNOW Knowledge Graph (KG) through a public RESTful API ²⁵.

The web application is produced using React, a JavaScript library to build user interfaces. It

²²<https://curl.haxx.se/>

²³<http://grlc.io/>

²⁴<http://grlc.io/api/silknow/api>

²⁵<https://github.com/silknow/adasilk>

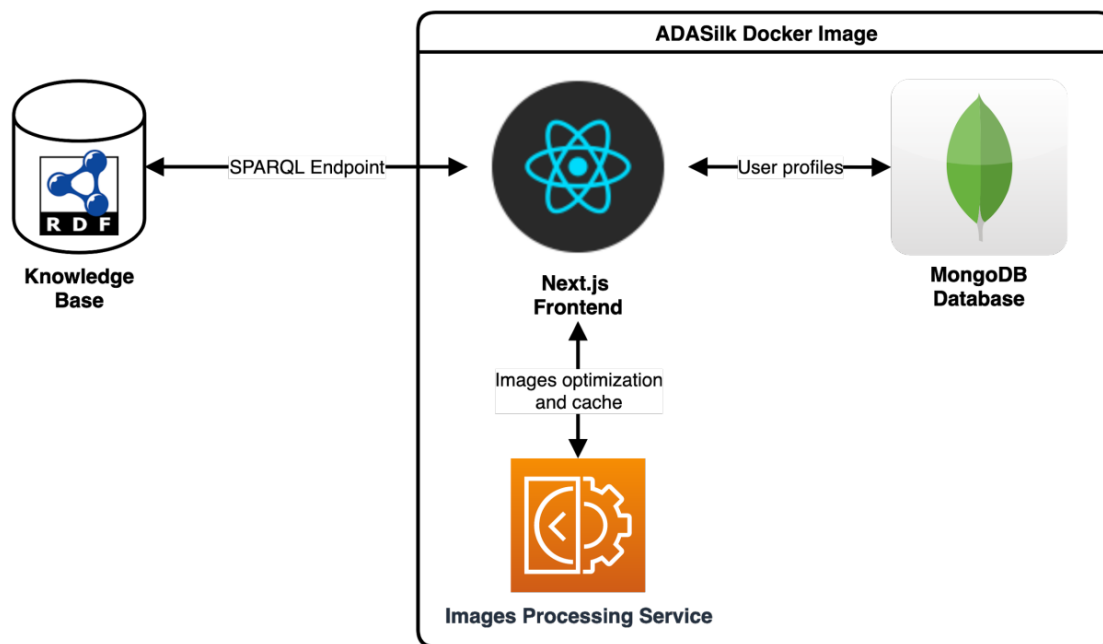


Figure 3.20: General architecture of ADASilk exploratory search User Interface

uses encapsulated components that manage their own state to help maximize code reusability. The application is also built on top of Next.js, a framework used for server-side rendering and page-based routing. Requests are made to the Knowledge Graph using SPARQL queries and the SILKNOW API, and the result is then rendered as HTML which is sent back to the user's browser (Figure 3.20). In addition to a triple store hosting the SILKNOW knowledge graph, ADASilk also uses a local MongoDB database in order to store user profiles and their saved lists. The image processing service enables to dynamically render the images illustrating the silk fabrics into a web-friendly resolution for optimizing the loading time of the pages in the web application.

The front-end uses several web technologies, namely:

- React²⁶, for components based rendering.
- styled-components²⁷, for styling React components using scoped CSS (Cascading Style Sheet).
- Next.js,²⁸ for server-side rendering.
- i18next²⁹, for the internationalization.

²⁶<https://facebook.github.io/react>

²⁷<https://styled-components.com/>

²⁸<https://zeit.co/blog/next>

²⁹<https://github.com/isaachinman/next-i18next>

- `next-auth`³⁰, for OAuth authentication.
- `sparql-transformer`³¹, for JSON based SPARQL requests.

ADASilk makes requests to the SILKNOW Knowledge Graph through the exposed SPARQL endpoint. Queries are generated using the `sparql-transformer` library and are defined in a configuration file.

The full architecture is developed in a microservice approach, implemented within the Docker framework. Thanks to the use of independent and self-sufficient containers, Docker enables the deployment of this architecture on any machine, without any particular software requirements.

3.4.5 Evaluation

In this subsection we will present an evaluation of the data access through our exploratory search engine ADASilk, aimed at determining its robustness in answering user requests [101]. This validation includes every part of the full stack that is fueling ADASilk:

1. ADASilk API (Internal) used by the ADASilk web application
2. SILKNOW's API (Public) powered by the SPARQL Transformer which can be used for third party integration
3. SPARQL API powered by a Virtuoso triple store.

To this end, we have prepared some experiments by simulating different numbers of concurrent users launching different requests, increasing the complexity of the query associated with the request. The experiments are described in subsection "Test description". During the experiments, we measured different parameters, explained in "Definition of the parameters measured", which are used to assess if the tool performs well, even when producing and interacting with complex requests and a large number of concurrent users. The data gathered during the testing is provided in section "Gathered data during the tests" and analysed in subsections "Results for ADASilk Internal API" to "Results for SPARQL API". Finally, a brief discussion is set out in the last subsection.

³⁰<https://github.com/iaincollins/next-auth>

³¹<https://github.com/D2KLab/sparql-transformer>

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

Number of users	Requests per user	Times repeated	Total requests
5	8	5	200
10	8	5	400
30	8	5	1200
50	8	5	2000

Table 3.5: Structure of the thread group of users and the number of requests performed for each stress test on one of the three access methods offered by SILKNOW

Test description

We organized the stress tests according to the different SILKNOW Knowledge Graph access methods:

- We define a thread group made up of different numbers of concurrent users: 5, 10, 30 and 50.
- Each user process launches a batch of requests, separated by a random timer.
- All users' processes associated with the thread group execute their requests concurrently.
- The whole process is repeated five times (once all users complete a batch).

Table 3.5 depicts the structure of the thread groups and the requests performed for the stress test executed. This structure was repeated for each access method.

The test configuration is the same as the one performed with the thesaurus stress test in order to clarify the process. Figure 3.21 shows a schema with the request execution process per request for three concurrent users.

Table 3.6 shows the timetable with the stress tests execution.

The hardware configuration which supports the three API services is described as follows:

- CPU: Intel Xeon L5640, 2.26 GHz, 12 cores (24 threads).
- RAM: 128 GB.
- Operating System: Linux Debian Buster, kernel 4.12.0.

The Knowledge Graph is hosted in a Virtuoso Docker replicated in a twin component. A load balancer between the two images is used to distribute the server load.

The client system where the tests were executed has the following characteristics:

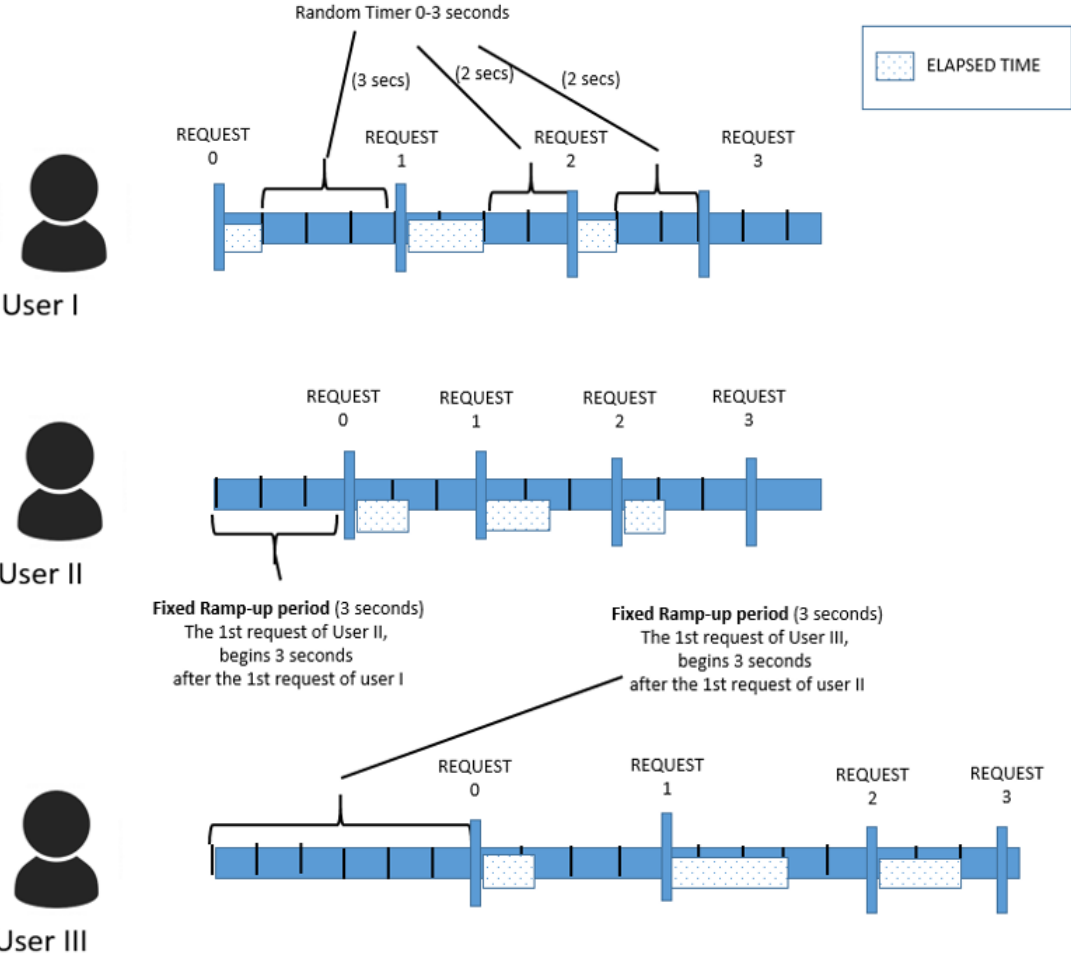


Figure 3.21: A request execution sample for users I, II and III in a thread group of 5 users

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

Time and date	Test launched
2021-06-25T04:00:00+0000	SILKNOW Public API (5 users)
2021-06-25T04:30:00+0000	ADASilk Internal API (5 users)
2021-06-25T05:00:00+0000	SPARQL API (5 users)
2021-06-25T06:00:00+0000	SILKNOW Public API (10 users)
2021-06-25T06:30:00+0000	ADASilk Internal API (10 users)
2021-06-25T07:00:00+0000	SPARQL API (10 users)
2021-06-24T10:00:00+0000	SILKNOW Public API (30 users)
2021-06-24T12:00:00+0000	ADASilk Internal API (30 users)
2021-06-24T14:00:00+0000	SPARQL API (30 users)
2021-06-24T16:00:00+0000	SILKNOW Public API (50 users)
2021-06-24T18:00:00+0000	ADASilk Internal API (50 users)
2021-06-24T20:00:00+0000	SPARQL API (50 users)

Table 3.6: Test stress execution timetable

- CPU: Intel i5-6400 CPU @ 2.70GHz.
- RAM: 8 GB.
- Operative System: Linux Fedora 7.0.

The tests were launched from the JMeter tool ³², using OpenJDK Java 1.8. We used this tool because of the API features and the test requirements, which makes Apache JMeter an adequate tool to perform such tests [2].

The tests are composed of a batch of requests. Each request has an associated query which is adapted to the request in order to be executed for the different evaluated APIs.

The queries are decomposed into two sets of four queries each. Inside a set, the queries have increased difficulty. The two different sets of queries, with different levels of difficulty, joined to the different and random timers per request execution, define a complex scenario which properly emulates a real situation.

The set of queries associated with the batch are:

Set 1:

- Production place: Italy,
- Text search: “damask”,
- Production time: eighteenth century (dates CE)
- Material: Metal thread

Set 2:

- Production Place: France
- Text search: “waistcoat”
- Technique: Velvet
- Material: silk thread

³²<https://jmeter.apache.org/>

Definition of the parameters measured

In order to evaluate the performance of the APIs, given the different requests made, a lot of data are gathered per JMeter tool, but the parameters analyzed are:

- Elapsed time: the time elapsed between the time a user is issuing a request and a response is received.
- Fails: the requests can fail for various reasons indicated by the different error codes returned by the server (401, 404, 5, etc.).

The elapsed time is a very important parameter to define user experience. Based on the operation, if the elapsed time is longer than what is usually observed in other similar applications, the user experience is impacted.

Fail is the most critical situation, because the user must repeat the process in order to get the required data.

Gathered data during the tests

In this paragraph we show the data gathered during the tests that will be then analyzed in "Results for ADASilk Internal API", "Results for SILKNOW Public API" and "Results for SPARQL API". Figure 3.22 gives a summary of the results of the tests, where the values presented are the average of the elapsed time per request and type of test, and the percentage of fails per request and type of test. Regarding the "fails percentage", we measure if the response expected per request finished without errors. This is summarized in Figure 3.22. Therefore, a result of "0" for this field means that all the response texts are received without errors.

In order to analyse the gathered data, we propose two graphics:

- A graphic with the average of the elapsed time per request and number of concurrent users.
- A graphic with the percentage of fails per request and number of concurrent users.

This methodology is slightly different from the one used in the thesaurus analysis. In these tests, the number of requests is twice as important as in the previous tests, and mixing up all the data in one chart would have made the graphic look overloaded.

	CONCURRENT USERS	DATA	REQUEST								AVERAGE	
			Italy	Italy + damask	Italy + damask + eighteen CE	Italy + damask + eighteen CE + metal thread	France	France + waistcoat	France + waistcoat + Velvet	France + waistcoat + Velvet + silk thread		
ADASilk Public	5	Average Elapsed time	1650	6089	30060	1609	1789	11304	30046	1512	10507	
		Fails percentage	0	0	100	0	0	0	100	0	25	
	10	Average Elapsed time	1711	5422	28852	3521	2354	5895	29459	2165	9922	
		Fails percentage	0	2	100	0	0	2	100	0	25.5	
	30	Average Elapsed time	2046	13159	29998	8425	1636	14604	29837	19547	14906	
		Fails percentage	20	28	100	66	30,6	17,3	100	70	54	
	50	Average Elapsed time	13003	13266	29705	12496	10901	9553	29709	12782	16427	
		Fails percentage	40	4.8	100	97.2	45.6	5.2	100	98.4	61.4	
	ADASilk Internal	5	Average Elapsed time	401	214	271	234	310	240	245	30163	4010
			Fails percentage	0	4	4	0	0	0	0	100	13.5
		10	Average Elapsed time	280	509	727	3393	163	237	1853	30068	4645
			Fails percentage	0	0.57	1.14	13.71	0	0	5.71	100	15.11
30		Average Elapsed time	463	315	297	307	313	267	277	28745	3873	
		Fails percentage	0	2.6	13	8	4	4	2.6	100	16.75	
50		Average Elapsed time	213	81	81	82	80	79	79	30007	3838	
		Fails percentage	0	0	0	0	0	0	0	100	12.5	
SPARQL		5	Average Elapsed time	184	55	93	8200	8377	136	174	208	2178
			Fails percentage	0	0	0	4	0	0	0	0	0.5
		10	Average Elapsed time	187	59	102	14457	10403	179	191	649	3278
			Fails percentage	0	2	0	50	34	0	0	0	10.75
	30	Average Elapsed time	177	51	95	12868	250	57	141	2777	2052	
		Fails percentage	0	0	0	44.6	0	0	0	0.6	5.65	
	50	Average Elapsed time	245	61	97	13040	272	57	149	4889	2351	
		Fails percentage	0	0	1.2	43.2	0	0.8	0	4.8	6.25	

Figure 3.22: Data gathered for the different access methods of the Knowledge Graph given the defined tests

Results for ADASilk Internal API

Figure 3.23 shows the average elapsed time per request on the tests with 5, 10, 30 and 50 concurrent users, using the ADASilk Internal API.

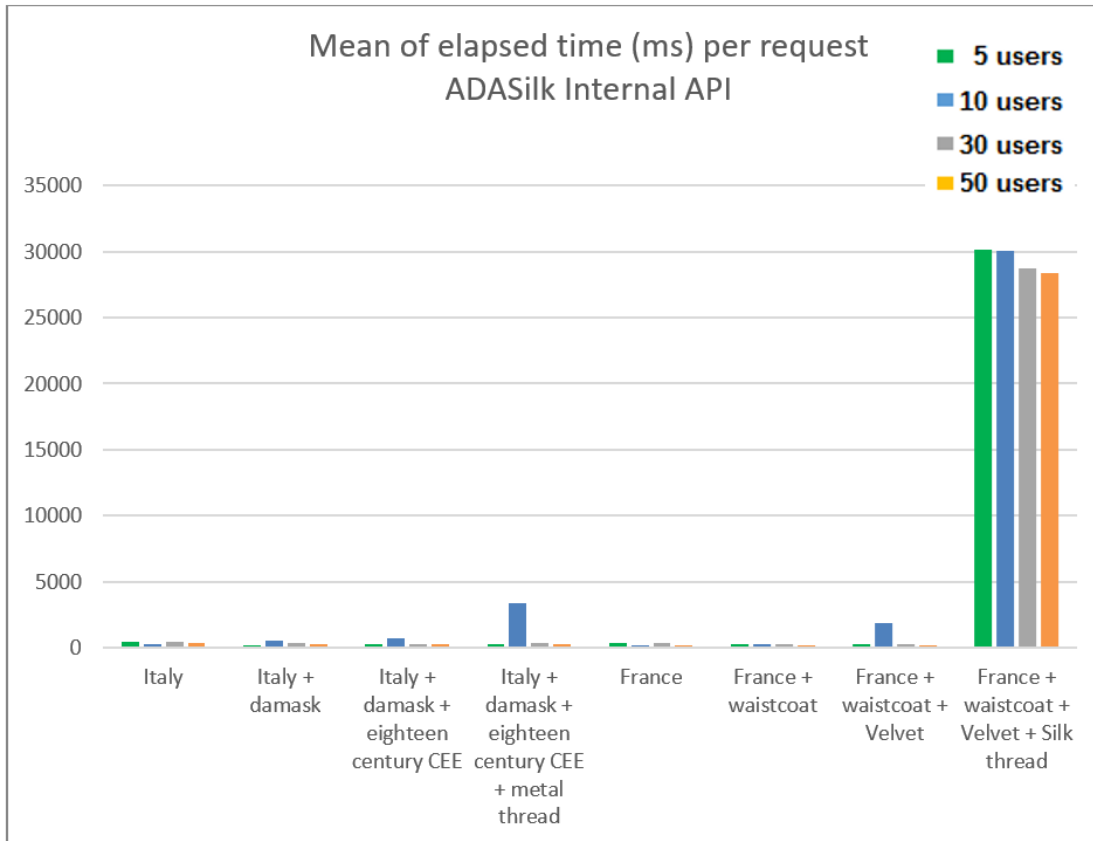


Figure 3.23: The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the internal ADASilk API

Figure 3.24 shows the percentage of fails per request on the tests with 5, 10, 30 and 50 concurrent users, using the ADASilk Internal API.

If the last request is not taken into consideration, the stress tests related to the requests performed on the ADASILK Internal API have ended up with very good results on the elapsed time and in the number of fails per request. We conclude that with the current server configuration this API can manage up to 50 concurrent users.

The main problem is that there is one request which seems to always fail: the last request of the second set (the 8th) has a similar complexity to the 4th request in the first set. We have yet to discover what causes this discrepancy in the results.

On the other hand, the number of concurrent users does not seem to cause a specific problem

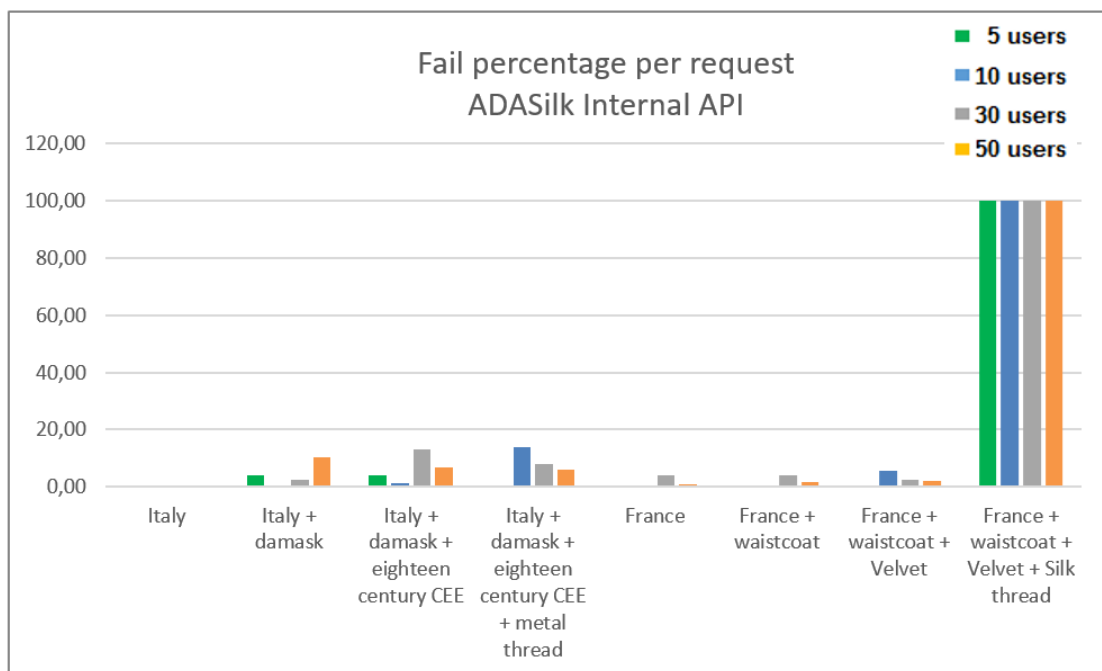


Figure 3.24: The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the internal ADASilk API

for any of the tests carried out.

Results for SILKNOW Public API

Figure 3.25 shows the average elapsed time per request on the tests with 5, 10, 30 and 50 concurrent users with the SILKNOW Public API.

Figure 46 shows the percentage of fails per request on the tests with 5, 10, 30 and 50 concurrent users with the SILKNOW public API.

The stress tests related to the requests performed on the SILKNOW Public API ended up with very good results on the elapsed time when the number of fails is not too large. The problem is that this number of fails is 100% in both the 3rd and the 7th request.

The 4th request has fewer fails than the 3rd one, but it is very high with tests executed with 30 concurrent users (greater than 60%) and the same situation occurs with tests executed with 50 concurrent users (almost 100%). This behaviour is abnormal since the query associated with the 4th request is more complex than the query associated with the 3rd request. We have yet to investigate why such a behaviour has been observed.

Given the results obtained, we recommend using this API with up to 10 concurrent users using

Chapter 3. Developing a Knowledge Graph about the production of silk artifacts

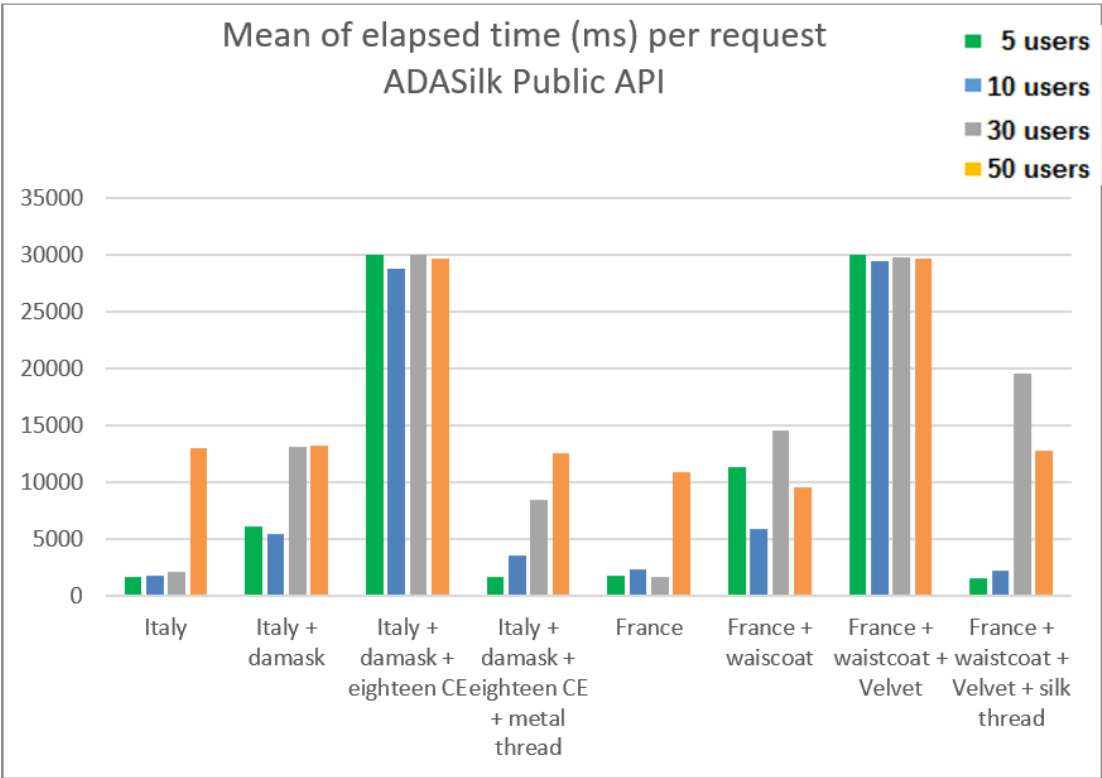


Figure 3.25: The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the public SILKNOW API

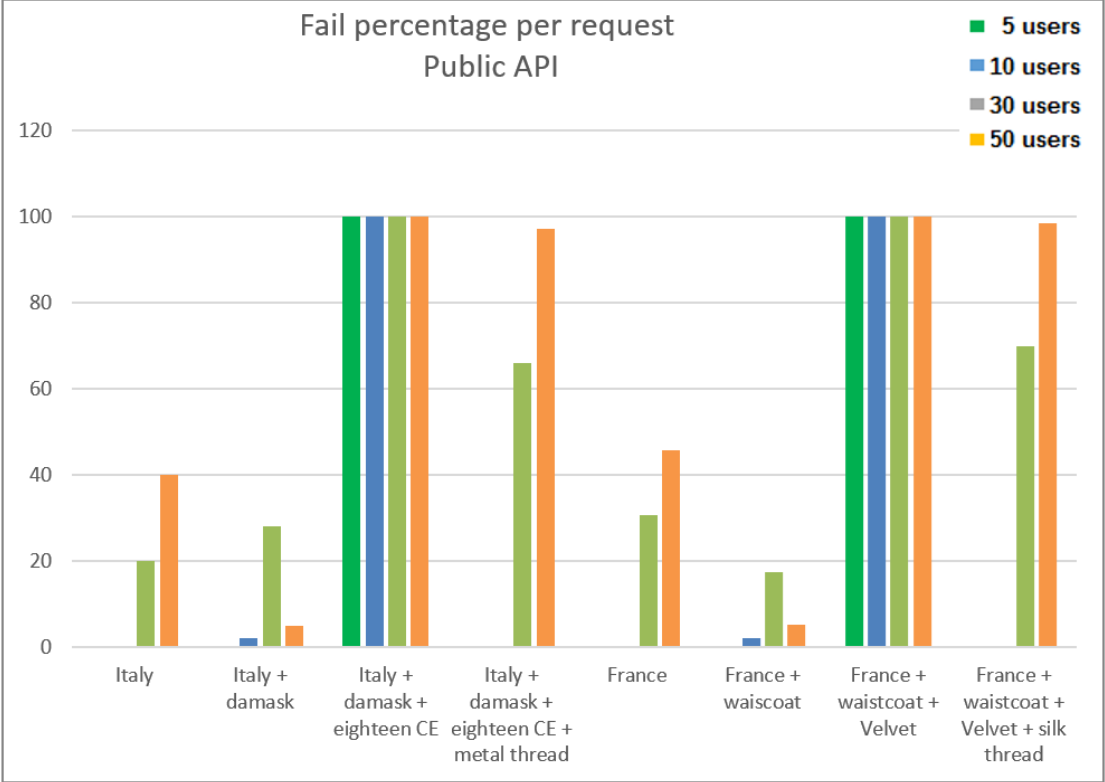


Figure 3.26: The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the SILKNOW Public API

this hardware configuration.

Results for SPARQL API

Figure 3.27 shows the average elapsed time per request on the tests executed with 5, 10, 30 and 50 concurrent users with the SPARQL API. Figure 3.28 shows the percentage of fails per request on the tests with 5, 10, 30 and 30 concurrent users with the SPARQL API.

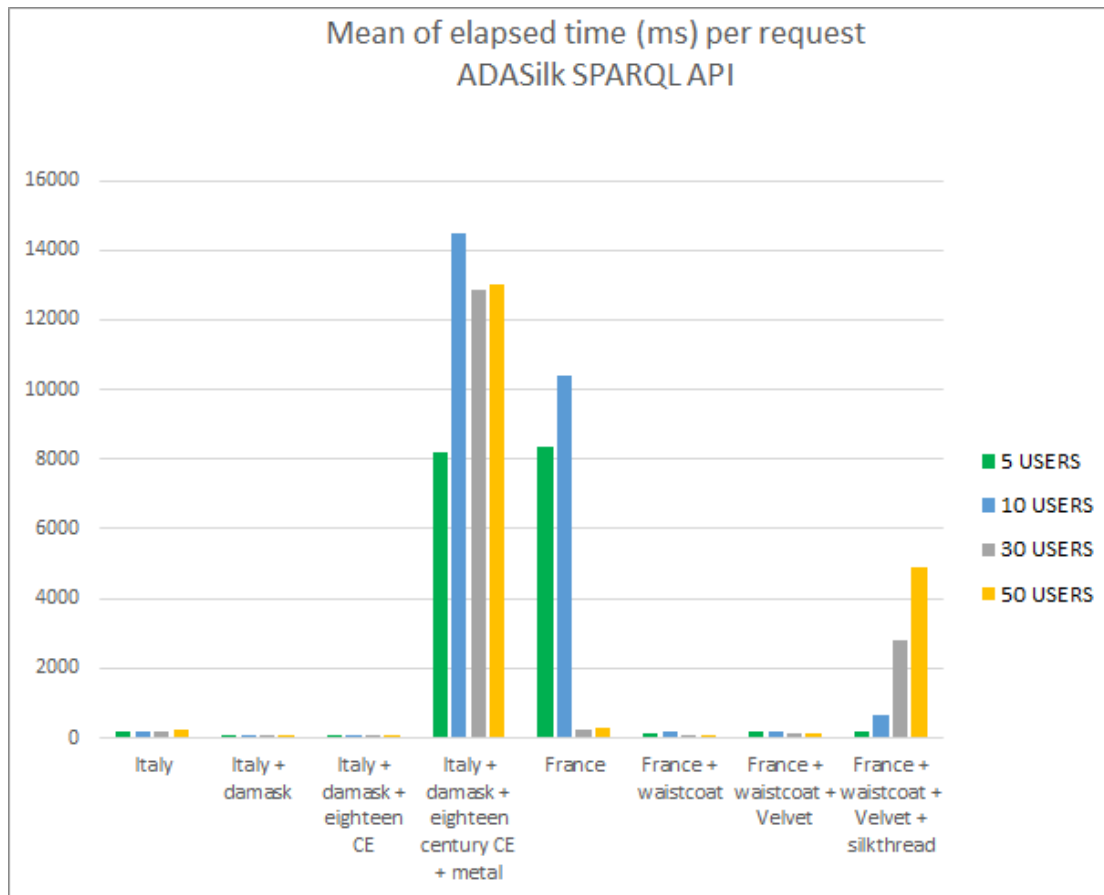


Figure 3.27: The mean of the elapsed time required per request in the tests performed with 5, 10, 30 and 50 concurrent users with the SPARQL API

The stress tests related to the requests performed on the SPARQL API ended up with very good results with regard to the elapsed time and the number of fails. The 4th request had a large number of fails in the tests executed with 10, 30 and 50 concurrent users (40%-50%) The 5th request also had a percentage of 35% of fails in the tests executed with 10 concurrent users. So, 4 requests had a significant number of fails, but always under 50% and only with tests executed with 10 or more concurrent users.

This API has the same problem with specific types of requests, but the problems are fewer

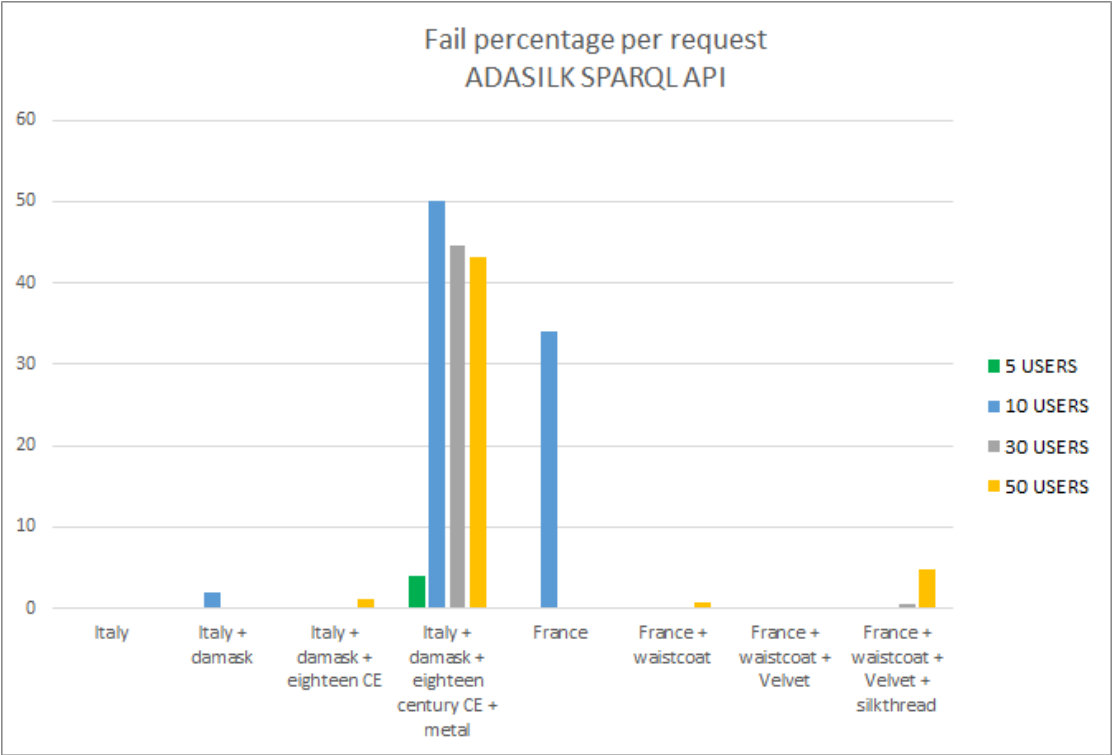


Figure 3.28: The percentage of fails per request in the tests performed with 5, 10, 30 and 50 concurrent users on the SPARQL API

than the ones with the SILKNOW Public API. In this case, the query never failed at a rate of 100%, and the number of requests affected is minimal.

Discussion

Given the hardware configuration, we conclude that the best performance is achieved using ADASilk. This is due, in part, to the smart caching we have implemented when developing this internal API. There is one specific query that fails, and it is necessary to find out why, in order to improve the system and to achieve an even better user experience. Considering the current hardware configuration, we conclude that the system can handle 50 concurrent users.

The performance is lower using the other two access methods, since they do not have the smart caching. We recommend that they are suitable for 10 concurrent users. On the one hand, the SPARQL API is slightly more efficient than the SILKNOW API. On the other hand, the SPARQL API is harder to use for a Web client as the response is not web developer friendly. The difference between the two is the overhead provided by the SPARQL Transformer component. While we generally expect that this overhead is minor in most circumstances. As to why this is so crucial to some queries should be the topic of future research.

3.5 Representing human knowledge based on multilingual museum records

Trying to "teach" a computer about something as specific as the production of silk artifacts is still not an automatic process. The Semantic Web provides many tools and methods to integrate and annotate heterogeneous Cultural Heritage metadata and images, which is why we rely on many tested steps from other projects to achieve our knowledge graph, our thesaurus and every tool necessary for it.

Many of these methods do, however, not work without adjustments and extensions, which we realized through a process that involved communication and discussions with domain experts and historians. The result at this stage is both an example of application and exploration. Our lessons are both documented and lead to the final implementations.

In this form it is not only an end in itself, but also a foundation for further research on the data and the knowledge itself. The more data we collected, the more it became evident, that many records had gaps which we could attempt to close with advanced NLP techniques.

In the next chapter, we will describe the use of information extraction and classification methods that we used to explore predicting the values of these gaps.

Chapter 4

Predicting metadata gaps

The mentioning of European silk textiles often evokes images of clothes and furniture of the old aristocracies and the lavish lifestyles of kings and queens. Nowadays, the knowledge about the occidental way of producing these expensive items is, however, more and more endangered.

Many museums and collections around the globe fortunately still have silk objects, or at least public records with metadata and images illustrating them. Such specific museum data, from many different sources about Cultural Heritage objects that are partly centuries old, have naturally some gaps: sometimes, the production year or place is unknown, but the material and technique used is described; sometimes, a rich textual description is provided with many little details about the object production and what it depicts, but categorical values informing about the exact material or technique used is not provided (Figure 4.1).

Figure 4.1 consists of three panels illustrating metadata gaps in museum records:

- a) Metropolitan Museum of Art:** Record 37.80.1. Object Details for "The Hunters Enter the Woods (from the Unicorn Tapestries)". Metadata includes Title, Date (1495-1505), Geography, and Culture. A missing subject depiction is noted.
- b) Musei di Venezia:** Record Cl. XXIV n. 1748. Record for "ultimo quarto - Diagonale". Metadata includes Author, Ambito Culturale, Denominazione, Oggetto, Descrizione, and Dati Tecnici. A missing material is noted.
- c) French Mobilier National:** Record GMMP-733-002. Record for "SOIERIE Bordures pour tenture et rideaux destinés à des salles de Versailles". Metadata includes Number of Inventory, Year of Conception, Type, and Materials. A missing technique is noted.

Figure 4.1: Examples from three different museums with missing categorical properties: a) no subject depiction for the record 37.80.1 from the Metropolitan Museum of Art; b) no material for the record Cl. XXIV n. 1748 from the Musei di Venezia; c) no technique for the record GMMP-733-002 from the French Mobilier National

However, the recent progress in natural language processing and more specifically in information extraction can help to address these problems. This chapter covers the following publications and (future) submissions:

- Luis Rei, Dunja Mladenić, Mareike Dorozynski, Franz Rottensteiner, Thomas Schleider, Raphaël Troncy, Jorge Sebastián and Mar Gaitán. **Multimodal Metadata Assignment for Cultural Heritage Artifacts**. In *Multimedia Systems (under review)*, 2022.
- Thomas Schleider and Raphaël Troncy. **Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles**. In *5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLFL) co-located with EMNLP, 2021*, Online.
- Ismail Harrando*, Alison Reboud*, Thomas Schleider*, Thibault Ehrhart and Raphael Troncy (*Equal contribution). **ProZe: Explainable and Prompt-guided Zero-Shot Text Classification**. In *IEEE Internet Computing: Special Issue on Knowledge-Infused Learning*, 2022. <https://doi.org/10.1109/MIC.2022.3187080>.

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

Some records have important information, like the production year or the weaving techniques, semantically annotated, others include it only in rich textual descriptions, and for some objects it is not available at all. These missing metadata can be considered as gaps that potentially could be filled in. Thanks to the progress in natural language processing, information extraction, and image processing, there are now techniques that can help to address such problems.

Digitization of culturally significant assets is a time-consuming process that requires experts and funding. This often forces a cultural institution to make a trade-off between the number of objects digitized and the effort per object. Less effort per object often implies a smaller number of details captured, less strict guidelines, and sometimes mistakes.

This section presents methods that enable further annotation of these museum objects through a multimodal classification approach that trains models to predict such missing metadata from images, text descriptions and other (available) metadata. The outcome is then further used to enrich an underlying knowledge graph. Domain experts can easily assess the quality of the automatically generated annotations through rich visualization and connections between the items.

Our first hypothesis is that we can predict, fairly accurately, a set of domain-relevant properties of cultural heritage objects (silk fabrics) from images and text descriptions. Our second hypothesis is that a multimodal approach involving both images, text descriptions and additional knowledge about other properties than those to be predicted will produce better results than any method relying on a single modality. In this context, the term "better" refers to both, the

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

quality of the results and the number of objects for which this information is inferred. That is, we expect the multimodal approach to result in more correct predictions and in predictions for a larger number of objects than the other methods. These hypotheses will be evaluated in the context of digitized metadata of silk fabric artifacts with data originating in multiple museums.

The main scientific contributions of this section are related to our research hypotheses. We introduce a multimodal machine learning approach, adapted to the cultural heritage domain, for predicting properties of digitized artifacts. We perform an in-depth analysis of the performance of our classification models, i.e. models based on individual modalities and the multimodal classifier. Additionally, we introduce a novel dataset to the cultural heritage and multimodal analysis domains that includes data for four different tasks and three different modalities. It consists of harmonized text and image data from heterogeneous, multilingual sources that went through different stages of preprocessing, cleaning and enrichments like domain expert-guided entity linking and grouping.

Finally, we show how our metadata predictions can be properly represented through classes and properties in our data model, which includes using information about a.o. their time stamp and or the used algorithm, and consequently integrated into existing Knowledge Graphs.

Challenges

The challenges faced in this work can be split broadly into those pertaining to the creation of the dataset and those related to the automated annotation. The latter ones can be further categorized according to the modality that is used for predicting the properties of the objects.

Data and Labels The data used in this work belongs to the cultural heritage domain. More specifically, it is related to silk textiles produced in Europe, primarily in the period between the 15th and the 19th centuries. In the domain of cultural heritage, we cannot expect all class labels to be equally likely or equally correlated. For example, in some locations, more silk fabric objects were produced than in others. Similarly, we know that the production of silk fabric objects in a given location likely started after a certain point in time and possibly subsided after a certain date. We also know that catalogs are curated by humans and often have strong thematic biases. For example, certain museums focus almost exclusively on objects created within one location.

The data we use in this work was aggregated from different sources. That is, it was crawled from 12 different museum or collection websites. Each museum may have different standards for how it collected the underlying objects and how it digitized the information related to

these objects. Importantly, this gives each museum its own standards for how to write text descriptions, how to create images, and how to annotate properties. Regarding these properties of digitized artifacts, accurately representing them requires adequate data modelling capabilities and considerable domain expert collaboration. This collaboration is also important in creating a dataset for machine learning. Labels need to be mapped from annotations made in different languages and grouped into domain relevant classes. Due to the partially automated nature required to create the dataset, challenges arise that are common in such processes: label text requires normalization such as correcting typos, unifying the styles of dates, and matching different locations to specific countries. Errors made in this process can often be systematic, for example, a failure to link a specific value of a property due to the form of writing it particular to that catalog will likely result in that value not being present in all records originating in that catalog.

Image Classification. In the context of this section, the classification of images aims to predict abstract properties of the silk fabrics depicted in the images. Whereas it may be relatively straightforward to learn to classify the material of a depicted piece of fabric, the prediction of semantic information such as the production place of the fabric, the period of time in which the fabric was produced or the technique used to manufacture the fabric is assumed to be much more challenging. Furthermore, it is assumed that there are interdependencies between these properties of silk fabrics, e.g. a certain production technique may only have been used in a certain period of time.

This is why multi-task learning is investigated for image classification. However, standard multi-task classification frameworks require one reference label for every task to be learned during training for every training sample. The challenge we have to face is that in real world data, as they were collected for the dataset presented in this section, there may be many training samples for which annotations are unavailable for some of the target variables to be predicted. Accordingly, this fact must be taken into account in the training of a multi-task classifier. Additionally, the available number of class labels constituting the class distribution of a variable is often imbalanced for real-world datasets. This constitutes a further challenge to supervised learning, which is addressed by utilizing a suitable training strategy for the image classification method.

Text Classification Supervised approaches are often challenging to perform with data from the cultural heritage domain for several reasons. Text descriptions are not present for the majority of objects in an archive. Many of the text descriptions that are available, in most museums, tend to be short sentences, almost title-like. In specific domains, such as the cultural heritage of silk production, many of the terms used in the text are very domain specific. Each museum has their own standard of how and what to write in a text description:

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

some may focus on the history of the objects and write very grammatical paragraph-length descriptions meant to be read by the public, others may focus on the properties of the object and write a single enumerating sentence, and others still, may focus solely on the depictions or visual patterns of an object. Finally, museums are spread geographically, and thus we can expect to deal with multiple languages, making our problem multilingual and cross-lingual. To summarize, we end up with a small collection of domain specific texts, written in different languages, with different content both semantically and syntactically, and wildly varying lengths. These texts are then associated with labels, based on the provided properties of the object. As already discussed, these labels are not all equally likely or correlated, and many of these accidental regularities are likely to interact with the language and the particularities of the text style of the museum.

Multimodal Classification One of the challenges in this work is that we want to integrate predictions made from images and text. Most work done in the literature, is exclusive to depictions or type of object: the image shows a scene or object and the text describes it. In our case, there may be no scene depicted in an object, and we do not consider describing the object beyond certain properties. For example, if we have a fabric that shows a certain pattern, describing the visual shapes of the pattern (e.g. triangles) is not a goal. Rather, we need to deduce, from the image, properties of how, when, where, and with what the object was made. Similarly, with text descriptions, there may be a good amount of words that describe visual patterns, scenes depicted, and historical facts associated with the object, but the goal is, again, to determine those same intrinsic properties of the object's making. Another challenge that is uncommon is the reduced and variable overlap between images and text descriptions. Not only is our work subject to a comparatively small dataset, restricted by historical reality and difficulties of data collection, but we must also deal with the fact that for most archives of culturally relevant objects, many objects that have been photographed have no corresponding textual description. In fact, we'll see that less than half of all objects have both these modalities. Another challenge, uncommon outside of retrieval scenarios, is that we can have multiple different images, with different angles and focus, per each individual object while it makes no sense to talk about multiple text descriptions per object. Yet another challenge we need to deal with, common to many real world applications but not to research datasets, is that we do not have all properties for all objects. For example, for a given object, we might know what material and techniques were used but not when or where it was made. Finally, our dataset, although drawn from several museums, contains under 30k objects and approximately 11k text descriptions. Effectively making it small compared to general datasets, but not uncommonly so for a dataset in the cultural heritage domain.

4.1.1 Datasets

Knowledge Graph

The SILKNOW Knowledge Graph (KG) [110]¹ lies at the center of all efforts to create a unified representation of the metadata of European silk textiles, particularly from the 15th to the 19th century. All the data used in our experiments was downloaded from 18 sources, most of them are public online museum records, for which we built crawling and harvesting software. In addition to that, we have data from the SILKNOW² project partners Garin and the University of Palermo (Sicily Cultural Heritage). The dataset used in the experiments was created from a full export of all objects in the knowledge graph, which consists of the metadata of 40,873 unique silk objects before any preprocessing steps. This export includes in total 74,527 unique image files.

In order to model this heterogeneous data from so many sources, we chose and relied strongly on the CIDOC Conceptual Reference Model (CRM). We also developed our own SILKNOW ontology³ to extend CIDOC-CRM with further classes and properties for cases where it did not cover some specifics of the silk textile domain and also for some extra information. For example, the confidence score for metadata predictions, once we started integrating the results of those predictions back to the KG.

In order to develop a converter⁴ that could unify all the original data with all these classes and properties into one knowledge base, mappings have been created by domain experts. And on a technical level, all museum records had to be harvested and were first converted into a common JSON file format through our crawler software⁵ but each array inside this format still had the original field labels from the museums before the final conversion. For example: the majority of museums have a field for describing the production time of a silk object, but in most cases museums use different names for their field. Moreover, the museums are from all over the world and we are facing different languages for both the field names and their values. This is why we created a mapping for, e.g, a field named "Date" (Metropolitan Museum of Arts) and the class `E12_Production` with the property `P4_has_time-span` and another class `E52_Time-Span`. Likewise, a mapping rule will be written for the field named "date_text" (API of the Victoria and Albert Museum) and for the (Spanish) field named "Datación" (Red Digital de Colecciones de Museos de España).

Another very central part of our knowledge representation is the SILKNOW Thesaurus⁶, a

¹<https://zenodo.org/record/5743090>

²[urlhttps://silknow.eu/](https://silknow.eu/)

³<https://ontome.net/namespace/36>

⁴<https://github.com/silknow/converter/>

⁵<https://github.com/silknow/crawler/>

⁶<https://skosmos.silknow.org/thesaurus/>

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

controlled vocabulary which contains many explicit and multilingual concept definitions for materials, techniques and motif depictions relevant for these silk textiles. Thanks to this thesaurus, a lot of information and entities from very explicit categorical fields of the original museum records could be linked, without any advanced machine learning techniques - the string literal could just be matched with the (multilingual) labels of the thesaurus and then replaced with a unique concept link. This explicit representation of knowledge forms the core of the dataset used to predict missing metadata. This includes cases where a categorical value is either not given at all or “hidden” in longer textual descriptions and not explicitly semantically annotated.

Once all the modelling, download, conversion and enrichment steps were taken, the final knowledge graph was uploaded onto a SPARQL endpoint from where all the data across languages and museums can be queried the same way. To make access easier, we also developed a RESTful API, so it is not necessary for web developers to write SPARQL queries, and an aforementioned exploratory search engine on top of this API, called ADASilk. It is aimed at users with only little technical background or little background knowledge about the domain of silk, to make them able to discover a lot of the data in the KG. ADASilk offers an advanced search with many filters, some topic suggestions, and in general a clean visual interface that shows all objects with their images and metadata.

Extracting and Normalizing Labels

The development of the SILKNOW Knowledge Graph is a combined effort of data processing that relies on a data modelling and annotation process created in collaboration with domain experts. This is especially true for the SILKNOW Thesaurus. The group labels used in the experiments in this section are based on the hierarchy and relations of concepts of the silk textile domain described in this controlled vocabulary. As described in 4.1.1, a big part of categorical property values could be easily extracted, linked and through the string replacement indirectly automatically normalized thanks to the SILKNOW Thesaurus. This means that many concepts are accessible even though there were originally different strings, including typos in some cases, synonyms, or translations. An example would be a weaving technique like "Damask", which would be "Damas" in French and "Damasco" in Spanish and Italian: for all these, we replace the string literal with one link to the same concept. In addition to the SILKNOW Thesaurus, we also use linked open data like such as GeoNames⁷ to normalize and link place names.

Matching strings with such thesaurus or other controlled vocabularies was not without challenges. As will be also explained in more detail in section 4.1.4, misspellings or unique punctuation could still cause the matching process not to work properly. To give an example:

⁷<https://www.geonames.org/>

Chapter 4. Predicting metadata gaps

If the string value of a record was "silk; gold thread" the latter would not have been linked, due to a bug that did not properly consider a semicolon as a separator. Other such cases existed as well, as the development of the SILKNOW Knowledge Graph is an ongoing process and concurrent with this work. See figure 4.2 for an illustration of a museum record in the knowledge graph.

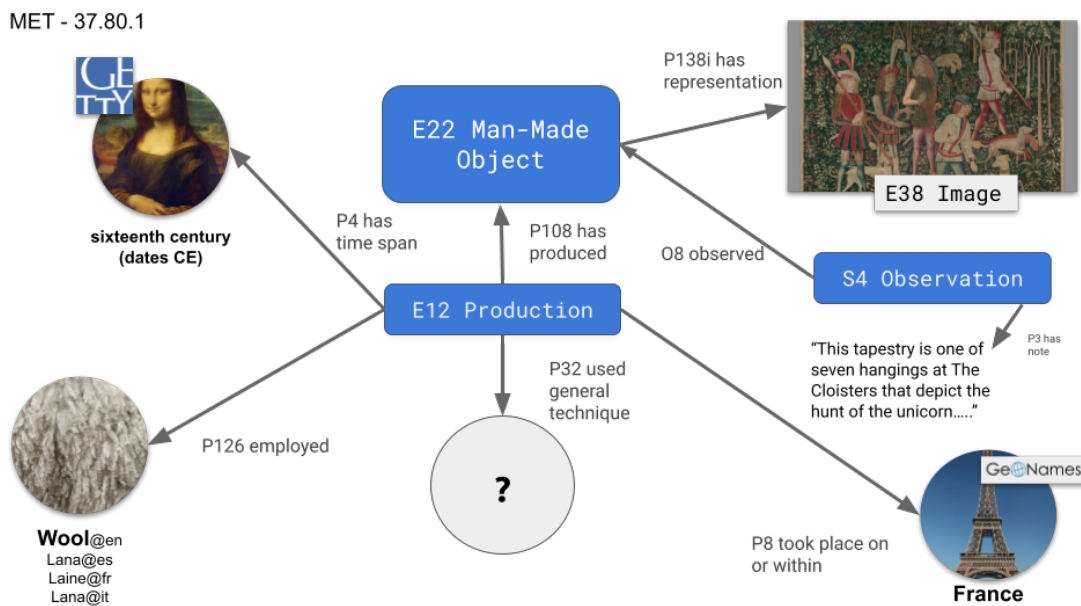


Figure 4.2: A record from the MET museum with a missing property represented in the knowledge graph using our ontology and controlled vocabularies

The aforementioned hierarchy defined in the SILKNOW Thesaurus can be used to select specific types or subtypes of properties. To refer back to the previous example, we could select only objects with the weaving technique "Damask", but also only objects made with "Two-coloured damask" which is even more specific. Based on the Thesaurus, we can also make sure that we only choose objects based on equivalent levels of this hierarchy.

Based on these enrichments and the linking process, we created a pipeline to extract the dataset based on pre-specified criteria. We first developed a comprehensive SPARQL query⁸ that outputs all museum objects described in the Knowledge Graph (KG) and includes, if available, the most relevant properties: the identifier of the object in the knowledge graph, the museum where the description comes from, the text description, and URL links to the images that illustrate the object. The results of this query were exported as a CSV file, which we then post-processed⁹ to make sure that we have a format of one row per object. In this final format, the CSV is used as the basis for all experiments.

⁸<https://github.com/silknow/converter/blob/master/jointtextimagemodule/total.sparql>

⁹<https://github.com/silknow/converter/blob/master/jointtextimagemodule/jointtextimagepost.py>

Label Grouping

In principle, the Knowledge Graph contents can be used to generate training and test samples for the classifiers described in section 4.1.2. One would just have to associate the images and/or the text given for a record with the annotations in the categorical variables of interest. The available annotations can be easily converted into class labels. However, a statistical analysis of these annotations revealed that most of them occur very rarely in the data, while for all categorical variables there were one or a few classes which were dominant in the sense that many records belonged to them. Supervised classifiers have problems with imbalanced training data sets, and it would seem very difficult for a classifier to successfully differentiate classes for which it has seen only a very small number of training samples, if on the other hand there are thousands of samples for some other classes. To still be able to extract meaningful information from the available modalities using supervised methods while at the same time having the chance of achieving a reasonably good classification performance, a simplified class structure was defined. Domain experts analyzed the class distributions and aggregated classes corresponding to different categories into compound classes. Care was taken for the aggregated classes to be consistent with the Thesaurus, and aggregated only if they were considered to be related according to the domain experts. At the same time, the aggregation was guided by the frequency of occurrence of class labels so that the compound classes would occur frequently enough to be used for training the supervised classifiers described in section 4.1.2.

The resultant simplified class structure was integrated into the Knowledge Graph in the form of so-called *group* fields, which were made available for all semantic properties of interest, principally, the ones corresponding to the different tasks in this work. Such grouping was applied to the following properties: Material, Technique, production place (with a country granularity), production time (with the century granularity) and the object type or object domain group, to be able to filter out non-textiles that use silk. Grouping was not an easy task, domain experts had to deal with more than 200 concepts that had to be grouped according to the aforementioned categories. Techniques were the most complex to group. To do so, domain experts grouped the concepts according to two fundamental criteria: 1) whether they belonged to the same hierarchy, for example, velvet and its types. In fact, there are many types of velvet, classified depending on the nature of the pile such as broderie velvet, ciselè velvet, cut velvet, pile-on-pile velvet, uncut velvet, etc. 2) If they were somehow related to a certain technique, for example, the effects obtained of applying differently warp and weft, that is, whenever a yarn is introduced into a fabric to produce an effect or pattern. On the other hand, materials were not complex as they were made in large groups according to their origin, that means according to the product obtained from the processing of one or more raw materials, in the course of which their structure has been chemically modified, e.g. animal fibres are distinguished from vegetable fibres.

Using a conversion table for aggregation prepared by the domain experts, the contents of the *group* fields could be derived automatically from the original semantic annotations. Having thus expanded the Knowledge Graph, training, and test samples could be easily generated from it by appropriate SPARQL queries that would export the contents of the *group* fields associated with each record.

Dataset Preparation and Properties

The goal of the dataset preparation is the conversion of the knowledge graph data with normalized and grouped labels described, respectively, in Sections 4.1.1, 4.1.1, and 4.1.1, into a dataset for the experiments in 4.1.3 using the classification methods described in Section 4.1.2.

The first step was to select the records in the knowledge graph that were relevant to the domain.

The second step was to select only records that contained a value for one of the variables to be predicted, i.e., labeled samples. Uncommon labels, with a total frequency below 150, were discarded. The final step was to randomly split the records into disjoint sets:

- a training set consisting of 60% of the data for supervised learning;
- a validation (or development) set, consisting of 20% of the data for hyperparameter tuning and multimodal supervised learning;
- a test set, consisting of 20% of the data, for evaluation of the proposed method.

Given that the objective is to train and evaluate a multimodal multitask approach on records, that regularities exist within each collection (i.e., museum) that comprises the data, that the text modality is also multilingual, and that both modalities and task specific labels may be missing from a record, we believe the most reasonable way to split the dataset is a random split of records. The distribution of the data per each set and class label can be seen in Table 4.1.

The distribution of samples over the museums can be found in table 4.2 and an overview of the modalities can be found in Table 4.3. We can see how 27,120 or 96,60% of the 28,077 records about annotated fabric objects contain at least one image, but only 11,034 or 39,29% of them contains a text description. The overlap consists of 10,664 or 37,98%. The proportion between training validation and test sets in each case corresponds roughly to the aforementioned 60-20-20 split.

Text data in our dataset consists of descriptions of fabrics or objects made mostly of fabrics. These descriptions range in length from a short sentences to multi-sentence paragraphs to

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

Table 4.1: Class structure and class distribution of the records.

Variable name	Class name	total	training	validation	test
<i>timespan</i>	19 th century	5,849	3,492	1,180	1,177
	18 th century	4,397	2,576	901	920
	20 th century	2,483	1,520	483	480
	17 th century	1,134	689	231	214
	16 th century	880	542	180	158
<i>place</i>	FR	5,265	3,156	1,037	1,072
	IT	3,205	1,853	687	665
	GB	2,837	1,721	562	554
	ES	2,630	1,605	521	504
	IN	1,190	735	231	224
	CN	699	426	127	146
	IR	671	409	142	120
	JP	533	325	92	116
	TR	331	205	57	69
<i>technique</i>	embroidery	3,123	1,814	657	652
	velvet	2,193	1,273	454	466
	damask	1,685	1,004	333	348
	other technique	1,150	722	219	209
<i>material</i>	animal fibre	17,382	10,387	3,445	3,550
	vegetal fibre	2,051	1,255	396	400
	metal thread	2,046	1,223	422	401

multi-paragraph texts with thousands of words. Some descriptions focus primarily on a single aspect, such as a scene depicted or the history of the object, while others focus on various properties of the object. Table 4.4 shows some examples of these descriptions. In order to eliminate some errors present in the data, we removed any text descriptions smaller than 60 characters. The resulting distribution of lengths is summarized in table 4.5. These descriptions are in 4 different languages: English, Spanish, French, and Catalan. The counts for each are shown in table 4.6

4.1.2 Methods

Image Classification

The goal of the image classification is to predict one class label per classification task, i.e., the prediction of a class label for each of the target variables *technique*, *timespan*, *material* and *place*, for an image that illustrates an object. For that purpose, an image classifier is trained using all images of all records contributing to the dataset described in Section 4.1.1. We propose

Chapter 4. Predicting metadata gaps

Table 4.2: Names of the museums contributing to the dataset with their identifiers (ID) used in this section, and distribution of the 28,077 records over the museums for the training (train.), validation (val.) and test sets.

Museum name	ID	total	train.	val.	test
Metropolitan Museum of Arts	met	6,524	3,835	1,325	1,364
CDMT Terrassa	imatex	6,119	3,690	1,204	1,225
Victoria and Albert Museum	vam	5,527	3,300	1,133	1,094
Rhode Island School of Design	risd	3,226	1,913	634	679
Boston Museum of Fine Arts	mfa	2,610	1,579	517	514
Garín 1820	garin	1,558	972	300	286
Collection du Mobilier National	mobilier	1,293	796	267	241
Red Digital de Colecciones de Museos de España	cer	781	490	142	149
Joconde Database of French Museum Collections	joconde	375	224	78	73
Smithsonian Museum	smithsonian	38	29	14	14
Versailles	versailles	18	8	5	5
Art Institute of Chicago	artic	8	4	2	2

to use a convolutional neural networks (CNN) for that purpose, motivated by the success of CNN in image classification. As there are many records with annotations for more than one of these variables, we propose to train the classifier to predict all classes simultaneously in a multitask framework, exploiting the inherent relations between the variables to learn a joint representation that is used by task-specific classification heads. A detailed description of the chosen network architectures can be found in Section 4.1.2, whereas the strategies used for training are presented in Section 4.1.2.

Network Architecture Figure 4.3 shows the structure of the CNN for multitask learning for the prediction of the four target variables. Its input consists of an RGB image scaled to a size of 224×224 pixels. This image is presented to the ResNet 152 network of [58] pre-trained on ImageNet [34], which serves as a generic feature extractor for the image [117] and produces a feature vector of 2048 dimensions. We apply dropout with a probability of 10% after this layer [122].

This is followed by $L_{fc} = 2$ fully connected layers, the first one having 1024 and the second one having 128 nodes, which are shared by all tasks. Rectified linear units [91] are used as nonlinearities in both of these joint layers. They produce a joint representation of the image of $N_r = 128$ dimensions. This representation is processed by four task-specific classification branches, each consisting of one additional softmax layer only, which delivers the class scores $y_{km}(\mathbf{x}, \mathbf{w})$ for the input image \mathbf{x} to belong to class k for variable m . The number of nodes of

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

Table 4.3: Modality statistics of all records in the dataset that provide a class label for at least one of the variables. The values are given for the training (train.), validation (val.) and test sets as well as for the total dataset.

dataset	total	with image	with text	with image and text	without images and text
train.	16,840	16,260	6,717	6,495	358
val.	5,602	5,419	2,184	2,101	100
test	5,635	5,441	2,133	2,068	129
total	28,077	27,120	11,034	10,664	587
	100.0%	96.6%	39.3%	38.0%	2.1%

Table 4.4: Examples of text descriptions present in our dataset.

Text Description
White and silver striped fabric with supplementary weft of flat silver strips whose floats form vertical stripes with leaves at intervals. White floats of the weft form outlines for serpentine floral sprays spread over the striped areas.
Furnishing fabric, woven, British, c. 1895, Alexander Morton & Co., red/brown plain silk weave
Dibujo Palma en color azul grisáceo Urdimbre: Trama: 36 pasadas Rapport: 65 cm ancho y 104 cm alto (incompleto)

the softmax layer corresponds to the number of classes to be differentiated for a specific task. The CNN architecture is shown in Figure 4.3.

The CNN predicts one class label per task for every image. In case of multiple images per record, one such class label is predicted for each one of the images and the prediction with the highest softmax score is chosen to be the prediction for the record.

Training In training, the parameters \mathbf{w} of the CNN described in Section 4.1.2 are learned by minimizing a loss function $E(\mathbf{w})$. The parameters of our network consist of the parameters \mathbf{w}_R of ResNet-152, which are initialized from a pre-trained model published by he2016, and the parameters \mathbf{w}_{FC} of the fully connected and softmax layers, which are initialized randomly by a variant of the Xavier initialization also described in [58]. In the training procedure, we determine the parameters \mathbf{w}_{Rt} of the last NL_{RT} layers of ResNet-152 considering exclusively entire residual blocks and the parameters \mathbf{w}_{FC} of the fully connected layers, whereas the parameters \mathbf{w}_{Rf} of the first $152 - NL_{RT}$ ResNet-152 layers are frozen [136]. Thus, the parameter vector consists of three subsets: $\mathbf{w} = \left(\mathbf{w}_{Rf}^T, \mathbf{w}_{Rt}^T, \mathbf{w}_{FC}^T \right)^T$. NL_{RT} is a hyperparameter to be tuned.

Two loss functions can be used for training the network. The first one, originally proposed in [44], is an extension of the standard softmax cross-entropy loss with weight decay [17]:

Table 4.5: Text length in characters and space delimited tokens.

	Min	Q1	Median	Mean	Q3	95th percentile	Max
Characters	60	173	343	693	856	2367	16333
Tokens	7	28	56	115	142	392	2826

Table 4.6: Language distribution of text descriptions based on language of the museum.

	English	Spanish	French	Catalan
Records	7271	1975	1126	680

$$E_{SCE}(\mathbf{w}) = - \sum_{n=1}^N \left(\sum_{m \in M_n} \sum_{k=1}^{K_m} t_{nmk} \cdot \ln(y_{km}(\mathbf{x}_n, \mathbf{w})) \right) + \omega_R \cdot R(\mathbf{w}_{RL}, \mathbf{w}_{FC}) \quad (4.1)$$

In eq. 4.1, $y_{km}(\mathbf{x}_n, \mathbf{w})$ is the softmax score for the n^{th} training image \mathbf{x}_n to belong to class k for variable m . The indicator variable t_{nmk} is one if the class label of sample n for variable m is k and zero otherwise. The sum is taken over all N training samples and K_m classes for task

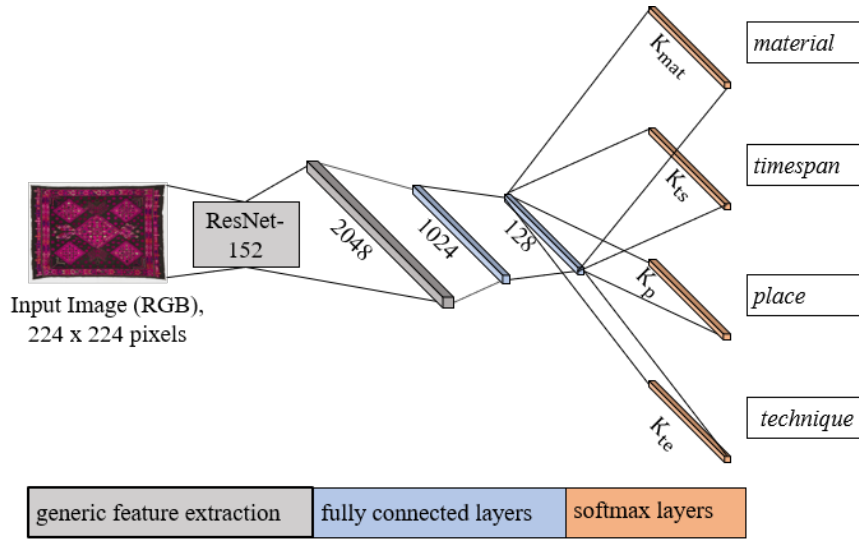


Figure 4.3: Network architecture of the CNN for multitask image classification. The input image scaled to 224 x 224 pixels is presented to a pre-trained ResNet-152 (grey) to extract generic features. The resulting 2048-dimensional feature vector is mapped to a domain-specific joint representation of 128 dimensions by two fully connected layers (blue). The task-specific classification branches consist of one softmax layer each (orange) that delivers the class scores for the corresponding variable. K_{mat} , K_{ts} , K_p , and K_{te} denote the number of class labels for the tasks *material*, *timespan*, *place*, and *technique*, respectively.

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

m . M_n is the set of tasks for which the true class label is known for the training sample n , so that the loss in eq. 4.1 considers exclusively samples x_n with $t_{nmk} = 1$ for learning task m . In this way, the fact that the annotations for most samples are incomplete, i.e. that annotations are only available for a subset of the variables to be predicted, can be considered. If multiple annotations are available, the corresponding classification losses will be back-propagated to the joint layers from multiple classification branches, thus supporting the learning of a joint representation for all variables. The outputs for variables for which the true class label is unknown will not contribute to the loss and to the parameter update. Finally, the term $R(\mathbf{w}_{RL}, \mathbf{w}_{FC})$ corresponds to regularization by weight decay, which is only applied to the parameters to be updated in training; ω_R is a hyperparameter defining the influence of this term on the result.

One problem of the data described in section 4.1.1 is its imbalanced class distribution. In this case, minimizing the cross-entropy loss in eq. 4.1 will favor the dominant classes, resulting in a poor performance for the underrepresented ones. In order to mitigate these problems, a multi-class extension of the focal loss [80, 84] with regularization is utilized for training:

$$E_F(\mathbf{w}) = - \sum_{m \in M_n} \left(\sum_{n=1}^N \sum_{k=1}^{K_m} (1 - y_{km}(x_n, \mathbf{w}))^\gamma \cdot t_{nmk} \cdot \ln(y_{km}(x_n, \mathbf{w})) \right) + \omega_R \cdot R(\mathbf{w}_{RL}, \mathbf{w}_{FC}) \quad (4.2)$$

The only difference between the loss functions in eqs. 4.1 and 4.2 is the penalty term $(1 - y_{km}(x_n, \mathbf{w}))^\gamma$, where γ is a hyperparameter modulating the influence of this term on the result. This penalty term forces the loss to put more emphasis on samples that are difficult to classify (having a small score y_{km} for the correct class). Assuming the samples of underrepresented classes to be hard to classify by the CNN, this loss is expected to improve the results for these classes.

Starting from initial values derived in the way described earlier, stochastic minibatch gradient descent based on the ADAM optimizer [67] is applied to determine the CNN parameters, using the default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) and a minibatch size of 300. The base learning rate η is another hyperparameter to be tuned. We use early stopping and use the model parameters leading to the lowest loss on the validation set.

Text Classification

Our problem is defined as value prediction for certain properties of an object, a silk fabric, given its text description, which can be written in any one of the four languages listed in

Table 4.6. We have 4 tasks, each denominated according to the property of the underlying fabric object we want to predict: the *technique* and *material* used to create it, the *timespan* or time period when it created, and the *place* where it was created. While some descriptions directly contain some of this information, as seen in Table 4.4, this is sufficiently uncommon to prevent a purely extractive approach from yielding good results. For example, of the 3 texts we showed, only one gives any indication as to where it was produced ("*British*"). We instead rely on regularities present in the text descriptions to make informed guesses. More technically, we frame our problem as a multiclass, multitask, multilingual text classification problem. That is, given a text description of a fabric, written in any language, we want to assign exactly one label out of a set of mutually exclusive class labels for each of the properties we wish to predict, i.e., the tasks.

The text classifier uses a hard parameter sharing based multitask architecture [107], shown in Figure 4.4. It consists of a shared encoder followed by task-specific classification heads. The encoder is the multilingual large pretrained transformer, XML-R [31]. The output embedding corresponding to the CLS token input is used to represent the encoded text and is the only transformer output forwarded to the classification heads. All classification heads are identical except for the output dimension of the last layer, the output projection layer, which depends on the number of classes of the task. A diagram of a classification head is shown in Figure 4.5. A softmax function can convert the output logits of the last layer to normalized probabilities.

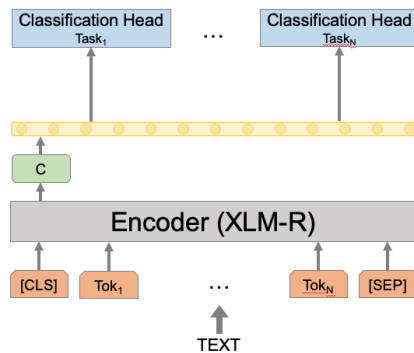


Figure 4.4: Multitask architecture: a shared XLM-R based encoder followed by task specific classification heads. The input to each classification head is the output of the transformer "C" corresponding to the input token "[CLS]".

To fine-tune our transformer-based classifier, at each step, a task is randomly selected using proportional sampling. A batch of examples for this task is then created and fed to the classifier. The cross entropy loss is then calculated and weights adjusted through backpropagation. Adam [68] is used as the optimizer with weight decay [86].

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

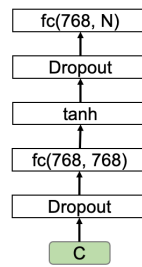


Figure 4.5: Task specific classification head: a fully connected (FC) layer followed by a tanh activation, followed by the output projection FC layer. Dropout is applied before both FC layers.

Tabular Classification

We use four separate task-specific classifiers to perform tabular classification. These all use the same learning algorithm, Gradient Boosted Decision Trees (GBDT) [47], implemented in XGBoost [24]. The input to the tabular classifier consists of the categorical values of non-target variables plus the identifier for the museum, as shown in Table 4.7. We replace missing values for a feature with a predefined value, represented by the symbol "[NA]" ("Not Available") in the table.

Table 4.7: Tabular Classification, one example input row per task. Note: time label format changed to roman numbers for ease of readability.

Target Variable	Target Value	museum	place	timespan	technique	material
place	FR	risd	-	[NA]	[NA]	animal fibre
timespan	XVIII	met	[NA]	-	embroidery	animal fibre
technique	other technique	garin	ES	XX	-	vegetal fibre
material	vegetable fibre	vam	GB	XIX	embroidery	-

Hyperparameters

While a detailed explanation of each hyperparameter that control the resulting model and learning of GBTs is beyond the scope of this work, we believe some contextualization is required. This is due to the relatively larger number of hyperparameters tuned for GBTs in Section 4.1.3 compared to the Neural Network based methods used for the other modalities, and for the convenience of the reader.

The hyperparameters *max_depth* (maximum depth of a tree), *min_child_weight* (minimum weight for tree partitioning), and *gamma* (minimum loss reduction for tree partition) all directly control model complexity, which in turn can have significant consequences in terms of fitting. The hyperparameters *subsample* (the percentage of data sampled per iteration) and *colsample_bytree* (the ratio of features sampled per iteration) can reduce overfitting by adding random noise to the iterative tree building process. Finally, the *learning_rate* and *number of rounds* control, respectively, the amount of learning per round and the total amount of learning (i.e., the total number of trees).

Multimodal Classification

Our approach to multimodal classification, shown in Figure 4.6, follows a decision level late fusion approach, in which the decision (prediction) from each of the 3 modalities serves as the input to a classifier that takes the final decision on which label to assign to the record. We opted to use the raw the tabular data as an additional input for this classifier on the hypothesis that the additional information would help it to take better decisions. We anticipated that the museum property would allow the fusion classifier to adapt the results to each museum, as the quality of each modality is highly dependent on the museum. We choose the GBDT algorithm for the multimodal classifier. The input is similar to the one for the tabular classifier described in Section 4.1.2 and Table 4.7; there is just one additional column for each of the three modalities, each column containing the class labels predicted by the corresponding classifier for all of the tasks. If a modality is missing, the values in the corresponding column are set to [NA], just in the way missing class labels are considered by the tabular classifier. Thus, the multimodal classifier can cope with incomplete records (i.e. records with missing modalities) by design. We created a separate multimodal classifier for each task, i.e. no multitask learning is applied in multimodal classification. .

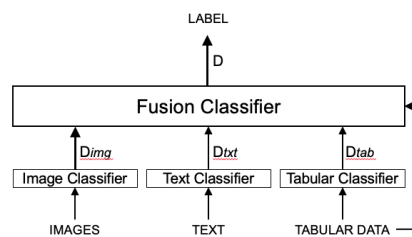


Figure 4.6: Architecture of the multimodal classifier. Each classifier based on a single modality takes its own independent decision, D_c , which serves as input to the multimodal classifier. The final decision D is taken by the multimodal classifier, predicting a task-specific label and assigning it to the record.

There are several advantages to late fusion over early or intermediate level fusion in our case. Firstly, each record may have multiple images but a single text description. Effectively, the input dimensionality is different. With late fusion, we allow the image classifier to deal with

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

it independently, e.g., by classifying multiple images for the same object and picking the decision with the highest confidence. Secondly, the decisions, represented by a one-hot class vector, have a smaller dimensionality than intermediate representations and thus are more appropriate for scenarios with few samples, which is a common problem in the context of our domain (cultural heritage).

4.1.3 Experiments and Results

Image Classification

For all experiments in the frame of image classification ¹⁰, we use the split of the dataset described in section 4.1.1 in order to train the CNN for image classification presented in section 4.1.2 by means of the training strategy described in section 4.1.2. We use all images that are assigned to a record for training and classification, assigning the class labels of the corresponding records to all images associated with it. As pointed out in section 4.1.2, for records associated with multiple images, all images are classified by the CNN at test time, and the image-based prediction having the highest class score is chosen to be the final result.

Experimental setup The workflow of our experiments is as follows: The training dataset is used to update the weights $(\mathbf{w}_{RT}^T, \mathbf{w}_{FC}^T)^T$ of the CNN with early stopping. The model parametrization and hyperparameters leading to the lowest loss are calculated on the validation set.

In this context, we tuned the hyperparameters listed in Table 4.8, choosing the values achieving the highest average F1 scores on the validation set. Table 4.8 also presents the selected hyperparameter values. Finally, all test set records for which at least one image is available are used for an independent evaluation, using the hyperparameters values tuned on the validation set.

¹⁰<https://github.com/silknow/image-classification>

Table 4.8: Hyperparameters tuned (image classification). An optimal variant is obtained with $\eta=1e-4$, $\omega_R=1e-3$, $NL_{RT}=30$ (i.e., 10 residual blocks), with the focal loss $E_F(\mathbf{w})$.

Hyperparameter	Range	Best
Learning Rate η	[1e-5, 1e-3]	1e-4
Weight Decay ω_R	[0.0, 1e-5]	1e-3
Degree of fine-tuning NL_{RT}	[0, 36]	30 (E_F) 15 (E_{SCE})
Loss $E(\mathbf{w})$	$\{E_{SCE}(\mathbf{w})$ (eq. 4.1), $E_F(\mathbf{w})$ (eq. 4.2)}	focal

We will report the overall accuracies as well as the average F1 scores of the best CNN variant in terms of the average F1 score obtained on the test and validation sets for two variants: the first CNN variant is trained by minimizing the softmax cross-entropy loss (equation 4.1), whereas the second variant is trained by minimizing the focal loss (equation 4.2). The overall accuracy *OA* describes the percentage of correctly classified images, denoted as true positives *TP*, among all classified images. As the *OA* is biased towards classes with more examples in an imbalanced class distribution, the classification performance of underrepresented classes is not reflected by the *OA*. In contrast, the class-specific F1 scores, being the harmonic means of precision (i.e., the percentage of the images assigned to a certain class that actually corresponds to that class in the reference) and recall (i.e., the percentage of the samples of a class according to the reference which is also assigned to that class by the CNN) reflect the classifier's ability to predict a certain class. We report the average F1 scores (also referred to as macro-averaged F1 score) per variable, i.e. the average values of all class-specific F1 scores of the classes for that variable.

Results The quality metrics obtained on the validation and test sets are listed in Table 4.9. These quality metrics are determined on the basis of the prediction results for records (i.e., not on the raw results for individual images in case of records having multiple images). In this section, some general observations and the conclusions drawn from them will be briefly described, where a more detailed analysis of the results can be found in section 4.1.4.

Comparing the F1 scores as well as the OAs obtained on the validation and the test set, respectively, shows that the hyperparameter tuning on the validation set did not result in overfitting as the order of magnitude of the quality metrics on the validation and the test set are en par. Furthermore, the average F1 scores and the OAs are higher in case of minimizing the focal loss in training. Accordingly, it can be concluded that the classifier is able to better predict the classes of the four tasks by focusing on harder training examples, as is realized in the case of the focal loss. In particular, underrepresented classes benefit more from the use of the focal loss, which is indicated by the larger improvements in terms of the F1 scores compared to the improvements in terms of OA. The average F1 scores over all variables is 3.7% higher in the evaluation for minimizing the focal loss compared to minimizing the softmax cross-entropy, whereas the improvement in terms of OA amounts to 0.9% on average.

Text Classification

Experimental setup. In the text classification experiment ¹¹ we use the method described in Section 4.1.2, implemented using PyTorch [99] and Transformers [132], and the data described in Section 4.1.1 split into training, validation (sometimes called development), and test subsets

¹¹<https://github.com/silknow/text-classification>

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

as described in Section 4.1.1. We used the base XLM-R architecture (125M parameters) with 12-layers, 768-hidden-state and its respective provided weights. The layers in the classification heads are initialized using the normal distribution $\mathcal{N}(0.0, 0.02)$ with bias parameters set to zero.

First, we performed a 50-trial random search hyperparameter tuning implemented using Optuna [3]. During hyperparameter tuning, the text classifier is trained on the train set, and we chose the hyperparameters that resulted in the highest macro F1 score obtained by evaluating on the validation set. These hyperparameters are detailed in Table 4.10. We then train a model on the train set with the previously selected hyperparameters and evaluate it on both the validation and test sets.

Results. The results of text classification are shown in Table 4.11, which presents the overall accuracy and the average F1 scores achieved on all records containing text in the validation and test sets.

Tabular Classification

Experimental setup. The experiments for tabular classification follow a similar protocol as those for the image and text classifiers, the main exception being that this classifier is not based on multitask learning. Thus, for each task we train an individual classifier with different parameters and hyperparameters, selected by task-specific hyperparameter tuning using grid search. We show the hyperparameters, the search space for tuning, and the selected values in Table 4.12. Note that the ranges selected were all within very reasonable intervals as an additional guard against overfitting.

Results. We show the evaluation results in Table 4.13. Given that it essentially relies on co-occurrences of very coarse labels, the results seem reasonable. In fact, in terms of F1 and accuracy, they almost match the image classifier. We also show feature importance by gain in Table 4.14. For every task, the tabular classifier’s most important feature is the museum. That could probably be expected, because museums are not random collections of objects.

Multimodal Classification

Experimental setup. For the experiments involving multimodal classifiers¹², we started by training the three classifiers based on single modalities (images, text, tabular, respectively) on the training set independently from each other in the way described in Sections 4.1.3 - 4.1.3.

¹²<https://github.com/silknow/text-classification/>

After that, these classifiers were used to classify the samples in the validation set. Finally, we used these predictions as inputs to train the multimodal classifier on the validation set. We used five-fold cross-validation on the validation set to perform hyperparameter tuning using grid search in the same space of hyperparameters that was used for tuning the tabular classifier. The details of hyperparameter tuning are shown in Table 4.15. We trained two versions of the multimodal classifier. The difference between the two versions is the way in which the tabular features are used. The first one corresponds to the multimodal classifier as presented in Section 4.1.2 and Figure 4.6, which used the raw tabular features as an additional input. In the second variant, the raw tabular features are omitted, i.e. that classifier is only based on the output of the three individual classifiers. We perform the evaluation on the test set, dealing with records for which one of the modalities is missing in the way described in Section 4.1.2. As described in the previous sections, we report overall accuracies and average F1 scores for all tasks.

We also performed an ablation study to assess the importance of the individual modalities for the classification results. The ablation study was performed by removing one of the modalities from the input of the fusion classifier, leaving only the other two modalities and, optionally, the tabular data as inputs. Again, overall accuracies and mean F1 scores achieved on the test set are reported for these variants of the classifier.

Results. The results of the experiments are shown in Table 4.16. In this table, we compare the results of the multimodal classifier with and without using the raw tabular data as an additional input. As expected, the variant of the classifier using these additional input features produces slightly better results than the one without these features. Table 4.17 gives the feature importance of the individual modalities for both variants of the classifier, The feature importance for the raw tabular data are presented in Table 4.18.

The overall accuracy and mean F1 score achieved by the multimodal classifier are better than those for the image classifier (Table 4.9) and slightly worse than those reported for the text classifier (Table 4.11), but this comparison is inconclusive because the results in Table 4.11 only consider records for which text is available, which is only about 39% of the test set, whereas the evaluation of the image classifier is based on about 96% of the test set and multimodal classification is based on the complete test set. For the majority of the samples, only images and / or tabular information are available, and thus the prediction would be based on these modalities. In order to be able to allow for a comparison of the results of all modalities, we carried out an evaluation of all modality-specific classifier and the multimodal classifier on the entire test set. In this evaluation, a record for which a modality was missing was considered a wrong prediction for that modality-specific classifier. For instance, a record without images was considered to be a false prediction for the image classifier. The resultant overall accuracy

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

values and mean F1 scores are shown in Table 4.19. In this comparison, the multimodal classifier significantly outperforms all of the classifiers based on a single modality only.

The results of the modality ablation study shown in Table 4.20 indicate that the best combination of any two modalities is text and image, suggesting they are the most complementary.

It comes at no surprise that combining image, text, and raw tabular features results in a classifier that performs almost on par with the complete multimodal classifier (75.1% vs. 75.6%), because it is based on the same information.

Comparing the results achieved by the classifier without images (text + tabular) to those achieved with images (image + text in Table 4.20 or complete classifier in Table 4.16), we can see that adding the images increases the F1 score by about 5%, which confirms the assumption that they provide meaningful information not present in the other modalities.

4.1.4 Discussion

Image classification

Here, we will provide a detailed analysis of the results of the CNN-based image classifier in Table 4.9.

The table shows that the classification performance strongly varies between tasks.

Comparing the OAs, one can see that the variable *material* achieves the highest OAs, followed by *technique* and *timespan*; the worst OA is achieved for the variable *place*. Taking the class structure shown in Table 4.1 into account, a connection can be made to the number of classes constituting a task's class structure. The larger the number of classes to be distinguished, the lower the achieved percentage of correctly classified images in the softmax experiment, where a similar behavior can be observed for the focal experiment; *material* having three classes has the highest OA of 80.7%, followed by *technique* having four classes with 76.8% correct predictions and *timespan* with five classes with a OA of 64.0%, whereas *place* with nine classes has the lowest OA of 62.2%.

An analysis of the task-specific F1 scores in connection with the class distributions of the respective task indicates a dependency of the F1 score on the degree of class imbalance. Taking the ratio of the number of image examples for the majority class, i.e., the class with the most labeled examples in the dataset, in relation to the number of image examples for the minority class, i.e., the class with the fewest examples, a negative correlation between this ratio and the achieved task-specific F1 score can be observed for the focal loss experiment, where a similar behavior can be observed for the softmax experiment. The majority class of *technique* has 2.5 times as many examples as the minority class and *technique* has the highest F1 score of

77.9%, followed by *timespan* with a ratio of 4.9 and a score of 57.5% and *material* with a ratio of 7.7 having a score of 51.2%. The lowest F1 score of 47.0% is obtained for *place* with a ratio of 8.3. We attempted to overcome this dependency of the F1 scores on the class distributions through focusing on hard training examples by means of the presented variant of the focal loss in equation 4.2. Analyzing the improvements of the F1 scores by utilizing the focal loss instead of the softmax cross-entropy loss shows that the focal loss indeed reduces this dependency: except for the variable *place*, there is an improvement of the task-specific F1 scores, and in these cases it is larger for tasks with a high class imbalance (indicated by a high ratio between the number of examples for the majority class and the minority class, respectively). The F1 score of *material* (ratio of 7.7) is improved by 7.8%, whereas the F1 score of *technique* (ratio of 2.5) is improved by 3.9%. The variable *place* with a ratio of 8.3 should have received the largest improvement in F1 score according to the general trend, but it actually is slightly worse (-0.2%). We assume this to be related to the large number of classes to be distinguished for *place*, which might make a correct prediction more complicated for this variable than for the other ones.

In summary, the utilization of the focal loss improves the performance of the trained classifier in correctly predicting the properties of silk based on images. Even though the variable-specific F1 score still seems to depend on the degree of imbalance of a task's class distribution, focusing on hard examples during training primarily improves the task-specific F1 scores of tasks with large class imbalances, as long as the number of classes to be differentiated is not too large. Solving the remaining challenge of predicting all classes of a task equally well may require more data, as not all aspects of all silk properties are equally well represented in the available images.

Text classification

We analyzed about 20 misclassified English language test set examples for each task. In around half of the cases, there was no direct information that could've allowed an accurate classification. E.g., no location mentioned when attempting to classify *place* or no year mentioned when attempting to classify *timespan*. This forces the classifier to rely on other statistical regularities present in the text to provide a classification.

The *material* task is particular. Its most common class, "animal fibre", is a de facto background class. All records in the dataset should be of silk fabrics, which means the material they are made of is an "animal fibre". Some have other materials too. These other materials can correspond to a vegetable fibre (e.g., cotton) or a thread with some metal (e.g., gold thread). While the problem of not having label specific information in the text is common in the examples we analyzed (6/20), obviously incorrectly labeled examples were even more common (9/20). This occurs when either the original record was missing the correct label or when the

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

automatic extraction and linking of the label failed. The high prevalence of this type of error within this task in the examples we analyzed, combined with its absence in other tasks, leads us to suspect that this is the main cause of the relatively lower accuracy and F1 scores for this task.

The *technique* task is also particular in terms of the examples we analyzed. A significant number of examples (5/19) contain information that would imply multiple labels, where usually a small part of the object was produced using a different technique from the main part of the object. A similar type of error occurs in the *timespan* task within a similar proportion of examples (5/10). In the *timespan* task, this can occur when an object was produced at a certain date but later altered or when the estimated date of production within the text crossed centuries.

We hypothesize that the somewhat better results for the place task are connected to regularities between the museum and an object's place of production. This connection is suggested in Table 4.14. Text descriptions are very indicative of the museum, not just in the language but usually also in style, length, and topics.

Tabular classification

Intuitively, from a domain perspective, we can expect that these variables to be associated. For example, a certain country is more active in the textile industries during a certain timespan than during others. Further, museums are typically curated and not random collections. However, given the limited number of features and the coarseness of the labels, we should not overestimate the strength of the association between variables, which we calculated as Cramer's V in Figure 4.7.

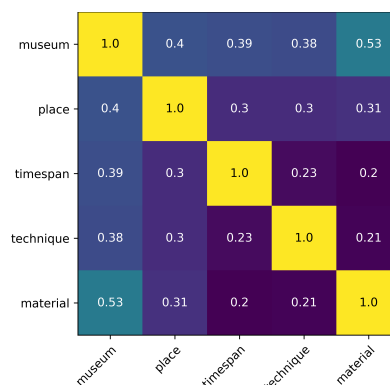


Figure 4.7: Association between features of the tabular classifier, measured using Cramer's V.

Multimodal classification

As pointed out in Section 4.1.3, the comparison of the classification accuracies indicates that the text classifier achieves the best performance of all modalities (Table 4.11), but of course it is only applicable when text is available, which is only the case for a relatively low number of records, (cf. Table 4.3). This confirms our hypothesis that multimodal classification results in a better classification performance if one of the aspects under consideration is to obtain correct predictions for a number of records that is as large as possible. When evaluated on samples having text, the text classifier might achieve higher accuracy metrics; however, a considerable percentage of samples cannot be classified in that way, and the total number of correct classifications is largest when using multimodal classification (cf. Table 4.19).

As Tables 4.19, 4.20, and 4.17 imply, each modality contributes significantly to the multimodal classifier. Looking at the feature importance of the multimodal classifier in Table 4.17, we can see that the output of the text classifier is the most important feature, except when it comes to predicting technique, almost certainly due to the relatively small number of records with an annotation for *technique* in the validation set for which text is available (487 records as opposed to about 1100-1600 for the other tasks). Table 4.18 indicates that the improvement achieved by considering the raw tabular data is almost entirely due to the information about the museum present in the tabular data, allowing the multimodal classifier to take better decisions; the other raw tabular features have a very low feature importance.

Figure 4.8 shows the confusion matrices. Most errors in the timespan task occur between chronologically similar dates. Most errors in the place task occur between countries that are geographically close to each other, e.g. Italy (IT) and France (FR). As far as material is concerned, there is no uniform distribution of errors: errors occur primarily between animal fiber and the other labels, because all objects are made of silk and due to the label imbalance. No clear trend can be observed for the prediction of technique.

Agreement between modalities

We calculated the agreement between the different classifiers on the subset of the test set for which all modalities are present, i.e., records that have the label for the task, a text description, and at least an image. The statistic used was Cohen's kappa [29], the common metric for inter-rater agreement [88]. Figure 4.9 shows the agreement as a heatmap. We can clearly see that in all cases, the text and multimodal classifiers have a very high agreement. This is expected because the text classifier has the highest accuracy and F1 of any individual modality, and Table 4.17 had already shown that the multimodal classifier heavily relies on the output of the text classifier. However, since we've restricted the data only to records that actually have a text description, the numbers here are much higher, even for the *technique* task, which does

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

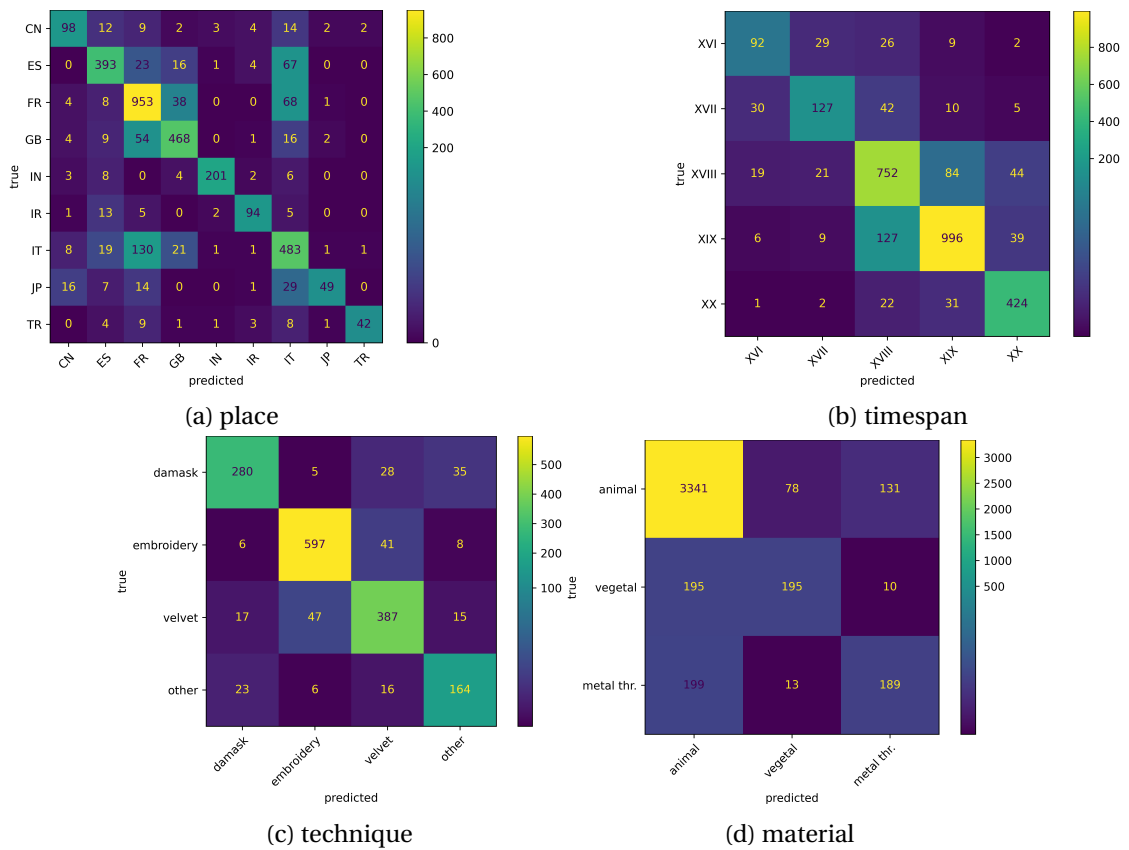


Figure 4.8: Multimodal classifier confusion matrices: predicted vs true labels.

imply that in the case of *technique*, the multimodal classifier does learn that, when present, the output of the text classifier is the best predictor of the result. That is also the only task where the multimodal classifier has a substantial agreement (>0.6) agreement with any other modality.

In all cases, the image and text classifiers agree more with each other than with the tabular classifier. The image and tabular classifiers have a similar level of agreement with the multimodal classifier in *place* and *time*, while the image classifier and multimodal have a higher agreement in *technique*, and, conversely, the tabular and multimodal have a higher agreement in *material*. This difference isn't very big as in the case of *material* both can be considerate as having only a "fair" agreement with the multimodal classifier and, in the case of *technique*, both land inside the threshold for "substantial" agreement. The agreements between each modality are the lowest in *material* and the highest in *technique*, perhaps explaining, in part, why they represent the lowest and highest F1 results for the multimodal classifier, respectively.

Chapter 4. Predicting metadata gaps

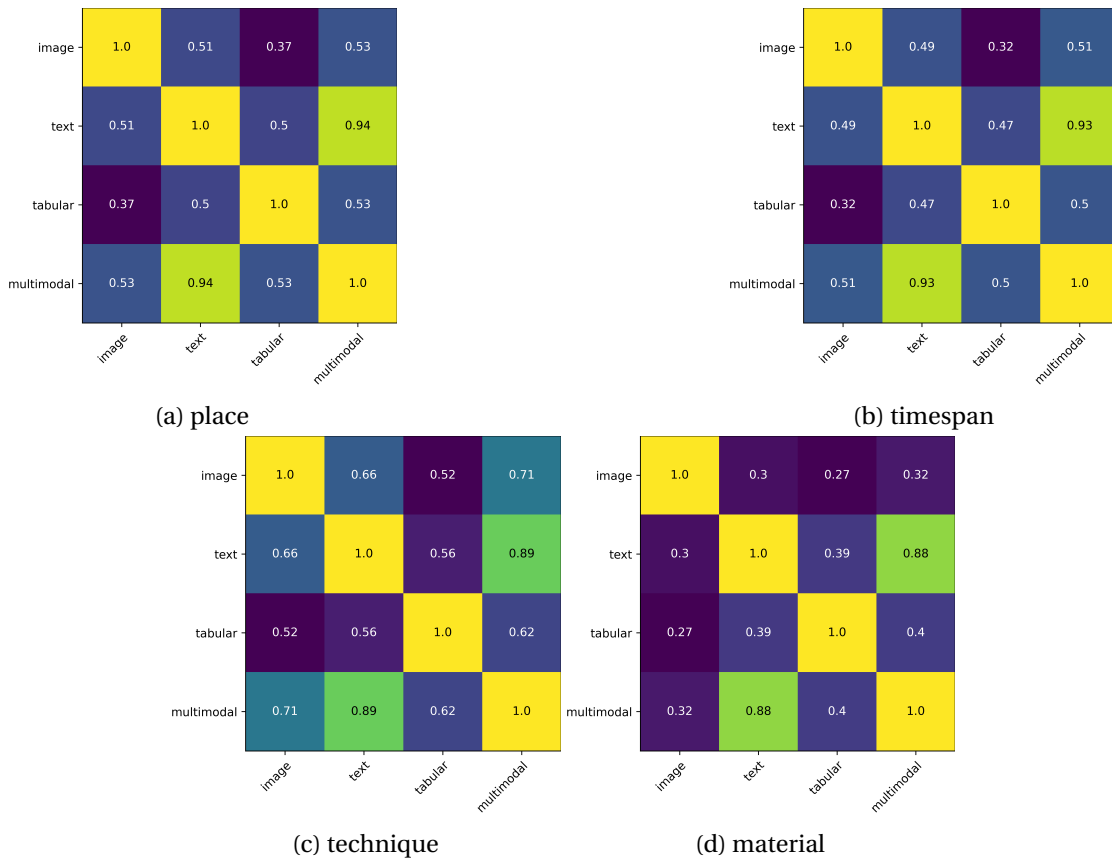


Figure 4.9: Agreement between modality predictions (Cohen's kappa).

Analysis of disagreements between modalities

In this analysis, we use the subset of the test set that contains both text and image descriptions. We also limit it to only English language text for ease of analysis and to present examples.

technique

In 10 / 12 cases where the image classifier is wrong, and the text classifier correct, the description includes the technique directly as a word (e.g., "embroidery", "velvet"); in all 5 cases where the text classifier is wrong, and the image classifier correct, the text description contains no useful information with regard to the technique used as in the example in Listing 4.1.

"This design consisting of an ogival framework enclosing floral motifs had a long period of popularity in Europe."

Listing 4.1: Example of incorrectly classified text description in the task "technique".

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

material

A material is hard to identify from a photo, and this becomes even more challenging since the task is not to identify a material but a mixture of materials – every object is, at least, partially, made from silk. In all instances of disagreement in this task, the text classifier was ultimately correct although in certain cases, the actual label present in the dataset was incorrect.



Figure 4.10: Example of disagreement between classifiers in the material task: the object includes cotton, hard to see in the image but clear in the description which includes the passage "filled with cotton".

place and timespan

All 21 disagreements in the place task between the text classifier and image classifier where the latter is correct follow from the lack of any information in the text regarding location or the presence, instead, of misleading information such as in Listing 4.2. There are 248 cases in this task where the image classifier made the wrong prediction but where the text classifier made the correct one, although some of these seem to be lucky guesses, in the majority there is a clear indication of the place of origin with sentences such as "IWWI Coulson Manufactes Lisburn Ireland." and "By 1500 the motif was popular on Ottoman Turkish textiles.". The situation for *timespan* was essentially the same.

(...) in French, is the name of the particular stitch demonstrated. (...)
French knot, gobelin, hem, laid work.

Listing 4.2: Example of incorrectly classified text description due to misleading text in the task "place".

Qualitative Analysis of Confusions

We now analyze the confusions of the different modality classifiers on selecting the tasks *timespan* in Figure 4.11 and *place* in Figure 4.12, both of which allow us to have an intuitive measure of distance. We can see that in both cases, the errors of the text classifier are mostly near, in time or location, to the correct label. This traces back, in part, to the fact that the text classifier is relatively accurate and the fact that many of the errors are due to a misleading text

Chapter 4. Predicting metadata gaps

description. These are often misleading in a way that creates this behavior, examples shown in Table 4.21.

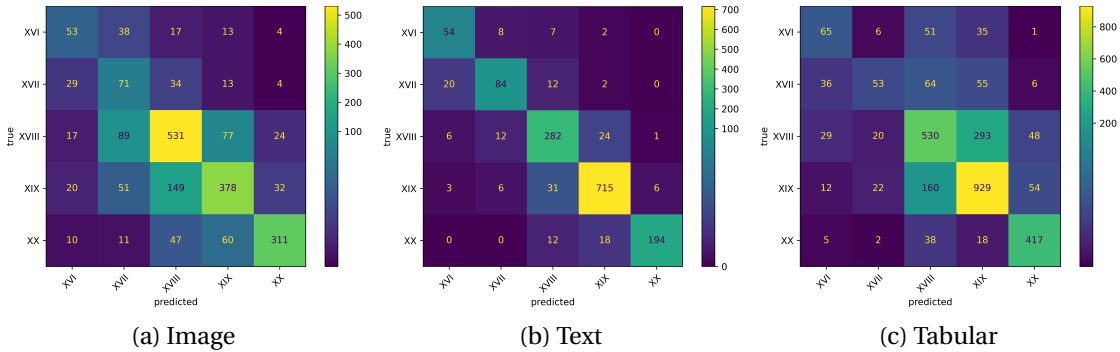


Figure 4.11: Confusion matrices for the different modalities, task: timespan.

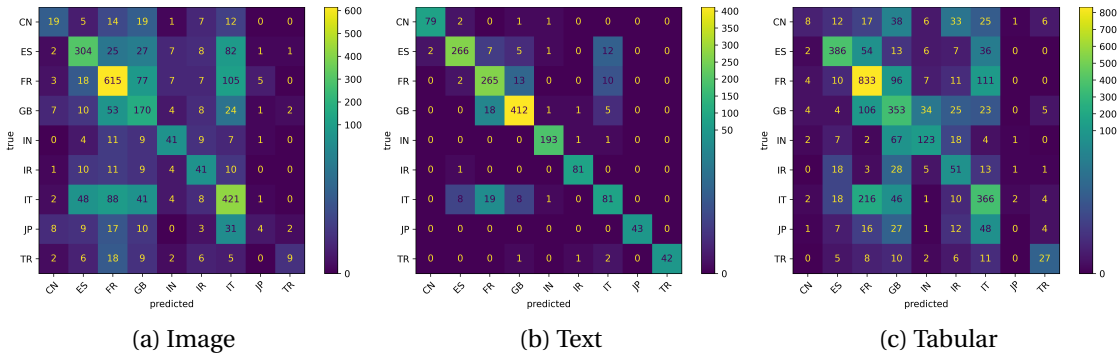


Figure 4.12: Confusion matrices for the different modalities, task: place.

4.1.5 Conclusion

We presented results for three individual modality classifiers, as well as multimodal results. In terms of our original hypothesis, presented in section 4.1, we showed that we were in fact able to accurately predict missing properties in the digitized silk fabric artifacts that made up our dataset. While the quality of predictions varied between individual modalities, we showed that the multimodal approach provided the best results. To recapitulate, our contributions included the already mentioned multimodal approach tailored specifically to the challenges we faced in the multimodal scenario, including the incomplete overlap of data across modalities. The individual approaches of each modality-specific classifier also provide a useful contribution to the automated classification of cultural heritage objects. The image and text classifier offer the possibility of being applied to data outside a Knowledge Graph (KG) or database, possibly even directly submitted by the user of a system. The tabular classifier, on the other hand, offers the possibility of classifying data in a KG or database when no text descriptions or images are present by relying on other properties. It is also important

4.1 Multimodal Metadata Assignment for the Cultural Heritage of Historical Silk Fabric Artifacts

to remember that in most practical situations, including inside a KG or other knowledge bases, images are more common than text descriptions of objects in the cultural heritage domain.

The data we used in our work originally comes from many different museum sources and is from a very specific Cultural Heritage domain: historical, European silk fabrics. We applied common methods to process such data and developed an ontology and a Knowledge Graph out of the original museum texts and images. Such an effort comes typically with challenges, which in our case consisted mostly of a small amount of (training) data, domain specificity, different styles of writing texts and capturing images of objects, different languages (in the case of texts) and finally simply annotation errors, typos and other errors that happened during the original digitization. Not all of these challenges can be completely overcome and some of them, like the metadata gaps, even constitute part of the motivation to conduct this research work. As some data imperfection could still not be totally excluded, some removal of data was necessary to ensure sufficiently clean and class balanced data for our supervised approaches. This could, however, be very much alleviated through the grouping of certain labels, which was also possible through our domain expert-designed thesaurus about silk fabric concepts. In the end we can present a cultural heritage dataset that can be used for automated classification or even multimodal approaches. In this section, we also provide the data modelling of how metadata predictions for data such as our can be represented within knowledge graphs or other knowledge bases.

We have shown that properties of silk fabrics can be predicted from images of these fabrics. In this context, we proposed to use the focal loss for training in order to compensate for the effects of class imbalance in the training set, a problem that is quite common in the cultural heritage domain. Our results indicate that the proposed strategy can mitigate this problem to a certain degree, in particular improving the classification performance for the underrepresented classes in terms of the F1 score. Image classification performs particularly well for the task to predict the technique used for producing a fabric. Nevertheless, there is still room for improvement, as indicated by the performance metrics for all variables.

When text descriptions are present, the text classifier provides the best results of any single modality. It seems, thus, that the text classifier was able to overcome the primary challenges it faced: small dataset, domain specificity, cross-linguality, and museum specific text styles. This was primarily achieved by the choice of XLM-R as the basis of the text classifier.

When all data is considered, we have shown that the multimodal approach is the best according to the macro F1 metric and that these results are largely due to combining the image and text classifiers, although there is still a clear benefit in including the tabular data. The advantage of having a separate tabular classifier over just including the same features during modality fusion is small in terms of F1 scores on our test set, but still significant. While most records contain images, not all do (3.4%) and a smaller number of records contain neither text nor

images (2.1%). On the other hand, if we had tried to implement a classifier using the text modality alone, we could only classify 40% of the records. While we can say that a multimodal approach does allow us to classify a greater number of records than using images alone, the primary practical benefit of the multimodal approach over performing just image classification is probably the qualitative improvement in classification results demonstrated.

We have given a detailed analysis of the errors and disagreements between classifiers and a description of common errors present in the dataset. Misleading text descriptions stand out as a challenge for text classification in this context. In terms of the dataset, a perhaps a better approach could be found for dealing with noisy labels, as well as finding better ways to deal with fine-grained labels and label ontology mismatches.

Future work on image classification could concentrate on improving the performance for underrepresented classes even more, e.g., by using methods for few-shot learning. Furthermore, as some experimental results indicated that some training labels might be incorrect, training methods that are robust against such errors ("label noise") could be investigated.

4.2 Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles

As shown in the first section of the chapter above it is possible to train text (and image) classification models from metadata records in order to predict missing categorical values in other records. However, such models do require a significant amount of annotated data for training, which is expensive to get when the domain is very specific.

Therefore, we propose a Zero-Shot Classification (ZSC) approach¹³ based on the ConceptNet common-sense knowledge graph [120], to predict the missing categorical metadata while avoiding to rely on training data. We compare our approach with supervised approaches and we show competitive results and demonstrate the ability to predict more fine-grained concepts despite the specificity of this domain.

The remainder of this section is structured as follows. We summarize the development of the Knowledge Graph and we describe the dataset being used in our experiments in Section 4.2. We detail our Zero Shot classification approach as well as two supervised learning baselines in Section 4.2.1. We analyze the classification results in Section 4.2.2. Finally, we provide conclusions and outline some future work in Section 4.2.3.

¹³<https://github.com/silkknow/ZSL-KG-silk>

Dataset and Preprocessing

The multilingual Knowledge Graph of silk textile productions consists of descriptions of 40,873 unique objects illustrated by 74527 images in four languages: English, Spanish, French and Italian. While the information integration process has been effective, one general problem of the KG is that many properties have missing values. In this section, we focus on three important properties describing the silk production namely: the material used, the weaving technique employed and a the subject depicted. Consequently, we extract from the knowledge graph three subsets corresponding to the set of objects having values for those properties.

The silk thesaurus which is being used to normalize the values of those properties contains a very exhaustive inventory of possible materials and techniques, organized in a hierarchy. While some of those materials or techniques are widely used in the data, others are niche and the knowledge graph includes only a very limited number of objects with some of them. One solution is to walk up the thesaurus hierarchy and only consider more general concepts. Ultimately, we need to find a trade off between using fine-grained concepts with the risk of having too sparse data, or too broad concepts with the risk of being non informative. This is a manual process informed by both the thesaurus hierarchy and the available data. Table 4.22 provides the list of the thesaurus concepts that we finally aim to predict for the three properties (material, technique, depiction) as well as the number of unique objects.

As a general preprocessing step, we also removed all records that have multi-valued properties. Some records have, for example, both “Gold Thread” and “Vegetable Fiber” set as material properties. Including such a record would make the training of a model capable of distinguishing these two concepts harder. We also create language specific subsets. We observe significant differences between the English and Spanish subsets, which highlight the heterogeneous nature of our sources. In particular, subject depiction sticks out as we only have objects from Spanish records having this property informed. The language specific subsets will be used by our Zero Shot classification approach (Section 4.2.1) while supervised learning methods will make use of the complete multilingual dataset in the 4 languages (Section 4.2.1).

4.2.1 Approach

Supervised Approaches

In order to be able to evaluate our approach, we propose to compare it with two supervised classification methods from section 4.1 that we use as baselines. For both of them, we will use the three sub-sets described in the Table 4.22 and perform a 80%-20% split in order to have training data. We will perform a five-fold cross validation for testing the models.

Classification based on textual descriptions. The goal of this approach is to predict missing

categorical values of museum records based on lengthy textual descriptions. This approach consists of a Convolutional Neural Network (CNN) built over cross-lingual pre-trained word embeddings which are the aligned fastText vectors trained on Wikipedia [62]. More precisely, a series of convolutional blocks with varying kernel sizes (2,3,4), each consisting of 100 filters, are applied to a sequence of such word-embeddings that got mapped from input description texts from the Knowledge Graph. These filters create an output for which a Gaussian Error Unit (GELU) non-linearity is used and a max-pooling operation is applied for each block. The idea is to, hopefully, select the best features of each block. Afterwards, they get concatenated into one single vector, regularised by a dropout layer and finally sent to a softmax classification layer to come up with the final predictions per input.

Classification based on images and metadata. The goal of this approach is also to predict missing categorical values but this time, using the images illustrating the objects with the other metadata values. The underlying assumption is, that it is to some degree visible on images what materials have been used or which technique was employed to produce a silk textile. For this model, a CNN was also used. More precisely, a pre-trained ResNet backbone network served as a generic feature extraction network. The output is then processed by several fully connected network layers that deliver a joint representation of the images and a final classification layer offers afterwards a probabilistic class score per variable and concept. The model is trained based on multi-task learning to perform predictions for the three properties simultaneously: material, technique and depiction. The training is based on stochastic minibatch gradient descent and using focal loss. For ResNet only, the last convolutional layers are fine-tuned.

Zero-Shot Prediction

The benefit of our approach is to perform a similar prediction without relying on any training data. The underlying assumption is that a textual document about a topic such as “Embroidery” will probably mention other words that are similar to this concept such as “fabric” or “stitch”.¹⁴ Our approach relies on the ConceptNet common sense knowledge graph. More precisely, we need to feed our approach with mappings between the targeted values we wish to predict and the ConceptNet network. These mappings have been manually established (Table 4.24).

ConceptNet is then used to produce a list of candidate words related to the concept of interest, which can be called “topic neighbourhood”. Each topic neighbourhood is created by querying every node that is N steps away from the label node. In our experiments, we set N=2. A score for each label is computed based on the content of the text document that we give as input. This score is calculated based on cosine similarity via ConceptNet Numberbatch, the graph embeddings of the network. Even if a word has several meanings, only one neighbourhood

¹⁴See for example the object described at <https://ada.silknow.org/object/c57358a7-c908-3110-b65d-70b09f5f4c4b>

4.2 Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles

per spelling is generated. The score is then supposed to represent the relevance of any other term to the main label inside a neighbourhood. Based on these scores, the whole document (museum record) gets also a score and, therefore, a document label too. This is done by quantifying the overlap between the document content as a list of tokens and the label neighbourhood nodes. Finally, as mentioned before, all predicted document labels can be explained by the model through showing the path between nodes or highlighting the words or n-grams that contributed to the final classification.

This approach has a number of limitations: the concepts that should be predicted must exist in ConceptNet. Furthermore, while ConceptNet is multilingual, the embeddings are language specific. Therefore, our zero shot classification approach will make use of language specific subsets described in Section 4.2.

4.2.2 Evaluation

In this section, we compare the results of our Zero-Shot Classification (ZSC) approach with the supervised methods described in the Section 4.2.1. We present the results alongside each of the three properties of interest: material, technique and depiction. It is worth to note that the precision, recall, F1 scores are obtained on 20% of the dataset following a 5-fold cross validation while the figures reported for the ZSC method concerns the entire language specific datasets described in Section 4.2.

Table 4.23 shows the results for predicting material concepts. The two baselines approaches can only predict whether the material used is “Metal” or “Vegetable Fiber” while our ZSC approach can predict more fine grained concepts than just “Metal”, such as “Gold” or “Silver”. On the English subset, the ZSC method shows promising results with F1-score of 71.6% and 64.4%) respectively. On the Spanish subset, the prediction results are lower for the ZSC method, in comparison to the supervised approach. The topic neighborhood in this language is also less elaborated. The Text CNN method benefits clearly from the multilingual embeddings.

The results for the prediction of technique concepts are presented in Table 4.25. On the Spanish subset, we again observe lower scores for the ZSC approach. The problem relies on both the quality of the Spanish textual descriptions and on the Spanish concept neighbourhoods in ConceptNet. On the English subset, the ZSC approach performs in par with respect to the Image CNN baseline but less well than the Text CNN one. We observe that “Tabby”, a very domain-specific concept, is particularly difficult to predict and it was discarded by the Text CNN approach since too little usable textual descriptions were available. The ZSC method performs reasonably well with a similar input and better than the Image CNN baseline.

Figure 4.13 provides the confusion matrix of the predictions made for the technique property with the ZSC method on the English subset. We can see that the true label is usually predicted

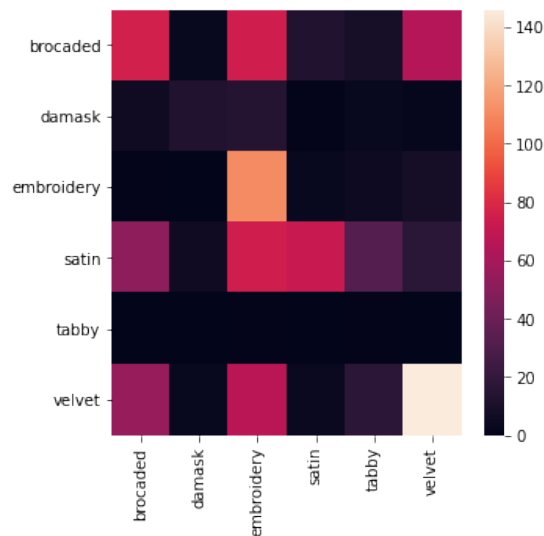


Figure 4.13: Confusion matrix for the property technique on the English subset for the ZSC method. The Y-axis represents the true labels and the X-axis the predicted ones.

the most per class, especially in the case of “embroidery” and “velvet”. We observe that “Brocaded” is often confused with “embroidery” or “velvet” while “Damask” and “tabby” are rarely predicted. In the case of “damask”, this just reflects a low amount of samples, whereas in the case of “tabby”, the predictions almost do not work at all (F1-score of 2.9%).

Figure 4.14 shows an example where “Embroidery” is correctly predicted, while Figure 4.15 depicts a counter-example (the correct technique should have been “Embroidery”). The graph highlights the most relevant words used for the predictions. The first example is based on the text: "*Spot samplers feature motifs that are scattered in a seemingly random fashion over the surface of the foundation fabric, usually linen. These samplers are rarely signed or dated, and often include motifs that are only partially worked, leading to the conclusion that this type of sampler was made as a personal stitch reference for its maker, and not for display, as band samplers were signed by student embroiderers. The sampler features flowers, obelisks on pedestals, and an "S" motif, in addition to geometric designs that are of the type that would have been used to decorate small purses, cushions, and other accessories.*", taken from a record from the MET Museum.¹⁵

The second text is "*An example of the kind of work [Catherine de Medici] appreciated is the Museum’s panel of yellow satin embroidered with silk threads. One of a set of three (the others are in the Musée Historique des Tissus, Lyon), it hung as a valence around the top of a four-poster bed. Various print sources were culled for the airy design of grotesques, while its five vignettes derived from Ovid’s Metamorphoses- based on the myths of Europa, Actaeon, Semele, Pyramus,*

¹⁵<https://ada.silknow.org/object/c57358a7-c908-3110-b65d-70b09f5f4c4b>

4.2 Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles

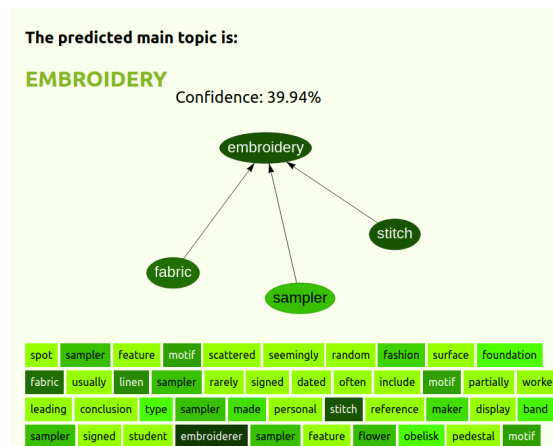


Figure 4.14: "Embroidery" was correctly predicted by our ZSC approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.

*and Salmacis- are adapted from woodcut illustrations published by Bernard Salomon in Lyon in 1557. Its brilliant colors, exquisite design, and sumptuous material would have suited the queen's taste perfectly.*¹⁶, also taken from the MET Museum.

The results for the prediction of subject depiction concepts are presented in Table 4.26. No results are reported for the Text CNN baseline due to the lack of data. We observe that our ZSC approach performs well for predicting the "Flower" concept as does the visual approach. The "Plant" and "Geometry" concepts are however more complicated to predict by the ZSC method. These concepts are general in ConceptNet and the topic neighborhood too broad for the narrower interpretation expected in the silk domain.

4.2.3 Conclusion and Future Work

For these methods, we hypothesize that a common sense knowledge graph such as ConceptNet can feed a Zero Shot classification method for enriching a domain specific knowledge graph such as one describing the silk textile production. Through extensive experiments, we have demonstrated promising results for such an approach in its ability to sometimes reliably predict fine-grained concepts without requiring any training data as supervised classification techniques do. Nevertheless, we observe several limitations: the concepts that should be predicted must exist in ConceptNet with an appropriate topic neighborhood. Our results can be reproduced using the code and datasets published at <https://github.com/silknow/ZSL-KG-silk>.

Even if supervised methods for metadata predictions perform generally better, ZSC remains an interesting method to get accurate predictions even in specific domains that often suffer

¹⁶<https://ada.silknow.org/object/e2f34144-9ce4-3bc4-b3e0-c67854cd994f>

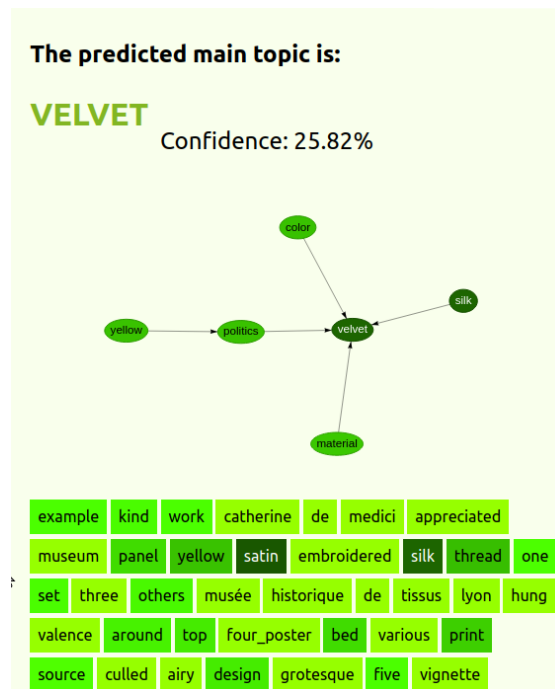


Figure 4.15: "Velvet" was predicted instead of "Embroidery" by our ZSC approach (English, Technique) in this case. Relevant words in the ConceptNet topic neighborhood are highlighted.

from data sparsity. We observe that it is also possible to bootstrap the predictions in using first the ZSC method and then applying supervised classification models to further increase performance. We aim to experiment with such hybrid approaches in the future.

4.3 Prompt-guided Zero-Shot Information Extraction (ProZe)

The Natural Language Processing (NLP) and Information Extraction (IE) fields have seen many recent breakthroughs, especially since the introduction of Transformer-based approaches and BERT [36], which has become the *de-facto* family of models to tackle most NLP tasks. Over the last years, few-shot and zero-shot learning approaches have gained momentum, particularly for the cases with little data and where uncommon or specialized vocabularies are being used. Fully zero-shot classification approaches do not require any training data and often show respectable performance. An interesting new paradigm is *prompt-based learning* which leverages pre-trained language models through prompts (i.e. input queries that are handcrafted to produce the desirable output) instead of training models on annotated datasets. However, a major downside of all these approaches based on transformer-based language models is that they suffer from a lack of explainability.

Recently, ZeSTE [57] tackled this lack of interpretability problem in text classification by departing from language models and relying instead on ConceptNet [121] and its explicit

4.3 Prompt-guided Zero-Shot Information Extraction (ProZe)

relations between words. With every word being a node in ConceptNet, ZeSTE can justify the relatedness between words in the document to classify its assigned label. While it shows state-of-the-art results in topic categorization, it does not offer ways to specialize the classifier beyond “common sense knowledge” (domain adaptation), nor does it offer the possibility to disambiguate labels. These challenges are important to solve for text classification of specific domains, especially since zero-shot classification is particularly useful for domain-specific use cases with little data to train a model. As a consequence, we propose *ProZe*, a Zero-Shot classification model which combines latent contextual information from pre-trained language models (via prompting) and explicit knowledge from ConceptNet¹⁷. This method keeps the explainability property of ZeSTE while still offering a step towards label disambiguation and domain adaptation.

The remainder of this section is structured as follows. First, we give an overview of the relevant state-of-the-art work. We then detail our proposed method called ProZE. Next, we present our results on common topic categorization datasets as well as on three challenging datasets from diverse domains: screenplay aspects for a crime TV series [46], historical silk textile descriptions [110], and the Situation Typing dataset [87]. We report and analyze the results of several empirical classification experiments, which includes a comparison to some state-of-the-art Zero-Shot approaches. Finally, we conclude and outline some future work.

4.3.1 Methods

Our model can be seen as a pipeline comprising several components. In this section, we explain each step of the process in further details.

Generating Label Neighborhoods

The first step of our approach is to manually create mappings between target class labels and their ConceptNet nodes. For instance, if we want our classifier to recognize documents for the class “sport”, we designate the node `/c/en/sport` as our starting node.¹⁸

Based on these mappings between target labels and concept nodes, we can then generate a list of candidate words (from ConceptNet) that are related to the respective concept. This list can be called the “label neighborhood”. Each of the candidate is produced by retrieving every node that is N-hops away from the class label node.

Afterwards, a score can be calculated for each label based on which words are present in the input text or document to classify. To this end, we score every word in the label neighborhood

¹⁷<https://gitlab.eurecom.fr/schleide/proze>

¹⁸In the remainder of this section, we will omit the prefix `/c/en/` as all labels in our datasets are in English.

based on its "similarity" to the class label.

Scoring a Document

Like ZeSTE, we proceed to score each document by first generating a score for each node in a label neighborhood. To do so, multiple approaches exist. In this section, we present and compare 3 such scoring methods (SM):

1. **ConceptNet embeddings similarity (SM1):** ConceptNet Numberbatch¹⁹ are graph embeddings computed for ConceptNet nodes. To quantify their similarity, we compute cosine similarity between the embedding of each node on the label neighborhood and the label node itself.
2. **Scoring through Inference (SM2):** for this scoring method, we use a model that is pre-trained on the task of Natural Language Inference. In a similar setting to the previous method, we prompt the model with a sentence related to the label or its domain, and then we ask it to score all the words from its neighborhood based on the logical entailment between the prompt (premise) and a template containing the word (hypothesis).
3. **Language Modeling Probability (SM3):** for this scoring method, we combine the predictive power of language models with the explicit relations that we can find on the label neighborhood. For each label, we supply the language model with a *prompt*, or a sentence that is likely to guide it towards a specific meaning of the label we target (for example, the definition of the label), and then, we ask it to predict the next word in a Cloze statement (a sentence where one word is removed and replaced by a blank). For example, to score words related to the label "sport", we can give the model a definition of the word, and then ask it to predict the blank word in the following Cloze statement: "*Sport is related to [blank].*". Given that language models, are pre-trained on predicting such blanks, we can use the scores they attribute to that blank to measure the similarity between our label and the candidate words from its neighborhood. For instance, when we give the dictionary definition of sport to the language model, the top predicted words are 'recreation', 'fitness' and 'exercise'. Because the language model outputs a probability for every word in its vocabulary, we score only the words that are originally on the label neighborhood. If a word in the neighborhood does not appear among the predictions of the model (i.e. out of the model's vocabulary), the score from SM1 is used.

Once the scores are computed by one of these methods, we can proceed to score any document given as input to the model. To score such document, we first tokenize it into separate words.

¹⁹<https://github.com/commonsense/conceptnet-numberbatch>

We then take all the nodes from the neighborhood of a label that appear in the tokenized document, and we add up their scores to produce a score for the label. We do so for each label we are targeting, and the final prediction of the model corresponds to the label with the highest score. Because all the nodes in the neighborhood are linked to the label node with explicit relations on ConceptNet, we can explain in the end how each word in the document contributed to the score and how it is related to our label.

Prompting Language Models

In this section, we explain how we leverage language models to score the label neighbors extracted from ConceptNet, as per the scoring methods SM2 and SM3 described above.

Both SM2 and SM3 methods rely on prompting the language model, i.e. to feed it a sentence that would function as a context to "query" its content (also known as *probing* [32]). As expressed in the related work, prompting language models is an open problem in the literature. In this work, we explore some potential ideas for prompting to serve our objective of measuring word-label relatedness.

The prompting follows the same scheme for both scoring methods. We vary both the premise and hypothesis templates and report the results for some proposals in the Evaluation section. For the premise, we experiment with two approaches:

1. Domain description: where we prime the model with the name or description of the domain of the datasets, i.e. "Silk Textile", "Crime series", etc.
2. Label definition: where we prime the model with the definition of the label, with the assumption that this will help it disambiguate the meaning of the label and thus come up with better related words. For instance, for the label "space", we provide the language model with the sentence "Space is the expanse that exists beyond Earth and between celestial bodies". We take the definitions from Wikipedia or a dictionary, we generate it using a NLG model etc.

We observed experimentally that using just the description of the domain as a prompts gives better overall performance. Therefore, we only report results on these prompts in the following sections. As for the hypothesis, we provide the model with a sentence like "*[blank]* is similar to *space*" or "*Space* is about *[blank]*" which we use in our reported results.

We note that, while the combination of premise and hypothesis can impact the overall performance of the model, the search space for a good prompt is quite wide. Thus, we only report the performance on some combinations, as we intend this section to only point out the use of such mechanism for this task rather than fully optimize the process.

Tool Demonstrator

To explain the decisions of the model, we follow the same method as ZeSTE [57], i.e. we highlight the words which contribute to the decision of the classification as shown in a graph that links them with semantic relations to the label node. The difference is that the scores in ProZe take also into account the scoring from the language model. To illustrate the contribution of the language model, we developed an interactive demonstrator enabling a user to test the effect of prompting the language model to improve the results of zero-shot classification (Figure 4.16). This demonstrator is available at <http://proze.tools.eurecom.fr/>.

After choosing a label to study, the user is asked to enter a prompt that can help the model to identify words related to the label (e.g. definition or domain). The user is then shown an abridged version of the prompt-enhanced label neighborhood: the connection between any node and the label node is omitted for clarity but it can be trivially retrieved from ConceptNet, and only the top 50 (based on the used scoring) words are shown to represent the new label neighborhood, with the intensity of the color reflecting higher scores.

The user can view in detail the updates happening before and after introducing the new scoring from the Language Model. For this demonstration, we use the SM3 method to score the nodes as it requires only one pass through the Language Model to generate a score for all words in its vocabulary, whereas the SM2 method requires an inference for every word in the label neighborhood. As a consequence, while the SM2 methods takes up to 7 minutes per label on our hardware, the SM3 method takes less than a second while still delivering good performance.

4.3.2 Datasets

In this section, we present three widely used topic categorization datasets in the news domain, as well as three other very different and domain-specific datasets making used of fine-grained labels.

News Topics Datasets Used to benchmark multiple text classification approaches, news datasets are often categorized by topic and are written in simple and common language. In our experiments, we report results on three such commonly-used datasets: AG News, BBC News and 20NG.

- **20 Newsgroups** [71]: a collection of 18000 user-generated forum posts arranged into 20 groups seen as topics such as “*Baseball*”, “*Space*”, “*Cryptography*”, and “*Middle East*”.
- **AG News** [53]: a news dataset containing 127600 English news articles from various

Crime Aspects The Crime Scene Investigation (CSI) dataset contains 39 CSI video episodes together with their screenplays segmented into 1544 scenes²⁰. An episode scene contains on average 21 sentences and 335 tokens. Originally, this dataset is used for screenplay summarization as each scene is annotated with a binary label denoting whether it should be part of a summary episode or not. Additionally, the three annotators had to justify their choice of their selected summary scenes with regards to it being about one/more or none of the following six aspects: i) victim, ii) the cause of death, iii) an autopsy report, iv) crucial evidence, v) the perpetrator, and vi) the motive/relation between perpetrator and victim.

We define the following labels to evaluate the ProZe system: victim, cause of death, crime scene, evidence, perpetrator, motive. For our classification task, we kept only the scenes which were associated to at least one aspect (449 scenes). In the case where one scene is associated to multiple labels, if the model predicts one of the labels, we consider it a success.

Silk Fabric Properties This dataset is an excerpt from our multilingual knowledge graph. Metadata about silk fabrics contains usually both explicit categorical information, like specific weaving techniques or their production years, but also rich and detailed textual descriptions. Our goal is to try to predict categorical values based on these text descriptions.

The SILKNOW Knowledge Graph dataset can be divided into using "material" and "weaving technique" subsets. More precisely, we slightly extend the dataset used in [111] (see section 4.2), and after removing objects with more than one value per property, we obtain 1429 object descriptions making use of 7 different labels for silk materials, and 833 object descriptions with 6 unique labels for silk techniques. The chosen labels have also to be mapped to ConceptNet entries to work with this approach. Table 4.27 shows the final selection of thesaurus concepts and their mapping to ConceptNet nodes.

4.3.3 Evaluation

We evaluate ProZe on these 6 datasets. In this section, we present the results of this evaluation.

Baselines

We compare our model with:

- *ZeSTE*: this approach solely relies on ConceptNet to perform Zero-Shot classification;
- *Entail*: this model was originally proposed in [135]. We use `bart-large-mnli` as the backend Transformer model, which it is a version of **BART** [77] that was been

²⁰<https://github.com/EdinburghNLP/csi-corpus>

4.3 Prompt-guided Zero-Shot Information Extraction (ProZe)

fine-tuned on the Multi-genre Natural Language Inference (MNLI) task, as per the implementation we use for our experiments (can be tested at <https://huggingface.co/zero-shot/>). Given a text acting as a *premise*, the task of Natural Language Inference (NLI) aims at predicting the relation it holds with an *hypothesis* sentence, labelling it either as false (contradiction), true (entailment), or undetermined (neutral). Generally, the labels are injected in a sentence such as “This text is about” + label, to form an *hypothesis*. The confidence score for the relation between the text to be labelled and the premise to be ‘entail’ is the confidence of the label to be correct. We use the implementation provided at <https://github.com/katanaml/sample-apps/tree/master/01>)

Quantitative Analysis

We limit the size of the label neighborhoods to 20k per label for each experiment, except in cases where querying ConceptNet returns less nodes than that. Then, we resize all the other neighborhoods to be all equal in size to the smallest one (by eliminating the nodes with the lowest similarity), as we found that having neighborhoods of different sizes skews the predictions towards the larger ones (by virtue of having more nodes to contribute to the score). This can be circumvented by increasing the number of hops (thus boosting the size of smaller neighborhoods before filtering), but according to our observations, this hurts the quality of the kept nodes as they get less semantically relevant as we hop further. Resizing the neighborhoods eliminate the bias against the in-domain labels that may not have so many related words in the first place.

Table 4.29 and Table 4.28 show a score comparison of the ProZe approaches to the baselines of ZeSTE and the Entail approach. **ProZe-A** refers to scoring the nodes using a combination of SM1 and SM2, whereas **ProZe-B** uses a combination of SM1 and SM3. We tested several ways to combine the scores from ConceptNet (SM1) and language models (SM2 and SM3), including taking the sum of the two scoring methods, their product, their max, or a weighted average. Empirically, we obtain the best empirical results by multiplying the two scores (both normalized to be between 0 and 1). The main advantage of multiplication is that it penalizes disagreement between the language model and the KG over how close two terms are. This also means that the explainability layer reflects accurately the decisions of the model, as words that are not scored well by the language model will not contribute significantly to the classification score.

Table 4.28 contains the accuracy and weighted average scores for the 3 news datasets that consist of general knowledge texts. ProZe has similar performance, but not beating ZeSTE, which is in line with our expectations: both approaches are based on the ConceptNet commonsense knowledge graph, and the vocabulary does not need or cannot be guided into a more fitting direction with the prompts. For all three news datasets, however, ProZe performs better than

Entail.

Table 4.29 shows the results for the 3 domain-specific datasets. We observe that ProZe is consistently outperforming ZeSTE, which we take as a confirmation that the guidance through the prompt is effective for specific domains. For two datasets, silk material and situations, ProZe even beats the non-explainable baseline scores of the Entail approach. This is not the case for the silk technique and the CSI screenplay datasets as some labels from these datasets have very limited neighborhoods in ConceptNet. Nevertheless, our approach is still close and retains in all cases its higher degree of explainability.

Qualitative Analysis

To illustrate why a re-ranking of related words induced by a domain prompt improves the score, we analyse a concrete example. Taken from the silk technique dataset, the top 10 candidate terms of the ConceptNet label neighborhood for the weaving technique "embroidery" are as follows: "Embroidery, overstitch, running stitch, picot, stumpwork, arresene, couture, fancywork, embroider, berlin work". While these words are clearly related to the concept of embroidery, they are not necessarily relevant in the context of silk textile. For example, "picot" is a dimensional embroidery related to crochet. The intuition is then that this neighborhood can be improved by specifying the domain.

In comparison, the top 10 candidate terms of the pre-trained BART language model, guided by a prompt that included the term "silk textile" are: "Craft artifact sewn, fabric, embroidery stitch, embroidery, detail, embroider, mending, embellishment, elaboration, filoselle". These terms are more general even if also related to silk textile. Words such as "detail", "mending", "elaboration" or "embellishment" seem useful for classifying texts that are not only consisting of details about different types of embroidery. When combining the scores from ConceptNet and the language model, the ProZe method increases its F1 score of circa 8%, from 61% to 69%.

4.3.4 Discussion

In this series of experiments, we demonstrated the potential of fusing knowledge about the world from two sources: First, a common-sense knowledge graph (ConceptNet), which explicitly encodes knowledge about words and their meaning. Second, pre-trained language models, which contain a lot of knowledge about language and word usage that is latently encoded into them. We explored several methods to extract this knowledge and leverage it for the use case of zero-shot classification. We also empirically demonstrated the efficiency of such combination on several diverse datasets from different domains.

4.4 Using transformer-based QA and CQ systems for metadata predictions

This work is experimental and does not fully explore all possibilities of this setup. As future work, we want to study the effect of prompt choice in more detail, and seeing how such choice impacts not only the quality of the predictions but also that of the explanations. Different language models can also be tried to measure how such choice can improve the overall classification, especially for specific domains such as e.g. medical documents.

Another potential improvement over this method is to filter out words unrelated to the label using the slot-filling predictions from the language model. From early experiments, this method seems to give good results by restricting the neighborhood nodes to ones that almost exclusively relate to the label in some way.

A natural direction of work is to involve the user in the creation of the label neighborhood (human-in-the-loop) by asking whether some words that only the Language Model and not ConceptNet suggests pertain to the target label. This allows to inject the extracted knowledge from the language model back into the zero-shot classifier, and fill in the gaps of knowledge from ConceptNet.

Finally, some existing limitations of the original work can be still improved upon such as letting the language model inform the label selection and expansion, handling multi-word labels, and integrating more informative concepts from ConceptNet beyond word tokenization (e.g. *'crime_scene'*, *'tear_gaz'*).

4.4 Using transformer-based QA and CQ systems for metadata predictions

How to assess text understanding in the field of Artificial Intelligence (AI) remains an open question. In order to evaluate if a human understands a concept, we often test their capacities to answer questions, but also to produce meaningful questions about the subject matter [25]. Question Generation (QG) [108] has been a relevant task inside the field of Natural Language Processing (NLP) for many years. Just as with human text comprehension, within Artificial Intelligence (AI) a model's ability to ask meaningful questions is considered to be central to evaluate its text comprehension ability [92].

In recent years, nearly all models for Question Generation were deep learning-based, particularly since the emergence of Seq2Seq [123]. Afterwards, a huge breakthrough in the whole field of NLP came with the emergence of Transformer-based models, particularly with the introduction of BERT [36]. Transfer learning is another sub-domain for which transformer-based approaches have been very relevant since years, for example Text-to-Text Transfer Transformer, or T5 [102]. T5 can not only easily be used for Question Generation, the performance of models based on it are also on par with other approaches.

In this section, we investigate if generating meaningful questions out of an input text could possibly imply a good text understanding and if it would be possible to leverage on this for other downstream tasks. We identified an applications with domain-specific texts to which these models could be particularly helpful. It consists of identifying the most central parts of rich textual descriptions of silk fabrics ²¹.

Given this very domain-specific dataset, the following question arises with respect to a possible leverage through question-answer generation models:

- How does such an approach compare with Zero-Shot classification for extracting specific type of information (e.g. about silk fabrics)?

The remainder of the section is structured as follows: In Section 4.4.1, we present our experiments and results for the task of silk information extraction. Finally, we conclude and outline some future work in Section 4.4.2.

4.4.1 Question generation for key information extraction from texts about silk fabrics

Dataset

The SILKNOW knowledge graph [112], which has already been partly enriched through mostly entity linking based on an expert-designed thesaurus about silk fabric concepts, consists of both structured data in form of precise explicit values and unstructured data in form of longer textual descriptions. One way of further enriching this knowledge graph can be done in form of text classification by training a model to associate existing categorical values with text descriptions in order to predict missing categorical values.

Although supervised training is usually giving the best results for such tasks, there are certain constraints when it comes to very domain-specific data, such as the one of SILKNOW: pre-trained NER models are trained on text corpora that are too general, e.g. Wikipedia, which makes them less powerful when it comes to disambiguation of precise terms related to silk fabrics. Another issue is the limited size of the dataset available to train our own models. Even if we conduct several experiments for each different properties, like one for all different material and technique values respectively, and further group possible values, we still need balanced training data for usually each of more than 3 labels per experiment.

For these reasons, we believe that alternatives or supplements need to be considered. With this section, we aim at exploring if using transformer-based question-(answer) generation models can lead to automatic highlighting of the most important parts of a rich textual description.

²¹<https://gitlab.eurecom.fr/schleide/qg4textunderstanding>

4.4 Using transformer-based QA and CQ systems for metadata predictions

The SILKNOW dataset used in this section consists of metadata about 1429 different objects making use of 7 different labels for silk material (Cotton, Wool, Textile, Metal Thread, Metal Silver Thread, Silver Thread, Gold Thread) and 955 objects with 6 unique labels (Damask, Embroidery, Velvet, Voided Velvet, Tabby, Muslin, Satin, Brocaded) for silk techniques. The whole knowledge graph consists of many more objects, but we decided to remove objects with more than one value per property. On the other hand, research with a classification or prediction setup has been motivated by the fact that many objects have no value at all. The code for this series of experiments can be found on GitLabe ²².

Method

Our approach consists in generating questions and answers from textual descriptions of silk made objects. For reasons of self-evaluation, we choose only objects with an associated property, such as the material, to see if this known categorical value is included in the generated output. To verify if this is the case, we both perform simple string matching and fuzzy string matching by first measuring an edit distance similar to the classical Levenshtein distance [75] between two tokens, the label and the tokenized version of the full output of one of the T5 models with an empirically defined threshold of 0.9.

Preliminary results have shown that this edit distance measurement is showing equal or better scores than calculate the semantic similarity after converting all input into word vectors with the most recent large English model of Spacy ²³, whose word vectors are trained with GloVe [100] on the Common Crawl ²⁴. This is why we do not include the latter results of the semantic similarity with word vectors in this section.

This dataset in this form is only a slight extension of one already used together with the model ZeSTe [111], which makes quantitative evaluation possible. ZeSTe [57] (or Zero-Shot Topic Extraction) uses ConceptNet and the nodes neighborhoods of the knowledge graph to compute similarity between the tokens of an input text and the target concept classifying the document. The candidate ranking of ZeSTe got hereby updated after prompting the language model BART [78] with a sentence related to the domain of the word (in this case, e.g. "silk textile"). We provide a comparison with both this updated and prompting-guided Zero-shot classification method as well as the results of ZeSTe itself on a very similar dataset. As a final baseline we provide the class distribution, which illustrates the multi-label classification setup.

To put the results into perspective, we compare our predictive scores with three baselines. Finally, we also qualitatively analyze if our approach has the ability to predict values that those other methods could not.

²²<https://gitlab.eurecom.fr/schleide/qg4textunderstanding>

²³https://github.com/explosion/spacy-models/releases/tag/en_core_web_lg-3.2.0

²⁴<https://nlp.stanford.edu/projects/glove/>

Quantitative Analysis

Table 4.30 contains the scores of our auto-evaluation experiments. We observe that both the scores obtained through the edit distance measurement and simple string matching in almost all cases beat the class distribution baseline, but are not coming very close to the two predictive performances of the Zero-Shot Classification models. The Multi-task model achieves the best performances between the three T5-based models, for both SILKNOW properties.

This might come from the added complexity of the multi-task model as it is more fine-tuned on the separate tasks of answer extraction, question generation and finally question answering. The single-task model is second best for materials, but behind end-to-end for the techniques condition. The end-to-end model is the only one that achieves a score lower than one of the class distribution baselines, but only when we apply simple string matching, which consistently yields worse results than edit distance measurement. On average the end-to-end model still produces output that scores mostly comparatively to the other models, despite it not producing any answers, but only questions.

Between all conditions we can also observe, that the scores for the SILKNOW Techniques are consistently higher, despite this condition having one more class and the baseline being accordingly lower. The reason for this may simply lie in the average length or quality of the original input texts that are ultimately coming from different museums. Potentially the difference could also stem from the semantic similarity or dissimilarity between the class labels inside of one condition. The two Zero-Shot models confirm this discrepancy between the data for the different properties as well.

Qualitative Analysis

For the SILKNOW dataset, we also investigate if the output of the question-answer generation models could be a supplement to the output of text classification models. For this, we compare if the output of the T5-based models explored in this study could be matched with the labels of objects for which the ZeSTe-based model could not predict them.

Table 4.31 shows some examples that are solely selected based on beating the prediction of our baseline model. In most cases, we still got proper English sentences that ask relevant questions. An exception is hereby the last row which shows the rather strange question "What scroll appears to have read 'Benedetto Ghalilei'?". We also have some examples of very technical question-answer pairs whose use might be quite limited, for example: "Question: How many threads per in? - Answer: 36-38". Nevertheless, this might be quite an interesting detail which is not yet explicitly available in the knowledge graph. An automatic extraction of it might be complex, but this output still emphasizes the highlighting abilities of the model which could further be leveraged in the future.

4.5 Predicting museum metadata gaps through classification

Next to that we have several examples of locations or time-spans in some of the answers that could easily be linked, like "Kerman, Iran", "India" or "17th century". Linking of such further properties can be used with several other applications of SILKNOW, for example a spatio-temporal map that an end-user could use. Given future expert evaluation we see great potential here for use of these question and answers not only for further enrichment of the knowledge graph, but directly in some web applications.

As a common pattern we could observe that the T5-based models described in this section were almost never better than the stated Zero-Shot methods at predicting one of the two respective majority labels (Textile for material, and embroidery for technique), but did occasionally so for some of the smaller labels. For example the one displayed in table. We could not find out a proper reason for this, but this also hints at a potentially useful complimentary function of these models next to other better performing (Zero-shot) classification methods for a future work.

4.4.2 Conclusion and Future Work

In this section, we explored the use of three different T5-based question(-answer) models for both information extraction and text summarization problems. We conducted a series of experiments with a very domain-specific dataset. We provided a quantitative analysis as the dataset provides labels that can be considered ground truth for the content of the generated output. Finally, we provided a qualitative analysis which is also directed at future applications.

For the application on a dataset with metadata about historical European silk fabric we can conclude that the output of the question(-answer) models is not directly surpassing the state-of-the-art of zero-shot classification, but showcases promising highlighting abilities when it comes to producing questions or answers from relevant text sections which goes beyond random selection. As far as we can qualitatively analyze without expert confirmation, we consider the output in most cases to be grammatically correct and useful. We can also observe that despite the lower classification performance of these models, some results appear to be complimentary to the main baseline model: some labels were matched that could not be predicted before. We therefore believe that further investigations into combining zero-shot classification approaches and question generation models bears high potential for use cases such as this one. We want to study if we can leverage question generation and question answer models even more for information extraction problems in the future.

4.5 Predicting museum metadata gaps through classification

Integrating heterogeneous, multilingual and domain-specific data is challenging, but doable, also thanks to many established tools and techniques. Such a process, however, often also

Chapter 4. Predicting metadata gaps

highlights the shortcomings of the original museum records and eventually also the initial digitization process of the respective metadata: From simply missing categorical information like the year to constraints of simple string matching-based linking like typos or inconsistencies. Also, in many cases an important information, like the production place, is hidden inside a rich textual description, but has never been explicitly annotated word by word.

The field of Natural Language Processing is nowadays advanced enough to offer many promising techniques to alleviate such issues. In the context of this thesis, we attempted several different of those, most of them from the sub-field of unsupervised or zero-shot classification. These different approaches can be further split into different ways of making the supervised training of a model unnecessary, most of them rely heavily on pre-trained language models. The reason for such a direction is mostly motivated by the high amount of classes combined with a high imbalance of those and only little data points in the data that we can export from our knowledge graph. Nevertheless, we also experimented with supervised approaches, but they require a reduction of classes.

Having explored ways to enrich our data by predicting missing information in this part of the thesis, the next and final chapter will be about the exploration of our knowledge graph.

4.5 Predicting museum metadata gaps through classification

Table 4.9: F1 scores (F1) and overall accuracies (OA) of the image classifier obtained by minimizing the Softmax loss (eq. 4.1) and the focal loss (eq. 4.2) both for the validation and the test sets (evaluated per record). Δ gives the difference between the quality metrics achieved using the focal loss and the softmax loss.

		validation set		test set	
Variable		F1 [%]	OA [%]	F1 [%]	OA [%]
Focal loss	place	49.2	62.5	47.0	63.1
	timespan	58.4	63.8	57.5	64.5
	technique	75.5	79.0	77.9	80.2
	material	52.2	80.6	51.2	80.6
	average	58.8	71.5	58.4	72.1
Softmax loss	place	48.2	61.0	47.2	62.2
	timespan	56.0	64.4	54.2	64.9
	technique	72.2	75.8	74.0	76.8
	material	45.0	79.4	43.4	80.7
	average	55.4	70.2	54.7	71.2
Δ	average	3.4	1.3	3.7	0.9

Table 4.10: Hyperparameter tuning. Hyperparameters, the investigated range range, and the value chosen to be the best in 50 random trials according to macro-F1 evaluated on the validation set.

Hyperparameter	Range	Best
Batch Size	4, 8, 32 64	64
Learning Rate	[1e-6, 1e-4]	3e-5
Weight Decay Coefficient	[0.0, 0.05]	0.02
Total Epochs	[4, 20]	12

Table 4.11: F1 scores (F1) and overall accuracies (OA) obtained in the multitask experiment both for the validation set and the test set (text classification)

Variable	validation set		test set	
	F1 [%]	OA [%]	F1 [%]	OA [%]
place	93.6	93.6	92.7	92.2
timespan	85.3	90.6	82.7	88.7
technique	83.5	86.4	84.0	86.9
material	79.7	85.6	77.3	83.9
average	85.5	89.1	84.2	87.9

Chapter 4. Predicting metadata gaps

Table 4.12: Hyperparameter tuning for the multimodal classifier: hyperparameters, the investigated range of values (Range) and interval of the search, and best values for each task, chosen by grid search according to macro-F1 evaluated on the validation set.

Hyperparameter	Range	Interval	place	timespan	technique	material
colsample_bytree	[0.6, 1.0]	0.2	0.8	0.8	0.6	0.8
gamma	[0.0, 0.4]	0.2	0.4	0.2	0.2	0.0
learning_rate	[0.1, 0.3]	0.1	0.3	0.3	0.3	0.2
max_depth	[2, 8]	2	4	4	4	8
min_child_weight	[1, 4]	1	1	2	4	2
n_round	[100, 500]	100	100	100	100	500
subsample	[0.6, 1.0]	0.2	0.6	1.0	0.6	0.8

Table 4.13: F1 (F1) and overall accuracies (OA) obtained in the experiment both for the validation set and the test set (tabular classification).

Variable	validation set		test set	
	F1 [%]	OA [%]	F1 [%]	OA [%]
place	47.9	62.4	46.2	61.9
timespan	57.4	65.1	58.6	67.6
technique	68.6	74.2	68.3	73.0
material	50.7	82.1	49.4	82.1
average	55.4	70.0	55.6	71.2

Table 4.14: Tabular classifier: feature importance per task (information gain).

Target Variable	Feature				
	museum	place	timespan	technique	material
place	0.49	-	0.20	0.12	0.19
timespan	0.41	0.31	-	0.16	0.12
technique	0.40	0.29	0.17	-	0.14
material	0.39	0.21	0.16	0.24	-

4.5 Predicting museum metadata gaps through classification

Table 4.15: Hyperparameter tuning for the multimodal classifier: hyperparameters, the investigated range of values (Range) and interval of the search, and best values chosen by grid search according to macro-F1 evaluated on the validation set. These hyperparameters apply to the multimodal classifier using the complete set of input modalities as shown in Figure 4.6.

Hyperparameter	Range	Interval	place	timespan	technique	material
colsample_bytree	[0.6, 1.0]	0.2	0.6	0.6	0.6	0.6
gamma	[0.0, 0.4]	0.2	0	0	0.4	0.4
learning_rate	[0.1, 0.3]	0.1	0.3	0.2	0.3	0.3
max_depth	[2, 8]	2	8	8	8	8
min_child_weight	[1, 4]	1	4	2	1	4
n_round	[100, 500]	100	100	100	100	100
subsample	[0.6, 1.0]	0.2	0.6	0.6	1.0	0.6

Table 4.16: F1 scores (F1) and overall accuracies (OA) on the test set of the multimodal classifier with and without the raw tabular data as additional input. The last two columns give the differences between OA and F1 scores of the two variants (Δ OA [%] and Δ F1, respectively).

Variable	without raw tabular data		with raw tabular data		Δ F1 [%]	Δ OA [%]
	F1 [%]	OA [%]	F1 [%]	OA [%]		
place	77.0	78.9	77.3	80.1	0.3	1.2
timespan	73.2	79.8	74.9	81.1	1.7	1.3
technique	82.1	84.1	83.3	85.3	1.2	1.2
material	61.4	85.2	66.8	85.6	5.4	0.4
average	73.4	82.0	75.6	83.0	2.2	1.0

Table 4.17: Feature importance (gain) for the multimodal classifier per modality for all tasks, both with and without raw tabular data.

Variable	Input Modality			Input Modality		
	without raw tabular data			with raw tabular data		
	Text	Image	Tabular	Text	Image	Tabular
place	0.42	0.24	0.34	0.36	0.08	0.25
timespan	0.43	0.25	0.32	0.35	0.16	0.19
technique	0.15	0.36	0.48	0.08	0.26	0.32
material	0.49	0.22	0.23	0.27	0.13	0.18

Table 4.18: Multimodal classifier feature importance (gain) per task per tabular data feature.

Target Variable	Feature				
	museum	place	timespan	technique	material
place	0.1	-	0.07	0.09	0.05
timespan	0.16	0.08	-	0.07	0.07
technique	0.21	0.04	0.04	-	0.04
material	0.17	0.09	0.07	0.08	-

Chapter 4. Predicting metadata gaps

Table 4.19: Mean F1 scores (F1) and overall accuracies (OA) of the different classifiers evaluated on the entire test set. Samples for which a modality was missing are considered as errors for the corresponding modality-specific classifier. In case of the multimodal classifier, the numbers are identical to those for the variant considering raw tabular data in Table 4.16.

Classifier	image		text		tabular		multimodal	
	F1	OA	F1	OA	F1	OA	F1	OA
Variable	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
place	38.0	46.8	64.5	42.1	46.2	61.9	77.3	80.1
timespan	49.2	45.6	54.4	45.1	58.6	67.6	74.9	81.1
technique	73.5	70.5	39.6	25.2	68.3	73.0	83.3	85.3
material	46.5	67.5	37.1	21.2	49.4	82.1	66.8	85.6
average	51.8	57.5	48.9	33.4	55.6	71.2	75.6	83.0

Table 4.20: Average F1 scores of the multimodal classifier using input modalities, with and without the raw tabular data (average over all tasks).

Input Modality	F1 score [%]	F1 score [%]
	with raw tabular data	without raw tabular data
image + text	75.1	70.25
image + tabular	63.4	57.8
text + tabular	70.6	69.9

Table 4.21: Examples of misleading text descriptions. Emphasis added to highlight the misleading text snippets.

Text Description (Snippet)	Predicted	True
includes the motifs found on seventeenth-century coverlets but must have been made in the early eighteenth century (...) "The Commercial Embroidery of Gujerat [sic] in the Seventeenth Century "	XVII	XVIII
derived from engravings after Maarten de Vos which first appeared in Gerard de Jode's 1579 illustrated bible	XVI	XVII
Center text reads " Vole vole mon coeur! "	FR	GB
Depiction from the Italian poem	IT	GB

4.5 Predicting museum metadata gaps through classification

Property	Concept	Nb of Objects	
		ES	EN
Material	Vegetal Fibre	225	-
Material	Gold Thread	34	125
Material	Silver Thread	10	176
Material	SUBTOTAL	269	301
Technique	Damask	512	25
Technique	Brocaded	-	190
Technique	Tabby	-	37
Technique	└ Muslin	2	37
Technique	└ Louisine	11	-
Technique	Satin	228	93
Technique	Embroidery	66	343
Technique	Velvet	111	236
Technique	└ Voided Velvet	-	1
Technique	└ Façonne cut velvet	102	-
Technique	└ Ciselé velvet	26	-
Technique	└ Plain cut velvet	23	-
Technique	└ Cut velvet	8	-
Technique	SUBTOTAL	1089	955
Depiction	Floral Motif	965	-
Depiction	└ Bunch	39	-
Depiction	└ Rose	2	-
Depiction	└ Fleur de lis	8	-
Depiction	Vegetal Motif	205	-
Depiction	└ Leaf	36	-
Depiction	└ Thistle	35	-
Depiction	└ Vine	4	-
Depiction	Geometrical Motif	125	-
Depiction	└ Rhombus	6	-
Depiction	SUBTOTAL	1425	-
All	TOTAL	3425	

Table 4.22: Number of objects exported for each property (material, technique, depiction)

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSC	ES	Vegetal Fiber	2.7%	100%	5.2%	6
ZSC	ES	Gold	70.6%	10.3%	17.9%	234
ZSC	ES	Silver	50.0%	17.2%	25.6%	29
ZSC	EN	Gold	96%	57.1%	71.6%	210
ZSC	EN	Silver	49.9%	95.5%	64.4%	91
CNN - Image	EN/ES/FR/IT	Vegetal Fiber	36.8%	24.1%	29.2%	
CNN - Image	EN/ES/FR/IT	Metal	32.1%	36.4%	34.1%	
CNN - Text	EN/ES/FR/IT	Vegetal Fiber	79.0%	81.0%	80.0%	229
CNN - Text	EN/ES/FR/IT	Metal	76.0%	61.0%	68.0%	125

Table 4.23: Results for the material property across approaches

Chapter 4. Predicting metadata gaps

Property	KG Concept	ConceptNet
Material	Vegetal Fibre	/c/es/vegetal
Material	Gold Thread	/c/es/oro, /c/en/gold
Material	Silver Thread	/c/es/plata, /c/en/silver
Technique	Damask	/c/es/damasco, /c/en/damask
Technique	Embroidery	/c/es/bordado, /c/en/embroidery
Technique	Velvet	/c/es/terciopelo, /c/en/velvet
Technique	Voided Velvet	/c/en/velvet
Technique	Cut Velvet	/c/es/terciopelo
Technique	Plain Cut Velvet	/c/es/terciopelo
Technique	Façonne Cut Velvet	/c/es/terciopelo
Technique	Ciselé Velvet	/c/es/terciopelo
Technique	Tabby (silk weave)	/c/en/tabby
Technique	Louisine	/c/es/tafetán
Technique	Muslin	/c/es/tafetán
Technique	Satin (Fabric)	/c/es/raso, /c/en/satin
Technique	Brocaded	/c/en/brocaded
Depiction	Vegetal Motif	/c/es/planta
Depiction	Vine	/c/es/planta
Depiction	Thistle	/c/es/planta
Depiction	Leaf	/c/es/planta
Depiction	Floral Motif	/c/es/flor
Depiction	Fleur-de-lis	/c/es/flor
Depiction	Rose	/c/es/flor
Depiction	Bunch	/c/es/flor
Depiction	Geometrical Motif	/c/es/geometría
Depiction	Rhombus	/c/es/geometría

Table 4.24: Mapping between the concepts used in our knowledge graph and ConceptNet

4.5 Predicting museum metadata gaps through classification

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSC	ES	Damask	6.8%	63.6%	11.2%	55
ZSC	ES	Embroidery	89.4%	6%	12.3%	987
ZSC	ES	Tabby	0%	0%	0%	6
ZSC	ES	Velvet	10.4%	100%	6.6%	28
ZSC	ES	Satin	3.5%	61.5%	18.8%	13
ZSC	EN	Damask	52.0%	33.3%	40.6%	39
ZSC	EN	Embroidery	32.4%	86.7%	47.1%	128
ZSC	EN	Velvet	61.6%	50.0%	55.2%	292
ZSC	EN	Tabby	1.5%	50.0%	2.9%	2
ZSC	EN	Satin	77.4%	28.5%	41.6%	253
ZSC	EN	Brocaded	40%	31.5%	35.3%	241
CNN - Image	EN/ES/FR/IT	Damask	70.3%	68.9%	68.9%	
CNN - Image	EN.ES/FR/IT	Embroidery	83.9%	83.2%	83.6%	
CNN - Image	EN/ES/FR/IT	Tabby	17.4%	30.8%	22.3%	
CNN - Image	EN/ES/FR/IT	Velvet	70.1%	67.8%	68.9%	
CNN - Text	EN/ES/FR/IT	Damask	98%	90%	94%	135
CNN - Text	EN/ES/FR/IT	Embroidery	96%	98%	97%	230
CNN - Text	EN/ES/FR/IT	Velvet	95%	84%	89%	62

Table 4.25: Results for the technique property across approaches

Approach	Language	Class	Precision	Recall	F1-score	Support
ZSC	ES	Flower	99.3%	72.1%	83.5%	1397
ZSC	ES	Plant	2.2%	26.1%	4.0%	23
ZSC	ES	Geometry	3.1%	100%	5.9%	4
CNN - Image	ES/FR	Flower	89.9%	88.8%	89.3%	
CNN - Image	ES/FR	Plant	45.1%	38.1%	41.3%	
CNN - Image	ES/FR	Geometry	35.8%	50.0%	41.3%	

Table 4.26: Results for subject depiction property across approaches

Chapter 4. Predicting metadata gaps

Property	SILKNOW Concept	ConceptNet
Material	Cotton	/c/en/cotton
Material	Wool	/c/en/wool
Material	Textile	/c/en/textile
Material	Metal thread	/c/en/metal
Material	Metal silver thread	/c/en/silver
Material	Silver thread	/c/en/silver
Material	Gold thread	/c/en/gold
Technique	Damask	/c/en/damask
Technique	Embroidery	/c/en/embroidery
Technique	Velvet	/c/en/velvet
Technique	Voided Velvet	/c/en/velvet
Technique	Tabby (silk weave)	/c/en/tabby
Technique	Muslin	/c/en/tabby
Technique	Satin (Fabric)	/c/en/satin
Technique	Brocaded	/c/en/brocaded

Table 4.27: Mapping between the concepts used in the SILKNOW knowledge graph and ConceptNet (ProZe and ZeSTE)

Datasets	20 Newsgroup		AG News		BBC News	
	Accuracy	Weighted Avg	Accuracy	Weighted Avg	Accuracy	Weighted Avg
ZeSTE	63.1%	63.0%	69.9%	70.3%	84.0%	84.6%
Entail	46.0%	43.3%	66.0%	64.4%	71.1%	71.5%
ProZe-A	62.7%	62.8%	68.5%	69.1%	83.2%	83.7%
ProZe-B	64.6%	64.6%	69.0%	69.6%	84.2%	84.8%

Table 4.28: Prediction scores for the news datasets (the top score in each metric is emboldened).

Datasets	Silk Material		Silk Technique		Crime aspects		Crisis situations	
	Accuracy	Weighted Avg	Accuracy	Weighted Avg	Accuracy	Weighted Avg	Accuracy	Weighted Avg
ZeSTE	34.3%	39.0%	46.9%	47.2%	31.2%	32.3%	46.3%	45.8%
Entail	29.0%	33.3%	64.0%	65.8%	43.7%	43.7%	46.7%	48.1%
ProZe-A	39.0%	40.1%	50.8%	57.6%	36.3%	37.6%	50.1%	49.7%
ProZe-B	37.4%	41.7%	48.5%	48.7%	29.8%	31.1%	50.1%	49.8%

Table 4.29: Prediction scores for the domain-specific datasets (the top score in each metric is emboldened).

4.5 Predicting museum metadata gaps through classification

Model	Measurement	SILKNOW Materials	SILKNOW Techniques
Single-Task T5	Edit Distance	19.40%	25.10%
Multi-Task T5		24.50%	28.80%
End-to-End T5		17.40%	26.30%
Single-Task T5	String Matching	15.30%	19.80%
Multi-Task T5		17.30%	23.00%
End-to-End T5		12.50%	20.00%
Prompt-guided ZS Classification	Accuracy	39.00%	50.80%
ZeSTE*		34.3%	46.9%
Baseline	Class Distribution	14.00%	12.50%

Table 4.30: Auto-evaluation scores based on matches between the target label and the generated question(-answers). Comparison with the label prediction accuracy of two Zero-Shot classification methods that have been performed on the same dataset (*For ZeSTE the results of its application on a minimally different, but comparable dataset are stated here). The baseline is representing the class distribution.

Matched Label	Prompt-guided ZS classification	Property	Selected Output
Single-Task T5			
Cotton	Wool	Material	{'answer': 'Kerman, Iran', 'question': 'Where was the 'Vase Carpet' lattice design located?'}, {'answer': 'silk', 'question': 'Along with cotton weft and wool knotted pile, what textile is used in Persian carpets?'}, {'answer': 'wool', 'question': 'What type of fiber is the carpet made of?'}
Velvet	Brocaded	Technique	{'answer': 'red', 'question': 'What color is the cut and uncut velvet?'}
Multi-Task T5			
Wool	Silver	Material	{'answer': '36-38', 'question': 'How many threads per in?'}, {'answer': '16', 'question': 'How many knots per in?'}, {'answer': 'wool', 'question': 'What is the Pile made of?'}
Muslin	Embroidery	Technique	{'answer': 'embroidered muslin', 'question': 'What is the girdle made of?'}, {'answer': 'India', 'question': 'In what country is the girdle of muslin embroidered with silk and silver threads?'}, {'answer': '17th century', 'question': 'When was the girdle of embroidered muslin made?'}
End-to-end T5			
Cotton	Silver	Material	'How many threads per inch does white cotton have?', 'How many shoots of weft do gold-coloured cotton and gold coloured silk have per inch?', 'What color are the lilies in the center of the present gragment?', 'Where do the white lily veins meet on the horizontal plane?', 'Which leaves form a square frame?'}
Velvet	Brocaded	Technique	'What is the coat of arms of the Galilei family of Florence represented by?', 'How many rungs are under the cross in the center of the velvet?', 'What scroll appears to have read "Benedetto Ghalilei?"', 'What may have been used in a set of ecclesiastical vestments for a family chapel?'

Table 4.31: Two generated output texts per T5-based model. All examples represent cases in which the target label could be matched with the output and the Prompt-guided ZS classification method used on the same dataset predicted a wrong label.

Chapter 5

Exploring the European Silk Heritage

Many cultural heritage domains consist of knowledge that is not broadly known by the public. Despite many objects having been digitized, even experts still struggle to find what they are looking for in online catalogs. The European production of silk fabrics is an example of one such domain. Already relatively obscure to the public, many descriptions and images of objects do exist in a digitized form, but are uploaded by many museums across the globe, in individual formats. They often give public access to the images and metadata of such silk objects through APIs or simply their websites, but originally there has been no worldwide harmonization and integration effort. Therefore, it is very hard for both the public, historical experts and industry (e.g. fashion) to access this knowledge.

Following up on how we developed a knowledge graph and explored how to fill metadata gaps, the following chapter will describe our efforts of making all the data easily accessible and to make it possible to further explore it even for non-experts. Section 5.1 presents our exploratory search engine ADASilk and section 5.2 is about a series of experiments for the effective creation of an Image Retrieval function based on the knowledge graph and domain expert-rules and has been published and presented as follows:.

- Thomas Schleider, Raphaël Troncy, Thibault Ehrhart, Mareike Dorozynski, Franz Rotensteiner, Jorge Sebastián and Georgia Lo Cicero. **Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules**. In 3rd workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC) co-located with ACM Multimedia, 2021. Online. **Best Paper Award**.

5.1 ADASilk

An exploratory search engine is a type of search engine that is typically suited for browsing collections of heterogeneous items and when the user has no precise search query in

mind, but rather wants to discover collections in a serendipitous way. Exploratory search engines typically make use of facets that enable to filter search results according to a number of dimensions. In the case of SILKNOW, those facets will typically be the production place and time of the silk fabrics, the material used, the technique employed or the main subject being depicted [126]. The SILKNOW exploratory search engine is named after Ada Lovelace (1815-1852), the mathematician who anticipated some of the main features of modern computing some 100 years before its advent. In her notes, she wrote that such a computation machine weaves algebraic patterns, just as the Jacquard loom weaves flowers and leaves. The application is available online at <https://ada.silknow.org>.

5.1.1 A user-friendly interface for non-experts

The exploratory search engine is a web application. We describe below the home page (or landing page), vocabularies pages that make use of the SILKNOW thesaurus, and the advanced search page that make use of facets. The primary goal of the exploratory search engine is to enable users to search for silk fabrics using complex queries while also discovering collections of objects that are today scattered across numerous museum web sites. Hence, one of the main view of ADASilk enables to add filters and to dynamically observe the results that match those filters as either a grid view or spatio-temporal map view. Clicking on individual objects enables to see the full set of metadata that has been collected for this museum object.

Homepage The home page (or landing page) contains a simple search box in the center of the page which allows the user to search for silk-related objects across the museums that have been integrated in the SILKNOW Knowledge Graph (Figure 5.1). When the user enters a search term, the exploratory search engine executes a SPARQL query with a REGEX filter in order to select all museum objects that have a label or a title that partially matches the search terms. The results are shown in an autocomplete box (Figure 5.2) which directly leads to the detailed view of this object. The home page provides also to switch of languages as ADASilk has been localized.

Advanced Search The advanced search page contains a faceted search engine which allows users to browse for objects within the SILKNOW Knowledge Graph. The sidebar on the left side contains facets (or filters). Each facet generates an extra condition to the main SPARQL query used for searching yielding a subset of results being displayed. (Figure 5.3)

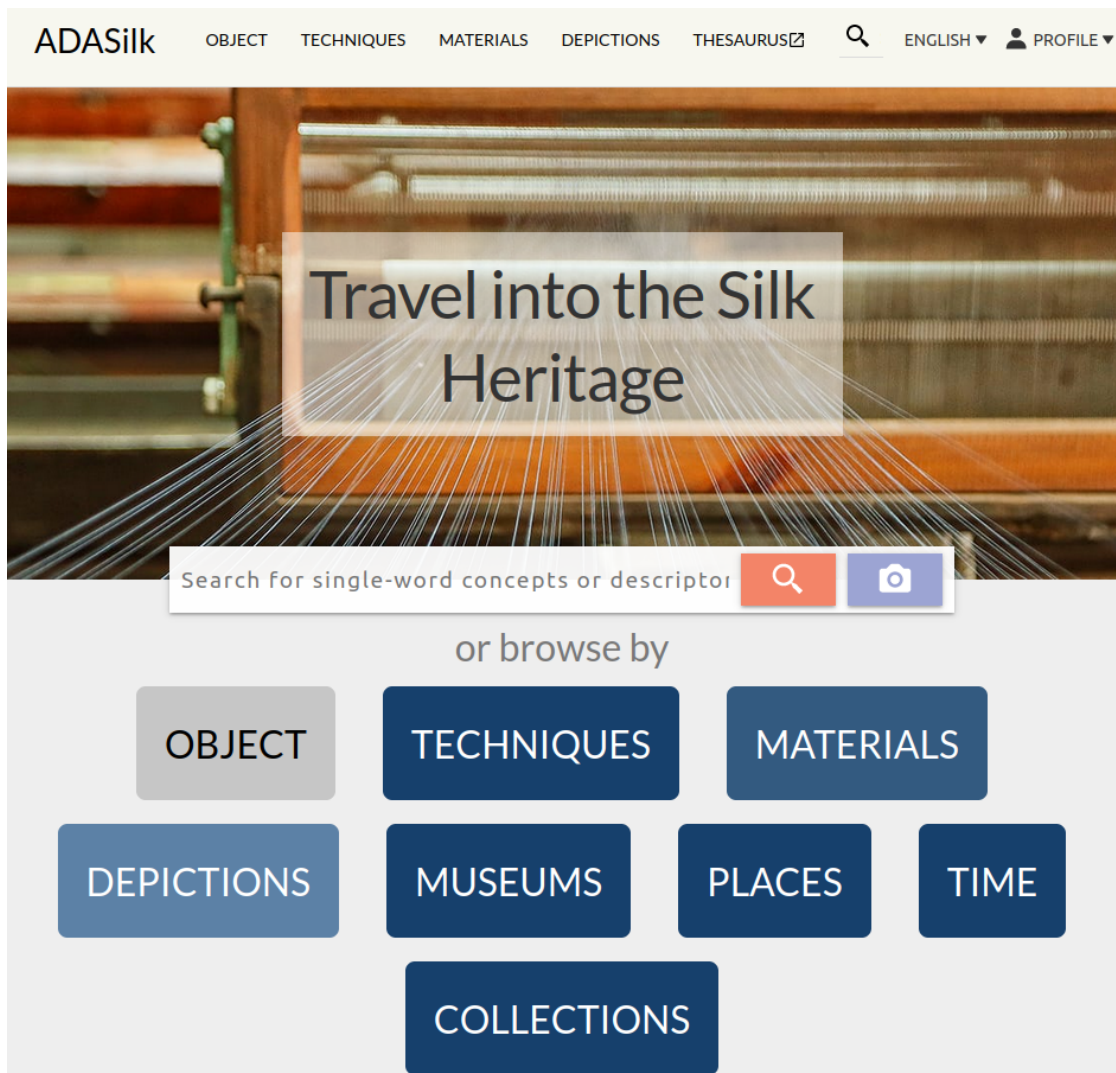


Figure 5.1: The ADASilk home page (<https://ada.silkknow.org/>): the user is invited to either enter a textual search term or to browse the collections of objects using shortcuts for the most common weaving techniques, materials and depicted subjects. The image-based search can be used through the camera icon

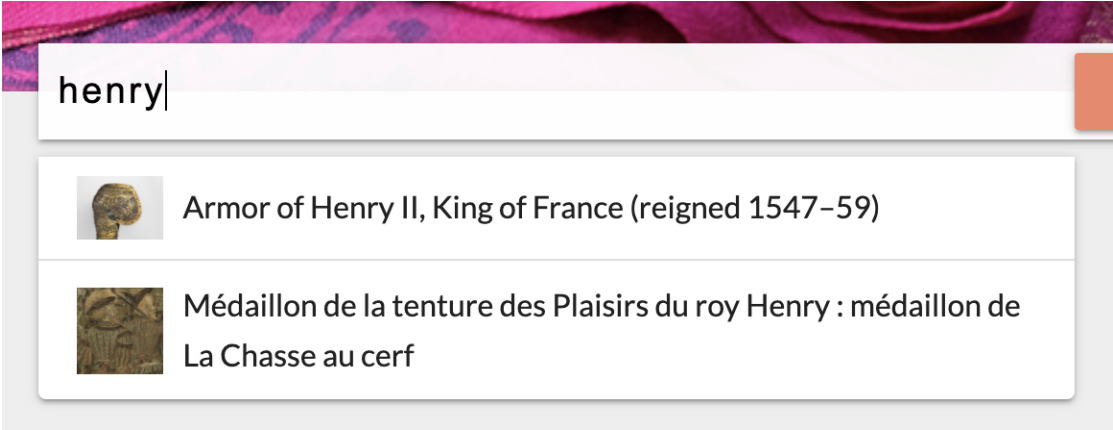


Figure 5.2: An example of a search using textual query terms with auto-completion.

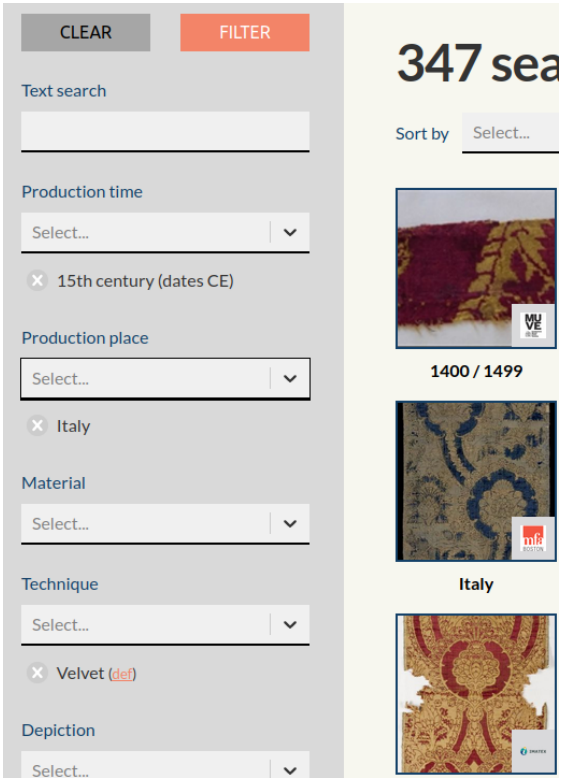


Figure 5.3: The advanced search page enables the user to refine a search using facets. Multiple values can be used for each facet.

5.1.2 Integration of research and engineering work of SILKNOW project partners

Virtual Loom

Virtual Loom is an application that deals with the 3D virtual representation of historical silk fabrics at the yarn level. Silk fabrics have specific characteristics, as they are nearly flat objects and very fragile. The documentation of their visual appearance has been traditionally done by means of imaging devices (e.g., RGB cameras, digital microscopes, etc.). However, within these devices, only the surface of the objects is documented, so the complex internal structure composed of a variety of yarns and their interlaces, remains undiscovered. To deal with this, in Virtual Loom we produce 3D models of silk fabrics at the yarn level, with the minimum information of an image as input data.

Spatio-Temporal map

STMaps is a visual tool implemented in Unity (Unity 2020.2.8.f1 is used to develop the last version of the tool). The use of Unity allows developing a cross-platform application with state-of-the-art graphics. The different releases of the tools are generated like a WebGL plugin. The WebGL technology allows the integration of a software module within a web application, the communication between the plugin and the web application is performed by invoking Javascript methods. STMaps allows the spatio temporal visualization of knowledge graph data. This software uses and expands on the Visualization ontology (VISO) [7] work in order to define how the knowledge graph data is going to be visualized. The functionalities, configuration and the design of the communication protocol between ADASilk and STMaps are detailed in deliverable D5.5. The main functionalities of STMaps are:

- Visualization in a 2D/3D navigational environment where the spatio-temporal data of the Knowledge Graph is displayed. This is performed by showing a map, where the user can navigate on it, by zooming and moving to every part of the map.
- Filtering the data according to the different properties of the data shown on the map.
- Visualization of the relationships between the different objects being displayed.
- Getting additional information about a displayed object.
- Visualize and analyse the variation of the data over the temporal dimension, using different techniques.

5.1.3 Evaluation

ADASilk is an exploratory search engine built on the Knowledge Graph. It is important to note that this tool was evaluated in its beta version [48]. In fact, it was frozen for user evaluations in February 2021. The results presented here correspond to this version, although it is true that this evaluation analyzes the overall system. Due to the pandemic situation and the delays caused, we could not perform a second evaluation to compare data. The activities to conduct the evaluation phase were:

- Preparation of questionnaires: one specific questionnaire were prepared for the tool. It consisted of three parts: the System Usability Scale (SUS), general questions related to user experience and the application of the tool according to SILKNOW's Target Audience (TA), and an open question concerning the respondents' general opinion about it.
- Selection of the evaluators: assessment sessions were conducted from January 1 to March 15, 2021.
- Preparing the training materials: both the questionnaire and the training session were loaded onto the SILKNOW website (<https://silknow.eu/index.php/evaluation/>).
- Submitting the questionnaire: it was disseminated via social media and SILKNOW's newsletter. In addition, some specific evaluations were carried out by our project partners University of Palermo, Jožef Stefan Institute and University of Valencia.
- Collecting and evaluating the results (SUS score): the evaluation lasted from mid- December 2020 to mid-March 2020. It was designed to be online and therefore not slowed down by the current health situation due to COVID-19. The University of Palermo conducted the analysis regarding the SUS score.
- Evaluating the results based on the joint strategies prepared by the University of Palermo. We took the cleaned data that the University of Palermo used in their joint strategy after deleting clearly inconsistent responses. These responses were recorded in Excel files, and the data was later uploaded and analysed in the SPSS v. 12.3 software package. Descriptive analyses were made of all the responses, without differentiation by audience, in order to have an overall view of each of the questions. Secondly, the responses were compared in three large audience blocks: Information and Communications Technologies, Cultural Heritage, and finally, the remaining audiences that answered the questionnaires but whose volume was too small to be analysed individually (namely media, tourism and Social Sciences and Humanities education).

As previously mentioned, the second group of questions aimed to investigate how much the user liked the tool. As can be seen in Table 5.1, the first 5 questions were asked using a Likert

scale, with 1 being “strongly disagree” and 5 “strongly agree”. The final one is an open question that needed to be codified; this was done by grouping similar responses, and following the so-called “closing open-ended questions” process, which responds to inductive coding. We manually coded the responses in a flat frame, meaning that all the codes have the same level of specificity and importance [65]. Also, in this table we show the results of the 115 respondents to these questions; we show the percentage of strongly disagree and disagree (1 and 2 on the Likert scale); the percentage of agree and strongly agree (4 and 5 on the Likert scale), and indifferent (3).

As can be seen from this set of questions, the majority of respondents rated them positively, although the SUS score was not high. This can be explained by the difference between questions focused on usability and others focused on whether the user liked the tool, i.e., a tool can be liked but difficult to use. However, it is striking that while on the Likert scale opinions tend to be favourable, when it comes to sharing the tool respondents are fairly split between those who would (42.6%), those who prefer to stay in the middle of the scale (21.7%) and those who would not share them (35.7%). In terms of general opinions, respondents think that ADASilk and STMaps are useful (21.2%) and attractive (20%), nevertheless it is worth mentioning that 16.6% of the respondents found that either ADASilk or STMaps crashed at some point of the evaluation, and another 16.6% of them stated that STMaps were slowly executed.

Regarding usability and execution questions, most of them are positive. 77.4% of the respondents could find objects by filtering in ADASilk, while 60.9% thought that the objects were optimally shown in STMaps. As regards information and understanding, we found positive answers, as 61.7% thought that STMaps allowed us to better understand relationships between objects, and the information provided there complemented the information provided by ADASilk. Both the information and the concepts used in ADASilk were positively ranked, the first one being the most liked with 86.4% of the respondents thinking that the information provided was appropriate, and regarding concepts, 53.1% of the respondents found the concepts used in ADASilk are appropriate to their background. As regards the relationships used in STMaps, the most appreciated ones are the temporal visualization of objects, with 67.9% of the respondents ranking them positively, and the least liked was the segment relationships, with 20% of the respondents stating that they did not like them. Finally, the linear relationships of STMaps were appreciated by 58.3% of the respondents.

The third group of questions aims to let the user express his/her opinion about the actual possible integration of the tools into existing domains. Questions 1 and 2 were multiple-choice questions, while the rest were asked using a Likert scale, 1 being “strongly disagree” and 5 “strongly agree”. Table 5.2 shows these questions and the results.

As can be seen, in general, respondents in the cultural heritage sector believe that ADASilk

Chapter 5. Exploring the European Silk Heritage

Second group of questions	Results
I managed to finish the training	15.7% did not manage to finish the training, while 73% managed and 11.3% seemed indifferent.
I could find the objects by using filtering options in ADASilk	10.4% could not find objects with the filtering options in ADASilk, while 77.4% could and 12.2% seemed indifferent.
The objects were optimally shown in the Map	16.5% found that objects were not optimally shown in STMaps, while 60.9% thought they were optimally shown, and 22.6% seemed indifferent.
I liked the "linear" relationships visualization of objects on the Map	19.10% did not like the linear relationships visualization of the objects on STMaps, while 58.3% liked them and 22.6% seemed indifferent.
I liked the "segment" relationships visualization of objects on the Map	20% did not like the linear relationships visualization of the objects on STMaps, while 52.2% liked them and 27.8% seemed indifferent.
I liked the temporal visualization of objects on the Map	14.8% did not like the temporal visualization of the objects on STMaps, while 67.9% liked them and 17.4% seemed indifferent.
The execution time is appropriate	31.3% found that the execution time was not appropriate, while 50.5% found it appropriate, and 18.3% were indifferent.
Information provided is appropriate	9.6% thought that the information given was not appropriate, while 86.4% thought it was, and 22.6% seemed indifferent.
Thanks to the Map visualization, it was easier to understand the relationships of the different objects	21.7% did not find it easier to understand the relationships of the different objects on STMaps, while 61.7% did understand them and 16.5% seemed indifferent.
STMap could complement the information given in ADASilk	13% thought that STMaps did not complement the information given in ADASilk, while 66.9% thought they did complement the information and 33% were indifferent.
The concepts used in ADASilk are appropriate to my background	27% did not find concepts used in ADASilk appropriate to their background, while 53.1% found them appropriate and 20% seemed indifferent
ADASilk met my expectations	21.7% of the respondents found that ADASilk did not meet their expectations, while for 60.9% ADASilk met their expectations and 17.4% were indifferent.
STMap could complement the information given in ADASilk	13% of the respondents think that STMaps could not complement the information, 67.8% could, 20% were indifferent.
Thanks to the Map visualization, it was easier to understand the relationships of the different objects	20% of the respondents think that the map visualization did not make it easier to understand the relationships of the object, 60% did, while 20% were indifferent.
I will share these tools among my colleagues and friends	35.7% would not share these tools, while 42.6% would share them and 21.7% seem indifferent
General opinion	Easy: 12.2% Attractive: 20% Didactic: 3.3% Useful: 21.2% STMaps slowly executed: 16.6% ADASilk and/or STMaps crashed: 16.6% Complicated: 5.5% Other: 4.4%

Table 5.1: Set of general questions and results to find out the feeling of users as regards efficiency and reliability of the promised functionalities and how comfortable they feel.

Third group of questions	Results
Where do you think ADASilk could be applied in a museum?	Conservation: 60.9% Education: 22.6% Research: 78% Exhibition: 4.3% Marketing: 2.6% Tourism: 0.9% Other: 0.9%
Where do you think STMap could be applied in a museum?	Conservation: 34.8% Education: 40.9% Research: 15.7% Exhibition: 5.2% Tourism: 0.9% Other: 2.6%
ADASilk could be useful for historical and / or artistic research	2.6% would not use ADASilk for historic or artistic research, while 81.8% would apply it for research, and 13.9% seemed indifferent.
STMap is useful for historical and / or artistic research	5.2% would not use STMap for historic or artistic research, while 81.8% would apply it for research, and 13% seemed indifferent.
ADASilk could be useful to enhance textile learning	7% did not find ADASilk useful to enhance textile learning, while 66.1% found it useful, and 27% seemed indifferent.
ADASilk could be useful to find inspiration in traditional designs	7.8% did not find ADASilk useful to enhance textile learning, while 65.3% found it useful, and 27% seemed indifferent.
I would like to use STMap on a museum/city tour	13% would not use STMap on a museum / city tour, while 68.7% would apply it, and 18.3% seemed indifferent.

Table 5.2: Set of general questions added to the SUS questionnaire to find out opinions about the integration of ADASilk and STMaps in specific domains

can be used for research (70%) followed by conservation (60.9%). This is corroborated when respondents were asked whether they considered ADASilk useful for research, with the majority agreeing to this fact (81.8%). The fact that after research, respondents think that ADASilk is useful for conservation is evident when one of the fundamental steps to conserve a cultural object is to research on it by contextualizing it in time and space, and comparing it with similar objects, a functionality that is provided by ADASilk. On the other hand, 40.9% of respondents would use STMaps in a museum for education purposes, followed by 34.8. It is worth mentioning that only a 15.7% would use them for research, while in the next question “STMap is useful for historical and / or artistic research”, 88.1% would apply them for research. This might be because in a museum it is more useful to find similar elements, such as those offered by ADASilk, while for academic research it is also useful to locate them in time and space, especially in research related to art history where spatial connections are fundamental. In this regard, 66.1% believe ADASilk to be useful for enhancing textile learning. With regard to the creative industries, 65.3% of the respondents answered that ADASilk could be useful to find inspiration in traditional designs. Finally, 68.7% of the respondents would use STMaps on a museum tour.

Having analyzed the respondents' answers globally, in Figure 5.4 we show the most significant responses by specific domain. Usability questions on STMaps were the ones that had the most differences between TAs. Firstly, regarding whether the objects were optimally shown on STMaps, as shown in Figure 5.4, both ICT and CH sectors agreed, or strongly agreed, that they were optimally shown, specifically the CH sector with almost 80%, while only a 37.5% of the other audiences thought that they were optimally shown. This can be explained by the fact that the first two audiences are more accustomed to reading maps, especially the cultural heritage sector.

As with the linear, segment and temporal relationships shown in STMaps, these also vary depending on the TAs, with the CH sector favouring these relationships the most, as shown in Figure 5.5 This can be explained by the fact that for both the professional and academic cultural heritage domain, establishing relationships between objects are fundamental to developing consistent research. A cultural property is never isolated from its historical, social, geographical and cultural context. A map of these characteristics, where objects are shown grouped by time, allows professionals in the sector to more easily establish their context, both for cataloguing and researching purposes. In this sense, it is evident that the relationships that the sector liked most are the linear ones (given that they allow relationships to be easily found at the click of a button) and temporal relationships. However, the ICT sector does not appreciate this as much, in particular the segment relationships with only 47.4% of respondents saying they liked this possibility. None of the other sectors (media, tourism and SSH education) answered this question positively.

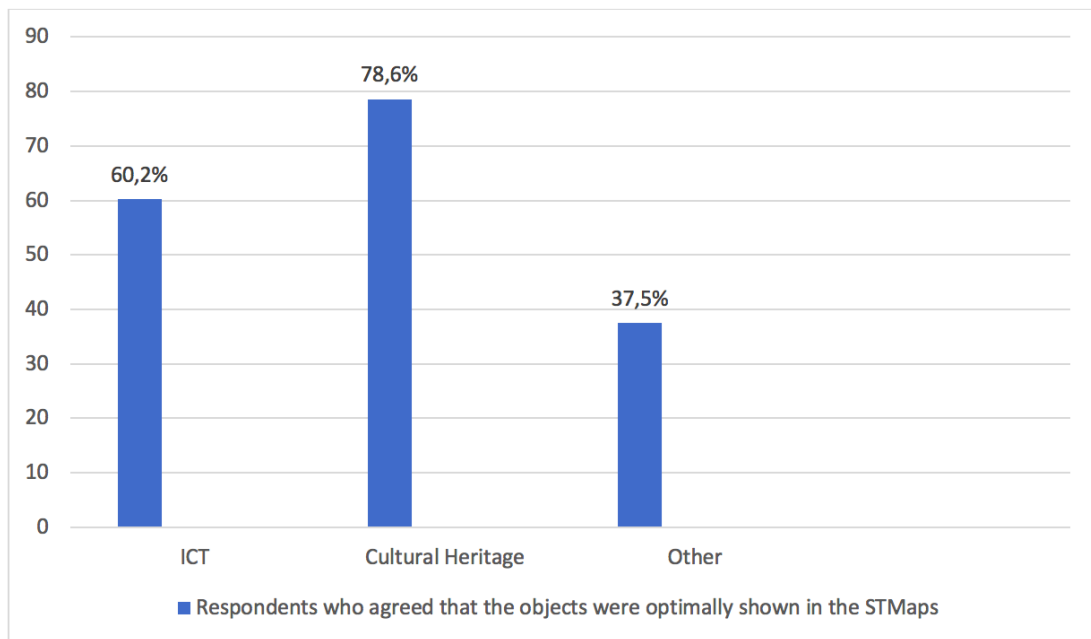


Figure 5.4: Respondents who agreed that the objects were optimally shown in STMaps. Percentages shown per TA; this graph allows us to compare them

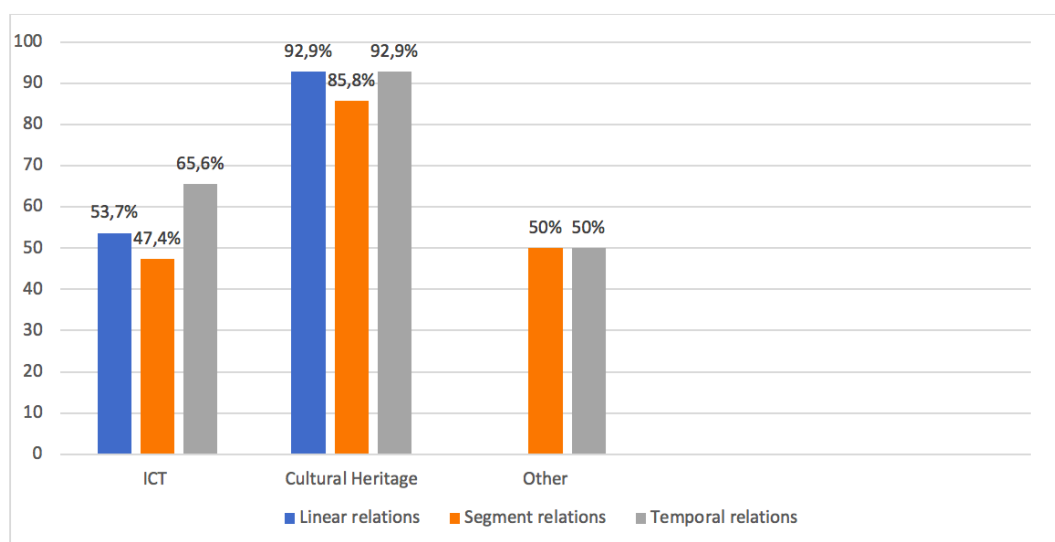


Figure 5.5: Comparison of STMaps concerning positive opinions among TAs. We show the positive answers to the questions. Percentages shown per TA; this graph allows us to compare them.

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

In this section, we describe how we make use of domain-expert rules and our knowledge graph (see figure 5.6 for the illustration of an excerpt) about European silk textiles to develop an image-based retrieval system to search and find related silk fabrics. Domain experts have helped to establish not only one of the definitions of image similarity based on formulated rules. They also designed an important part of an evaluation framework, which strongly shaped the assessment process towards good results for other actual experts, but more importantly the public and other users.

The aim of all developed tools is to make the history of the silk heritage accessible to everyone: From domain experts to enthusiasts and historians, it shall be possible to overcome physical distances and learn about, see images of and study textile artefacts held in many collections about European silk textiles. Using digital images and metadata makes it possible to make quick comparisons between different search results and to study even those fragile fabrics that are impossible to manipulate physically¹.

The remainder of this section is structured as follows. In Section 5.2.1, we detail our approach. We evaluate our method in Section 5.2.2 and we illustrate the integration of this component into an exploratory search engine in Section 5.2.3. Finally, we conclude and outline some future work in Section 5.2.4.

5.2.1 Approach

Domain Experts Rules for Image Similarity

Training of the method for image-based retrieval requires pairs of images (x_n, x_o) for which it is known whether they are similar or not, as it will be described in Section 5.2.1. One way of obtaining these data is to provide rules defining sets of images that should be similar and sets of images that should be dissimilar based on the content of the knowledge graph.

Such rules have been formulated by cultural heritage experts on the basis of an analysis of early image retrieval results. The rules are used to formulate SPARQL queries to the European silk textile knowledge graph. The results of these queries are transformed into a set T_{CE} of image pairs (x_n, x_o) with known similarity status. This state is binary, i.e. a pair can be similar according to the rules formulated by the experts or not. An overview of the rules that are used is given in Table 5.3.

These rules correspond to different aspects of similarity. Rule 1 corresponds to self-similarity

¹<https://github.com/silkknow/image-retrieval>

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

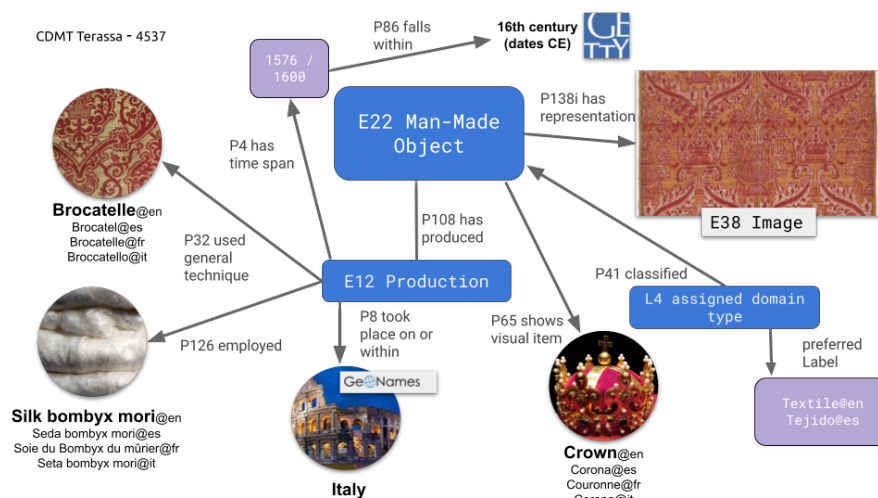


Figure 5.6: Excerpt of the knowledge graph: a textile object coming from the CDMT Terasse museum which has been produced in Italy in the 16th century, with the Brocatelle technique, using silk bombyx mori as material and showing the motif of a crown.

but is based on real images; for the image pairs affected by it, the loss in eq. 5.11 is equivalent to the one in eq. 5.8 (see Section 5.2.1). Rules 2-6 consider semantic properties of silk fabrics and can be seen as variants of semantic similarity. However, they only consider one or two semantic properties and disregard all of the others, and a binary concept of similarity is used. Finally, rule 7 considers the colour distribution and, thus, an aspect of visual similarity. This rule has been designed based on the results of a cluster analysis of colour feature vectors as follows. First, a large set of images has been exported from the knowledge graph. From these images, colour feature vectors $h(x_n)$ were computed in the way described below (Section 5.2.1) in the context of the colour similarity loss. After that, k-means clustering using $k = 30$ clusters was carried out. Some clusters, identified by their cluster indices in Tab. 5.3, were found to contain images which should be considered similar.

Originally, the domain experts have also defined one dissimilarity rule (i.e. a negative rule example). However, this rule did not produce a sufficient number of examples to be useful and only positive rules (i.e. pairs of images considered to be similar) have been considered. Consequently, as there are no dissimilar pairs, the loss in eq. 5.11 can only be used in combination with other loss functions in training. However, the principle could be expanded by additional rules to produce dissimilar pairs in the future.

Image-Based Retrieval

The goal of image retrieval is to use images as input to search for records in the knowledge graph. The core of the method is a convolutional neural network (CNN) [70] that converts

Nr.	Rule: Two images are supposed to be similar if ...
1	they belong to the same record in the knowledge graph, i.e. if they show the same fabric
2	they both correspond to objects in the Garín dataset and the production material is “graph paper” or the technique is “gouache sobre papel” or “gouache sobre papel milimetrado”
3	for both corresponding fabrics, the information about production material or production technique is “pile-on-pile velvet”
4	for both corresponding fabrics, the information about production material or production technique is “ciselé velvet”
5	for both corresponding fabrics, the information about production material or production technique is “ciselé velvet” and the sub-depiction is “pomegranate”
6	anywhere in the corresponding records, “plain fabric” is mentioned
7	for both corresponding fabrics, the colour feature vectors belong to the same colour cluster among clusters identified to be relevant for defining similarity because the corresponding objects were found to be similar according to the cultural heritage experts. These clusters are clusters 9 and 11 (saturated red), cluster 5 (blue), cluster 22 (blue damasks) and cluster 27 (green damasks), see also Section 5.2.1.

Table 5.3: Rules defined by the cultural heritage experts to define pairs of similar images. Nr: number of the rules.

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

images into feature vectors (*descriptors*) so that the descriptors of similar image pairs have a small Euclidean distance and descriptors of dissimilar pairs have a large one. Using the CNN to compute descriptors for all images in the knowledge graph and using a k-d tree as a spatial index, image retrieval itself becomes a k nearest neighbour (knn) search in the k-d tree [14]. The prerequisite of our method is a knowledge graph with records containing both images and annotations in one or more semantic variables. In this work, the knowledge graph containing images of silk fabrics with annotations in the five variables *production timespan*, *production place*, *production material*, *production technique* and *subject depicted* is used for that purpose. This section describes the CNN used to compute descriptors, focusing on the training procedure, which leverages the contents of the knowledge graph to generate training samples automatically without any human intervention.

CNN architecture

The CNN architecture used for image retrieval is based on [28]. Using an RGB image x scaled to 224 x 224 pixels as an input, a ResNet-152 [58] backbone is applied to generate a 2048-dimensional feature vector. This is followed by two fully connected (FC) layers with ReLU (Rectified Linear Unit) activations [91] of 1028 and 128 dimensions, respectively. The output of the last layer is normalized to unit length, resulting in the 128-dimensional feature vector $f(x)$ which represents the input image.

Training

For the ResNet-152 backbone, the pre-trained parameters from [58] are used and they remain constant during training. Thus, the only parameters that are determined in training are those of the two FC layers of the network. Training is based on standard stochastic minibatch gradient descent (SGD) using backpropagation for computing gradients. The training procedure requires pairs of images (x_n, x_o) for which it is known whether they are similar or not. It is based on the assumption that descriptors for similar image pairs should have a small Euclidean distance, whereas for dissimilar images this distance should be large [20, 50]. The loss function $E(\mathbf{x}, w)$ minimized in training to determine the parameters w of the network using the data \mathbf{x} available for training, which can be derived automatically from the contents of the knowledge graph, is the weighted sum of four loss terms:

$$E(\mathbf{x}, w) = \alpha_t \cdot E_t(\mathbf{x}, w) + \alpha_s \cdot E_s(\mathbf{x}, w) + \alpha_c \cdot E_c(\mathbf{x}, w) + \alpha_r \cdot E_r(\mathbf{x}, w). \quad (5.1)$$

The four loss terms (E_t, E_s, E_c, E_r) in eq. 5.1 correspond to different definitions of similarity and are explained in the subsequent paragraphs. The weights $(\alpha_t, \alpha_s, \alpha_c, \alpha_r)$, which have to sum to 1, can be modified to define different similarity scenarios. Compared to [28], the

innovation of our method lies in an improved formulation of the semantic similarity loss E_t and the integration of the other three loss terms.

E_t : Semantic Similarity. This loss term considers two images (x_n, x_o) to be similar if the semantic information associated with them is similar. Thus, "similarity" becomes a gradual concept: the more annotations are shared, the more similar a pair of images is considered to be. This definition of *semantic similarity* Y_s of a pair of images (x_n, x_o) has to consider the fact that a sample might have annotations for a subset of the considered variables only:

$$Y_s(x_n, x_o) = \frac{1}{M} \cdot \sum_{m=1}^M v_m \cdot d_m(x_n, x_o) \cdot \pi_m^n \cdot \pi_m^o. \quad (5.2)$$

In eq. 5.2, m is the index of a semantic variable and M denotes the number of variables considered. The binary variables π_m^i indicate whether for the image $i \in \{n, o\}$ the annotation for variable m is available ($\pi_m^i = 1$) or not ($\pi_m^i = 0$). Thus, the term

$$u(x_n, x_o) = 1 - \frac{1}{M} \cdot \sum_{m=1}^M \pi_m^n \cdot \pi_m^o, \quad (5.3)$$

the percentage of variables for which there is no annotation in at least one of the images (x_n, x_o) , expresses the level of uncertainty of the similarity. The weight v_m of a variable m can be used to give more or less importance to certain variables. In accordance with cultural heritage experts, these weights were set to 0.30, 0.25, 0.20, 0.15, 0.10 for the variables *subject depicted*, *production material*, *production place*, *production technique* and *production timespan*, respectively, i.e. the depicted subject was considered to be most relevant. Finally, the function $d_m(x_n, x_o)$ computes the level of agreement between the annotations of (x_n, x_o) for variable m :

$$d_m(x_n, x_o) = \frac{1}{\max(K_n, K_o, \varepsilon)} \cdot \sum_{k=1}^K \delta(l_{mk}(x_n) = l_{mk}(x_o)), \quad (5.4)$$

where $l_{mk}(x_i)$ is an indicator variable with $l_{mk} = 1$ if for variable m , the class label k applies to image x_i with $i \in \{n, o\}$, K is the number of class labels for variable m , and $\delta(\cdot)$ denotes the Kronecker delta which returns 1 if the argument is true and 0 otherwise. K_i , $i \in \{n, o\}$, is the sum of all values $l_{mk}(x_i)$ for the image x_i and ε is a small constant to avoid division by zero. For most semantic variables m , $K_i = 1$, i.e. the class labels are mutually exclusive. However, for some classes, multiple class labels are permitted for a sample, e.g. a sample may consist of multiple *production materials*.

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

The loss term E_t is based on the the *triplet loss* of [28, 113]:

$$E_t(\mathbf{x}, w) = \frac{1}{N_t} \cdot \sum_{n_1=1}^{N_t} \max(M(x_a^{n_1}, x_{ps}^{n_1}, x_{ng}^{n_1}) + \Delta_{a,ps,w}^{n_1} - \Delta_{a,ng,w}^{n_1}, 0). \quad (5.5)$$

The sum in eq. 5.5 is taken over N_t triplets of images $x_a^{n_1}, x_{ps}^{n_1}, x_{ng}^{n_1}$, where n_1 is the index of a triplet and each triplet consists of an anchor sample $x_a^{n_1}$, a positive sample $x_{ps}^{n_1}$ (i.e., a sample considered to be similar to $x_a^{n_1}$), and a negative sample $x_{ng}^{n_1}$ (a sample considered to be dissimilar from $x_a^{n_1}$). The term

$$\Delta_{a,i,w}^{n_1} = \|f_w(x_i^{n_1}) - f_w(x_a^{n_1})\|_2 \quad (5.6)$$

denotes the Euclidean distance of the feature vectors $f_w(x_i^{n_1})$ computed for the image x_i , $i \in \{ps, ng\}$, of triplet n_1 and the feature vector $f_w(x_a^{n_1})$ of the anchor pixel. $M(x_a^{n_1}, x_{ps}^{n_1}, x_{ng}^{n_1})$ is a margin:

$$M(x_a^{n_1}, x_{ps}^{n_1}, x_{ng}^{n_1}) = Y_s(x_a^{n_1}, x_{ps}^{n_1}) - (Y_s(x_a^{n_1}, x_{ng}^{n_1}) + u(x_a^{n_1}, x_{ng}^{n_1})) > !0 \quad (5.7)$$

Minimizing this loss forces the learned descriptors of x_a and x_{ps} to be close together in feature space and the descriptor of x_{ng} to have a larger distance from x_a than x_{ps} . The restriction expressed by $M(x_a^{n_1}, x_{ps}^{n_1}, x_{ng}^{n_1}) > !0$ in eq. 5.7 is used to select triplets. For each minibatch consisting of a set S of images with annotations, all possible triplets are considered as potential training triplets. For each such triplet, the margin is computed, and all the N_t triplets fulfilling the constraint are used to compute the loss and, consequently, to update the parameters. The constraint implies that the similarity $Y_s(x_a, x_{ps})$ of the anchor and the positive sample has to be larger than the sum of the similarity $Y_s(x_a, x_{ng})$ of the anchor and the negative sample and the potential positive similarity according to the unknown properties expressed by the uncertainty term $u(x_a, x_{ng})$.

E_s : Self-Similarity. This loss function considers a visual aspect of similarity: an image should be considered similar to a synthetically adapted version of itself. It should help the CNN to learn that images of the same fabric that were captured, e.g. from different perspectives should be considered to be very similar. For every image x_n in a minibatch consisting of N_{MB} images, a synthetic image x'_n is generated by applying random rotations by 90° , a random horizontal and vertical flipping, and the cropping of a window containing a random percentage $b_{crop} \in [0.7, 1]$ of the pixels of x_n . Furthermore, a random zero mean Gaussian noise with a standard deviation $\sigma_G = 0.1$ is added to the grey values. The loss is forces the Euclidean distance between the feature vectors generated by the CNN for an image x_n and its synthetic

partner x'_n to be close to zero:

$$E_s(\mathbf{x}, w) = \frac{1}{N_{MB}} \cdot \sum_{n=1}^{N_{MB}} \|f_w(x_n) - f_w(x'_n)\|_2. \quad (5.8)$$

E_c : Colour Similarity. This loss takes into account a visual aspect of similarity: two fabrics should be considered similar if the corresponding images have a similar colour distribution. To avoid dependencies on the intensity, the images to be compared are transformed into the *HSV* (hue H , saturation S , value V) colour space, with $H \in [0, 1]$ and $S \in [0, 1]$. In order to compensate for the periodic definition of H , which is usually interpreted as an angle, H and S are considered to be polar coordinates and used to determine Cartesian coordinates (x^c, y^c) , both in the interval $[0, r]$:

$$\begin{aligned} x^c(H, S) &= \frac{r}{2} + \frac{r}{2} \cdot S \cdot \cos(2 \cdot \pi \cdot H) \\ y^c(H, S) &= \frac{r}{2} + \frac{r}{2} \cdot S \cdot \sin(2 \cdot \pi \cdot H), \end{aligned} \quad (5.9)$$

where r defines the scale of the transformation. In this Cartesian coordinate system, a 2D grid of $r \times r$ cells and grid size 1 is defined. A 2D histogram is determined by assigning each transformed point to the grid cell in which it is situated and counting the number of points per grid cell. The histogram obtained for an input image x_n is converted into a *colour feature vector* $h(x_n)$ having r^2 components by stacking the columns of the 2D histogram on top of each other; it represents the colour distribution of x_n . Unless otherwise noted, we used $r=5$ in all experiments involving the colour loss, i.e. each colour vector had 25 elements. Using the N_{MB} images of a minibatch, $N_c = N_{MB} \cdot (N_{MB} - 1)/2$ pairs of images $(x_1^{n_2}, x_2^{n_2})$ can be generated, where n_2 is the index of a pair, and the colour feature vectors $h(x_1^{n_2})$ and $h(x_2^{n_2})$ can be computed. Using the symbol Δ^{n_2} to denote the Euclidean distances of the feature vectors $f_w(x_1^{n_2})$ and $f_w(x_2^{n_2})$ delivered by the CNN for the two images of pair n_2 and $\rho^{n_2} \in [-1, 1]$ to denote the normalized cross correlation coefficient of the corresponding colour feature vectors $h(x_1^{n_2})$ and $h(x_2^{n_2})$, the colour similarity loss is formulated as:

$$E_c(\mathbf{x}, w) = \frac{1}{N_c} \cdot \sum_{n_2=1}^{N_c} \max(0, |\Delta^{n_2} - (1 - \rho^{n_2})|). \quad (5.10)$$

For image pairs having a similar colour distribution, i.e. a value of ρ^{n_2} close to 1, this loss will force the Euclidean distance to be close to 0, i.e. the feature vectors to be similar. The smaller the correlation coefficient, the more the Euclidean distance will be pushed away from 0; for $\rho^{n_2} = -1$, the distance will be pushed to 2, the maximum possible value because of the normalization (section 5.2.1).

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

E_r : Similarity Rules The last loss function is based on the rules for defining sets of images that should be similar and sets of images that should be dissimilar described in section 5.2.1 (cf. Tab. 5.3. Assuming that a minibatch contains N_r such pairs $(x_1^{n_3}, x_2^{n_3}) \in T_{CE}$ (cf. section 5.2), where n_3 is an index of such a pair, and denoting the Euclidean distances of the feature vectors $f_w(x_1^{n_3})$ and $f_w(x_2^{n_3})$ by Δ^{n_3} , a standard loss to train the CNN to produce similar descriptors for similar images and dissimilar descriptors for dissimilar images can be applied:

$$E_r(\mathbf{x}, w) = \frac{1}{N_r} \cdot \sum_{n_3=1}^{N_r} \delta_s^{n_3} \cdot \Delta^{n_3} + (1 - \delta_s^{n_3}) \cdot \max(2 - \Delta^{n_3}, 0). \quad (5.11)$$

In eq. 5.11, the variable $\delta_s^{n_3}$ indicates whether the pair $(x_1^{n_3}, x_2^{n_3}) \in T_{CE}$ is similar ($\delta_s^{n_3} = 1$) or not ($\delta_s^{n_3} = 0$). For pairs which are similar according to the rules defined in section 5.2.1, only the first term is active, and the loss will try to minimize the Euclidean distance of the descriptors of the two images. For dissimilar pairs, only the second term is active, and the loss will try to push the Euclidean distance close to the maximum possible distance of 2.

Minibatch generation. The training data \mathbf{x} consist of images with annotations exported from the knowledge graph. In each training iteration, N_{MB} images are randomly selected from these data to form a minibatch (we used $N_{MB} = 150$ in training). If $\alpha_r \neq 0$, i.e. if the rule-based loss E_r (eq. 5.11) is used, 50% of the samples in the minibatch are drawn from the subset of images found to be affected by one of the rules described in Section 5.2.1 to ensure that the number N_r of pairs considered in E_r is sufficiently high. Note that the loss function terms are based on a comparison of different numbers of images. If the semantic similarity loss E_t (eq. 5.2) is used, all triplets fulfilling the constraint expressed by eq. 5.7 will be considered. If colour similarity E_c (eq. 5.10) is used, all possible pairs of images will be considered. For the self-similarity loss E_s (eq. 5.8), every image in the minibatch and a synthetically modified version of it will be considered. Finally, if the loss E_r (eq. 5.11) is to be used, all pairs of images in the minibatch affected by one of the rules will be considered. Note that the formulation of the total loss E_{total} (eq. 5.1) is flexible w.r.t. the combination of the loss terms. However, at least one of the two terms E_t and E_c has to be considered, because E_s and E_r do not contribute for dissimilar pairs, in the first case by design and in the second case because the rules in Tab. 5.3 do not define any dissimilar pairs.

One iteration of SGD starts by extracting a minibatch from the training data and defining the required sets of image pairs and triplets in the way just described. Afterwards, all images are propagated through the network, and the loss E_{total} is computed and back-propagated through the network to compute the gradient of the loss with respect to the unknown parameters w . Finally, these gradients are used to update the parameter values.

5.2.2 Evaluation

The evaluation was carried out in three steps and all test data was exported from the knowledge graph. The test data consisted of 25,825 images with annotations in at least one of five semantic variables mentioned in Section 5.2.1 and an additional set of records affected by at least one of the rules described in Section 5.2.1. The first step involved a set of experiments for finding the optimal set of hyperparameters training and classification. These experiments were based on semantic similarity only. As we involved domain experts in the development of definitions of similarity, we also wanted to make sure we do not only evaluate the model with regards to a general similarity. Based on the different defined types of similarity, 5 different scenarios have been created together with the cultural heritage domain experts of our project in the second step:

- **Scenario A:** Semantic similarity and self-similarity.
- **Scenario B:** Colour similarity and self-similarity. Only scenario with exclusively visual similarities.
- **Scenario C:** Augmentation of semantic similarity with the rules defined by cultural heritage domain experts.
- **Scenario D:** Augmentation of colour similarity with the rules defined by cultural heritage domain experts.
- **Scenario E:** Combination of all concepts of similarity, which is meant to be a compromise between semantic and visual aspects of similarity.

As part of the second step, a purely technical evaluation has been performed based on five-fold cross validation and performing a k-nearest neighbour classification based on the optimal hyperparameters identified in step 1. This part of the evaluation focused on the ability to find images having similar semantic properties. The average accuracies and F1 scores of step 2 can be seen in Tables 5.4 and 5.5. As can be seen in these results, the overall F1 scores and accuracies are relatively similar, with Scenario E being altogether the best case.

The third step relied on these five scenarios, but the evaluation was performed by cultural heritage experts through an interactive analysis of the results. This type of expert evaluation is very time consuming, therefore only a limited amount of test data has been chosen and a fixed split into training and test data was used. 100 images were selected to be retrieved as test images, for which the $k = 10$ most similar images should be retrieved by the image retrieval tool. All remaining samples were used for training. Images of objects that contribute to the test set were excluded from training. This is especially important as one object can have several images.

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

The evaluation criteria used by the domain expert were based on the following concepts:

- *Pattern*: This concept is about decorative motives, for example flowers or birds. Therefore it is related to aspects of semantic similarity, as some records have explicit textual metadata descriptions about those.
- *Colour*: The perception of the colour of an image is relatively easy for most users. This term represents a visual type of similarity here.
- *Appearance*: The domain experts use this term for a concept of a generic evaluation of the outward form of the silk fabric in the image. This includes shape, the geometric form, but also colour again. The domain experts consider this to be a characteristic that can also be easily perceived by a typical user.

If a pair of images matches at least two criteria it was considered a meaningful pair, otherwise not. A graphical representation of the top-k-scores and the percentage of meaningful images for values k between 1 and 10 can be seen in Figure 5.7 and 5.8. For Step 3, we can see Scenario B performing by far the best, with Scenario E actually being mostly second best, but with a significant distance.

Based on these results two best scenarios have been chosen: Scenario E led to the highest F1 scores and overall accuracies based on semantic similarity, whereas Scenario B proved to be the best one according to the evaluations by the domain experts.

Variable	α_s	Material	Production Place	Technique	Timespan	Depiction	Average
Scenario A	1/2	78.2 / 73.6	44.4	61.8	54.0	88.0	65.3 / 64.4
Scenario B	0	77.1 / 72.6	40.6	57.3	52.5	89.6	63.4 / 62.5
Scenario C	1/3	77.9 / 73.3	43.3	60.2	54.4	90.1	65.2 / 64.3
Scenario D	0	77.9 / 73.2	43.3	61.1	54.1	89.4	65.2 / 64.2
Scenario E	1/4	78.2 / 73.4	44.1	61.6	53.8	89.1	65.4 / 64.4
SIR_LR_4	1	78.2 / 73.9	44.6	61.6	55.3	88.9	65.7 / 64.9

Table 5.4: Overall accuracies [%] per variable for the different scenarios of similarity as well as the best performing experiment of test step 1 (SIR_LR_4). The highest score per variable is highlighted in bold font. The second column contains the weight α_s of the loss function term related to semantic similarity and, thus, indicates whether semantic similarity is considered ($\alpha_s > 0$) or not ($\alpha_s = 0$); the last column gives average values over all variables. In case of the variable Production Material, the first value refers to the classification results based on a binary classification procedure; the second value refers to the results including the most probable class of samples assigned to the background for all classes.

The investigated scenarios for similarity are based on different definitions of the loss function; they indicate that the consideration of the additional loss terms beyond those used in Scenario

Variable	α_s	Material	Production Place	Technique	Timespan	Depiction	Average
Scenario A	1/2	28.3 / 29.6	27.0	57.8	42.8	63.1	43.8 / 44.1
Scenario B	0	25.1 / 26.7	22.8	52.8	41.2	62.0	40.8 / 41.1
Scenario C	1/3	27.7 / 29.0	26.3	56.1	43.4	66.3	44.0 / 44.1
Scenario D	0	28.1 / 29.3	26.0	57.0	43.0	63.3	43.5 / 43.7
Scenario E	1/4	29.6 / 29.7	26.7	57.3	42.9	65.6	44.4 / 44.4
SIR_LR_4	1	29.2 / 30.2	26.8	56.9	43.0	58.0	42.8 / 43.0

Table 5.5: Average F1-Scores [%] per variable for different scenarios of similarity as well as the best performing experiment of test step 1 (SIR_LR_4). For more details, see the caption of Table 5.4

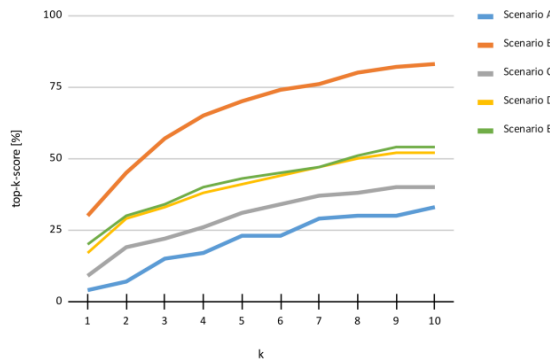


Figure 5.7: Top-k-scores as a function of k for all evaluated scenarios. The score gives the percentage of query images for which there was at least one meaningful result among the k most similar images delivered by the image retrieval module.

A do indeed contribute to a better performance if the focus of retrieval is on semantic aspects of similarity, whereas the new colour loss is essential for retrieving meaningful results according to the evaluation by domain experts. A full ablation study considering the contributions of all loss terms is beyond the scope of this study. Note that the method described in [28] is very similar to Scenario A; the difference is in the use of an improved version of the semantic loss and in the self-similarity loss.

5.2.3 An Exploratory Search Engine for Finding Similar Objects

The knowledge graph that we used to train the models is accessible via a RESTful API that has been developed using the grlc framework and the SPARQL Transformers library [82]. A web based application has been developed using this API to provide an exploratory search engine for silk textiles. It offers a user-friendly interface with faceted search to apply filters corresponding to the different properties of the silk textiles, like the material or technique being used or the production place and time [45].

5.2 Retrieval of images of silk textiles through domain expert rules and our knowledge graph

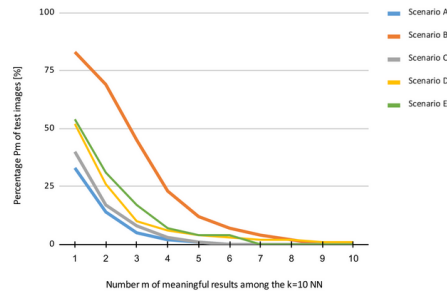


Figure 5.8: Percentage P_m [%] of query images for which the image retrieval module delivered at least m meaningful images among the $k=10$ nearest neighbours for Scenarios A-E.

We have integrated in this exploratory search engine the image-based retrieval module described in this section ². More precisely, we have integrated the two scenarios B and E described above under two buttons named "visually similar images" and "objects with similar properties". The user can upload any image of his choice (preferably depicting a silk textile) and invoke one of these two methods to retrieve up to 20 similar objects from the knowledge graph. Similarly, when browsing the knowledge graph, the user can request what are the similar objects (either visually or semantically) with respect to the one being viewed (Figure 5.9).

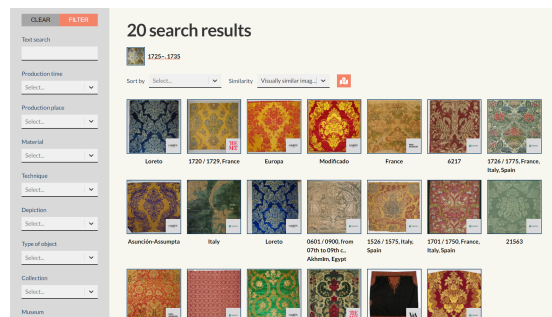


Figure 5.9: Objects that are visually similar with respect to an object produced in 1725-1735 in France using the embroidery technique and coming from the Art Institute of Chicago (ARTIC) museum.

5.2.4 Conclusion

In this section, we have presented an image retrieval module that considers different aspect of similarity between cultural heritage objects that silk textiles are. One of our contribution is to use a knowledge graph in order to convert domain-expert similarity rules into queries that generate vast amount of training data. The design and the evaluation of the image retrieval models benefit from the knowledge of domain experts. The code of the image retrieval method

²<https://github.com/silknow/image-retrieval-server>

is available under GitHub ³.

While exploring different scenarios, we observe that the simplest visual only similarity provides the best accuracy: At least one meaningful image was retrieved per query image in 83% of all cases. The semantic similarity proves also to be useful for domain experts who appreciate to switch from one to the other and observe the differences. The integration of this module in a user friendly interface, an exploratory search engine, enables to conduct additional human evaluations.

5.3 Enabling the exploration of the cultural heritage of silk artifacts

Many Cultural Heritage topics are very niche and it often requires the consultation of expert literature or the visit of a museum to understand the vocabulary and the granularities of a such a field. The historical production of silk items in Europe is no exception to this, maybe it is even one of the more obscure ones. This fact, however, motivates the exploration of means to offer not only access, but also tools to facilitate the exploration of knowledge graph and all its data that we have developed.

An exploratory search covers a broad array of activities to address the needs of a search without or with a target in mind, and a searcher familiar or unfamiliar with the topic. ADASilk offers for example topic-based exploration, but also an advanced search with many fine-grained filters, on top of our integrated data and developed knowledge graph. It features a graphical web interface that is supposed to be easy to use and sufficiently fast.

One of the most advanced features of ADASilk is its retrieval of similar silk object images. This is based on a model that we trained by leveraging domain expert rules of image similarity. With this function, our exploratory search engine is even usable without any text input. This feature also represents a way of measuring similarity between two objects inside the knowledge graph, or even between an external photo of a silk item and one in our dataset. There is a potential of future research work exploring further ways of measuring and establishing similarity between objects represented in the knowledge graph.

³<https://github.com/silkknow/image-retrieval>

Conclusion

The field of knowledge modeling and representation remains very active and is located at the center of many other concurrent directions in Artificial Intelligence and Natural Language Processing research. The research work presented in this thesis constitutes a combination of knowledge modeling and information extraction: the SILKNOW Knowledge Graph was both the data source and the target for many experiments that aimed at enriching multilingual real-world data from a very specific Cultural Heritage domain: the history of European silk objects.

Crucial parts of the this thesis are built upon an application of established methods to build a knowledge graph from museum data. However, such an endeavour is not straight-forward and is not comparable with either the construction or use of most dataset used for many benchmarks in recent years: the amount of available data has been small, the vocabulary multi-lingual, the domain niche and the data sources were heterogeneous. Not only is it required to deal with details at many ends that are currently not possible to automate, it is also in all cases imperative to adjust and sometimes further develop aforementioned existing tools.

Thanks to many breakthroughs specifically in information extraction it is nowadays possible to go much further with a knowledge graph than to work solely on data integration and data modeling. From the beginning on, it has therefore been the goal, to apply advanced NLP methods in order to enrich its metadata. In our cases, we focused a lot on rich textual description of museum records, which contained a lot of information that was originally not annotated and on methods to automate categorizing some of this information inside the knowledge graph.

6.1 Summary of the Research

This thesis represents a contribution in research application to Cultural Heritage domain of historical, European silk objects, knowledge representation and information extraction. There was a special focus on zero-shot and unsupervised metadata predictions, and a more general one on different knowledge graph enrichments. Finally, we focused on an exploratory search engine, with which all these efforts are visualized, presented and accessible for everybody.

In more detail, the main outcomes of this thesis are:

- Modelling the SILKNOW Ontology, a semantic network, heavily based on CIDOC-CRM with several extension. It has been extended with classes and properties that go beyond the scope of a generic museum ontology, e.g. to express what type of object a silk item is. It has also been extended through both PROV-O and our own classes to model metadata predictions to prepare the ontology for further experimental enrichments.
- The SILKNOW Thesaurus, a multilingual controlled vocabulary for concepts related to the domain of silk objects. It makes enrichment through linking of entities from the original museum records possible and helps with their multilingual disambiguation. Our thesaurus consists also of a rich hierarchy and relations between concepts that enable many advanced search filter function of our exploratory search engine.
- Designing and implementing the SILKNOW Knowledge Graph, a rich and unified resource on silk items, resulting from many different, multilingual museum sources across the world. This included the development of a range of smaller software tools, especially a web crawler and harvesting software and a converter. It required the design of expressive mapping tables by historians and domain experts based on classes and properties of our SILKNOW Ontology.
- Exploration of ways to predict gaps of categorical values, like production material or motives, that existed in the original records. This ranges from computationally intensive supervised text and image classification approaches, one of them multimodal, to several zero-shot text classification methods aimed at dealing with some of the specific challenges of our dataset and domain. There have also been experiments with induced knowledge, either through prompting (ProZe) or question/answer generation.
- ADASilk, coined after Ada Lovelace, a web application for exploratory search. It offers simple access to the uploaded SILKNOW Knowledge Graph and all its data. The includes also all images and offers a way to experience the several enrichments of the data, for example an advanced search fueled by the facets and categories of the SILKNOW Thesaurus as well as entity linking of concepts with it. It also offers a way to present some of the metadata predictions that we could generate.

- Finally, we implemented an Image Retrieval function into ADASilk for finding images similar to a selected one inside the knowledge graph. It is also able to retrieve images from the knowledge graph that are similar to any uploaded picture. This functionality has been made possible by experimenting with ways of training image classification models effectively with the help of explicit rules postulated by domain experts.

6.2 Limitations and Further Perspectives

The work presented in this manuscript is experimental in nature and could be further followed up and fleshed out in many ways. In this final, section we will show up some of them that we could identify.

- **Multimodal Supervised Classification** Despite promising results for this series of experiment, some results hint at incorrect training labels. As much as such problems could be solved by making bigger efforts towards data cleaning and pre-processing, we believe looking at training methods that are more robust against such label noise should be a next step for work with data such as ours. Another problem with supervised approaches in general is that our data contained few data points, but many different categories or classes. Supervised methods rely strongly on class balance and sufficient training data. At least to solve any class imbalances, we needed to work on the reduction of classes, either by discarding some of them or through grouping certain classes. We tried to address these latter problems, by investigating into alternatives to the more common and usually more accurate supervised classification methods, which are also presented in chapter 4.
- **Explainable Zero-Shot Text Classification** Dataless or zero-shot classification is a method that enabled us to retain more classes than for the supervised approaches. Despite, in our opinion, being a good choice for sparse data such as in our cases it comes almost naturally with lower prediction scores. However, we believe there should be potential in combining our zero-shot classification approaches and the supervised text classification approaches through bootstrapping, which could alleviate their downsides. Another limitation that needs to be considered is that our zero-shot methods rely strongly on ConceptNet, both in order to not rely on training data, but also to be explainable. ConceptNet does not have the detail as, e.g., our SILKNOW thesaurus when it comes to silk-related vocabulary. It also does not have the same degree of multilingualism, as it has more vocabulary in English than in any other language. There is potential in future work that aims at working around or solving these inherent limitations as well as trying to experiment with combining it with other external resources. Next to ConceptNet, we think it is also worth to look at the used language models, concretely

letting it also inform the label selection and expansion and to handle multi-word labels.

- **Prompt-guided Information Extraction**

Using prompting to further leverage the implicit knowledge of language models is an relatively new and exciting direction, which we tried to combine with zero-shot text classification. One of the limitations is that this method is naturally very dependent on finding the "perfect" prompt. This is a very indirect trial-and-error process, and we are sure there is more potential in investigating which changes to the prompt have which effects. We think that trying out more different and specialized language models, such as one specifically designed for a domain such as ours or e.g. the healthcare domain based on medical documents, would also improve the quality of the classification based on this approach.

- **Automatic Question-Answer Generation**

The experiments based on models that generate question and answers based on text input show promise by making predictions of certain labels possible that our other zero-shot methods could not produce. Although overall scores are relatively low, output depends a lot on the way input is given to the model and in which way the text is pre-processed and selected. Parallel to the prompt-based methods, finding out what works best is hereby very indirect and the options are certainly not yet exhausted by us. As we worked solely with pre-trained T5 models, there is also a promising direction in training and fine-tuning own, more domain-specific ones as well.

- **Expert Rules to improve the Training of Domain-specific Image Retrieval**

Our approach to let domain expert formulate rules of similarity in order to maximize the training process of an image retrieval model with relatively few image files has shown satisfying first results. We still believe, there is more future work necessary to compare this method with the state-of-the-art. Thanks to this model being easily accessible through our exploratory search engine ADASilk, a logical next step would be to invite volunteers to perform additional human evaluation.

6.2 Limitations and Further Perspectives

GitHub / GitLab repositories	Chapters	Publications
https://github.com/silknow/crawler	3	The SILKNOW Knowledge Graph
https://github.com/silknow/converter	3	The SILKNOW Knowledge Graph
https://github.com/silknow/thesaurus	3	The SILKNOW Knowledge Graph
https://github.com/silknow/knowledge-base	3	The SILKNOW Knowledge Graph
https://github.com/silknow/skosmos	3	The SILKNOW Knowledge Graph
https://github.com/silknow/api	3	
https://github.com/silknow/adasilk	3, 5	
https://github.com/silknow/image-classification	4	Multimodal Metadata Assignment for Cultural Heritage Artifacts
https://github.com/silknow/text-classification	4	Multimodal Metadata Assignment for Cultural Heritage Artifacts
https://github.com/silknow/ZSL-KG-silk	4	Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles
https://gitlab.eurecom.fr/schleide/proze	4	ProZe: Explainable and Prompt-guided Zero-Shot Text Classification
https://gitlab.eurecom.fr/schleide/qg4textunderstanding	4	
https://github.com/silknow/image-retrieval	5	Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules
https://github.com/silknow/image-retrieval-server	5	Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules

Table 6.1: Overview of all GitHub and GitLab repositories that contains code that has been at least partial relevant for our work in this thesis and related publications.

Publications list

Journal

- Ismail Harrando*, Alison Reboud*, Thomas Schleider*, Thibault Ehrhart and Raphael Troncy (*Equal contribution). **ProZe: Explainable and Prompt-guided Zero-Shot Text Classification**. In IEEE Internet Computing: Special Issue on Knowledge-Infused Learning, 2022. <https://doi.org/10.1109/MIC.2022.3187080>.
- Luis Rei, Dunja Mladenić, Mareike Dorozynski, Franz Rottensteiner, Thomas Schleider, Raphaël Troncy, Jorge Sebastián and Mar Gaitán. **Multimodal Metadata Assignment for Cultural Heritage Artifacts**. In Multimedia Systems (*under review*), 2022.
- Thomas Schleider, Raphaël Troncy, Mar Gaitán, Ester Alba, Jorge Sebastián, Dunja Mladenić, Avguštin Kastelic, M. Beshar Massri, Arabella León, Marie Puren, Pierre Vernus, Dominic Clermont, Franz Rottensteiner, Maurizio Vitella, Georgia Lo Cicero. **The SILKNOW Knowledge Graph**. In Semantic Web Journal: Special Issue on Cultural Heritage and Semantic Web (*under revision*), 2022.

Conferences and Workshops

- Thomas Schleider and Raphaël Troncy. **Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles**. In 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLFL) co-located with EMNLP, 2021, Online.
- Thomas Schleider, Raphaël Troncy, Thibault Ehrhart, Mareike Dorozynski, Franz Rottensteiner, Jorge Sebastián and Georgia Lo Cicero. **Searching Silk Fabrics by Images Leveraging on Knowledge Graph and Domain Expert Rules**. In 3rd workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC) co-located with ACM Multimedia, 2021. Online. **Best Paper Award**.
- Thomas Schleider and Raphaël Troncy. **Exploring the European Silk Cultural Heritage**

through the SILKNOW Knowledge Graph. In International Conference on Silk heritage and Digital Technologies (Weaving Europe), 2020. Online.

Technical Reports

- Nacira Abbas, Kholoud Alghamdi, Mortaza Alinam, Francesca Alloatti, Glenda Amaral, Claudia d’Amato, Luigi Asprino, Martin Beno, Felix Bensmann, Russa Biswas, Ling Cai, Riley Capshaw, Valentina Anita Carriero, Irene Celino, Amine Dadoun, Stefano De Giorgis, Harm Delva, John Domingue, Michel Dumontier, Vincent Emonet, Marieke van Erp, Paola Espinoza Arias, Omaira Fallatah, Sebastián Ferrada, Marc Gallofré Ocaña, Michalis Georgiou, Genet Asefa Gesese, Frances Gillis-Webber, Francesca Giovannetti, María Granados Buey, Ismail Harrando, Ivan Heibi, Vitor Horta, Laurine Huber, Federico Igne, Mohamad Yaser Jaradeh, Neha Keshan, Aneta Koleva, Bilal Koteich, Kabul Kurniawan, Mengya Liu, Chuangtao Ma, Lientje Maas, Martin Mansfield, Fabio Mariani, Eleonora Marzi, Sepideh Mesbah, Maheshkumar Mistry, Alba Catalina Morales Tirado, Anna Nguyen, Viet Bach Nguyen, Allard Oelen, Valentina Pasqual, Heiko Paulheim, Axel Polleres, Margherita Porena, Jan Portisch, Valentina Presutti, Kader Pustu-Iren, Ariam Rivas Mendez, Soheil Roshankish, Sebastian Rudolph, Harald Sack, Ahmad Sakor, Jaime Salas, Thomas Schleider, Meilin Shi, Gianmarco Spinaci, Chang Sun, Tabea Tietz, Molka Tounsi Dhoub, Alessandro Umbrico, Wouter van den Berg and Weiqin Xu. **Knowledge Graphs Evolution and Preservation – A Technical Report from ISWS 2019.** arxiv:2012.11936. <https://arxiv.org/abs/2012.11936>.

Résumé en français

6.1 Introduction

6.1.1 Motivation

De tous les patrimoines culturels de l'Europe, la connaissance historique de la fabrication des tissus de soie est probablement l'une des plus obscures. Par le passé, les tissus de soie et les objets fabriqués à partir de ces tissus ont fait partie des biens commerciaux les plus chers et les plus recherchés dans le monde pendant de nombreux siècles. L'existence historique de ce que l'on appelle la "route de la soie" laisse supposer une origine lointaine, en Chine et en Asie du Sud-Est. Les connaissances sur le tissage des articles en soie et l'utilisation des vers à soie nécessaires se sont répandues en Europe au moins des siècles, voire des millénaires, après sa découverte initiale.

En Europe, les textiles en soie ont toujours été associés au luxe, notamment aux vêtements, aux meubles et aux décorations des aristocrates, et de nombreux articles en soie peuvent également être considérés comme des œuvres d'art. Derrière chaque objet en soie se cache également le savoir-faire et les compétences des artisans ou, pour des exemples plus récents, les progrès technologiques et scientifiques ayant permis la construction de meilleurs métiers à tisser. Enfin, l'histoire de la production européenne de soie est liée à un jalon très important de l'histoire du matériel informatique : le maître tisserand et marchand de soie français Joseph Marie Jacquard a inventé le premier métier à tisser programmable. Ce métier, appelé métier Jacquard, était contrôlé par des cartes perforées, des morceaux de papier contenant des données numériques par la présence ou l'absence de trous dans des positions prédéfinies. Il était utilisé pour la production d'objets en soie et permettait d'utiliser les techniques courantes de tissage de la soie - brocart, damas, etc.

De nombreux musées à travers le monde possèdent encore des objets historiques en soie européenne, qu'il s'agisse de drapeaux, de marquises, de tapisseries, de costumes, d'éventails ou de fourreaux d'épée, et plus particulièrement des objets du 15^e siècle et des siècles suivants. Heureusement, l'accès public à leurs métadonnées et à leurs photos est souvent possible. La connaissance de l'ensemble de leur domaine historique et des spécificités des objets en soie,

en particulier ceux de fabrication européenne, est cependant aujourd'hui dispersée, inconnue de beaucoup et peut donc être considérée comme menacée.

L'identification et la conservation des objets du patrimoine culturel nécessitent un inventaire et un archivage cohérents de leurs métadonnées et des autres informations environnantes, comme les images et leurs propres métadonnées. De nombreux musées et bibliothèques, y compris ceux qui possèdent ou exposent des objets historiques de la soie européenne, ont déjà numérisé la plupart des parties de leurs collections et les ont rendues accessibles au public - soit par le biais d'une interface web, soit même en fournissant des API spécifiques. Ce que nous pouvons cependant observer, c'est qu'à côté de ces efforts de catalogage numérique de base, de nombreux outils numériques possibles ne sont pas appliqués, en particulier lorsqu'il s'agit d'intégrer de manière congruente toutes les métadonnées pertinentes des objets existants. En outre, en ce qui concerne les connaissances spécialisées dans le domaine des tissus de soie européens : Il n'existe tout simplement pas de lieu unique, dans le monde physique ou en ligne, où un public peut accéder à des informations sur tous ces articles et toutes les informations de base pertinentes, par exemple sur la façon dont ils ont été tissés ou sur les motifs qu'ils présentent.

6.1.2 Contexte de la recherche : le projet SILKNOW

SILKNOW est un projet de recherche financé par H2020 (2018-2021) visant à comprendre, conserver et diffuser le patrimoine européen de la soie du 15e au 19e siècle. Il s'agit d'un projet pluridisciplinaire dont l'un des objectifs est d'appliquer des méthodes de recherche informatique aux besoins des musées, de l'éducation, du tourisme, des médias et des industries créatives.

6.1.3 Les questions de recherche

Comment représenter les connaissances spécifiques à un domaine provenant des archives des musées ?

La création de tout type de base de connaissances, ou plus précisément d'un graphe de connaissances qui représente les connaissances des experts, nécessite un flux de travail spécifique. Il faut d'abord développer ou décider d'une ontologie spécifique. Les experts du domaine doivent être en mesure de faire correspondre les données sémantiquement hétérogènes des champs d'enregistrement des musées originaux avec les classes et les propriétés du modèle d'ontologie cible. Ces règles de mise en correspondance doivent être appliquées par un logiciel et, si nécessaire, réajustées en fonction de leur applicabilité réelle. À ce stade, l'utilisation ou la conception de vocabulaires contrôlés doit également être envisagée. La mise en œuvre de ces règles de mise en correspondance sémantique et de la correspondance des chaînes de

caractères avec les concepts d'un vocabulaire contrôlé n'est pas triviale et le succès n'est pas garanti. La qualité d'un graphe de connaissances développé et enrichi est difficile à évaluer et l'utilisation de questions de compétences est devenue une norme, mais il s'agit toujours d'un processus très manuel et subjectif pour lequel une plus grande automatisation pourrait être envisagée. Enfin, donner un accès approprié à un graphe de connaissances à différents types d'utilisateurs finaux est un autre défi qui doit être relevé.

Comment extraire le plus efficacement possible des informations structurées dans différentes langues à partir de descriptions textuelles sans avoir besoin de grandes quantités de données d'entraînement annotées ?

Les progrès récents dans le domaine du traitement du langage naturel, et plus particulièrement dans celui de l'extraction d'informations, peuvent nous aider à résoudre des problèmes tels que les métadonnées manquantes dans les systèmes basés sur la connaissance qui reposent sur des données sources hétérogènes et multilingues. En particulier, il est courant d'entraîner des modèles de classification de texte à partir d'enregistrements complets de métadonnées afin de prédire les valeurs catégorielles manquantes dans d'autres enregistrements. d'autres enregistrements. Cependant, de tels modèles nécessitent une quantité importante de données annotées pour l'entraînement, ce qui est coûteux à obtenir pour ces domaines d'expertise spécifiques. Une alternative pour de tels cas est l'utilisation d'approches non supervisées pour la prédiction de métadonnées, en s'appuyant sur l'apprentissage à zéro, l'apprentissage par transfert et les modèles de langage.

Comment faciliter l'exploration d'un graphe de connaissances sur le patrimoine culturel ?

Les nœuds ou les objets à l'intérieur d'un graphe de connaissances sur le patrimoine culturel peuvent être considérés comme similaires de différentes manières, sur la base de leurs métadonnées ou (le cas échéant) des images disponibles d'un objet. Même au sein de cette division, différentes méthodes de mesure peuvent être établies, par exemple par la sélection ou la pondération de différentes propriétés de métadonnées textuelles ou de propriétés visuelles. Une façon de décider d'une mesure de similarité est de recourir à des règles d'experts du domaine ou à une autre forme d'évaluation humaine.

6.1.4 Résumé des contributions

Cette thèse a contribué à la recherche avec les résultats suivants :

- Un modèle de données et un thésaurus pour et sur les objets historiques en soie d'Europe qui sont stockés et exposés dans les musées. Ces contributions ont été mises

en œuvre en s'appuyant fortement sur l'apport, les connaissances et la conception d'experts du domaine et d'historiens.

- Le développement du graphe de connaissances SILKNOW avec lequel les métadonnées et les images du musée ont finalement été intégrées. Il a été mis en œuvre et téléchargé avec d'autres technologies et outils du Web sémantique, comme SPARQL et un triple-store, et est publié dans le Web des données pour rendre tout notre travail accessible à tous.
- Un ensemble d'outils pour l'exploration du Web, la récolte, le téléchargement et la conversion des données de musées avec notre modèle de données vers notre format de graphe, en remplaçant les chaînes de caractères par des URI de concepts et en reliant ainsi les entités à notre thésaurus. En outre, nous pouvons contribuer aux outils qui offrent un accès API pour les développeurs web à notre point de terminaison SPARQL.
- Exploration de plusieurs approches, la plupart d'entre elles effectuant une classification à zéro, pour combler les lacunes des métadonnées en prédisant les valeurs manquantes.
- Un moteur de recherche exploratoire appelé ADASilk, pour offrir une interface web graphique simple pour les non-experts, qui offre une recherche avancée basée sur plusieurs de nos enrichissements de données dans le KG et aussi des intégrations de plusieurs autres outils logiciels de nos partenaires du projet SILKNOW. Enfin, nous avons intégré à ADASilk un modèle de recherche d'images qui peut être utilisé pour trouver des images similaires d'objets en soie. Nous avons formé ce modèle en tirant parti des connaissances des experts du domaine en formulant notamment des règles de similarité.

6.1.5 Plan de la thèse

Le reste de cette thèse est organisé comme suit :

6.2 Développement d'un graphe de connaissances sur la production d'objets en soie

Le graphe de connaissances SILKNOW (KG) ¹ se trouve au centre de tous les efforts visant à créer une représentation unifiée des métadonnées des textiles de soie européens, en particulier du XVe au XIXe siècle. Toutes les données utilisées dans les expériences de cet article ont été extraites ou téléchargées à partir de 20 sources, la plupart d'entre elles étant des archives publiques de musées en ligne, pour lesquelles nous avons construit un logiciel de crawling

¹<https://zenodo.org/record/5743090>

6.2 Développement d'un graphe de connaissances sur la production d'objets en soie

et de harvesting. En outre, nous disposons de données provenant des partenaires du projet SILKNOW, Garin 1820 et l'Université de Palerme (Patrimoine culturel de la Sicile). Pour le jeu de données utilisé dans les expériences de cet article, un export complet de tous les objets du graphe de connaissances a été effectué, ce qui consiste en les métadonnées de 40 873 objets de soie uniques avant toute étape de prétraitement. Cette exportation comprend au total 74 527 fichiers d'images uniques.

6.2.1 Le modèle de données

Un modèle de données est généralement un modèle abstrait pour l'organisation d'éléments de données et une normalisation de leurs relations tant entre eux qu'avec les propriétés des entités en mots réels.

Exigences

Chaque modèle de données doit répondre à un ensemble d'exigences pour être utile dans son application. Compte tenu de notre domaine et de l'origine de nos données, nous devons nous assurer que nous pouvons prendre en charge non seulement les textes multilingues, mais aussi les langues utilisées dans les documents originaux, à savoir l'anglais, l'espagnol, le français et l'italien. Comme nos données ne concernent qu'un sujet humain très spécifique, comme l'histoire ou la chimie, dans notre cas les tissus de soie historiques, nous avons également besoin principalement d'une ontologie de domaine. Avec une telle spécificité, il est également nécessaire d'être capable d'adapter et d'étendre un tel modèle de données.

Les métadonnées sur les tissus de soie européens proviennent toujours de collections ou de musées, ce qui rend nécessaire une ontologie qui offre des classes et des propriétés pour représenter un objet et éventuellement des photos ou des images de celui-ci de manière appropriée. Un aspect important des objets de musée, surtout lorsqu'ils sont historiques, est qu'ils sont souvent à la fois des produits uniques qui ont été créés à la main et, dans presque tous les cas, ont été trouvés ailleurs que dans le musée. Par conséquent, ils ont au moins une fois changé de lieu et de propriétaire.

L'ontologie de SILKNOW

L'ontologie SILKNOW sur laquelle notre graphe de connaissances est construit est fortement basée sur le modèle de référence conceptuel CIDOC (CIDOC-CRM).

De petites parties de l'ontologie totale pour le graphe de connaissances SILKNOW sont basées

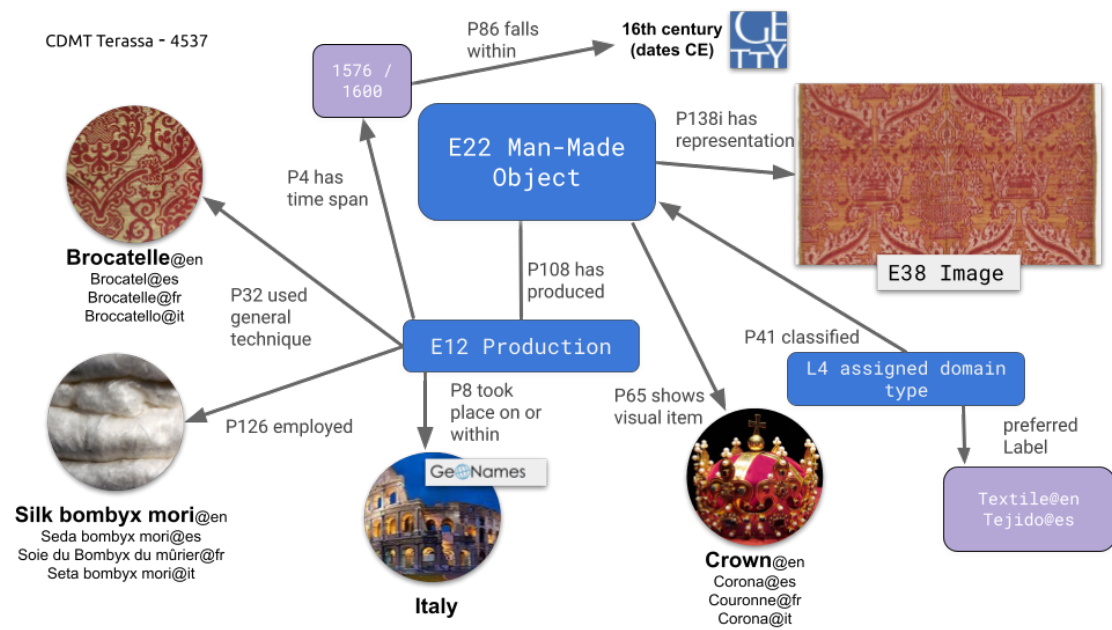


Figure 6.1: Illustration de la représentation du CDMT Terassa / IMATEX record 4537 dans le SILKNOW knowledge graph

sur plusieurs propriétés de schema.org² et de l'ontologie du temps du W3³. La majorité des classes et des propriétés utilisées dans SILKNOW proviennent de la version publiée actuelle de CIDOC-CRM (6.2) et de ses extensions, le modèle d'observation scientifique (CRMsci) [41] et le CRM numérique (CRMdig) [42]. Le premier est une ontologie formelle permettant d'intégrer des métadonnées sur l'observation scientifique, et le second est destiné à encoder des métadonnées sur les étapes et les méthodes de production des produits de numérisation. L'utilisation et l'implémentation complètes de ces ontologies et modèles de données peuvent être récupérées sur GitHub où elles font partie du logiciel du convertisseur.⁴

6.2.2 Modélisation des prédictions de métadonnées

Un aspect de notre contribution à la recherche consistait non seulement à intégrer des données dans un graphe de connaissances, mais aussi à expérimenter l'enrichissement des données par diverses méthodes. En gardant cela à l'esprit, notre ontologie devait être capable de représenter ces enrichissements en conséquence. Une grande partie de ces enrichissements consiste en des prédictions de diverses lacunes de métadonnées dans les données : par exemple, des dates de production ou des techniques de tissage manquantes.

Pour modéliser la prédiction dans le cadre de l'ontologie SILKNOW Knowledge Graph, nous

²<https://schema.org/>

³<https://www.w3.org/TR/owl-time/>

⁴<https://github.com/silknow/converter/tree/master/src/main/java/org/silknow/converter/ontologies>

6.2 Développement d'un graphe de connaissances sur la production d'objets en soie

avons ajouté des classes et des propriétés du modèle de données de provenance (Prov-DM), plus précisément l'ontologie PROV (PROV-O). ontologie PROV (PROV-O)⁵, une ontologie OWL2. Il permet de mapper PROV-DM en RDF. S'agissant d'une recommandation du W3C, elle permet l'expression d'éléments importants des prédictions, aussi bien pour celles basées sur des images que pour celles basées sur des descriptions textuelles et sur des valeurs catégorielles.

6.2.3 Évaluation avec des questions sur les compétences

La formulation des questions de compétence était une partie importante pour contraindre correctement notre ontologie et préparer une évaluation de notre modèle de données. Nous avons publié toutes les questions de compétence, qui ont été formulées par des experts du domaine au début du projet SILKNOW, sur GitHub ⁶, où une requête correspondante et une liste de résultats peuvent être directement récupérées. Bien que nous disposions non seulement de questions en anglais, mais aussi en espagnol, nous n'avons pas encore utilisé ces dernières pour l'évaluation de notre modèle de données.

6.2.4 Controlled vocabularies

Le Thésaurus SILKNOW

Méthode de développement impliquant des experts Des experts du patrimoine de la soie ont été impliqués afin de développer le thésaurus SILKNOW. Ces experts comprenaient des historiens de l'art, des historiens, des tisserands, des ingénieurs et des philologues. La pluridisciplinarité était essentielle pour sélectionner les termes, retracer leur évolution, leur utilisation historique et actuelle, et la manière dont certains termes ont évolué dans le temps et l'espace (par exemple, les variations locales). Le thésaurus SILKNOW étant symétrique, tous les termes devaient être traduits, les spécialistes du textile ont eu recours à des sources spécialisées, qui dans certains cas ont fourni des traductions dans d'autres langues (comme le dictionnaire Castany Saladrigas, 1949). Dans d'autres cas, des traductions directes étaient nécessaires, une note d'application était ajoutée si nécessaire ou la langue source était utilisée comme prêt. Néanmoins, chaque traduction a été effectuée en suivant les directives ISO pour un thésaurus [33].

Couverture du thésaurus Le thésaurus SILKNOW a été validé sur des données textuelles des musées sélectionnés dans plusieurs langues naturelles. La fréquence des différents concepts du thésaurus présents dans le musée en question a été calculée. Les traductions espagnole,

⁵<https://www.w3.org/TR/prov-dm/>

⁶https://github.com/silknow/converter/tree/master/competency_questions

Sujet	Nb de Questions	Requête Possible	Résultats effectif
Lieu	9	33,3%	33,3%
Temps	6	50%	40%
Temps et Lieu	4	25%	25%
Matériau	8	12,5%	37,5%
Artistes	8	25%	37,5%
Artistes et Temps	3	66,7%	33,3%
Artistes et Lieu	3	0%	0%
Style	7	28,6%	28,6%
Type d'objets	4	75%	25%
Type d'objets et Matériau	2	100%	100%
Type d'objets, Matériau et Style	2	0%	0%
Type d'objets et Lieu	2	50%	50%
Type d'objets et Temps	1	100%	100%
Type d'objets, Temps et Lieu	2	50%	50%
Type d'objets, Temps, Lieu et Matériau	2	0%	0%
TOTAL	64	39,1%	28,1%

Table 6.2: Résumé de l'évaluation du modèle de données à travers les questions de compétences (à l'exception des questions espagnoles). La couverture est donnée à la fois pour les questions pour lesquelles toute sorte d'interrogation utile était possible et pour les questions auxquelles on pouvait répondre par au moins un résultat.

6.2 Développement d'un graphe de connaissances sur la production d'objets en soie

anglaise et française du thésaurus ont chacune été comparées aux ressources de la langue correspondante. Le programme pour le calcul de la couverture a été écrit en Python. Le prétraitement a été effectué à l'aide de la bibliothèque Natural Language Toolkit (NLTK) [16] qui contient le Stemmer Snowball. Il a été utilisé sur tous les termes et leurs synonymes du thésaurus, ainsi que sur tous les mots provenant de ressources en ligne.

Une façon d'enrichir les données matérialisées dans le KG est de transformer des chaînes (expressions littérales) en choses (objets identifiés par des URI dans le paradigme des données liées). Pour cela, nous utilisons l'outil `string2vocabulary`⁷ et nous prenons des vocabulaires contrôlés existants qui fournissent déjà des identités aux choses en ligne ou nous créons manuellement un tel vocabulaire.

6.2.5 Collecte et conversion des données

Développement d'un crawler et d'un scraper web pour les musées publics

Avec notre crawler⁸, nous sommes en mesure de télécharger des ensembles de données à partir de 18 sources, soit via l'API, soit via l'exploration manuelle de sites Web. Le robot d'exploration est réalisé en Node.js. Il utilise Axios pour les requêtes HTTP et Cheerio pour l'analyse du DOM si nécessaire. Toutes les données sont mises à la disposition du public par les musées ou collections respectifs. Nous recevons deux autres ensembles de données directement de Garin 1820 et de l'Université de Palerme (UNIPA), car ils font partie de SILKNOW.

Logiciel de conversion

Dans tous les cas, à l'exception de celui de Garin 1820 et de l'Université de Palerme / Patrimoine culturel de la Sicile, ce format JSON commun est ensuite utilisé comme base pour le logiciel de conversion⁹ afin de produire des fichiers Terse RDF Triple Language (Turtle) / TTL qui peuvent finalement être téléchargés vers un Triplestore basé sur le serveur universel Virtuoso.

La conversion des données consiste principalement en une traduction des règles de mise en correspondance en algorithmes qui attribuent les classes et les propriétés que nous avons définies sur la base de l'ontologie SILKNOW en fonction des champs originaux des métadonnées du musée. Par exemple, une règle de mappage peut stipuler que les valeurs du champ du musée "Place" doivent être mappées comme une classe `E53_Place`, qui est également la valeur de la propriété `P8_took_place_on_or_within` attachée à la classe `E12_Production`. La figure 6.2 montre une capture d'écran d'une table de correspondance contenant plusieurs règles qui ont été mises en œuvre avec le convertisseur et la figure 6.3.

⁷<https://github.com/DOREMUS-ANR/string2vocabulary>

⁸<https://github.com/silknow/crawler>

⁹<https://github.com/silknow/converter>

Résumé en français

Field	Example	Main Class	Path	Comments
Title	Passementerie		<p>E22_Man-Made-Object P102 has title E35_Title (Passementerie)</p> <p>E17_Type Assignment P41 classified E22_Man-Made-Object</p> <p>E17_Type Assignment P42 assigned E55_Type (Passementerie)</p> <p>E17_Type Assignment P14 carried out by E40_Legal-Body (Victoria and Albert Museum)</p> <p>E17_Type Assignment P2 has type E55_Type (Title)</p>	
Description	'Empty'	S4_Observation	<p>S4_Observation 08 observed E22_Man-Made-Object</p> <p>S4_Observation P3 has note E62_String</p> <p>S4_Observation P2 has type E55_Type (Description)</p>	
Date	18th century	E12_Production	E12_Production P4 has time-span E52_Time-Span P78 is identified by E49_Time Appellation	
Culture	Italian	E12_Production	E12_Production P7 took place at E53_Place	
teaserText	Date: 18th century Accession Number: 08.48.46	WON'T BE MAPPED		
url	https://www.metmuseum.org/art/collection/search/213382?sortBy=AccessionNumber&deptids=12&what=Silk%7cTextil	WON'T BE MAPPED		

Figure 6.2: Partie de la table de correspondance pour l'enregistrement exemplaire "08.48.46" du MET

Mapping rules - created by domain experts

Field label	Value	Class / Property	Full path	Annotation
Medium	Ribbed silk and wool ground embellished with metallic and silk yarn embroidery	E12_Production	E12_Production P3 has note E62_String	This field gives information on the material and the technique. If they are extracted : E12_Production P126 employed E57_Material (silk, wool, metal) E12_Production P32 use general technique E55_Type (ribbed ground, embroidery)
Materials	silk, metallic yarn	E12_Production	E12_Production P126 employed E57_Material	
Techniques	plain weave, embroidery, embroidering, embroidered, appliqué (preferred spelling), applique, ribbed	E12_Production	E12_Production P32 use general technique E55_Type E55_Type P3 has note E62_String	

Mapping excerpt for a record from the Rhode Island School of Design (RISD) museum

3

Figure 6.3: Illustration de la manière dont les règles de cartographie sont mises en œuvre avec le logiciel du convertisseur.

6.2.6 Accès aux données

Intégrer les requêtes SPARQL et leurs résultats dans le développement Web peut être un défi, même lorsque le format de sortie est JSON : il contient des métadonnées inutiles, chaque valeur a un type de données et fait partie d'un tableau plus grand avec son propre nom et les attributs "type" et "valeur" ou des liaisons identiques qui, par exemple, ne diffèrent que par la balise de langue ne sont pas automatiquement fusionnées et affichées plusieurs fois. La mise en correspondance des résultats avec une autre structure peut être difficile, surtout si l'on évite de coder en dur les requêtes dans le code de l'application.

Accès par des requêtes sémantiques - SPARQL Endpoint

Une fois exprimées en RDF, nous téléchargeons toutes les données vers un point de terminaison SPARQL à partir duquel elles peuvent être interrogées. Le graphe de connaissances constitue la base de tous les autres travaux et outils axés sur les données qui font partie de SILKNOW en général. En outre, nous proposons un navigateur à facettes, une API RESTful ainsi qu'un moteur de recherche exploratoire pour faciliter l'accès aux données, qui sont détaillés dans les sous-sections suivantes.

Le langage d'interrogation SPARQL est un langage d'interrogation déclaratif (comme SQL) permettant d'effectuer des opérations de manipulation et de définition de données sur des données représentées comme une collection d'énoncés RDF [125].

Une requête SPARQL comporte un modificateur de solution (ou tête) et un corps de requête. Le modificateur de solution fournit la base pour catégoriser les différents types de solutions de requêtes SPARQL. Le corps de la requête comprend une collection de modèles d'énoncés RDF qui représentent les relations entre les entités auxquelles la requête s'applique. Le modificateur de solution comprend l'accès aux données en lecture (SELECT, ASK, DESCRIBE, CONSTRUCT) et l'accès aux données en écriture (CREATE, INSERT, UPDATE, DELETE, CLEAR, DROP).

Accès pour les développeurs web - SPARQL Transformer

Grâce à une combinaison de grlc¹⁰ et de SPARQL Transformer [81], nous avons pu créer un accès API facile pour le graphe de connaissances SILKNOW, qui permet aux développeurs web de travailler directement avec un format plus adapté¹¹. SPARQL Transformer s'appuie sur un seul objet JSON pour définir quelles données doivent être extraites du point de terminaison et sous quelle forme. Les liaisons SPARQL sont fusionnées sur la base des identifiants et du cadre API grlc. Grâce à son interface graphique, le graphe de connaissances peut également faire l'objet d'une recherche avec SPARQL Transformer pour toute chaîne de caractères du type temps, lieu, matériau ou technique et le résultat est affiché dans un format JSON plus simple.

Accès pour le développement web - RESTful API

Le transfert d'état représentationnel (REST) est un style architectural pour les systèmes hypermédia distribués. C'est aujourd'hui une orientation et un style largement acceptés pour les API web. Les API qui respectent les contraintes REST sont appelées API RESTful. Une partie de leur définition, lorsqu'elles sont basées sur HTTP, comprend les méthodes suivantes pour effectuer des actions sur les ressources : GET, POST, PUT, DELETE.

6.2.7 Représentation de la connaissance humaine à partir de documents muséographiques multilingues

Essayer d'"enseigner" à un ordinateur quelque chose d'aussi spécifique que la production d'objets en soie n'est toujours pas un processus automatique. Le Web sémantique fournit de nombreux outils et méthodes pour intégrer et annoter des métadonnées et des images hétérogènes du patrimoine culturel, c'est pourquoi nous nous appuyons sur de nombreuses étapes testées dans le cadre d'autres projets pour réaliser notre graphe de connaissances, notre thésaurus et tous les outils nécessaires.

Toutefois, nombre de ces méthodes ne fonctionnent pas sans ajustements et extensions, que nous avons réalisés au cours d'un processus impliquant des communications et des

¹⁰<http://grlc.io/>

¹¹<http://grlc.io/api/silknow/api>

discussions avec des experts du domaine et des historiens. Le résultat à ce stade est à la fois un exemple d'application et d'exploration. Nos leçons sont à la fois documentées et conduisent aux implémentations finales.

Sous cette forme, il ne s'agit pas seulement d'une fin en soi, mais aussi d'une base pour de nouvelles recherches sur les données et les connaissances elles-mêmes. Plus nous recueillions de données, plus il devenait évident que de nombreux enregistrements présentaient des lacunes que nous pouvions tenter de combler à l'aide de techniques NLP avancées.

Dans le chapitre suivant, nous décrirons l'utilisation des méthodes d'extraction d'informations et de classification que nous avons utilisées pour explorer la prédiction des valeurs de ces lacunes.

6.3 Préviation des lacunes en matière de métadonnées

L'évocation des textiles en soie européens évoque souvent des images de vêtements et de meubles des anciennes aristocraties et du style de vie somptueux des rois et des reines. De nos jours, les connaissances sur la manière occidentale de produire ces articles coûteux sont toutefois de plus en plus menacées.

Heureusement, de nombreux musées et collections du monde entier possèdent encore des objets en soie, ou du moins des documents publics contenant des métadonnées et des images les illustrant. Ces données muséales spécifiques, provenant de nombreuses sources différentes et concernant des objets du patrimoine culturel en partie vieux de plusieurs siècles, présentent naturellement des lacunes : parfois, l'année ou le lieu de production est inconnu, mais le matériau et la technique utilisés sont décrits ; parfois, une description textuelle riche est fournie avec de nombreux petits détails sur la production de l'objet et ce qu'il représente, mais les valeurs catégorielles informant sur le matériau ou la technique exacte utilisés ne sont pas fournies (Figure 6.4).

Cependant, les progrès récents dans le traitement du langage naturel et plus particulièrement dans l'extraction d'informations peuvent aider à résoudre ces problèmes.

6.3.1 Prédire les lacunes des métadonnées des musées par la classification

L'intégration de données hétérogènes, multilingues et spécifiques à un domaine est un défi, mais réalisable, notamment grâce à de nombreux outils et techniques établis. Cependant, un tel processus met souvent en évidence les lacunes des documents originaux des musées et, éventuellement, du processus de numérisation initial des métadonnées correspondantes : Qu'il s'agisse simplement d'informations catégorielles manquantes, comme l'année, ou de

The figure shows three museum record interfaces.
a) MET (Metropolitan Museum of Art): Record for 'The Hunters Enter the Woods (from the Unicorn Tapestries)'. It includes fields for Title, Date (1495-1505), Geography (Made in Paris, France (cartoon); Made in Southern Netherlands (woven)), and Culture (French (cartoon); South Netherlandish (woven)). It is missing a subject description.
b) MUVE (Musei di Venezia): Record for 'ultimo quarto - Diagonale'. It includes fields for Ambito Culturale (ultimo quart), Denominazione (ultimo quart), Oggetto (Corpetto), and Quantità (1). It is missing material information.
c) SOIERIE (Mobilier National français): Record for 'Bordures pour tenture et rideaux destinés à des salons de Versailles'. It includes fields for Numero d'inventaire (GMMP-733-002), Année de conception (1812), Type (Textile, Tissu précieux), and Matière (Soie, Textile). It is missing technique information.

Figure 6.4: exemples de trois musées différents avec des propriétés catégorielles manquantes : a) pas de description du sujet pour l'enregistrement 37.80.1 du Metropolitan Museum of Art ; b) pas de matériel pour l'enregistrement Cl. XXIV n. 1748 du Musei di Venezia ; c) pas de technique pour l'enregistrement GMMP-733-002 du Mobilier National français.

contraintes liées à la simple mise en correspondance de chaînes de caractères, comme des coquilles ou des incohérences. En outre, dans de nombreux cas, une information importante, comme le lieu de production, est cachée dans une description textuelle riche, mais n'a jamais été explicitement annotée mot par mot.

Le domaine du traitement du langage naturel est aujourd'hui suffisamment avancé pour offrir de nombreuses techniques prometteuses pour pallier à ces problèmes. Dans le contexte de cette thèse, nous avons essayé plusieurs de ces techniques, la plupart d'entre elles étant issues du sous-domaine de la classification non supervisée ou " zero-shot ". Ces différentes approches peuvent être divisées en différentes manières de rendre inutile l'entraînement supervisé d'un modèle, la plupart d'entre elles s'appuyant fortement sur des modèles de langage pré-entraînés. La raison d'une telle orientation est principalement motivée par la grande quantité de classes combinée à un fort déséquilibre de celles-ci et à seulement quelques points de données dans les données que nous pouvons exporter de notre graphe de connaissances. Néanmoins, nous avons également expérimenté des approches supervisées, mais elles nécessitent une réduction des classes.

Après avoir exploré les moyens d'enrichir nos données en prédisant les informations manquantes dans cette partie de la thèse, le prochain et dernier chapitre portera sur l'exploration de notre graphe de connaissances.

6.4 Explorer le patrimoine européen de la soie

De nombreux domaines du patrimoine culturel sont constitués de connaissances qui ne sont pas largement connues du public. Bien que de nombreux objets aient été numérisés, même les experts ont encore du mal à trouver ce qu'ils cherchent dans les catalogues en ligne. La production européenne de tissus en soie est un exemple d'un tel domaine. Déjà relativement obscures pour le public, de nombreuses descriptions et images d'objets existent sous une

forme numérisée, mais sont mises en ligne par de nombreux musées à travers le monde, dans des formats individuels. Ils donnent souvent un accès public aux images et aux métadonnées de ces objets de soie par le biais d'API ou simplement de leurs sites web, mais à l'origine, il n'y a pas eu d'effort d'harmonisation et d'intégration au niveau mondial. Par conséquent, il est très difficile pour le public, les experts historiques et l'industrie (par exemple, la mode) d'accéder à ces connaissances.

Après avoir développé un graphe de connaissances et exploré comment combler les lacunes en matière de métadonnées, le chapitre suivant décrit nos efforts pour rendre toutes les données facilement accessibles et permettre une exploration plus approfondie, même pour les non-experts.

6.4.1 Permettre l'exploration du patrimoine culturel des objets en soie

De nombreux sujets liés au patrimoine culturel sont très spécialisés et il faut souvent consulter des ouvrages spécialisés ou visiter un musée pour comprendre le vocabulaire et la granularité d'un tel domaine. La production historique d'articles en soie en Europe ne fait pas exception à la règle, c'est peut-être même l'un des sujets les plus obscurs. Ce fait, cependant, motive l'exploration de moyens pour offrir non seulement l'accès, mais aussi des outils pour faciliter l'exploration du graphe de connaissances et de toutes ses données que nous avons développé.

Une recherche exploratoire couvre un large éventail d'activités pour répondre aux besoins d'une recherche sans ou avec une cible en tête, et d'un chercheur familier ou non du sujet. ADASilk offre par exemple une exploration basée sur le sujet, mais aussi une recherche avancée avec de nombreux filtres à grain fin, sur la base de nos données intégrées et du graphe de connaissances développé. Il est doté d'une interface web graphique qui est censée être facile à utiliser et suffisamment rapide.

L'une des fonctions les plus avancées d'ADASilk est la recherche d'images d'objets en soie similaires. Cette fonction est basée sur un modèle que nous avons formé en tirant parti des règles de similarité des images établies par les experts du domaine. Grâce à cette fonction, notre moteur de recherche exploratoire est même utilisable sans aucune saisie de texte. Cette fonction représente également un moyen de mesurer la similarité entre deux objets dans le graphe de connaissances, ou même entre une photo externe d'un objet en soie et une photo de notre ensemble de données. De futurs travaux de recherche pourraient explorer d'autres moyens de mesurer et d'établir la similarité entre les objets représentés dans le graphe de connaissances.



Appendix: Full list of competency questions

Location

- Which items were produced in Spain?
- Where were Mudejar-style fabrics produced?
- Where was the production center called Tiraz?
- What was la Fabrique Lyonnaise ?
- Which items have been produced in Italy and are now preserved in France ?
- Give me all the items that are preserved in the Musée des Tissus de Lyon
- What Valencian fabrics are located in the Spanish royal collections?
- In which museums and collections around the world are located the Spanish textiles?
- Give me a list of textile factories in a Florence

Time

- Which items were produced during the 16th century?
- What are the common decorative elements in 16th century fabrics?
- Which fabric became popular in Italy in the fifteenth century?
- What kinds of fabrics / weaving techniques / designs were most frequent in 18th-century France? Please give me a list of the top 5 (or 10, 15. . .) occurrences in a particular field.
- Which items have been produced in 1815?
- What are the most common decorative motifs in the Hispanic Middle Ages?

Time and location

- Which items were produced in France during the 18th century?
- Give me all the items that have been produced after 1750 in France.
- Give me all the items that are preserved in the Musée des Tissus de Lyon, and that have been produced between 1650 and 1750.
- Who (person, institution ...) was the main textile French producer during the XVII?

Materials

- Which items were produced with silk and silver?
- When does the "a pizzo" design become popular?
- When does the "bizarre" design become popular?
- What is the Blonda?
- What is the Buratto?
- Where does the name of the Batista fabric come from?
- Give me the objects that involve at most silk, silver and wool.
- Give me the objects that involve silk, silver and wool, except those that involve gold.

Artists

- Which items have been created by Philippe de la Salle ?
- Give me all the information you have on Philippe de la Salle
- Give me all the items inspired by a work of Giambologna
- Give me all the items designed by François Boucher
- Give me all the items designed by Italian artists
- Are there items designed by French artists in the 17th century?
- Give all the items for which the designer has been influenced by Philippe de la Salle
- Who were the printers or engravers that produced graph paper for making mise-en-cartes?

Artist and time

- Give me all the items designed by François Bouchez in the 18th century
- Give me all the items created by Philippe de la Salle in the last 5 years of his life.
- Give me a list of designers from a Valencia during the 19th century

Artists and location

- Give me all the designers who were born in England
- Give me all the designers who were trained in Italy
- Give me all the designers who were trained in Italy and in France

Style

- Who is the Revel style name after?
- Give me all the items that have been influenced by oriental fashion.
- Give me all the items with flowers on them.
- Give me all the items with hearts and flowers on them
- Give me all the items with purple
- Who was the introducer of the realistic style in textiles?
- Give me examples of textile designs that appear in paintings.

Type of items

- Give me all the scarves
- Give me all the dresses that have been worn with a petticoat
- Give examples of textiles that conserve both the fabric and the mise-en-carte
- When do the first mise-en-carte appeared?

Type of items and materials

- Give me all the ribbons with cotton

Résumé en français

- Give me all the dresses with silk, cotton and gold

Type of items, materials and style

- Give me all the scarves with cotton and with hearts on them
- Give me examples of imitations or revivals of textiles during the 18th century

Type of items and location

- Give me the religious clothing produced in Spain
- What textiles belonged to the collector Mariano Fortuny?

Type of items and time

- Give me all the dresses produced during the Victorian era

Type of items, time and location

- Give me all the clothes produced in Spain during the Renaissance.
- Give me all the scarves that have been produced in England between 1800 and 1850.

Type of items, time, location and material

- Give me all the ribbon involving silver and produced in Italy during the Renaissance
- Give me those ornamental motifs from classical antiquity that appear in fabrics, mises-en-carte and designs ... Organized by chronology, location, place of origin ...

Questions in Spanish

- ¿Cuáles son los motivos decorativos más habituales en la Edad Media hispánica?
- ¿Qué tejidos valencianos hay en las colecciones reales españolas?
- ¿Qué tejidos españoles hay en diferentes museos y colecciones?
- Dame ejemplos de piezas en los que se conserva tejido y puesta en carta.
- Dime todos los tejidos que pertenecieron al coleccionista Mariano Fortuny (provenance)

6.4 Explorer le patrimoine européen de la soie

- Dime motivos ornamentales de la antigüedad clásica que aparecen en tejidos, puestas en carta, diseños... Organizados por cronología, ubicación, lugar de origen...
- ¿Quién fue el introductor del estilo realista en tejidos?
- ¿Quién (persona, institución...) es el principal productor francés de tejidos en el XVII?
- ¿Cuándo aparecen los espolinados?
- ¿Cuándo aparecieron las primeras puestas en carta sobre papel milimetrado impreso?
- ¿Qué impresores o grabadores produjeron papel milimetrado para puestas en carta?
- Dame una lista de talleres o fábricas textiles de una ciudad.
- Dime una lista de diseñadores de una ciudad o región durante un periodo.
- Dame ejemplos de diseños textiles que aparecen en pinturas.
- Dame ejemplos de imitaciones, revivals, copias, falsificaciones, ... (copias de modelos antiguos hechas mucho tiempo después)

Bibliography

- [1] *QMiner: Data Analytics Platform for Processing Streams of Structured and Unstructured Data*, Montreal, Canada, 2014.
- [2] Rabiya Abbas, Zainab Sultan, and Shahid Nazir Bhatti. Comparative analysis of automated load testing tools: Apache jmeter, microsoft visual studio (tfs), loadrunner, siege. In *2017 International Conference on Communication Technologies (ComTech)*, pages 39–44, 2017.
- [3] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [4] E. Alba. Catálogo e inventario como instrumentos para la gestión del patrimonio cultural. *Educación y entorno territorial de la Universitat de València*, pages 67–93, 2014.
- [5] Chris Alberti, Daniel Andor, Emily Pitler, Jacob Devlin, and Michael Collins. Synthetic qa corpora generation with roundtrip consistency. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6168–6173, 2019.
- [6] Peggy M. Andersen, Philip J. Hayes, Steven P. Weinstein, Alison K. Huettner, Linda M. Schmandt, and Irene B. Nirenburg. Automatic extraction of facts from press releases to generate news stories. In *Third Conference on Applied Natural Language Processing*, pages 170–177, Trento, Italy, March 1992. Association for Computational Linguistics.
- [7] Pepa Atanasova, Jakob Grue Simonsen, Christina Lioma, and Isabelle Augenstein. A diagnostic study of explainability techniques for text classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3256–3274, 2020.
- [8] M. Baca. Fear of authority? authority control and thesaurus building for art and material culture information. *Cataloguing & Classification Quarterly*, 38:143–151, 2004.

Bibliography

- [9] Alexei Baevski, Sergey Edunov, Yinhan Liu, Luke Zettlemoyer, and Michael Auli. Cloze-driven pretraining of self-attention networks. *CoRR*, abs/1903.07785, 2019.
- [10] M. Barroso-Ruiz. La normalización terminológica en los museos. el tesoro. *Revista General de Información y Documentación*, 2(4):121, 1994.
- [11] Abdelhak Belhi, Abdelaziz Bouras, and Sebti Foufou. Leveraging known data for missing label prediction in cultural heritage context. *Applied Sciences*, 8(10), 2018.
- [12] Eyal Ben-David, Nadav Oved, and Roi Reichart. PADA: A prompt-based autoregressive approach for adaptation to unseen domains. *CoRR*, abs/2102.12206, 2021.
- [13] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *J. Mach. Learn. Res.*, 3(null):1137–1155, mar 2003.
- [14] J.L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [15] C. Bezerra, F. Freitas, and F. Santana. Evaluating Ontologies with Competency Questions. In *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pages 284–285, 2013.
- [16] Steven Bird, Edward Loper, and Ewan Klein. *Natural Language Processing with Python*. O’Reilly Media Inc., 2009.
- [17] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York (NY), USA, 1st edition, 2006.
- [18] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data: The Story so Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.
- [19] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.
- [20] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a "Siamese" time delay neural network. In *Advances in Neural Information Processing Systems (NIPS)*, pages 737–744, 1994.
- [21] A. Carretero. *Normalización documental de museos: elementos para una aplicación informática de gestión informática*. Ministerio de Educación y Cultura, Dirección General de Bellas Artes y Bienes Culturales, Madrid, 1st edition, 1998.

-
- [22] Valentina Anita Carriero, Aldo Gangemi, M. Mancinelli, L. Marinucci, Andrea Giovanni Nuzzolese, V. Presutti, and Chiara Veninata. ArCo: the Italian Cultural Heritage Knowledge Graph. In *International Semantic Web Conference (ISWC)*, 2019.
- [23] Ying-Hong Chan and Yao-Chung Fan. A recurrent bert-based model for question generation. In *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, pages 154–162, 2019.
- [24] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.
- [25] Angelo V Ciardiello. Did you ask a good question today? alternative cognitive and metacognitive strategies. *Journal of adolescent & adult literacy*, 42(3):210–219, 1998.
- [26] Dan Claudiu Cireșan, Ueli Meier, Luca Maria Gambardella, and Jürgen Schmidhuber. Convolutional neural network committees for handwritten character classification. In *2011 International conference on document analysis and recognition*, pages 1135–1139. IEEE, 2011.
- [27] D. Clermont, M. Dorozynski, D. Wittich, and F. Rottensteiner. Assessing the Semantic Similarity of Images of Silk Fabrics Using Convolutional Neural Networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:641–648, 2020.
- [28] D. Clermont, M. Dorozynski, D. Wittich, and F. Rottensteiner. Assessing the semantic similarity of images of silk fabrics using convolutional neural networks. In *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, volume V-2, page 641–648, 2020.
- [29] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [30] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. ICML '08, page 160–167, New York, NY, USA, 2008. Association for Computing Machinery.
- [31] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July 2020. Association for Computational Linguistics.

Bibliography

- [32] Alexis Conneau, German Kruszewski, Guillaume Lample, Loïc Barrault, and Marco Baroni. What you can cram into a single \mathbb{R}^d vector: Probing sentence embeddings for linguistic properties. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2126–2136, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [33] Asociación Española de Normalización y Certificación. *UNE 50124: Documentación: Directrices para la creación y desarrollo de tesauros multilingües*. AENOR, Madrid, 1st edition, 1997.
- [34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [35] Michael J. Denney, Dustin M. Long, Matthew G. Armistead, Jamie L. Anderson, and Baqiyyah N. Conway. Validating the extract, transform, load process used to populate a large clinical research database. *International Journal of Medical Informatics*, 94:271–274, 2016.
- [36] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [37] Bhuwan Dhingra, Danish Danish, and Dheeraj Rajagopal. Simple and effective semi-supervised question answering. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 582–587, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [38] Chris Dijkshoorn, Lizzy Jongma, Lora Aroyo, J. V. Ossenbruggen, G. Schreiber, Wesley ter Weele, and J. Wielemaker. The Rijksmuseum collection as Linked Data. *Semantic Web*, 9:221–230, 2018.
- [39] Jesse Dodge, Gabriel Ilharco, Roy Schwartz, A. Farhadi, Hannaneh Hajishirzi, and Noah A. Smith. Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping, 2020.
- [40] M. Doerr. The cidoc crm, an ontological approach to schema heterogeneity. In *Semantic Interoperability and Integration*, 2005.

-
- [41] Doerr, Martin, and Kritsotaki, Athina, and Rousakis, Yannis, and Hiebel, Gerald, and Theodoridou, Maria, and others. Version 1.2 | CRMsci, 2014.
- [42] Doerr, Martin, and Stead, Stephen, and Theodoridou, Maria, and others. Version 3.2 | CRMdig, 2014.
- [43] Doerr, Martin and Theodoridou, Maria. D14.1: Extended CRM – ARIADNE Infrastructure, 2016.
- [44] M Dorozynski, D Clermont, and F Rottensteiner. Multi-task deep learning with incomplete training samples for the image-based prediction of variables describing silk fabrics. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4 (2019), Nr. 2/W6, 4(2/W6):47–54*, 2019.
- [45] Thibault Ehrhart, Pasquale Lisena, and Raphaël Troncy. KG Explorer: a Customisable Exploration Tool for Knowledge Graphs. In *6th International Workshop on Visualization and Interaction for Ontologies and Linked Data (VOILA)*, Online, 2021.
- [46] Lea Frermann, Shay B Cohen, and Mirella Lapata. Whodunnit? crime drama as a case for natural language understanding. *Transactions of the Association for Computational Linguistics*, 6:1–15, 2018.
- [47] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [48] Mar Gaitán and Arabella León. Silknow deliverable 7.6 - silknow system evaluation. Technical report, University of Valencia, Garín 1820, 2021.
- [49] F. García-Marco. Normas y estándares para la elaboración de tesauros de patrimonio cultural. In *Cultura y Deporte Secretaría General Técnica. Centro de Publicaciones, Ministerio de Educación*, editor, *El lenguaje sobre el patrimonio: estándares documentales para la descripción y gestión de colecciones*, pages 29–46. Ministerio de Educación, Cultura y Deporte, Madrid, 2016.
- [50] A. Gordo, J. Almazán, J. Revaud, and D. Larlus. Deep image retrieval: Learning global representations for image search. In *European Conference on Computer Vision (ECCV)*, pages 241–257, 2016.
- [51] Derek Greene and Pádraig Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *23rd International Conference on Machine learning (ICML)*, pages 377–384, 2006.
- [52] G.Schreiber, A. Amin, M. Van Assem, V. Der Voer, L. Hardman, M. Hildebrand, B. Ome-layenko, J. Van Ossembruggen, A. Today, J. Wielemaker, and B. Vielinga. Semantic

Bibliography

- annotation and search of cultural-heritage collections: The multimedial e-culture demonstrator. *Journal of Web Semantics*, 6(4):243–249, 2008.
- [53] Antonio Gulli. *AG's corpus of news articles*, 2005.
- [54] C. Gunzburger. *The Textile Museum Thesaurus*. Textile Museum, Washington, 1st edition, 2005.
- [55] Himanshu Gupta, Amogh Badugu, Tamanna Agrawal, and Himanshu Sharad Bhatt. Zero-shot open information extraction using question generation and reading comprehension, 2021.
- [56] P. Harpring and M. Baca. *Introduction to Controlled Vocabularies: terminology for art, architecture and other cultural works*. The Getty Research Institute, Los Angeles, 1st edition, 2015.
- [57] Ismail Harrando and Raphael Troncy. Explainable Zero-Shot Topic Extraction Using a Common-Sense Knowledge Graph. In *3rd Conference on Language, Data and Knowledge (LDK)*, Zaragoza, Spain, 2021.
- [58] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645, 2016.
- [59] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
- [60] E. Hyvönen. *Publishing and using cultural heritage linked data on the semantic web. Synthesis lectures on the semantic web: theory and technology*. Morgan & Cleypool publishers, Espoo, 1st edition, 2012.
- [61] Alon Jacovi and Yoav Goldberg. Towards faithfully interpretable NLP systems: How should we define and evaluate faithfulness? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4198–4205, Online, July 2020. Association for Computational Linguistics.
- [62] Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.
- [63] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain, April 2017. Association for Computational Linguistics.

-
- [64] Ridwan Andi Kambau, Zainal Arifin Hasibuan, and M. Octaviano Pratama. Classification for Multiformat Object of Cultural Heritage using Deep Learning. In *3rd International Conference on Informatics and Computing (ICIC)*, 2018.
- [65] Kenneth C. W. Kammeyer and Julius A. Roth. Coding responses to open-ended questions. *Sociological Methodology*, 3:60–78, 1971.
- [66] Ralph Kimball and Joe Caserta. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley, Indianapolis, IN, 2004.
- [67] D. P Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR 2015)*, 2015.
- [68] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.
- [69] Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. Text classification algorithms: A survey. *Information*, 10(4), 2019.
- [70] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS'12)*, volume 1, pages 1097–1105, 2012.
- [71] Ken Lang. Newsweeder: Learning to filter netnews. In *12th International Conference on Machine Learning (ICML)*, pages 331–339, 1995.
- [72] LARHRA. OntoME SILKNOW project, 2019.
- [73] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1188–1196, Beijing, China, 22–24 Jun 2014. PMLR.
- [74] M. Lecron-Foster. Symbolisms: the foundation of culture. In Tim Ingold, editor, *Companion Encyclopedia of Anthropology. Culture and Social Life*, pages 366–394. Routledge, London, 1997.
- [75] Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, 10:707–710, 1965.
- [76] Omer Levy, Yoav Goldberg, and Ido Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225, 2015.

Bibliography

- [77] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *ACL*, 2020.
- [78] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July 2020. Association for Computational Linguistics.
- [79] Xiaoya Li, Xiaofei Sun, Yuxian Meng, Junjun Liang, Fei Wu, and Jiwei Li. Dice loss for data-imbalanced nlp tasks, 2019.
- [80] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, Oct 2017.
- [81] Pasquale Lisena, Albert Meroño-Peñuela, Tobias Kuhn, and Raphaël Troncy. Easy Web API Development with SPARQL Transformer. In *18th International Semantic Web Conference (ISWC), In-Use Track*, Auckland, New Zealand, 2019.
- [82] Pasquale Lisena, Albert Meroño-Peñuela, Tobias Kuhn, and Raphaël Troncy. Easy Web API Development with SPARQL Transformer. In *18th International Semantic Web Conference (ISWC), In-use Track*, Auckland, New Zealand, 2019.
- [83] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv e-prints*, 2021.
- [84] Wei Liu, Lin Chen, and Yajun Chen. Age classification using convolutional neural networks with the multi-class focal loss. *IOP Conference Series: Materials Science and Engineering*, 428:012043, oct 2018.
- [85] Luis Enrico Lopez, Diane Kathryn Cruz, Jan Christian Blaise Cruz, and Charibeth Cheng. Simplifying paragraph-level question generation via transformer language models. In Duc Nghia Pham, Thanaruk Theeramunkong, Guido Governatori, and Fenrong Liu, editors, *PRICAI 2021: Trends in Artificial Intelligence*, pages 323–334, Cham, 2021. Springer International Publishing.
- [86] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019.

- [87] Stephen Mayhew, Tatiana Tsygankova, Francesca Marini, Zihan Wang, Jane Lee, Xiaodong Yu, Xingyu Fu, Weijia Shi, Zian Zhao, and Wenpeng Yin. Karthikeyan k, jamaal hay, michael shur, jennifer sheffield, and dan roth. 2019b. university of pennsylvania lorehlt 2019 submission. Technical report, Technical report, 2019.
- [88] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.
- [89] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- [90] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [91] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [92] Judith S Nappi. The importance of questioning in developing critical thinking skills. *Delta Kappa Gamma Bulletin*, 84(1):30, 2017.
- [93] Cheikh Niang, Claudia Marinica, Beatrice Markhoff, Elise Leboucher, Olivier Malavergne, Luc Bouiller, Claude Darrieumerlou, and Francois Laissus. Supporting semantic interoperability in conservation-restoration domain: The parcours project. *J. Comput. Cult. Herit.*, 10(3), jul 2017.
- [94] M. Nielsen. Thesaurus construction: key issues and selected readings. *Cataloguing and Classification quarterly*, pages 57–64, 2004.
- [95] Christian Emil Ore and Martin Doerr. Version 6.2.1 | CIDOC CRM, 2015.
- [96] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [97] Orestis Papakyriakopoulos, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. Bias in Word Embeddings. In *Conference on Fairness, Accountability, and Transparency (FAT)*, pages 446–457, 2020.
- [98] Bhargavi Paranjape, Julian Michael, Marjan Ghazvininejad, Hannaneh Hajishirzi, and Luke Zettlemoyer. Prompting contrastive explanations for commonsense reasoning tasks. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4179–4192, Online, August 2021. Association for Computational Linguistics.

Bibliography

- [99] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [100] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.
- [101] Cristina Portalés, Javier Sevilla, Pablo Casanova, Thibault Ehrhart, and Raphaël Troncy. Silknow deliverable 6.6 - functional evaluation report. Technical report, University of Valencia, EURECOM, 2021.
- [102] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67, 2020.
- [103] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100, 000+ questions for machine comprehension of text. In *EMNLP*, 2016.
- [104] N. Rodríguez-Ortega. Construcción y uso de terminologías, categorías de descripción y estructuras semanticas vinculadas al patrimonio en la sociedad global de datos. *El lenguaje sobre el patrimonio. Estándares documentales para la descripción y gestión de colecciones*, pages 115–130, 2016.
- [105] Peter Rogiest. Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works, patricia harpring, introduction to series. los angeles: Getty research institute, 2010. 245 p. ill. isbn 9781606060186. 50.00 (paper). *Art Libraries Journal*, 36:39–41, 01 2011.
- [106] Dongyu Ru, Zhenghui Wang, Lin Qiu, Hao Zhou, Lei Li, Weinan Zhang, and Yong Yu. *QuAChIE: Question Answering Based Chinese Information Extraction System*, page 2177–2180. Association for Computing Machinery, New York, NY, USA, 2020.
- [107] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *CoRR*, abs/1706.05098, 2017.
- [108] Vasile Rus, Zhiqiang Cai, and Art Graesser. Question generation: Example of a multi-year evaluation campaign. *Proc WS on the QGSTEC*, 2008.

- [109] Timo Schick and Hinrich Schütze. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 255–269, 2021.
- [110] Thomas Schleider, Thibault Ehrhart, Pasquale Lisena, and Raphaël Troncy. Silkknow knowledge graph, November 2021.
- [111] Thomas Schleider and Raphael Troncy. Zero-shot information extraction to enhance a knowledge graph describing silk textiles. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 138–146, Punta Cana, Dominican Republic (online), November 2021. Association for Computational Linguistics.
- [112] Thomas Schleider, Raphaël Troncy, Mar Gaitan, Ester Alba, and et al. The silkknow knowledge graph. *Semantic Web Journal, Special Issue on Cultural Heritage and Semantic Web, March 2021, IOS Press*, 2021. © IOS Press. Personal use of this material is permitted. The definitive version of this paper was published in *Semantic Web Journal, Special Issue on Cultural Heritage and Semantic Web, March 2021, IOS Press* and is available at :.
- [113] F. Schroff, D. Kalenichenko, and J. Philbin. A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.
- [114] Hinrich Schütze. Word space. In S. Hanson, J. Cowan, and C. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, 1992.
- [115] Holger Schwenk. Continuous space language models. *Computer Speech & Language*, 21(3):492–518, 2007.
- [116] Shaban Shabani, Maria Sokhn, and Heiko Schuldt. Hybrid Human-Machine Classification System for Cultural Heritage Data. In *2nd Workshop on Structuring and Understanding of Multimedia HeritAge Contents (SUMAC)*, pages 49—56, 2020.
- [117] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.
- [118] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *NAACL-HLT (2)*, 2018.
- [119] Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts . In *Empirical Methods in Natural Language Processing (EMNLP)*, page 4222–4235, 2020.

Bibliography

- [120] R. Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *31st AAAI Conference on Artificial Intelligence*, pages 4444–4451, 2017.
- [121] Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *31st AAAI Conference on Artificial Intelligence*, 2017.
- [122] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(2014):1929–1958, 2014.
- [123] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS’14*, page 3104–3112, Cambridge, MA, USA, 2014. MIT Press.
- [124] Sebastian Thrun and Lorien Pratt. Learning to learn: Introduction and overview. In *Learning to learn*, pages 3–17. Springer, 1998.
- [125] Raphaël Troncy. Silknow deliverable 6.3 - ontology web server. Technical report, EURECOM, 2018.
- [126] Raphaël Troncy, Thibault Ehrhart, Pasquale Lisena, Thomas Schleider, Dunja Mladenic, Javier Sevilla, Pablo Casanova, and Manolo Pérez. Silknow deliverable 6.5 - integrated system. Technical report, EURECOM, Jožef Stefan Institute, University of Valencia, 2020.
- [127] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [128] Maurizio Vitella, Valeria Seidita, Georgia Lo Cicero, Dunja Mladenic, Beshar Massri, Franz Rottensteiner, and Raphael Troncy. Silknow deliverable 7.1 - testing report in a controlled scenario. Technical report, University of Palermo, Jožef Stefan Institute, Leibniz University Hannover, ERecordsURECOM, 2020.
- [129] Orion Weller, Nicholas Lourie, Matt Gardner, and Matthew E. Peters. Learning from Task Descriptions. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1361–1375, 2020.
- [130] Sarah Wiegrefe, Ana Marasović, and Noah A. Smith. Measuring association between labels and free-text rationales. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10266–10284, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

- [131] Kenneth Wilhelmsson. Automatic question generation from Swedish documents as a tool for information extraction. In *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*, pages 323–326, Riga, Latvia, May 2011. Northern European Association for Language Technology (NEALT).
- [132] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics.
- [133] Zhiquan Ye, Yuxia Geng, Jiaoyan Chen, Jingmin Chen, Xiaoxiao Xu, Suhang Zheng, F. Wang, J. Zhang, and Huajun Chen. Zero-shot Text Classification via Reinforced Self-training. In *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.
- [134] Zhiquan Ye, Yuxia Geng, Jiaoyan Chen, Jingmin Chen, Xiaoxiao Xu, SuHang Zheng, Feng Wang, Jun Zhang, and Huajun Chen. Zero-shot text classification via reinforced self-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3014–3024, 2020.
- [135] Wenpeng Yin, Jamaal Hay, and Dan Roth. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In *International Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3914–3923, 2019.
- [136] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems 27 (NIPS'14)*, volume 2, pages 3320–3328, 2014.
- [137] Michelle Yuan, Hsuan-Tien Lin, and Jordan L. Boyd-Graber. Cold-start active learning through self-supervised language modeling. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.
- [138] Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. Integrating semantic knowledge to tackle zero-shot text classification. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1031–1040, Minneapolis, Minnesota, 2019.
- [139] Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. Integrating semantic knowledge to tackle zero-shot text classification. In *NAACL-HLT (1)*, 2019.

Bibliography

- [140] Tianyang Zhao, Zhao Yan, Yunbo Cao, and Zhoujun Li. A unified multi-task learning framework for joint extraction of entities and relations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14524–14531, May 2021.
- [141] Ruiqi Zhong, Kristy Lee, Zheng Zhang, and Dan Klein. Adapting language models for zero-shot learning by meta-tuning on dataset and prompt collections. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2856–2878, 2021.