



HAL
open science

Memory Island : Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang

► **To cite this version:**

Bin Yang. Memory Island : Visualizing Hierarchical Knowledge as Insightful Islands. Other [cs.OH]. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066137 . tel-01241135

HAL Id: tel-01241135

<https://theses.hal.science/tel-01241135>

Submitted on 10 Dec 2015

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.

Université Pierre et Marie Curie

École doctorale Informatique, Télécommunications et Électronique (Paris)

(ED130)

Lip6 ACASA Group

MEMORY ISLAND: VISUALIZING HIERARCHICAL KNOWLEDGE AS INSIGHTFUL ISLANDS

Présentée Par

Bin Yang

Thèse de doctorat d'Informatique

Dirigée par Monsieur le professeur **Jean-Gabriel Ganascia**

Présentée et soutenue publiquement le 08 Juin 2015

Devant un jury composé de:

Monsieur Dragicevic, Pierre, Chargé de Recherche, Examineur

Professeur Ganascia, Jean-Gabriel, Professeur, Directeur de these

Professeur Kuntz-Cosperec, Pascale, Professeur, Rapporteur

Professeur Labat, Jean-Marc, Professeur, Examineur

Monsieur Lecolinet, Eric, Maître de Conférences, Examineur

Professeur Venturini, Gilles, Professeur, Rapporteur

*This thesis is dedicated to
my parent and my wife Xiang.
Thank you for the wealth of love.*

ABSTRACT

This thesis is devoted to the study of an original cartographic visualization approach named Memory Island. We discuss how hierarchical knowledge can be meaningfully mapped and visualized as an insightful island. Our technique is inspired by the "*loci*" (plural of Latin "*locus*" for places or locations) method of the ancient "Art of Memory" technique. A well-designed map in mind can make sense of knowledge, which leads to the accomplishment of one's information seeking tasks, and helps to extend one's knowledge. To this end, Memory Island technique consists of associating each entity of knowledge to a designated area on a created virtual island. With the geographic visual metaphors we define, Memory Island can present phenomena found in knowledge, which is often difficult to understand.

In this thesis, we discuss how we design our visualization technique to make it achieve the great features of visualization: automatically generate a truthful, functional, beautiful, insightful, and enlightening island with its technical details. In order to make Memory Island more convenient for its users, we present our "overview+detail" interface, to support them with visual exploration and knowledge analysis. We also demonstrate how to create knowledge maps using Memory Island technique, by giving some example on different datasets of Digital Humanities (Project OBVIL), e-books (Project LOCUPLETO) and other domains.

Then, we propose our validation and evaluation protocols with two preliminary user experiments. The results from these studies indicate that the use of Memory Island provides advantages for non-experienced users tackling realistic browsing, helps them improve their performances in knowledge navigation and memorization tasks, and that most of them choose to use it for navigation and knowledge discovery.

We end up by concluding our researches and listing some perspectives and future works that can be based on our Memory Island technique.

TITRE EN FRANÇAISE « ÎLES DE MEMOIRES: UNE NOUVELLE APPROCHE POUR LA VISUALISATION INTUITIVE DES CONNAISSANCES HIERARCHIQUES»

RÉSUMÉ

Dans cette thèse nous étudions une nouvelle approche de visualisation cartographique appelée « îles de mémoires ». Le terme « îles de mémoires » a été inspiré par la méthode des «*loci*» (pluriel de «*locus*» en latin qui signifie « endroit » ou « lieu ») de l'ancien « Art de la mémoire ». Une carte bien représentée dans l'esprit peut donner un sens à la connaissance, ce qui améliore une de recherche d'information (une recherche intuitive), et contribue à enrichir les connaissances issues de cette carte. Pour cela, la technique « îles de mémoires » consiste à associer chaque entité de connaissance à un endroit désigné sur une île virtuelle. Grâce aux métaphores géographiques que nous avons définies, une représentation en « îles de mémoires » peut inférer des phénomènes souvent difficile à identifier et comprendre dans la connaissance.

Dans une première partie, nous détaillons notre approche de visualisation d'une hiérarchie de connaissances en île de mémoire. Nous présentons les algorithmes que nous avons définis pour générer automatiquement une belle carte réaliste, fonctionnelle, intuitive et inspirante. Nous présentons aussi l'interface de visualisation "overview+detail" qui permet de naviguer dans les îles de mémoire.

Dans une deuxième partie, nous détaillons les expérimentations réalisées avec notre outil dans le cadre du projet LOCUPLETO et des exemples issus du domaine des humanités numériques (Projet OBVIL, InPhO, etc.). Les résultats obtenus avec notre approche de visualisation sont prometteuses. En effet, les résultats démontrent que la navigation est intuitive et est capable d'augmenter la mémorisation des connaissances chez les utilisateurs de l'outil.

Nous concluons notre thèse par le bilan des travaux menées et nous proposons un ensemble de travaux futurs basé sur notre approche de visualisation « îles de mémoires ».

ACKNOWLEDGEMENTS

This work is supported by the French National Research Project (investissement d'avenir) LOCUPLETO. This work has been possible only because of the French National Research Project (investissement d'avenir) LOCUPLETO. I thank all the participants of the Project LOCUPLETO and the participants of our user studies.

Professor Jean-Gabriel Ganascia has been the ideal thesis supervisor. His sage advice, insightful criticisms, and patient encouragement aided the working and writing of this thesis in innumerable ways. I want to thank Raluca Tocmag, Suradej Panich, Marie Laxenaire, Soroosh Nalchigar and Fabio Pardo for their interns and supports for the evaluation experiments.

I would like to thank for Professor Gilles Venturini and Professor Pascale Kuntz-Cosperec for their interests they have shown with this thesis, and their agreement to examine this thesis in detail as reviewers. I also appreciate Professor Jean-Marc Labat, Dr. Pierre Dragicevic and Dr. Eric Lecolinet for accepting to examine this work and it's my pleasure that they can participate in my thesis committee. I also want to give my specific thanks to Dr. Jean-Daniel Fekete, even unfortunately, he cannot by my thesis committee as he is on sabbatical year on the USA. He suggest me to invite Dr. Pierre Dragicevic who is an ideal thesis committee.

I have had the good fortune to be able to present my research in several conferences and workshops, and would like to thank everyone who made this possible, including anonymous reviewers who offered valuable input. I would like to thank also the Association EGC, and AFIHM (L'Association Francophone d'Interaction Homme-Machine), who have organized the workshops on InfoVis, give me a chance to discuss our idea with others. I thank the Cost Action TD1210, who bring me the opportunities and support me to present my research in many places in Europe. I also thanks to the members of Labex OBVIL, who give me the rights to apply Memory Island on their dataset and the encouragements from them. I also thank the editors, Aida Slavic and Almila Akdag Salah of the International UDC Seminar 2013 proceedings for valuable comments and suggestions for improving our works.

I would like to thank for Professor Colin Allen of Indiana University Bloomington for his comments and helps when we work on his InPhO ontology. I would like also thanks the Research Scientist Michaël AUPETIT of Qatar Computing Research Institute, he give me some reference on how to design a visualization tools.

I appreciate the encouragement and useful suggestions of Professor Andre Skupin of San Diego State University, Professor Charles van den Heuvel of University of Amsterdam and Xia Lin of Drexel University for the valuable discussion on Memory Island.

I want to give my specific thanks to the people in ACASA Team at Lip6: Christophe Jouis, Gauvain Bourgne, Mihnea Tufis, Alexandre Bazin, Mohamed Amine Boukhaled, Suzanne

Mpouli Njanga Seh, Fiona Berreby, Alaa Abi Haidar Zied Sellami and Francesca Frontini. I have passed wonderful three years with all of you for both my research and my stays in Paris.

Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK and the IFD department of University Pierre and Marie Curie for producing their Microsoft Word thesis templates used to produce this document.

In the end, I would like to give my specific thanks to Marco Quaggiotto, Martin Rosvall and Andrea Scharnhorst who allow me to reproduce your images in my thesis.

CONTENTS

PART I. BACKGROUND AND LITERATURE REVIEW	1
CHAPTER 1 GENERAL INTRODUCTION	3
1.1 <i>Overview and Research Context</i>	5
1.2 <i>Industrial Context</i>	6
1.3 <i>Contributions and Related Research Areas</i>	7
1.4 <i>Organization</i>	8
CHAPTER 2 REVIEW OF LITERATURE.....	11
2.1 <i>Information Visualization and knowledge visualization</i>	13
2.2 <i>InfoVis Toolkit and InfoVis Tools</i>	13
2.3 <i>Design and Evaluation Visualization Systems</i>	15
2.4 <i>Tree and Hierarchic Knowledge Visualizations</i>	16
2.4.1 <i>Tree, Hierarchic knowledge and Skeleton</i>	16
2.4.2 <i>Classical Tree Visualization Techniques</i>	17
2.5 <i>Ontology and Ontology Visualization</i>	21
2.6 <i>Knowledge Maps</i>	23
2.7 <i>Cartography and Knowledge Cartography</i>	25
2.7.1 <i>Cartography</i>	25
2.7.2 <i>Knowledge Cartography</i>	26
2.8 <i>Cartographic Visualization</i>	28
2.8.1 <i>Treemapping Approaches</i>	28
2.8.2 <i>Spatialization (Layout) and Tree drawing algorithms</i>	30
2.8.3 <i>Map-like Visualization Approaches</i>	30
2.9 <i>Schematization</i>	32
2.10 <i>Spatial Cognition</i>	33
PART II. MEMORY ISLAND TECHNIQUE	35
CHAPTER 3 MEMORY ISLAND IDEA AND DESIGN METHODOLOGY.....	37
3.1 <i>The arts of Memory technique and the notion of Memory Islands</i>	39
3.2 <i>The Objectives of Memory Island technique</i>	41
3.2.1 <i>Recognition (truthful)</i>	41
3.2.2 <i>Discovery (functional and beautiful)</i>	41
3.2.3 <i>Surprise (insightful and enlightening)</i>	42
3.3 <i>The idea of Memory Island</i>	42
3.4 <i>Map (Landscape) Metaphor</i>	43
3.5 <i>Why do we choose 2D traditional map representation</i>	43
3.6 <i>Geographic Metaphors and Cartographic Means</i>	44
3.6.1 <i>Proportion Metaphor</i>	44
3.6.2 <i>Distance/centrality Metaphor</i>	45
3.6.3 <i>Font Attributes and Point Attributes</i>	46

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

3.6.4 Colors	48
3.6.5 Paths and Gaps	49
3.7 <i>Memory Island Prototype Algorithm</i>	50
CHAPTER 4 HIERARCHICAL REORGANIZATION BY SEMANTIC SIMILARITY	53
4.1.1 Related works.....	55
4.1.2 Hierarchical Reorganization by Semantic Similarity	56
CHAPTER 5 VISUALIZING HIERARCHICAL DATA AS ISLANDS	59
5.1 <i>Island Generation</i>	61
5.1.1 Polyle II Algorithm.....	61
5.1.2 Memory Island Algorithm	64
5.1.3 Discussion on Memory Island Algorithm.....	67
5.2 <i>Reshaping the Resulting Island</i>	68
CHAPTER 6 LABELLING AND MAP GENERATION	71
6.1 <i>Related Works</i>	73
6.1.1 PFLP Problem	74
6.1.2 PFLP Algorithms.....	75
6.1.3 Google Map Mechanism.....	79
6.2 <i>Labeling and Map Generation in Memory Island</i>	80
6.3 <i>Discussion</i>	82
6.3.1 Improvement of the Label placement algorithm.....	82
6.3.2 Apply Area-features label placement algorithm	83
CHAPTER 7 MEMORY ISLAND INTERFACE.....	85
7.1 <i>Overview+Detail interactive interface</i>	87
7.1.1 Map-like Focus + Context and Element highlighting technique:.....	87
7.1.2 Map interactive functions.....	89
7.1.3 Design of Memory Island Interface.....	90
7.2 <i>Discussion</i>	93
PART III. IMPLEMENTATION AND APPLICATIONS	95
CHAPTER 8 MEMORY ISLAND IMPLEMENTATION.....	97
8.1 <i>Memory Island Application's sub-system components</i>	99
8.1.1 Knowledge Extraction.....	100
8.1.2 Island Generation.....	102
8.1.3 Label Placement and Map Generation.....	103
8.1.4 Memory Island Interface Generation.....	104
8.2 <i>The run-time of Memory Island Application</i>	104
8.3 <i>Discussion</i>	105
CHAPTER 9 CREATING KNOWLEDGE MAPS USING MEMORY ISLAND.....	107
9.1.1 Text and Document Data.....	109
9.1.2 Ontologies	110
9.1.3 Hierarchical Dataset	112
9.1.4 Project OBVIL – Digital Humanities Dataset.....	113

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

9.2 Discussion.....	114
PART IV. VALIDATION AND EVALUATION OF MEMORY ISLAND	115
CHAPTER 10 VALIDATING MEMORY ISLAND.....	117
10.1 Visualization mantras	119
10.2 User Study.....	119
10.2.1 Methods and Visualization Set-up	120
10.2.2 Subjects	120
10.2.3 Procedure.....	121
10.2.4 Result	122
10.3 Discussion.....	123
CHAPTER 11 PSYCHOLOGICAL EVALUATION.....	125
11.1 Related work.....	127
11.1.1 Users' requirements analysis.....	129
11.1.2 Evaluations of Ontology Visualization tools	129
11.1.3 Evaluations of Tree Visualization tools	131
11.1.4 Evaluations of Map-like Visualization tools.....	131
11.2 Psychological experimental protocol.....	132
11.2.1 Ontology Browsing Task.....	132
11.2.2 Ontology Understanding Task.....	133
11.2.3 Ontology Remembering Task	133
11.2.4 The Subjective Task	134
11.2.5 Suggested Evaluation Procedure.....	135
11.3 A Preliminary Users Study	136
11.3.1 Methods and Experiment Set-up	136
11.3.2 Subjects	136
11.3.3 Procedure.....	136
11.3.4 Result	137
11.4 Discussion with the past evaluation experiment	139
PART VI. CONCLUSION AND PERSPECTIVES.....	141
CHAPTER 12 CONCLUSIONS.....	143
CHAPTER 13 FUTURE WORKS	145
13.1 Develop the Visualization Tools for Digital Humanities	145
13.2 Improve the Usage of space.....	145
13.3 Evaluation of the long-term memorization.....	146
13.4 User Interface and interactive functions.....	147
13.5 Integrate Memory Islands to platforms of DH and DL.....	148
REFERENCES	151
APPENDIX.....	167

LIST OF TABLES

TABLE 1 THE EXISTING WORKS ON EVALUATION AND USERS' REQUIREMENTS ANALYSIS	127
TABLE 2 SUGGESTED EXPERIMENT PROCEDURE FOR OUR PSYCHOLOGICAL EXPERIMENTAL PROTOCOL	135
TABLE 3 SELECTED COLOR SCHEMES FOR MEMORY ISLAND APPLICATION.....	171
TABLE 4 THE COMPLETE LIST OF SUBJECTS FOR VALIDATION OF MEMORY ISLAND.....	173

LIST OF ALGORITHMS

ALGORITHM 1 MEMORY ISLAND PROTOTYPE ALGORITHM	50
ALGORITHM 2 POLYLE II ALGORITHM	61
ALGORITHM 3 MEMORY ISLAND ALGORITHM	64
ALGORITHM 4 GENERATECROWN.....	65

LIST OF FIGURES

FIGURE 1.1 THE LOGO OF PROJECT LOCUPLETO.....	6
FIGURE 1.2 MEMORY ISLAND FOR A CHILDREN’S BOOK, THIS VISUALIZATION HELP THE CHILDREN WITH THEIR LEARNING, THEY CAN ALSO SHARE THEIR VISIT WITH OTHERS.	7
FIGURE 1.3 THE RELATED RESEARCH AREAS OF THE MEMORY ISLAND TECHNIQUE (THIS MAP WAS CREATED BY USING MINDJET MIND MANAGER SOFTWARE).	8
FIGURE 1.4 THE FIVE PARTS OF THIS THESIS. (THIS MAP WAS CREATED BY USING MINDJET MIND MANAGER SOFTWARE).....	9
FIGURE 2.1 VISUAL REPRESENTATION CREATED BY GEPHI, FOR INPHO ONTOLOGY (LEFT) AND ROCK ONTOLOGY (RIGHT). THE FIRST ONE CREATED BY THE SPATIALIZATION ALGORITHMS PROVIDED IN GEPHI. IN THE SECOND ONE, WE MANUALLY CONSTRUCT THE SPATIALIZATION.....	14
FIGURE 2.2 THE FOUR-LEVEL NESTED DESIGN MODEL PROPOSED BY MUNZNER ET AL.[21] [22] FOR DESIGNING AN INFOVIS TECHNIQUE.....	15
FIGURE 2.3: THE PROCEDURE OF DESIGNING AN INFOVIS TECHNIQUE, EACH IMPROVEMENT OF THE INFOVIS TECHNIQUE NEEDS TO BE EVALUATED BY THE USER EXPERIMENTS.	16
FIGURE 2.4 A CLASSIC NODE-LINK DIAGRAM VISUALIZATION OF A TREE STRUCTURE DATASET.	18
FIGURE 2.5 CLASS BROWSER IN THE RECENT WEB-PROTÉGÉ FOR THE ONTOLOGY OF MERCURE GALANT ONTOLOGY OF LABEX OBVIL.....	19
FIGURE 2.6 INDENTED OUTLINES OF AN ENCYCLOPEDIA OF PHILOSOPHY SEP (INPHO ONTOLOGY’S SKELETON).	20
FIGURE 2.7 A NODE-LINK DIAGRAM OF A SIMPLE TREE STRUCTURE (NODE SIZE AND COLOR USED TO EMPHASIZE A SPECIFICATION OF A NODE).	21
FIGURE 2.8 A KNOWLEDGE MAP FOR EXPLORING THE MECHANISMS OF MAP EQUATION[40], THE INFOMAP CODE AVAILABLE ON MAPEQUATION.ORG (IMAGE CREATED BY MARTIN ROSVALL , REPRODUCED IN THIS THESIS WITH PERMISSION).	24
FIGURE 2.9 A KNOWLEDGE MAP DESIGNED TO VISUALIZE THE KNOWLEDGE ORDERS. MAP CREATED BY ALMILA AKDAG, CHENG GAO, KRZYSZTOF SUCHECKI AND ANDREA SCHARNHORST. THIS KNOWLEDGE MAP REPRODUCED IN THIS THESIS WITH PERMISSION.....	25
FIGURE 2.10 LEONARDO DA VINCI'S MAPPAMUNDI (APPROXIMATELY IN 1514).....	27

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

FIGURE 2.11 THE SCREENSHOTS TAKEN FROM MARCO QUAGGIOTTO'S KNOWLEDGE ATLAS [44] [45] (IMAGE REPRODUCED FROM HTTP://KNOWLEDGECARTOGRAPHY.ORG/#IMAGES WITH PERMISSION).....	27
FIGURE 2.12 TREE MAP AND ITS EQUIVALENT NODE-LINK DIAGRAM REPRESENTATION (WITH NODE'S WEIGHT AND NODE'S TYPE INFORMATION).....	28
FIGURE 2.13 USING TREE MAP TO DISPLAY THE AMOUNT OF SPACE USED BY FILES ON A DISK PARTITION. GENERATE BY THE FREE SOFTWARE WINDIRSTAT.....	29
FIGURE 2.14 A SCHEMA FOR MOST EXISTING MAP-LIKE VISUALIZATIONS, APPLY AN EXISTING CLUSTERING TECHNIQUE ON THE UNDERLYING DATA TO CREATE THE 2D DISTRIBUTION, THEN BUILD A MAP BASED ON THIS DISTRIBUTIONS.....	31
FIGURE 2.15 THE MOCS WITH ITS BASE MAP GENERATED BY GMAP ALGORITHM USED THE TOPIC OF IJCAI 1990-2014 OVERLAY WITH THE HEAT-MAP GENERATED USED THE TOPICS OF ECAI 1950-2014. THE OPEN-SOURCE PROGRAM MAP OF COMPUTER SCIENCE PROGRAM IS USED TO GENERATE THIS FIGURE.....	32
FIGURE 3.1: TWO EXAMPLES OF USING THE NOTION OF MEMORY ISLANDS FOR THE E-BOOKS AND THE DTV.	40
FIGURE 3.2 THE OBJECTIVES OF MEMORY ISLAND TECHNIQUE.....	41
FIGURE 3.3 MEMORY ISLANDS IDEA: TRANSFORM HIERARCHICAL KNOWLEDGE (E.G., ONTOLOGY) INTO AN INSIGHTFUL 2D ISLAND; EACH CONCEPT HAS ITS OWN SUB-ISLAND. THE ISLAND IS GENERATED BY USING THE POWER OF KNOWLEDGE SUCH AS ONTOLOGIES, WHICH SIMULATE THE "ART OF MEMORY" USED BY GREAT SCIENTIFIC PERSONALITIES..	42
FIGURE 3.4 TWO EXAMPLES GENERATED FOR THE SAME ONTOLOGY INPHO WITH DIFFERENT PROPORTION METAPHOR WITH MEMORY ISLAND.	45
FIGURE 3.5 AN EXAMPLE ABOUT THE DISTANCE/CENTRALITY METAPHOR, THE HIGH LEVEL IS LOCATED IN THE CENTER OF SUB-ISLAND, THE DISTANCE FROM A LOCATION OF CONCEPT TO ITS PARENT CONCEPT'S LOCATION IS CALCULATED WITH NUMBER OF INSTANCES, AND SUB-CONCEPTS (WEIGHT) ASSOCIATED WITH THIS CONCEPT.....	46
FIGURE 3.6 THE FONTS WE USED WITH MEMORY ISLANDS.	47
FIGURE 3.7 POINTS (LOCATIONS) ATTRIBUTES IN MEMORY ISLANDS, POINTS MAY HAVE DIFFERENT TYPES AND DIFFERENT SIZES.....	47
FIGURE 3.8 THE COLOR WHEEL SHOWS OUR COLOR SELECTION STRATEGY, AND AN EXAMPLE OF MEMORY ISLAND GENERATED FROM TABLE OF CONTENTS OF A PAPER; THE MAIN PART OF THE ISLAND IS COLORED BY USING THE COLORS IN-GROUP A (IN THE WHEEL), THE SUB-	

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

ISLAND FOR CHAPTER 4 (ABOUT TECHNICAL DETAILS) IS COLORED BY USING THE COLORS OF GROUP B.....	48
FIGURE 3.9 THE BLUE AND WHITE DATA-LINES DESCRIBE THE RELATIONS (SUBSUMPTION) ASSOCIATED WITH THE SELECTED CONCEPT PHILOSOPHY OF MIND.	49
FIGURE 3.10 CARTOGRAPHIC PATHS (VIOLET) BETWEEN TWO CONCEPTS INDICATED THE TWO CONCEPT IS MORE CLOSED.	50
FIGURE 5.1 AN EXAMPLE OF MEMORY ISLAND GENERATED WITH POLYLE II ALGORITHM.	63
FIGURE 5.2 AN EXAMPLE GENERATED BY POLYLE II ALGORITHM, IT HAS THE OVERLAPPING BETWEEN ITS SUB-ISLANDS.	64
FIGURE 5.3 AN ILLUSTRATION OF THE WEDGES OF THE CONCEPTS WITH THE SAME PARENT.....	67
FIGURE 5.4 THE ISLAND GENERATED WITH MEMORY ISLAND ALGORITHM (WITHOUT LABELING AND CARTOGRAPHIC MEANS).	67
FIGURE 5.5 LEFT: VISUALIZATION SEP WITH INPHO ONTOLOGY WITH THE METAPHORS WE DEFINED. RIGHT: WE GENERATE A RADIAL-LAYOUT MEMORY ISLAND FROM THE SAME DATASET.	68
FIGURE 5.6 A MEMORY ISLAND WITHOUT RESHAPING PROCESS.	69
FIGURE 6.1 A MAP WITH HEAVY LABEL OVERLAPS. IN THIS EXAMPLE, WE RANDOMLY PLACED THE LABELS IN THIS MAP. THE HUMAN USERS CANNOT DISCOVERY THROUGH A MAP WITH HEAVY LABEL OVERLAPS.	73
FIGURE 6.2 A SET OF POTENTIAL LABEL POSITIONS WITH THEIR RELATIVE DESIRABILITY. THIS SET OF POTENTIAL LABEL POSITIONS WAS PROPOSED BY CHRISTENSEN ET AL. [110], BASED ON THE WORK OF YOELI [102]. THE NUMBER OF EACH POSITION WAS USED TO DESCRIBE THE CARTOGRAPHIC PREFERENCES FOR REPOSITIONING THE LABEL. (LABEL POSITIONS' PREFERENCES: 1>2>3>4>5>6>7>8) [107].	75
FIGURE 6.3 FIRST ZOOM LEVEL MAP IN GOOGLE MAP. CAPTURE FROM THE WEB SERVICE APPLICATION GOOGLE MAP.....	79
FIGURE 6.4 THE SECOND ZOOM-LEVEL MAP FOR THE MAP OF EARTH, CAPTURE FROM GOOGLE MAP APPLICATION.	80
FIGURE 6.5 MEMORY ISLAND'S LABELING AND MAP GENERATION MECHANISM.	82
FIGURE 6.6 TWO DIFFERENT LABELS' CONFIGURATIONS FOUND IN MEMORY ISLAND.....	83
FIGURE 7.1 NAVIGATING THE SEP'S MEMORY ISLAND WITH INTERACTIVE FUNCTION: (A) (B): THE CARTOGRAPHIC GAPS AND VIOLET PATHS ARE USED TO EMPHASIZE THE ORDER OF	

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

CONCEPTS. FOR EXAMPLE, THE PHILOSOPHY OF MATHEMATICS IS CLOSE TO THE LOGIC AND PHILOSOPHY OF SCIENCE AND THE SCIENCE. (C): A DETAIL-ON-DEMAND WINDOW PROVIDES DETAILED INFORMATION ABOUT THE CONCEPT. USERS CAN ACCESS TO THE SEP. (D): SHOWS THE VISITED TRACE IN THE MAP FOR A USER NAMED “EUROVIS15”. 87

FIGURE 7.2 THE OVERVIEW PANEL SHOWS THE CURRENT POSITION IN THE OVERVIEW STRUCTURE; THE USERS CAN ALSO USE THIS PANEL TO CHANGE THE DETAIL CONTEXTS TO SELECTED PART OF MEMORY ISLAND. THIS FUNCTION IS FRIENDLY FOR MOST OF USERS, BECAUSE THEY ARE ALREADY FAMILIAR WITH THE GEOGRAPHICAL INFORMATION APPLICATION, SUCH AS THE GOOGLE MAP. 88

FIGURE 7.3 THE SEARCH FUNCTION IN MEMORY ISLAND INTERFACE..... 89

FIGURE 7.4 DRAW THE VISITED TRACE ON THE MEMORY ISLAND. 90

FIGURE 7.5 THE MEMORY ISLAND INTERFACE WITH NOTATIONS. 91

FIGURE 7.6 THE MORE GENERAL CONCEPTS WILL HAVE THE PRIORITY TO DISPLAY IN THE RESULT PANEL, IN THIS EXAMPLE, WE HIDDEN THE OVERVIEW MAP TO HAVE MORE DISPLAYING SPACE FOR THE KNOWLEDGE MAP. 92

FIGURE 7.7 AN EXAMPLE OF THE (LOCAL) MODIFICATION FUNCTION AND THE PERSPECTIVE ON RE-GENERATE KNOWLEDGE FROM DIFFERENT MODIFIED MEMORY ISLANDS. 93

FIGURE 8.1 THE 4 SUB-SYSTEM COMPONENTS OF MEMORY ISLAND APPLICATION 99

FIGURE 8.2 THE ACTIVITY DIAGRAM OF OUR MEMORY ISLAND APPLICATION..... 100

FIGURE 8.3 LEFT: AN EXAMPLE OF A TREE DATASET (WITH A ROOT AND FOUR CHILDREN). RIGHT: THE FULLY CONNECT GRAPH OF CHILDREN NODES BASED ON THEIR SIMILARITY SCORE. 101

FIGURE 8.4 THE SYMMETRIC REORGANIZING OF MEMORY ISLAND APPLICATION. 102

FIGURE 8.5 WE RE-SHAPE THE POLYGON BY USING BEZIER CURVE; IT REPLACES THE STRAIGHT LINE WITH THE CURVE TO MAKE MORE A NATURAL VISUAL REPRESENTATION..... 103

FIGURE 8.6 THE RUNTIME FOR OUR IMPLEMENTATION OF MEMORY ISLAND APPLICATION.... 105

FIG. 8.7. TWO MAPS GENERATED FOR THE INPhO ONTOLOGY WITH REORGANIZATION..... 106

FIGURE 9.1 A CHILDREN BOOK “OEDIPE LE MAUDIT” (COPYRIGHT BELONGS TO SEJER) VISUALIZED WITH THE MEMORY ISLAND TECHNIQUE. WE IDENTIFIED PEOPLE ENTITIES FROM THE BOOKS, AND DISPLAYED THEM WITH THE CORRESPONDING CHAPTER. 110

FIGURE 9.2 USERS COULD NAVIGATE THROUGH THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY WITH THE HELP OF MEMORY ISLAND. IN THIS EXAMPLE, WE FOCUS ON THE

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

INSTANCE “ATOMISM FROM THE 17TH TO THE 20TH CENTURY” ASSOCIATED WITH CONCEPT “PHILOSOPHY OF SCIENCE AND THE SCIENCES”	111
FIGURE 9.3 THE MEMORY ISLAND FOR A LARGE TABLE OF CONTENTS CREATED FOR <i>ÉCOLE SUPÉRIEURE D'ART ET DESIGN GRENOBLE-VALENCE</i>	112
FIG. 9.4. A PUBLIC DEBATE (TOPICS) IN THE FRENCH PARLIAMENT VISUALIZED WITH MEMORY ISLANDS.	113
FIGURE 9.5 THE MEMORY ISLANDS FOR THE <i>LE MERCURE GALANT</i> JOURNALS. THE TOP IMAGE SHOWS THE ISLANDS FOR ALL THE DOCUMENTS IN THE WEBSITE OF OBVIL AND WE CAN ACCESS THE JOURNAL BY THIS VISUALIZATION. THE SECOND ISLAND IS VISUALIZATION OF AN ONTOLOGY WITH THE TOPICS (NEARLY 1000) OF <i>LE MERCURE GALANT</i> JOURNALS.	114
FIGURE 10.1 THE MEAN RESPONSE TIME OF DIFFERENT GROUPS OF USERS, FOR DIFFERENT TASKS.	122
FIGURE 10.2 USERS' PREFERRED TOOL FOR NAVIGATING ONTOLOGY.	123
FIGURE 11.1 AN EXAMPLE OF THE TASK T3.1 IN OUR PRELIMINARY IMPLEMENTATION OF USER STUDY SOFTWARE.	134
FIGURE 11.2 THE RESPONDS TIME (MEAN) FOR TASK BROWSING AND REMEMBERING, DATA FROM [10].	137
FIGURE 11.3 THE CHANGES IN RT FROM BROWSING TO REMEMBERING, DATA FROM [10].	138
FIGURE 11.4 OVERALL CORRECT RATE FOR EACH VISUALIZATION TOOL ON THE THREE TASKS (DATA FROM [10]).	138
FIGURE 13.1 AN EXAMPLE OF ZOOM FUNCTION OF GOOGLE (WHEN WE ZOOM-IN TO DEEPER LEVELS, WE CAN SEE MORE DETAILED GEOGRAPHIC INFORMATION), SCREEN SHOT FROM THE GOOGLE MAP WEB APPLICATION.	146
FIGURE 13.2 EVALUATE THE LONG-TERM MEMORIZATION EFFECT OF THE MEMORY ISLAND. IN THIS FIGURE, WE ASK THE PARTICIPANTS TO PUT SOME NAME OF CONCEPTS TO ITS LOCATIONS.	147
FIGURE 13.3 AN EXAMPLE FOR USING THE THIRD DIMENSION (2.5D DESIGN) WITH MEMORY ISLAND.	148
FIGURE 13.4 IMAGINARY RENDERINGS FOR THE FUTURE APPLICATION INTERFACE. WITH ICONS ASSOCIATED WITH EACH CONCEPTS.	148

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

FIGURE 13.5 AN EXAMPLE OF THE PRÉVU 3D PLATFORM. IMAGE CREATED BY DONATIEN
AUBERT OF ENSADLAB EN-ER, AND REPRODUCED IN THIS THESIS WITH PERMISSION. . 149

FIGURE 0.1 AN EXAMPLE OF OUR ON-LINE USERS' SURVEY PROGRAM. 173

LIST OF ABBREVIATIONS AND ACRONYMS

BFS (Breadth-First Search)

DH (Digital Humanities)

DL (Digital Libraries)

DTV (Digital Television)

FSN (File System Navigator)

GA (Genetic Algorithm)

GIS (Geographic Information Systems)

HDS (Hierarchical Data Set)

InfoVis (Information Visualization)

InPhO (Indiana Philosophy Ontology)

KNOWeSCAPE (analyzing the dynamics of information and KNOWledge landSCAPEs)

MOCS (Maps of Computer Science)

Nmap (Neighborhood Map)

OBVIL (*Observatoire de la vie littéraire*, Observatory of literary life)

PFLP (Point-Futures Label Placements)

SA (Simulated Annealing)

SEP (Stanford Encyclopedia of Philosophy)

SOM (Self-Organizing Map)

TS (Tabu Search)

UDC (Universal Decimal Classification)

LIST OF APPENDICES

RELATED PUBLICATIONS AND COMMUNICATIONS	169
COLOR SCHEMES.....	171
QUESTIONS FOR VALIDATION EXPERIMENTS	173

Part I. Background and Literature Review

"The only new thing in the world is the history you don't know."

Harry S. Truman

quoted by David McCulloch in "Truman"

*"...the two operations of our understanding, intuition and deduction, on which
alone we have said we must rely in the acquisition of knowledge."*

René Descartes [1628]

From Rules for the direction of the mind

Chapter 1 GENERAL INTRODUCTION

CONTENTS

1.1 OVERVIEW AND RESEARCH CONTEXT	5
1.2 INDUSTRIAL CONTEXT	6
1.3 CONTRIBUTIONS AND RELATED RESEARCH AREAS.....	7
1.4 ORGANIZATION	8

1.1 Overview and Research Context

Recently many researchers have started interdisciplinary cooperation in emerging scientific research areas, such as Digital Humanities, relying on visualization techniques for their domain knowledge. A large number of users need the interactive information visualization techniques to help them with their information seeking tasks [1]. Users which are unfamiliar with a domain have difficulties in understanding and discovering its knowledge, even when provided with descriptive metaphors [2, 3]. According to Ware (in 2000) [4] and Auber (2003) [5], “in the human brain, over 70% of the receptors and 40% of the cortex are implicated in vision processing.” The visual representations with visual metaphors are more efficient than the verbal representation even with descriptive metaphors: it is easier for the users to achieve their information seeking tasks by using a visualization technique.

However knowledge visualizations (e.g. ontologies) are still a challenge [6]. Users are still waiting for beautiful, insightful ontology visualizations, bringing them the power of ontologies to support visual knowledge discovery [7].

The problem we study in this thesis is: how can we meaningfully and insightfully visualize knowledge, such as the ontologies, to help the human users improve their performance in information seeking tasks and help them improve their memorability?

Recently, knowledge maps are useful tools for managing and sharing the large-scale hierarchical knowledge, that have recently started to be widely applied. Beside the advantage brought by the map metaphor, using knowledge maps people from different domains can collaborate with each other, thus leading to additional benefits. For example, the famous Torrance’s experiments [8] show that working in pairs facilitates creativity.

From antiquity to Middle-Age, the "Art of Memory" was a popular technique [9] to retrieve knowledge. Based on the "loci" method from the "Art of Memory," knowledge is stored within a virtual map, such as an island, in the memory. Distinguished people efficiently used this technique to improve their capacity of knowledge retrieval and memorization. Meaningfully represented ontologies, in the form of interactive knowledge maps can provide with powerful use cases for human users. To this end, we introduce the Memory Island technique, inspired from the "Art of Memory.”

In this thesis, we show how to generate an insightful knowledge island, for such knowledge as ontology by using the notion of Memory Islands. With these Memory Islands, the users can visually navigate through the information contents based on the knowledge skeleton, and improve their performance of information seeking tasks. We then in the end of this thesis, present some user studies to evaluate and validate the Memory Island technique.

The works presented in this thesis completed within Lip6 between February 2012 and January 2015. This thesis is also the result of collaborations with the InPhOrmers¹ team (directed by Professor Colin Allen) of Indiana University, the Labex OBVIL² (Project of Paris Sorbonne University and University Pierre and Marie Curie) for Digital Humanities and the European COST Action TD1210³ analyzing the dynamics of information and knowledge landscapes (KNOWeSCAPE).

1.2 Industrial Context

This thesis is supported by the project LOCUPLETO (February 2012 - January 2015, Figure 1.1), and its partners the Publishers Jouve, Sejer and *les Editions des Braques*, the company Tralalere, and the School of Animation *Ecole les Gobelins*. In this project, we develop a new platform for generating and editing e-books. We apply the Memory Island technique proposed in this thesis to the text and documents data for visualizing children's books (Figure 1.2); to this end, we introduce the trace function, which keeps the trace of a child's visits to help them with their extracurricular book and after-school learning. The publishing houses SEJER and Jouve are particularly interested with the Memory Islands generated for their books and are considering their future use.



Figure 1.1 The logo of project LOCUPLETO.

¹ <https://inpho.cogs.indiana.edu/about/>, The Indiana Philosophy Ontology (InPhO) is a project on modeling the discipline of philosophy created by Indiana University of Bloomington. InPhO is a dynamic ontology generated from over 13 million words in the Stanford Encyclopedia of Philosophy (SEP).

² <http://obvil.paris-sorbonne.fr/>, Labex OBVIL (Observatory of literary life) intends to develop the resources offered by computer applications to examine the French literature of the past.

³ <http://knowescape.org/>, . The COST Action KNOWeSCAPE aims to create interactive knowledge maps for the end-users.

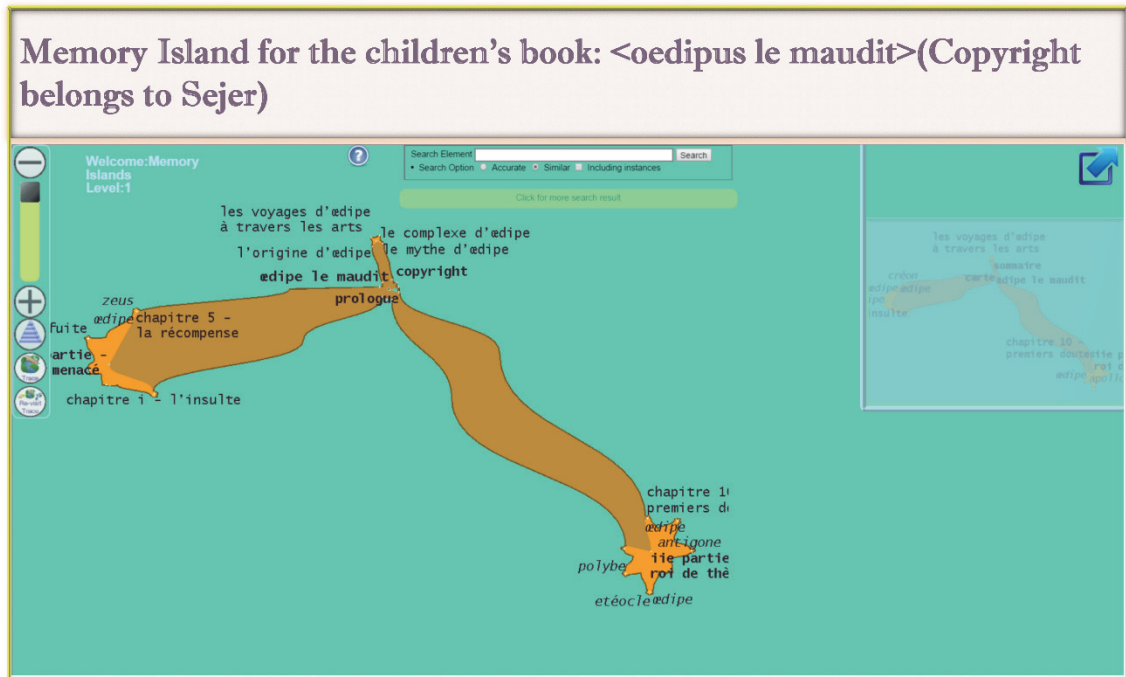


Figure 1.2 Memory Island for a children's book, this visualization help the children with their learning, they can also share their visit with others.

1.3 Contributions and Related Research Areas

This thesis proposes a novel cartographic visualization technique – Memory Island – for the hierarchical knowledge. It uses the notion of Memory Island, which is inspired by the Art of Memory technique. We present our methodology for designing the Memory Island, and then we propose an architecture to implement this methodology. We discuss how Memory Island, a novel form of visual representation, helps its users who look for discovering through the knowledge, such as relations between concepts.

We introduce our Memory Island algorithms; design our Memory Island interface and its interactive functions. Then we discuss how to apply our technique on different knowledge, text, and documents data, to solve real-world problems.

Once we propose a technique, we need to evaluate and validate its ability to transform hierarchical knowledge into insightful and meaningful memory islands. In order to evaluate the relevance of the technique, we propose some evaluation protocols. The first task evaluates the Response times (RT) of the users of Memory Island and compares them to those of classic visualization tools: indented list and node-link diagrams.

We then propose a protocol to evaluate the hierarchical knowledge visualizations. This protocol evaluates the users' performances with the visualizations considering the correct response rate and response times for the three main tasks: Ontology Browsing, Ontology Understanding, and Ontology Memorization. We ask the users to answer some questions with

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang - June 2015

different InfoVis tools on different knowledge (ontologies). From the result of a user study [10], we found that using Memory Island provides advantages (with high correct rate) for the users, in terms of both remembering and navigating the knowledge.

We also discuss the future directions of this work. In each part of this thesis, we show the future direction to extend our work, and we discuss how the researchers in different domains can use our technique to help their own researches and derive meaningful visualizations. We believe this technique could become more powerful with the new generation HMI techniques.

The main research area of this work is Information Visualization; however, it has many related research areas as presented in Figure 1.3.

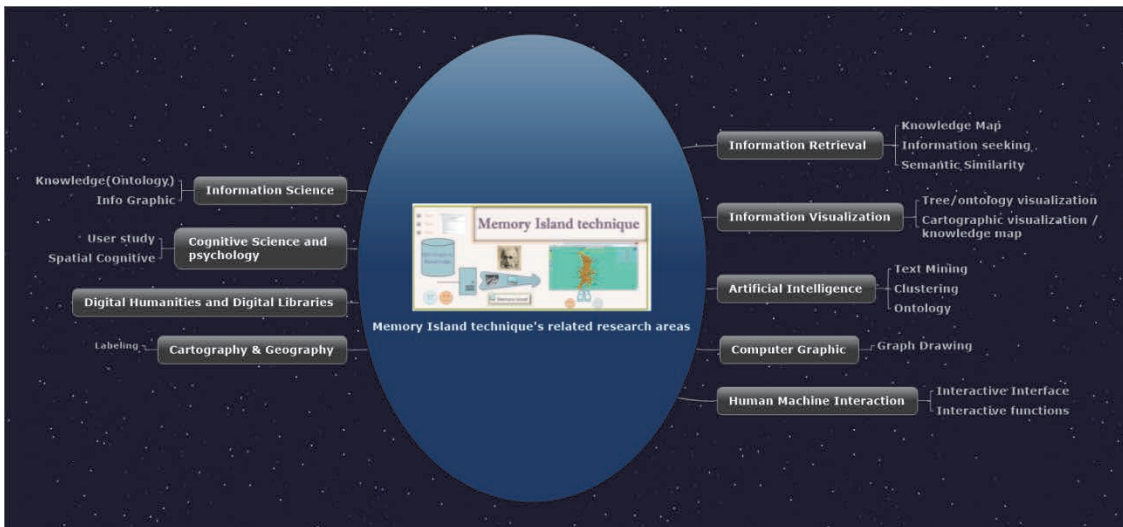


Figure 1.3 The related research areas of the Memory Island technique (This map was created by using Mindjet Mind Manager Software).

1.4 Organization

This thesis is divided into five parts (Figure 1.4). The remainder of this thesis is organized as follows:

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang- June 2015

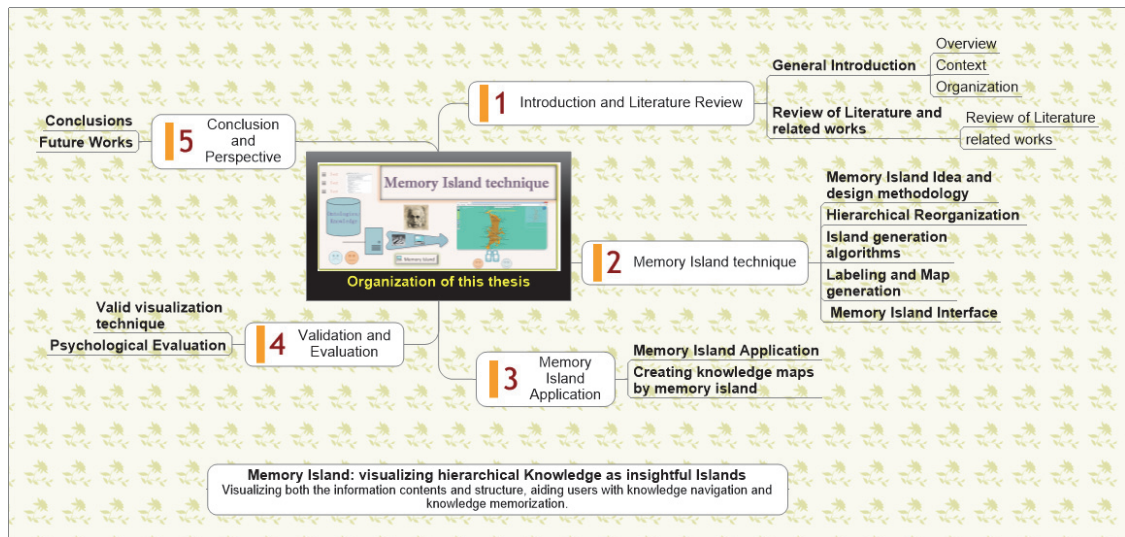


Figure 1.4 The five parts of this thesis. (This map was created by using Mindjet Mind Manager Software).

Chapter 1 gives an overview of this thesis as well as a general introduction, including the purposes and motivations of this works, and provides a roadmap for the rest of the thesis. **Chapter 2** contains the basic notions related to our research; we review background concepts and the state of the art in cartographic visualization and map-like visualization. We also introduce viewpoints from research on spatial cognition, as well as a methodology for visualization design, helping us with the design of the Memory Island technique.

Thereafter, **Part II** introduces the principal research contributions of this thesis: the Memory Island technique with all its details, including its algorithms and interfaces for exploratory visualization of data.

In particular, **Chapter 3** describes the basic idea behind the Memory Island technique, we discuss why we believe our technique can make sense of the data by describing our design methodology, including the geographic metaphors we used, and express the cartographic means we introduced to our approach, and discuss why we choose 2D traditional representation for this technique. In the end of this chapter, we present our Memory Island prototype algorithm.

Then **Chapter 4 – 7** discuss each aspect in the prototype algorithm with all technical details. **Chapter 4** discusses the hierarchical reorganization using the semantic similarity measure (knowledge orders) for Memory Island technique, and then in **Chapter 5** we present our island generation algorithms for hierarchical data. We discuss the important issues of cartographic labelling and map generation in **Chapter 6**, while in **Chapter 7**; we talk about our user-friendly interactive Memory Island interface.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

Part III is about the technique implementation and case studies with Memory Island. It introduces the Memory Island application in **Chapter 8**, and then in **Chapter 9** we give some case studies on different datasets to show how to apply this technique to different domains.

Then, **Part IV** addresses the validation and evaluation of our technique. The experiment for validating the Memory Island technique are presented in **Chapter 10**; we present some preliminary experimental studies to verify the effects of the visual metaphors used by our technique, as well as a study about the users' preferences on visualization tools for navigation large ontologies.

In **Chapter 11**, we review some important works on the evaluations of visualization techniques, we introduce our psychological evaluation protocols, and we describe a user study based on this protocol, we discuss the result from this experiment.

In the last part of this thesis, we present our conclusions in **Chapter 12**, and discuss directions for the future research in the **Chapter 13**.

Chapter 2 REVIEW OF LITERATURE

CONTENTS

2.1 INFORMATION VISUALIZATION AND KNOWLEDGE VISUALIZATION	13
2.2 INFOVIS TOOLKIT AND INFOVIS TOOLS	13
2.3 DESIGN AND EVALUATION VISUALIZATION SYSTEMS	15
2.4 TREE AND HIERARCHIC KNOWLEDGE VISUALIZATIONS	16
2.4.1 <i>Tree, Hierarchic knowledge and Skeleton</i>	16
2.4.2 <i>Classical Tree Visualization Techniques</i>	17
2.4.2.1 The Indented List	18
2.4.2.2 Node-link Diagram.....	20
2.5 ONTOLOGY AND ONTOLOGY VISUALIZATION	21
2.6 KNOWLEDGE MAPS.....	23
2.7 CARTOGRAPHY AND KNOWLEDGE CARTOGRAPHY	25
2.7.1 <i>Cartography</i>	25
2.7.2 <i>Knowledge Cartography</i>	26
2.8 CARTOGRAPHIC VISUALIZATION	28
2.8.1 <i>Treemapping Approaches</i>	28
2.8.2 <i>Spatialization (Layout) and Tree drawing algorithms</i>	30
2.8.3 <i>Map-like Visualization Approaches</i>	30
2.9 SCHEMATIZATION.....	32
2.10 SPATIAL COGNITION.....	33

We review some important concepts and related works in this chapter.

2.1 Information Visualization and knowledge visualization

Information visualization (InfoVis) tools function as intermediaries between information (such as recorded knowledge) and the users of it, like teachers, whose role is to interpret an area of knowledge to the student. In order to fulfill this function, InfoVis techniques are aiming at supporting the cognitive system of their users. Researchers in Visualization are trying to develop and use tools that foster the access to information resources. They want their visualization tools to help users in their information tasks, such as searching, browsing, learning, and exploration.

The term InfoVis and Visualization can be referenced in a variety of contexts, the most definition was given by Card, Shneiderman, and Mackinlay [11], they defined InfoVis as “the use of computer-supported, interactive, visual representations of abstract non-physically based data to amplify cognition”. In 2009, Friendly and Denis [12] provided a graphic overview of the events in the history of visualization, they illustrated the milestones (important works and events in the domain of visualization) within a timeline (chronology of innovations) from 1600 to 2009.

There are numerous attempts to define the two basic concepts of Information and Knowledge. For example, Keller and Tergan give a definition: “Information is data that has been given meaning through interpretation by the way of relational connection and pragmatic context. Knowledge is information, which has been cognitively processed and integrated into an existing human knowledge structure” [13] (another widely-used definition by Chen et al can be found in [14]). Schank and Abelson have defined the concept knowledge structures in [15]. In 2012, Van Biljon has given a summary of the similarities and differences between these two types of Visualizations on the existed visualization works [16].

Nowadays, some InfoVis works handle with the abstract data structures have been introduced to visualize knowledge structures[6], most of InfoVis and knowledge visualization researchers see no more obvious differences between knowledge visualization and InfoVis. For this reason, in the rest of thesis, I no longer intentionally distinguish the terms InfoVis and knowledge visualization.

2.2 InfoVis Toolkit and InfoVis Tools

In 2014, Fekete [17] created an InfoVis ToolKit with Java Swing, it helped the researches to create and extend the 2D InfoVis technique, then different InfoVis ToolKit were created, for instance, the open-source InfoVis ToolKit tool in JavaScript.

The visualization toolkit also becomes an indispensable part of many AI tools. For example, in the ontological engineering tool Protégé [18], there are many InfoVis plug-ins providing the

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

visual representation to help the users to build their ontology. In addition, with the famous (in data mining) Orange Software, users can select different visualizations to visualize the results of data/text mining. Some other well-known visualization tools for graph and network are PHYLOViZ⁴ [19], Gephi⁵ [20] and Tulip⁶ [5]. Gephi is an InfoVis tool, which provides different existed spatialization algorithms for visualizing network and other datasets. With the semantic plug-in SemanticWebImport developed by Erwan Demairy of Inria DREAM team, Gephi can be used for visualizing ontologies. In Figure 2.1, two visualizations created with Gephi for InPhO and Rock ontologies are presented.

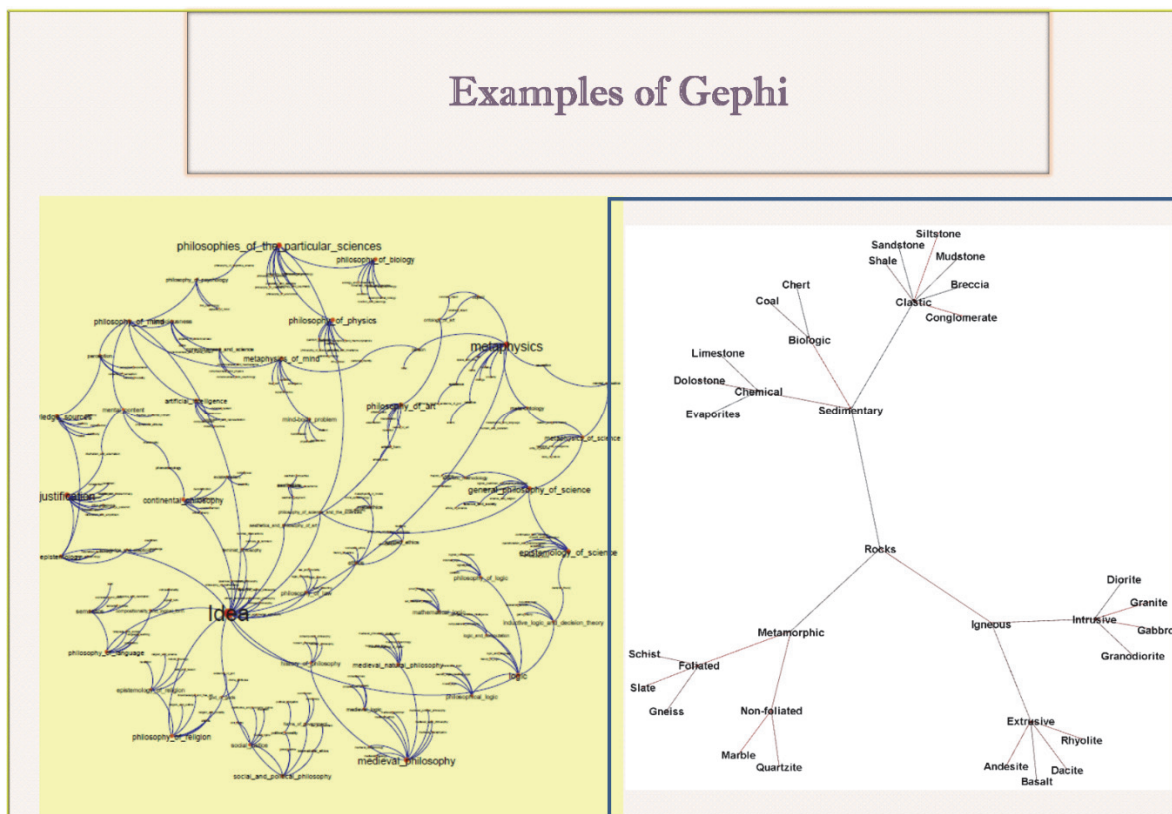


Figure 2.1 Visual representation created by Gephi, for InPhO ontology (left) and Rock ontology (Right). The first one created by the spatialization algorithms provided in Gephi. In the second one, we manually construct the spatialization.

⁴ Available at <http://www.phyloviz.net/wiki/>

⁵ Available at <https://gephi.github.io/>

⁶ Available at <http://tulip.labri.fr/TulipDrupal/>

2.3 Design and Evaluation Visualization Systems

How effectively design, valid visualization systems, and evaluation evaluate the different InfoVis techniques is one of the most important part of the InfoVis research. Munzner [21] proposed the four-level nested design model about the design and validation of visualization systems. They suggest that for designing a visualization system, the visualization researchers need to first design the domain problem characterization, then do the data/task abstraction design and encoding/interaction technique design, and at last design the algorithms of the visualization technique. Meyer et al. then extends this works and they proposed a four-level nested model which contains the blocks and guidelines at each level.[22].

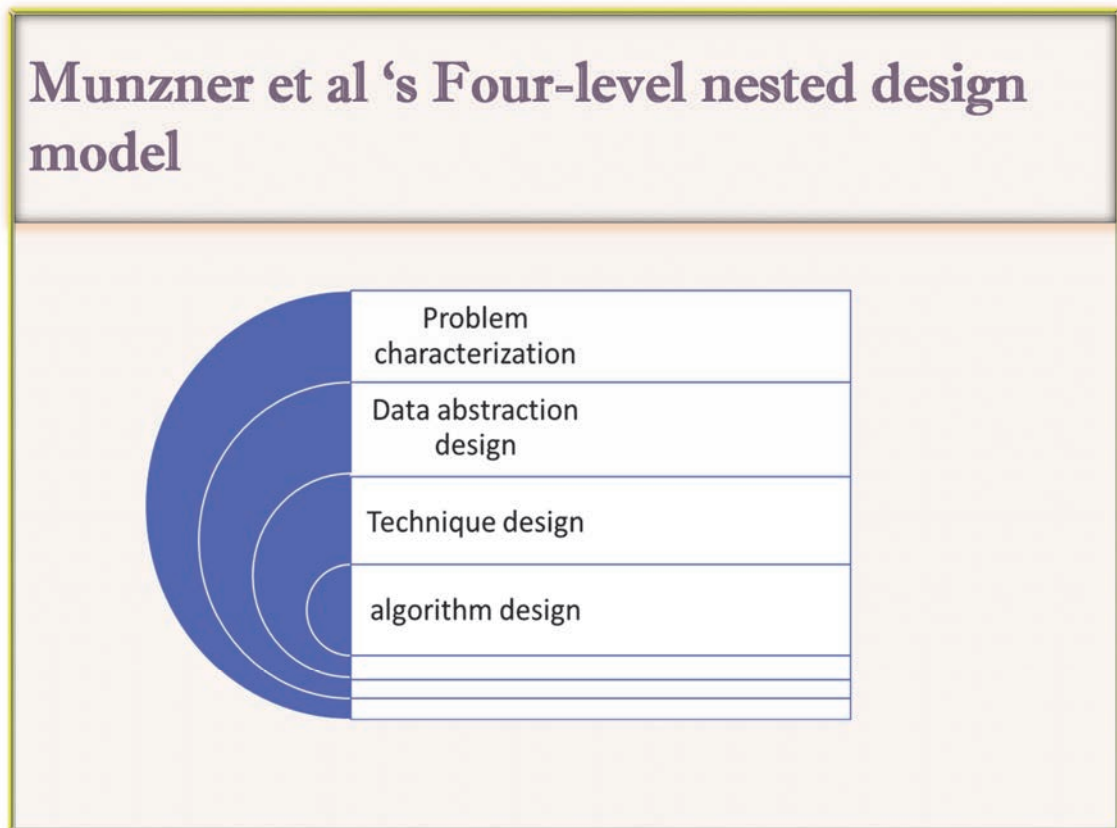


Figure 2.2 the four-level nested design model proposed by Munzner et al.[21] [22] for designing an InfoVis technique.

Beside the model of Munzner et al., we believe the discussion on the basic idea behind technique and its objectives (hypothesis) is important, and the evaluation protocol and users study with different tools is also an indispensable part for designing InfoVis techniques (in Figure 2.3).

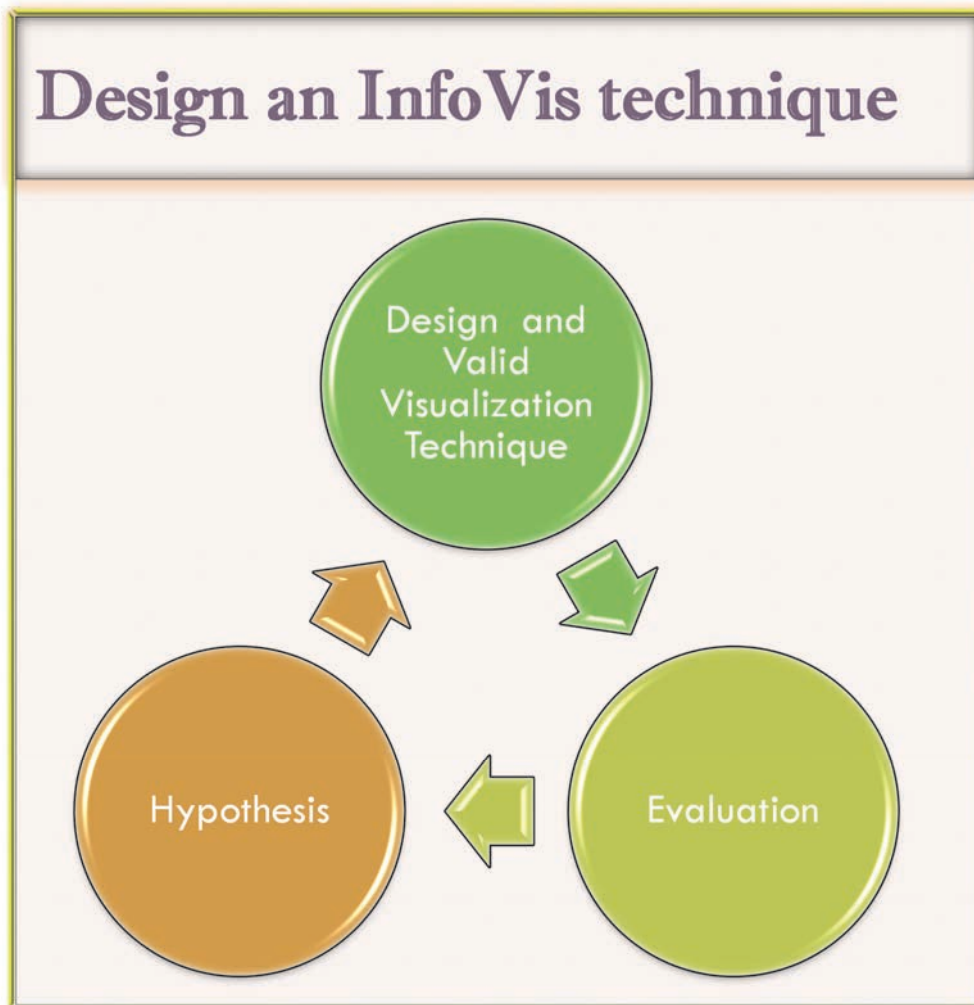


Figure 2.3: The procedure of designing an InfoVis technique, each improvement of the InfoVis technique needs to be evaluated by the user experiments.

2.4 Tree and Hierarchic Knowledge Visualizations

In this section, we begin with the definition of hierarchic knowledge and its skeleton – tree structure, and then we review the different hierarchic knowledge visualization techniques and the graph-drawing algorithms and theories behind these visualizations.

2.4.1 Tree, Hierarchic knowledge and Skeleton

A Tree is a hierarchical data set (HDS); it organizes data or knowledge (information records) into a hierarchy. Although hierarchical knowledge is a data type richer than a tree, its information records are organized around a hierarchy, such as, the taxonomy of an ontology. As this hierarchy is the most important knowledge structure, we call it as the skeleton of the knowledge. To visualize a hierarchical knowledge, representing knowledge's information contents with this knowledge skeleton is necessary and it is still an challenge.[6]

2.4.2 Classical Tree Visualization Techniques

Visualizing the tree structure is one of the most essential and important task, a network can be also abstracted as a tree-structure by using the Spanning tree algorithm (e.g., Minimal Spanning tree algorithms [23, 24]) or based on a clustering technique (e.g., the hierarchical clustering technique of networks, such as the technique proposed in [25]).

Tree hierarchies or tree structures are collections of items (e.g., information contents or concepts) with each item having a link (relation) to one parent item (except the root). The items of a tree structure are also called as tree nodes. Items and the links between parent and child can have multiple attributes (see basic tree structure in Figure 2.4).

To create a visual representations of tree structure dataset , we can use an outline style of indented labels likes the tables of contents of a book [26], a node-link diagram, or a Treemap (the space slit into nested regions). In this section, we discuss the basic approach for tree visualization: the indented List and Node-link diagram. We discuss the Treemap (tilling algorithms) later within the map-like visualization approaches in section 2.8.1.

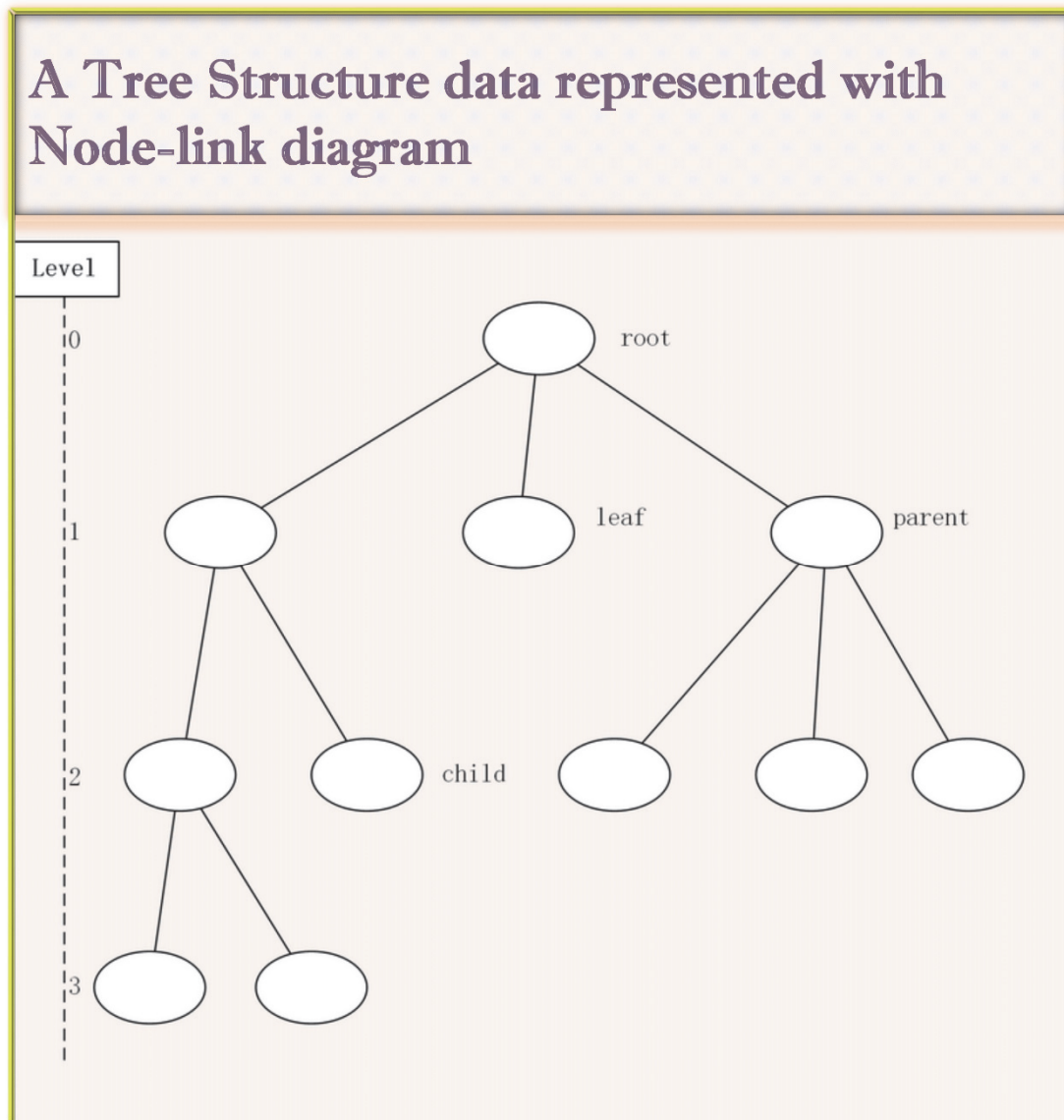


Figure 2.4 A classic node-link diagram visualization of a tree structure dataset.

2.4.2.1 The Indented List

Tree-structured data has long been displayed with indented outlines [27], then Kumar et al. in 1997 [28] proposed the PDQ (Pruning with Dynamic Queries) Tree-browser visualization tool to visualize the data with tree hierarchies with interactive functions.

The indented list approach used the classic file system (navigation) metaphor, where clicking on a folder opens up its sub-folders. This tree visualization was also called indented list, it allows the users to focus on a specified part of the tree structure and to hide the others by simply clicking on the label (identifier) of a concept. Figure 2.6 is an example of an indented list for the encyclopedia of Philosophy SEP. This approach is widely applied and ubiquitous in both file system (interfaces) and ontology engineering tools.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

When researchers compared this method with some of the others sophisticated visualization techniques, they found that the indented list approach could be surprisingly effective. For instance, the evaluation reported in [28] shows that users who are using Protégé Class Browser (Figure 2.5) performed better than those who are using alternative visualization plug-ins (node-link displays) in various ontology engineering tasks.

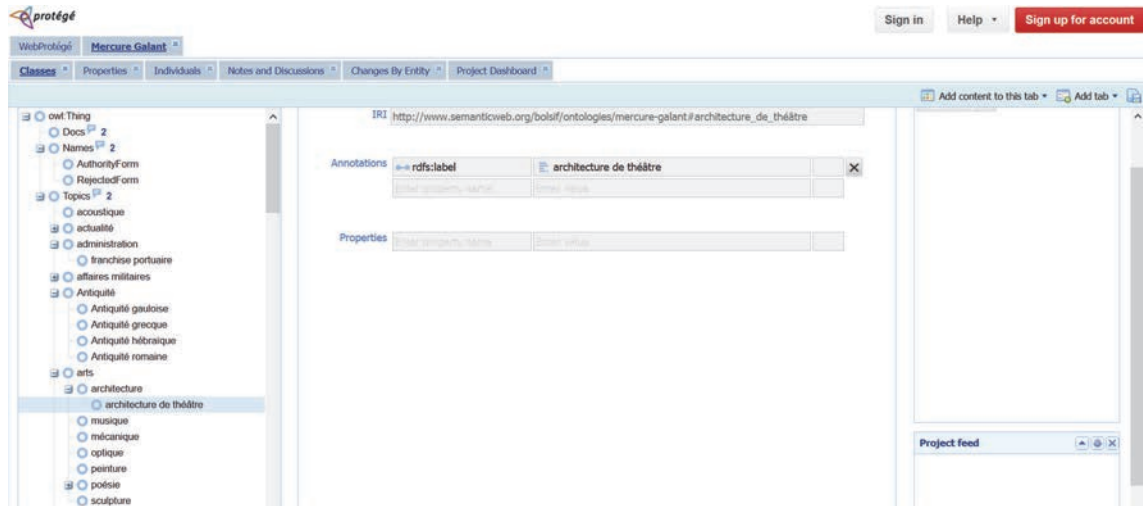


Figure 2.5 Class Browser in the recent web-Protégé for the ontology of Mercure Galant ontology of Labex Obvil.

The reason of the success of this visualization technique is simple: The indented list and the classic file system navigation metaphor is very familiar to the end-users, and it makes possible to display quite a lot of information in a rather small amount of space, in contrast with node-link diagrams, which need huge space for displaying. As a result, it is not much of surprising that these visualization schemes often perform better in evaluation scenarios than the graphical alternatives. However, when the list becomes large, it is difficult for human users to learn and use this list visualization, even with the help of search function. It is hard to discover through the information contents for this indented list approach.

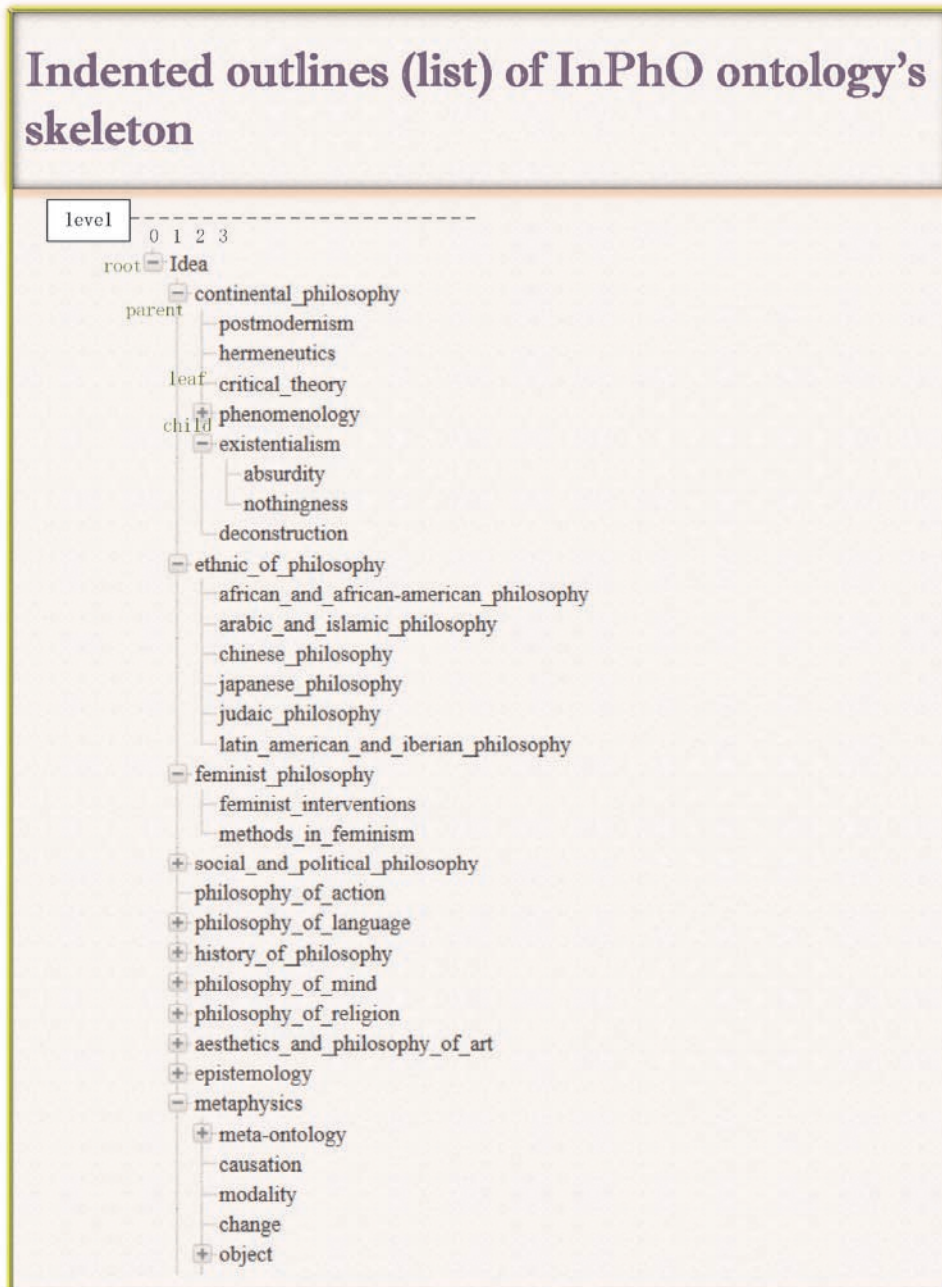


Figure 2.6 Indented outlines of an encyclopedia of Philosophy SEP (InPhO ontology's skeleton).

2.4.2.2 Node-link Diagram

The classic node-link diagram is another most popular tree visualization technique. Most of the users are already familiar with the mapping of structured relationships and they can easily understand this metaphor. This makes the node-link diagram visualizations wildly applied in many domains. They can also display attributes of links by color or size if required and the node-link diagram was used to visualize the set-dataset (using colors metaphor to display the group information, for example the Figure 2.7 and Figure 2.8). The sizes of nodes are used to emphasize the importance of the concepts, an example shown in Figure 2.7.

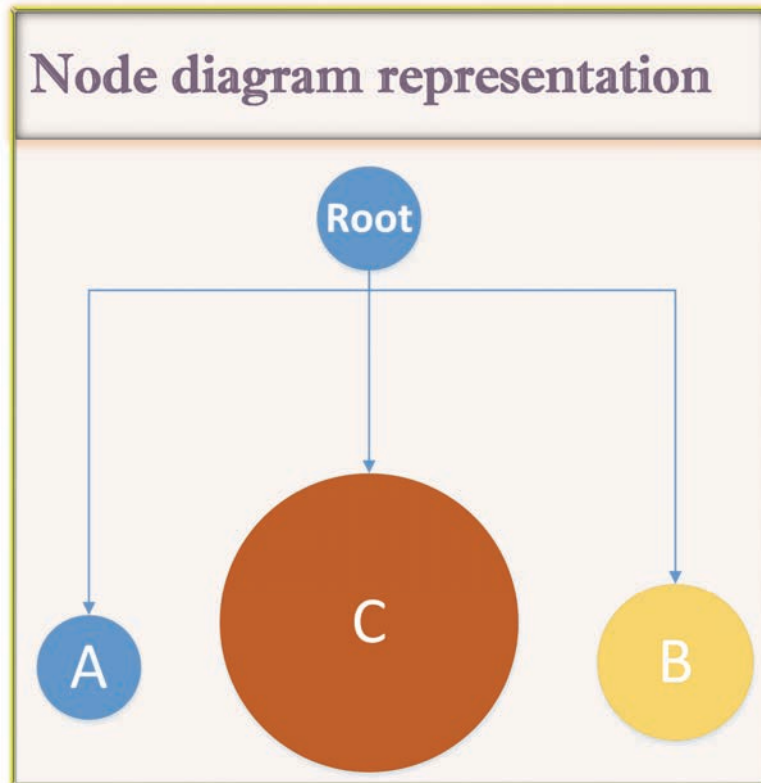


Figure 2.7 A node-link diagram of a simple tree structure (Node size and color used to emphasize a specification of a node).

However, classic node-link diagrams make inefficient use of screen space, and even trees of medium size need multiple screens to be completely displayed. This necessitates scrolling of the diagram and global context may be lost, because only a sub-part of the diagram is visible at any given time. Labeling of the node-link diagram is still a challenge, the users cannot trust on a node-link diagram visualization where there are heavy label overlapping or some labels of key concepts (may be different for different users) were selected to delete during the process called selective omission. This is the common limitation of almost all techniques in the early period of InfoVis (except the Treemap and Indented list).

2.5 Ontology and Ontology Visualization

Ontology is usually referred to as a formal and explicit description of concepts (classes) in a domain of discourse [29]. It contains the objects, concepts and other entities that are presumed to exist in some areas of interest and the relations that exist between them [30-32]. There are many mathematical definitions of ontology, such as those by Amann and Fundulaki [33] that can help in understanding how ontology can be processed by programs, and already wildly applied in the domain of Database.

Ontologies are useful to effectively present knowledge. The main reason ontologies reach outside the AI domain is their ability to support semantic linking, user interaction and

visualization. For example, with the power of InPhO ontology[34], we can generate an insightful island for a large encyclopedia. Ontology enables many complex semantic relationships, associations, and interactions in a knowledge system to be formalized for processing by machines, which provides multiple ways of presenting or operating on the same set of data. For this reason, ontology visualization has attracted much interest with many research projects developing and testing methods, trying to find the best way of visualizing ontologies in order to achieve favorable outcomes for end-users.

The ontology visualization is not an easy task, because ontology is a data type richer than a tree structure dataset. The complexity of ontology involved including a hierarchy of concepts, concept attributes, concept relationships, and relationship roles. This is further complicated when concepts have thousands of instances attached to the concepts.

This problem is usually addressed in ontology visualization by reducing ontologies to an approximation of a hierarchical structure (tree-structure) that constitutes what is sometimes termed as “skeleton”. Usually, this skeleton gives a useful approximation of the ontology. However, low levels of user satisfaction in relation to the support of ontology visualization and exploration provided by current ontology visualizing tools [35].

In 2007, Katifori et al. [35] reviewed the existing works(published before July 2006) on ontology(taxonomy) visualization. They presented the techniques and methods and categorized their characteristics and features in order to assist method selection and promote future research in the area of ontology visualization. Besides ontology visualization techniques, they also included some tree or network (graph) visualization techniques, which are not created specifically for ontologies in their survey. They categorized existing techniques into the following six categories:

- Indented list: The windows explorer-like (file-explorer metaphor) tree view of the ontology (taxonomy), for example, the Protégé Class Browser [18]. See section 2.4.1 for details.
- Node-link diagram: represents taxonomy of ontologies as a set of interconnected nodes, see section 2.4.2.2. Normally these node-link diagrams allow their users to expand and retract nodes and their subtrees, in order to adjust the detail of the information shown and avoid display clutter (nodes' overlaps).
- Zoomable: These techniques allow the user to zoom-in to the child nodes in order to enlarge them, making them readable in the viewing level. Grokker [36] is an example of this group.

- Space filling: The treemapping approaches. The classical Treemap[3], is an example from this category, see section 2.8.1 for more details about space-filling algorithms for creating visualizations.
- Focus + context or distortion: These approaches based on the notion of distorting the view of the presented graph in order to combine context and focus. The node (highlighted) on focus is usually the central one and the rest of the nodes placed around it. The 2D hyperbolic tree[37] is an example of this group of methods.
- 3D visualizations: These approaches placed the documents or classes on a plane as color-coded and size-coded 3D objects. For example, the data mountains[38] is an instance of 3D visualization.

They concluded that (until 2006)“there is not one specific method that seems to be the most appropriate for all applications”.[35].

One of the reasons why ontology visualization is not useful as we think is that many methods (e.g., most InfoVis plug-ins with Protégé[39]) dumb ontologies down to simple hierarchies (the skeleton of ontology), and simply visualize this tree structure with existing visualizations technique (e.g., node-link diagrams).

That may completely miss the purpose of knowledge visualization and the power of ontology is not present in these visualizations. Moreover, when ontology becomes large (more than hundreds of concepts), the users are not interested in exploring it with a large node-link diagrams, which just hold the same information as the Protégé classes browser. Additional once ontology grows large enough it becomes difficult to show its entire structure on a limited presentation space provided by a computer or a tablet.

The uses need a tool, who provides an overview of the ontology; it helps its user to maintain an overall mental model of the ontology. On the same time, an exploration process (function) needs to be supported, where the user can effectively focus on a part of the ontology, thus the users can change the level of analysis as they wish during their visual discovery, and meanwhile they do not lose the track of the overall organization of the ontology.

2.6 Knowledge Maps

In the meantime, knowledge maps is one of the most promising tools for visualizing knowledge, as it could help the users to access the knowledge contents with an overview of its structures (skeleton). Figure 2.8 is a knowledge map created by Martin Rosvall to Explore the mechanisms of map equation[40]. The knowledge map designed with cartographic means and geographic metaphors can effectively present the information besides that skeleton to make

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

sense of the data. Furthermore, users can visually navigate through knowledge on map with "intuitions" created by experts or visualization techniques, and benefit from these "intuitions" (visual metaphors). For example, in Figure 2.9, Scharnhorst et al created a knowledge map of knowledge orders with the data of Wikipedia and UDC (Universal Decimal Classification).

The reason why knowledge maps become useful can be explained clearly, because they provide the same utility of a map in our daily lives for the abstract information records. With the help of technology, a map is not prerequisite for the traditional tasks likes finding a path; one simple query to an information system can achieve these tasks more effectively. Nevertheless, the map is still indispensable, especially for the tasks of discovery and enlightening, or cooperation with others. For example, when we want to visit a city to discover the unknown area, and find the places that we want to visit with its detailed information, the map functions as a visualization of city's contents with its structure (geographic positions), and helps the users with tasks like comparison of different areas, discovery and sharing of geographic information. In the same way, we do not need a knowledge map that just presents the result of queries. We want to design a truthful, beautiful, insightful, and enlightening knowledge map benefited from power of knowledge, such as ontology, for the use of humans.

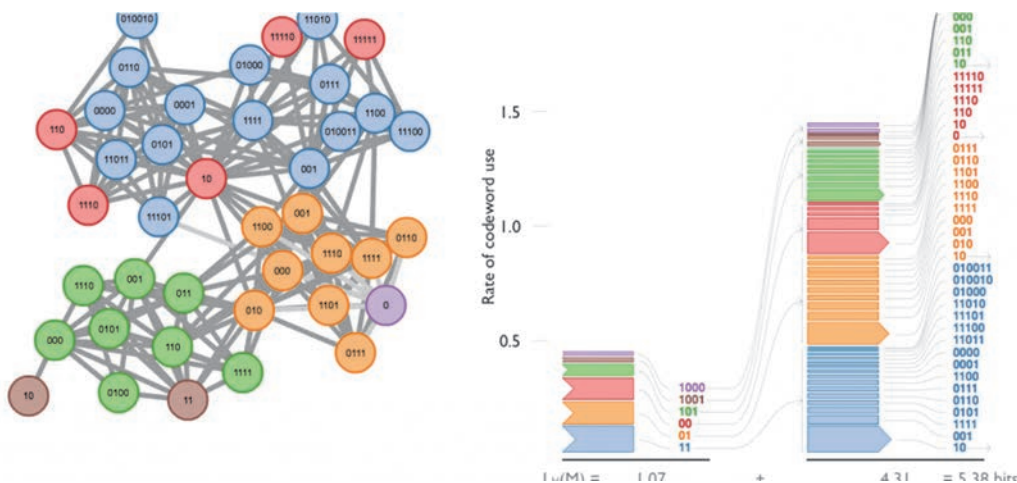


Figure 2.8 A knowledge map for exploring the mechanisms of map equation[40], the Infomap code available on mapequation.org (Image created by Martin Rosvall , reproduced in this thesis with permission).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang- June 2015

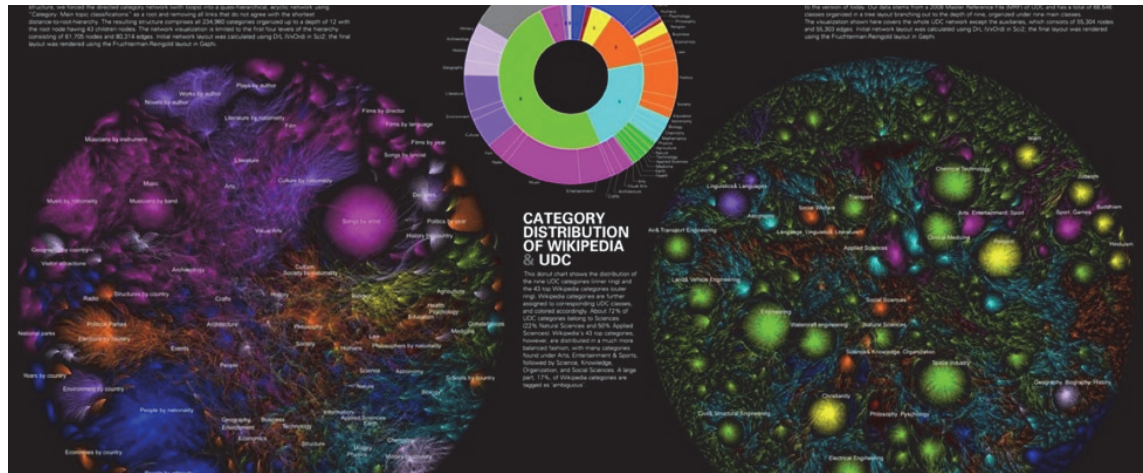


Figure 2.9 A knowledge map designed to visualize the knowledge orders. Map created by Almila Akdag, Cheng Gao, Krzysztof Suchecki and Andrea Scharnhorst. This knowledge map reproduced in this thesis with permission.

2.7 Cartography and Knowledge Cartography

Many cartographic principle and technique have been introduced to create the knowledge map; these cartographic InfoVis approaches are also called Map-like visualizations, as they created a knowledge map as the visual representation of the dataset. In this section, we briefly review the basic concepts of Cartography, the Knowledge Cartographic and the existing map-like visualization techniques.

2.7.1 Cartography

Cartography, as its name, is the study on making maps. It is combining science, aesthetics, and technique. The main topic of cartography is how reality can be effectively modeled.

The fundamental problems of traditional cartography are to:

- Set the map's agenda and select traits of the object to be mapped. This is the concern of map editing. Traits may be physical, such as roads or landmasses, or may be abstract, such as the (political) boundaries.
- Map projections: Represent the terrain of the mapped object on flat media (a 2D plane).
- Map generalization: The concern of generalization, how to make the map.
 - Eliminate characteristics of the mapped object that are not relevant to the map's purpose.
 - Reduce the complexity of the characteristics that will be mapped.
- Map design: Orchestrate the elements of the map to convey best its message to its audience.

These traditional fundamental problems need to consider when we design a cartographic/geographic InfoVis technique.

Modern cartography is largely integrated with geographic information science (GIScience) and constitutes many theoretical and practical foundations of Geographic Information Systems (GIS). The cartographies also worked on the evaluation of geographic visualization, and compared the map with the verbal. For example, Eide [41] has studied relationship between verbal and map-based expressions, and study why the people do not like to learn the geographical information from the map in Europe.

2.7.2 Knowledge Cartography

In recent years, some cartographers and researchers in InfoVis have tried to extend cartographic techniques to InfoVis for non-geographic information. These works are divided into two: Knowledge Cartography and cartographic/geographic visualization. Similar to the difference between Infographics (specific, handcrafted) and visualization (general, automatic)⁷, Knowledge Cartography is the discipline about mapping intellectual landscapes. It focus on how manually make an interactive, hyper-textual map for a knowledge, with one's own understanding, and facilitates the communication process. Okada et al. wrote a book on the topic of knowledge cartography's approaches [43].

Some outstanding maps have been created for a long time, such as Leonardo da Vinci's Mappamundi (Figure 2.10) for geographic knowledge. This map has many exceptional properties, such as the earliest map showing that America is not connected to Asia. Recently, Marco Quaggiotto has proposed a knowledge cartographic tool : Knowledge atlas(an example shown in Figure 2.11) [44] [45] to help the knowledge experts manually craft their knowledge map.

⁷ More detail about the different between Infographics and visualization can find in the blog of Robert Kosara 42 <http://eagereyes.org/blog/2010/the-difference-between-infographics-and-visualization>.



Figure 2.10 Leonardo da Vinci's Mappamundi (approximately in 1514).

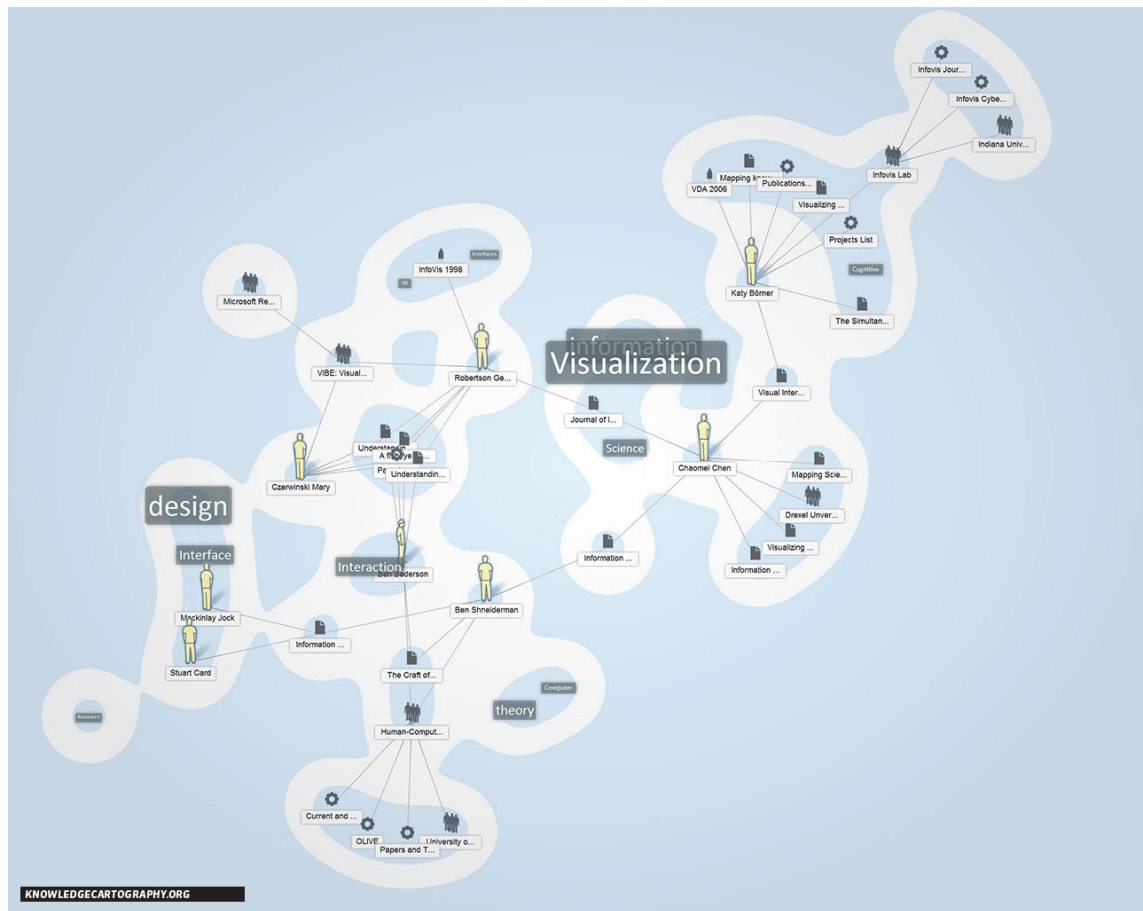


Figure 2.11 The screenshots taken from Marco Quaggiotto's Knowledge ATLAS [44] [45] (Image reproduced from <http://knowledgecartography.org/#images> with permission).

2.8 Cartographic Visualization

Memory Island technique creates a knowledge island (map) as the visual representation of the given knowledge. Visualizing a tree structure has been studied for many years, before we discuss our Memory Island technique; we review some important works on creating a map from the hierarchic data including the Tree-Maps, the approaches based on the spatialization, and the map-like approaches.

2.8.1 Treemapping Approaches

Tree mapping approaches apply the space-filling algorithms for creating maps. They display hierarchical data as a map of nested regions (rectangles or non-rectangular regions). For example, the famous classic Tree-Map uses a space-filling algorithm to create the maps. In the beginning of 90s, the existing tree-drawing algorithms have problem for display a large tree-structure in a limited display space. This type of space filling algorithms also been considered as tiling algorithms, which try to fill the display space. The first treemapping approach was proposed by Johnson and Shneiderman [3]. They proposed a space-filling algorithm inspired by the idea of mosaic for tree-structure data. As this visualization technique creates a map (e.g., Figure 2.12) for the hierarchical data set (tree-structure dataset), Johnson and Shneiderman named their InfoVis technique as Treemaps.

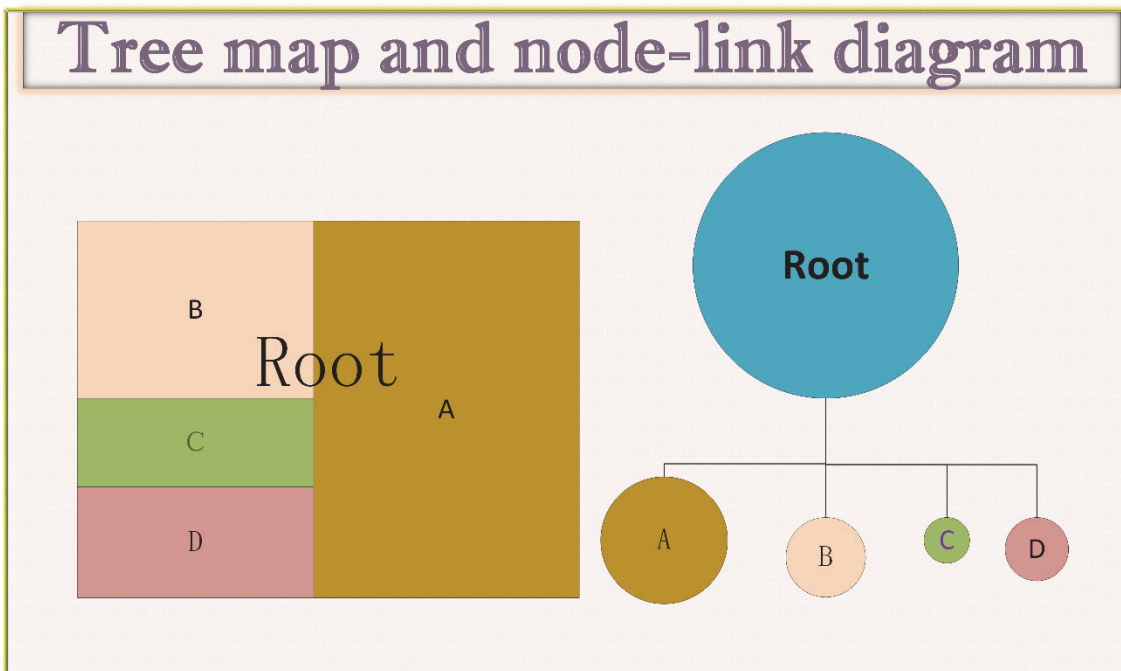


Figure 2.12 Tree map and its equivalent node-link diagram representation (with node's weight and node's type information).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang- June 2015

Then there are many other space-filling (tiling) algorithms proposed after the work of Johnson and Shneiderman. All these algorithms tried to create a map by using nested rectangular regions or non-rectangular regions. In recent time, in 2013, Auber et al. [46] used the geometry of Goseper Curve(non-rectangular regions) to create Goseper maps. In 2014, Duarte et al. proposed their Nmap (Neighborhood Map) space-filling algorithm [47]. This Treemapping technique tried to keep the distance-similarity metaphor (between the concepts in the hierarchy) in its result Map.

These approaches were widely applied in many domain, such as using in Disk space visualization tools for different operator systems. For example, in Figure 2.13, WinDirStat software applied a Tree Map for graphically displaying the amount of space used by files on a disk partition. With this interactive tree map, the end users can easily achieve the tasks of space managements of their disk. However, with these tiling algorithms, the relations between the concepts in the hierarchical knowledge become less evident.

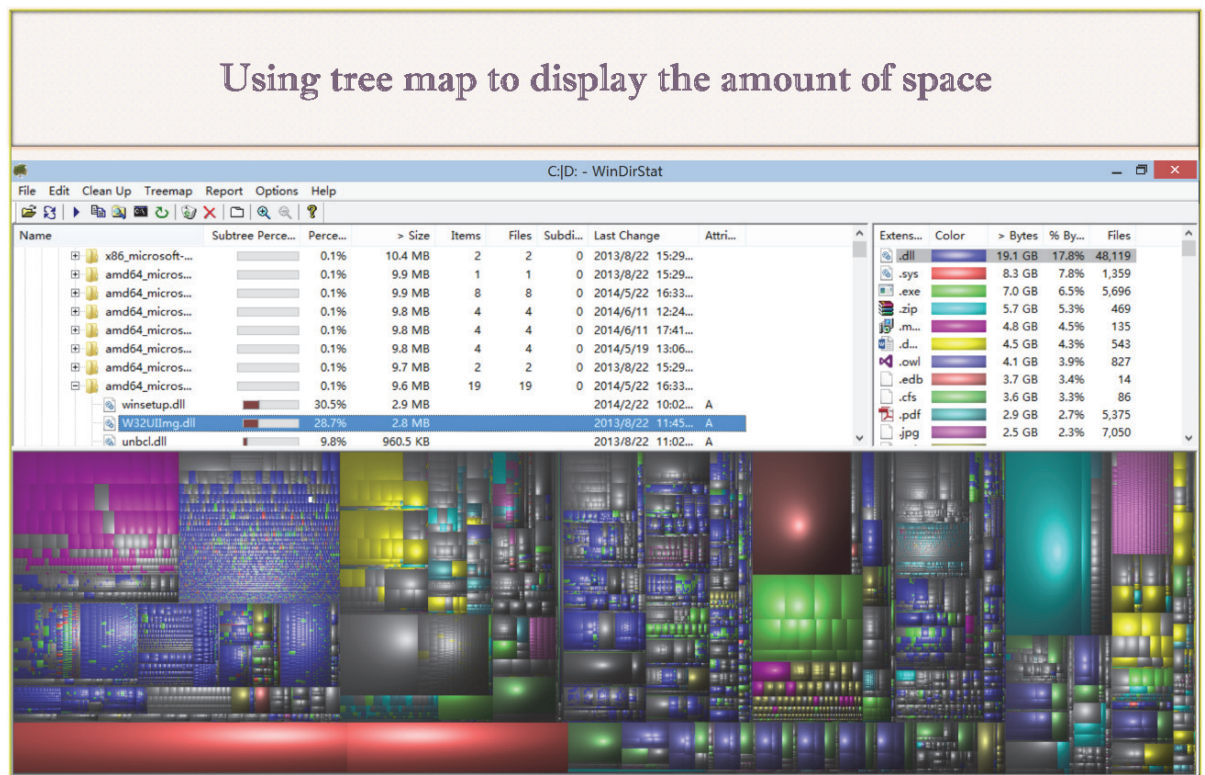


Figure 2.13 Using tree map to display the amount of space used by files on a disk partition. Generate by the free software WinDirStat⁸.

⁸ WindirStat software is available at <http://windirstat.info/>.

2.8.2 Spatialization (Layout) and Tree drawing algorithms

Graph drawing algorithms create the spatialization of Graph datasets, the tree drawing algorithms are the algorithms that create the visual representations for a given hierarchical structure. Tree drawing is a specific part of the Graph Drawing. The Graph drawing algorithms are also called layout algorithm, as the output of these algorithms are the proxy elements distribution (layout). We can automatically generate a geographic representation by directly applying those algorithms for the given relational information.

In 2013, Rusu [48] summarized in detail the different tree drawing algorithms. To build a node-link graph visualization, one of the layout algorithms in Tree Drawing or Graph Drawing can be applied to create the 2D or multidimensional representation for the concepts or clusters. For example, the BubbleSets [49] create for Sets visualization, it build a map representation by drawing on the existing spatialization (tree or network spatial layout) use either the traditional convex hull or implicit surfaces(draw contiguous contours around nodes). Hong et al. summarized the classic layout algorithms for visualizations in [50]. Skupin and Fabrikant [51] summarized the existing spatialization for visualizing non-cartographic data.

2.8.3 Map-like Visualization Approaches

Beside the Tree Map, the GMap algorithm[52] enclose group members with map metaphors(countries, seas and lakes) and the Self-Organizing Map(SOM) [53] approaches based on the clustering technique to build the 2D distribution for underlying data, are the most popular map-like visualization. One usefulness of this map-like visualization is that their result map can be used as basic maps to create many visualizations. For example, in the works of "Maps of Computer Science (Mocs)[54]" (an example is shown in Figure 2.15), Fried et al. used the GMap algorithm to generate a map and then overlap it with a heatmaps to create a map of computer science from different database.

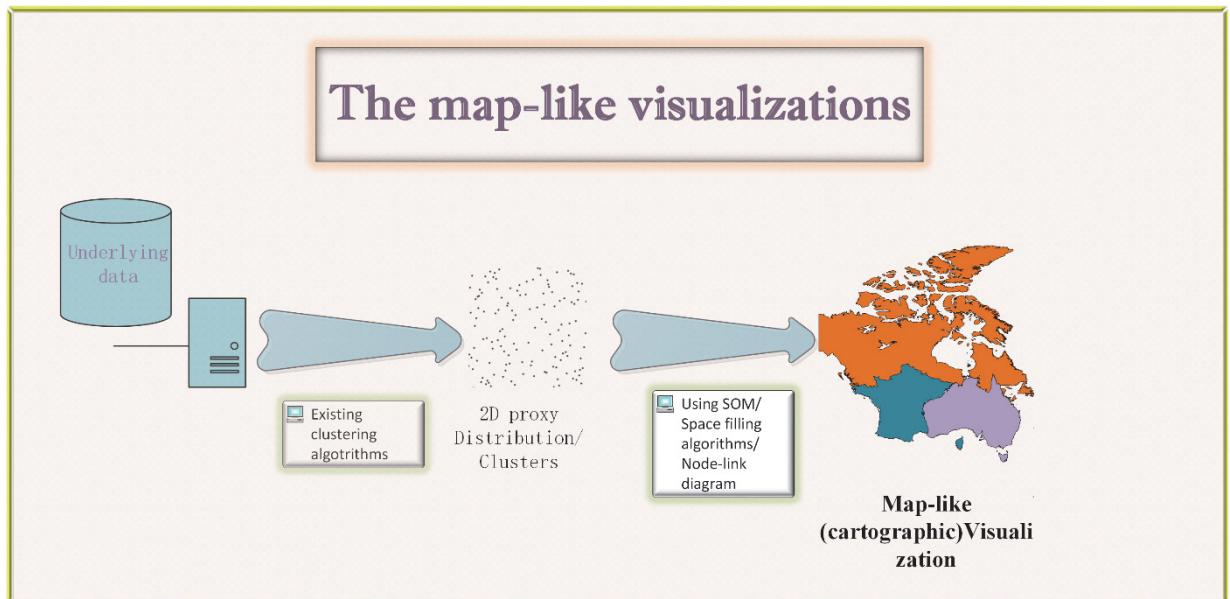


Figure 2.14 A schema for most existing map-like visualizations, apply an existing clustering technique on the underlying data to create the 2D distribution, then build a map based on this distributions.

In 2000, Skupin published his famous paper on cartographic perspectives on InfoVis[55]. Then in 2004, Skupin presents a map-like visualization with cartographic means[56], where he firstly introduced the geographic metaphor to the map generated by clustering technique. In his work, he used Self-Organizing Maps [53] (SOM, SOM is a clustering method considered as an unsupervised variation of the Artificial Neural Network, also called as Kohonen Map) proposed by Kohonen to generated the 2D distribution of the points. Other researchers introduced many different map-like visualizations using Skupin's method to visualize a specific dataset. For example, the island of music (based on psychoacoustics models and self-organizing maps) [57] for visualizing the music archives.

In 2005, Tu et al. [58] proposed a ontology visualization tool, which produces a holistic image of the ontology. It tries to arrange the classes in a semantic layout (distance between classes and/or instances is based on semantic similarity), their result map was generated by spreading the $n \times n$ (n is number of classes) network generated by SOM to form a grid.

Most of existing map-like visualizations are creating maps from data to derive knowledge. However, it does not means that representation of knowledge is useless, as many information scientists work on creating the info-graphics that make sense of the knowledge (e.g., the works of Knowledge Cartography). It is still a challenge to present knowledge into an interactive knowledge map with the own understanding of a domain expert. Our Memory Islands is a visualization technique designed to generate an insightful knowledge island from knowledge presented in ontology or clusters, by using the geographic metaphors and cartographic means to

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang - June 2015

reinforce human cognition and help the users with their knowledge understanding and memorizing, and help them to share knowledge and invent new ideas.

Gansner et al. [59] summarized a number of existing map-based approaches. For most of these works, the map generation(the proxy elements) have been done by applying clustering techniques[60] directly to the original data space or based on previously created spatial 2D space filling algorithm, e.g., by applying tessellation for the Tree-Maps[3]. For example, Balzer et al. used the Voronoi tessellation to build their Voronoi Treemap[61].



Figure 2.15 The Mocs with its base map generated by GMap algorithm used the topic of IJCAI 1990-2014 overlay with the Heat-Map generated used the topics of ECAI 1950-2014. The open-source program Map of Computer Science Program is used to generate this figure⁹.

2.9 Schematization

Many researchers worked on the schematization techniques for the domain of visualization and cartography. These schematizing techniques arrange or represent object in a schematic form. They can help to build various educational applications by emphasizing some key aspects and deemphasizing others. For example, Van Goethem et al. introduce a technique to obtain the shorthand for shapes [62]. In this work, they used the schematization technique then applied

⁹ This program is available at <http://mocs.cs.arizona.edu/code.php>.

Area-to-line transformation achieved via medial axis generation; for instance, countries like France are schematized to a simple tree-structure.

Another typical example of schematization technique is the subway/metro maps. In these schematic maps, the topological information between the stations are emphasizing in the maps, the other information about the cities are selected to deemphasize in these maps.

The Memory Island we proposed in this thesis is an inverse process of schematization, which tries to build an island representation for the hierarchical structure to make sense of hierarchical knowledge using our island generation algorithm. With Memory Island, a rich dataset, such as ontology, can be insightful visual represented as an interactive insightful island.

2.10 Spatial Cognition

Spatial cognition is concerned with the knowledge acquisition about spatial environments. The conclusions of spatial cognition have become a set of guidelines for designers of visualizations. Thorndyke et al. [63-65] studied the spatial knowledge acquired from maps. According to them, the knowledge presented in maps can reside in memory in the form of images, just like a physical map. The knowledge maps benefit from this to help their users' knowledge acquisition and memorization. Some of their conclusions have already become a common knowledge for the InfoVis designers, such as the use different geographic metaphors like distance in a map.

Their works were then extended to virtual space, for instance, Darken and Sibert [66] investigate navigation of large virtual spaces, stating "adding real world landmarks, like borders, paths, boundaries and directional cues, can greatly benefit navigation performance in virtual reality". The visualization technique Data Mountain was inspired by this conclusion[38]. Based on the researches on spatial cognition, we designed our own geographic metaphors and cartographic means to use the power of knowledge.

Part II.

Memory Island technique

[New ideas would come about] “by a connexion and transferring of the observations of one Arte, to the uses of another, when the experience of several misteries shall fall under consideration of on mans minds.”

--Sir Francis Bacon [1605],

From The two books of

<The proficience and advancement of learning>.

“The Purpose of Visualization is Insight, Not Pictures.”

-- Ben Shneiderman[2008][67]

"Everything is related to everything else, but closer things are more closely related" (First Law of Geography)

W. Tobler[1970][68]

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

Chapter 3 MEMORY ISLAND

IDEA AND DESIGN

METHODOLOGY

CONTENTS

3.1 THE ARTS OF MEMORY TECHNIQUE AND THE NOTION OF MEMORY ISLANDS.....	39
3.2 THE OBJECTIVES OF MEMORY ISLAND TECHNIQUE.....	41
3.2.1 Recognition (truthful).....	41
3.2.2 Discovery (functional and beautiful).....	41
3.2.3 Surprise (insightful and enlightening).....	42
3.3 THE IDEA OF MEMORY ISLAND	42
3.4 MAP (LANDSCAPE) METAPHOR	43
3.5 WHY DO WE CHOOSE 2D TRADITIONAL MAP REPRESENTATION.....	43
3.6 GEOGRAPHIC METAPHORS AND CARTOGRAPHIC MEANS.....	44
3.6.1 Proportion Metaphor.....	44
3.6.2 Distance/centrality Metaphor.....	45
3.6.3 Font Attributes and Point Attributes	46
3.6.4 Colors.....	48
3.6.5 Paths and Gaps.....	49
3.7 MEMORY ISLAND PROTOTYPE ALGORITHM.....	50

According to Sir Francis Bacon (see the cover page of Part II), we can learn a lot by taking lessons learned from one discipline and looking to see if we can apply them in another. Our Memory Island technique is inspired by the ancient Art of Memory technique. Here in this chapter, we start with the basic concept about the art of Memory technique and the notion of Memory Islands. We discuss in detail the basic idea of Memory Island technique and we discuss why this technique can give its users insights but not only pictures with its objectives and hypothesis. In the end of this chapter, we present our general prototype algorithm technique for visualizing hierarchical knowledge using the notion of Memory.

3.1 The arts of Memory technique and the notion of Memory Islands

The notion of Memory Islands for visualization was introduced by Ganascia to visualize the electronic books [69] by using a radial layout algorithm, then Ganascia et al. used this approach to developed an application of memory Island to the description of the EPG (Electronic Program Guide) for the DTV(Digital TV) content [70]. Two examples of this work are shown in Figure 3.1. This notion was inspired by the ancient “Arts of Memory” technique, that why it was named Memory Islands. According to Ganascia et al., “the representation corresponds (by using the Memory Island) aim to an increase of dimensions, which is quite unusual in InfoVis, since the general aim (of InfoVis) is to reduce data dimensions. The main goal of the notion of Memory Islands is not to focus attention on a particular item, but to represent a wide variety of contents and to stimulate human memories with an easy to remember picture, which facilitates user interactions with the contents” [70].

In this thesis, we propose a new InfoVis technique by using this notion of Memory Islands. In this chapter, we start by discussing the objectives of Memory Island technique.

The notion of Memory Islands

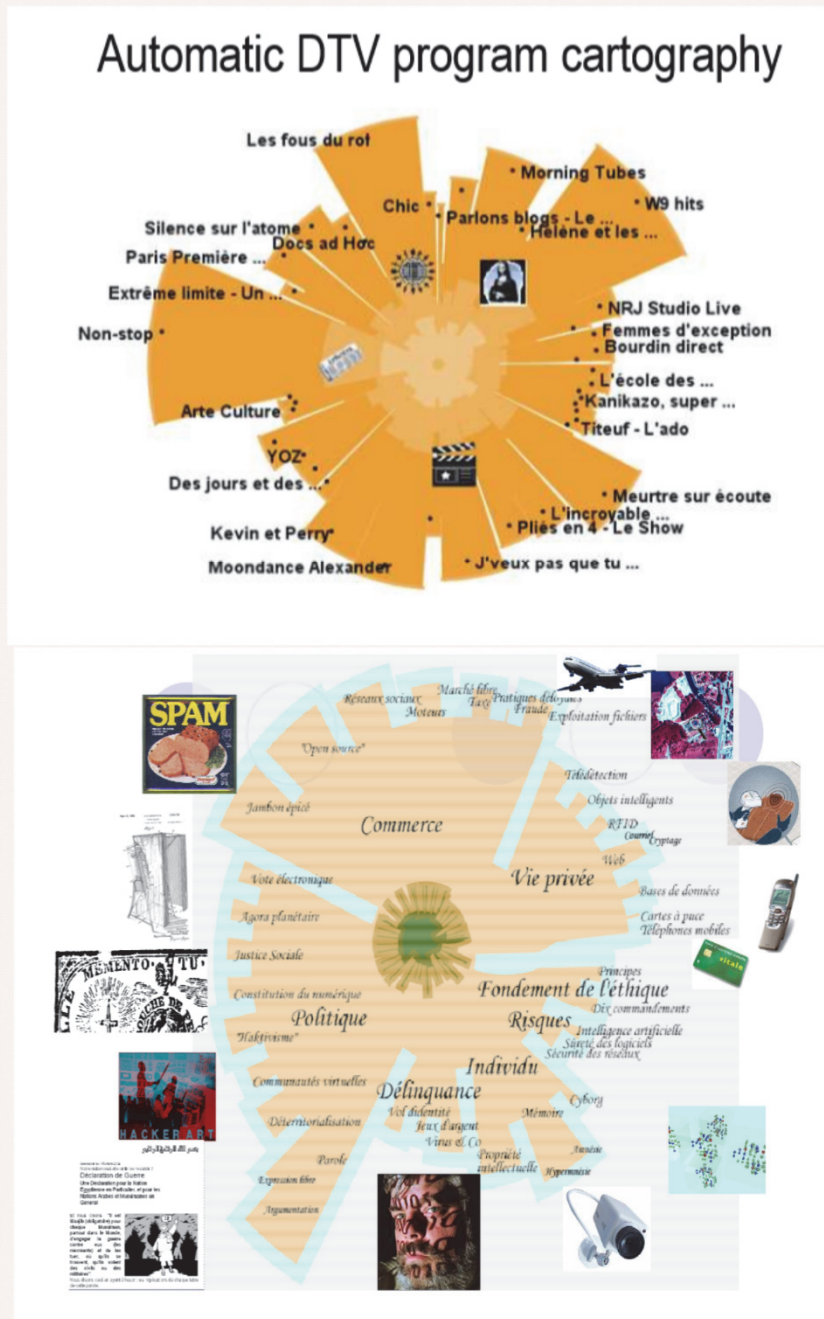


Figure 3.1: Two examples of using the notion of Memory Islands for the e-books and the DTV.

3.2 The Objectives of Memory Island technique

According to Skupin[71], a knowledge map need to fulfill three main tasks: Recognition, Discovery, and Surprise. I would like to use these three tasks with Cairo's 5 great features [72] for visualization (truthful, functional, beautiful, insightful and enlightening) to explain Memory Island technique's objective (Figure 3.2), as it generates an interactive insightful knowledge island as output.

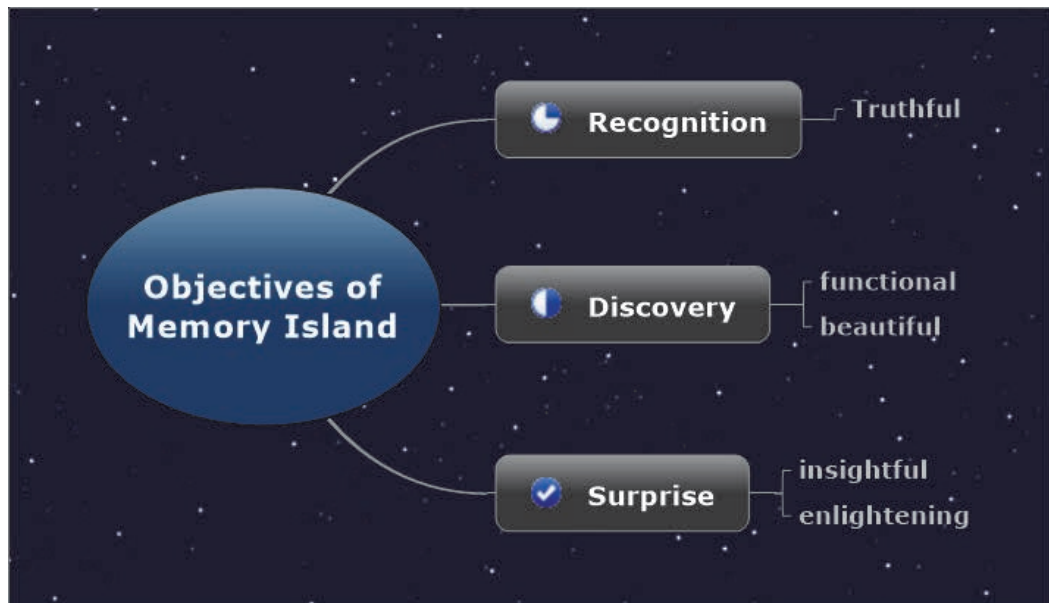


Figure 3.2 The Objectives of Memory Island technique

3.2.1 Recognition (truthful)

Recognition means that we need to provide some facts, which the users already know, then they will trust the Memory Island; in order to fulfill this function, we create our island generation algorithm based on the skeletons of knowledge (tree-structure), which the users can easily recognize. Then they can trust on our resulting knowledge maps, and enjoy their visual knowledge discovery and information seeking tasks with Memory Islands.

3.2.2 Discovery (functional and beautiful)

When the users trust the Memory Island, they should be able to navigate through the information space, and be willing to explore and discover by using the visualization. We need to design the Memory Island interactive functions and interface, to make most users easily achieve the task of navigation even without training, and support them to focus on any part that arouses their interest.

3.2.3 Surprise (insightful and enlightening)

Surprises are insights and phenomena, which are not easy to find in a large dataset by users. The giants and experts used "intuition" to prepare such surprises. With Memory Island, we would like to use the power of knowledge to provide them. These features are presented with geographic metaphors and cartographic means. The resulting map may even surprise and enlighten the experts, because when the dataset becomes large, the resulting knowledge map cannot be completely foreseen by human users.

3.3 The idea of Memory Island

The basic idea behind Memory Islands was inspired by the ancient "Art of Memory" technique, more specifically, it is inspired by its method of "loci" (Latin for place or location), that consists of associating each entity to a designated area and creating a virtual map (island) for them to learn and memorize knowledge. A well-designed map in mind can make sense of knowledge, and help users to extend their knowledge.

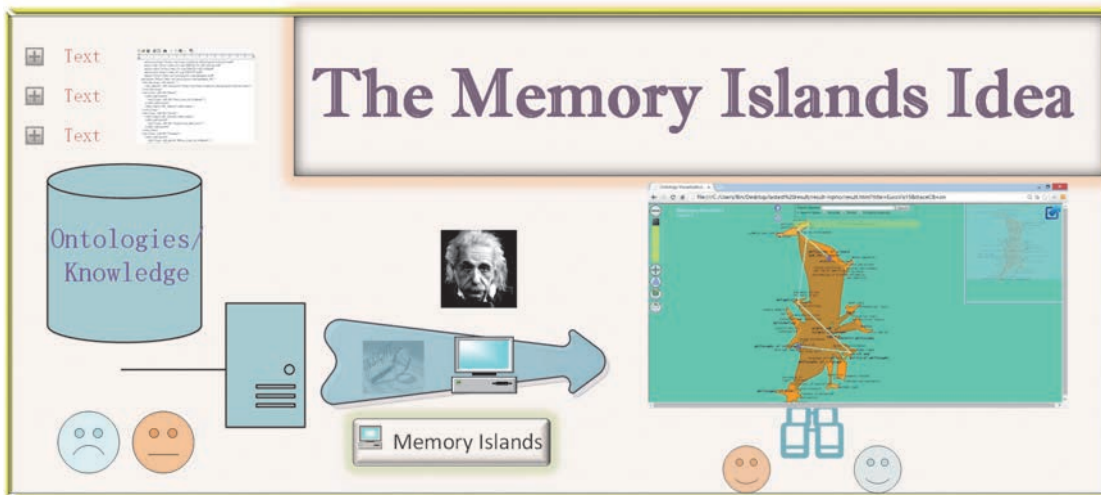


Figure 3.3 Memory Islands Idea: transform hierarchical knowledge (e.g., ontology) into an insightful 2D island; each concept has its own sub-island. The island is generated by using the power of knowledge such as ontologies, which simulate the "Art of Memory" used by great scientific personalities.

The recent study called "island of knowledge" [2] shows that using the "intuition" is very important for the tasks with knowledge, and even scientific personalities such as Einstein, used their "intuition" to extend their knowledge. Many studies, such as in the work of Dobrowolski et al. [73] and Westerman et al. [74], confirmed that the ability to observe the different thing between concepts(e.g., the (dis)similarities) by navigating is crucial for the creative use of information. An interactive knowledge map such as infographics together with the own understanding of a domain expert can help the users as a cognitive aid. The knowledge such as

ontology has the power to support us to create such "intuition". That was the underlying idea of Memory Island technique, to generate an insightful interactive island for the users, to help with the visual knowledge discovery and learning (Figure 3.3), and improve their memorability.

3.4 Map (Landscape) Metaphor

Map or landscape metaphors are well used in situations where numbers of data items are to be displayed. Map metaphor have been widely used in InfoVis, such as the Topic islands [75]. With a Map-like visualization, we could easily overlay another map to provide more insights and make sense of data. For example, the "Map of Computer Science"[54]. Beside the advantages of a map representation in spatial cognition, one reason I want to use the map metaphor to represent knowledge is: it could open a door for extending visualization of geographic information to the domain of Visual Analytics and Knowledge Discovery.

3.5 Why do we choose 2D traditional map representation

To design a visualization with map metaphor, the choices need to be decided between a 2D and 3D map, as well as between a static or dynamic map. When we design a 2D/3D visualization, we must consider the overall affordance, cognitive perceptual and interaction costs. I have the following reasons to stay with 2D map representation: First, the users are familiar with the 2D map (like the Google Map), and the advantages of a 2D map have already been proved. Both knowledge structure and the information contents can be presented in a 2D map. In addition, the 2D map can be easily used by the researchers in InfoVis to create a meaningful visualization by overlaying it with other maps. With a 3D map, it will be more difficult to effective overlap it, for instance, with heatmaps. Secondly, Cockburn et al. [76, 77] have investigated the usability of 2D and 3D representation for visualization, and they strongly suggest that "the effectiveness of spatial memory is unaffected by the presence or absence of three-dimensional perspective effects in monocular static displays".

A fully dynamic visualization may help the experts in Information Science to create their own knowledge map. However according to some works and from our own user experiments, for most users without background knowledge, the full dynamic interactive function may become a disaster. While bringing more options for navigating, it increases the burden for users. The users perform better with a visualization if they are more familiar with the data. Additionally with a dynamic knowledge map, the effect of visualization depends on the user itself that may missing the purpose of a knowledge map: there is no surprise from the map. For example, a full dynamic visualization cannot benefit from the power of ontology to support the visual knowledge discovery. That will make knowledge visualization dumb to a full dynamic visualization of the knowledge's skeleton. In addition, a good knowledge map generated by the

visualization can help the users to open the door for the future question to extent the knowledge[72].

3.6 Geographic Metaphors and Cartographic Means

Beside the landscape metaphor, we also introduce some geographic metaphors and cartographic means to express the information that we found inside the knowledge dataset, such as an ontology(InPhO [34]) generated from the Stanford Encyclopedia of Philosophy (SEP). Most people have been already using the geographic maps and can recognize and understand geographic metaphors easily. These metaphors are indispensable when we visualize ontologies: ontologies are rich data structures for a profound representation of knowledge. They are not only exclusively trees (nodes can have multiple parents) and the nodes as well as the edges are typed, many resources can be associated with the concepts, etc. Memory Islands use the geographic metaphors to describe this information. In order to explain more clearly, we used the InPhO ontology for the SEP as a typical example. In InPhO (version 2014.Nov.) ontology, there are 265 ideas (277 nodes in its skeleton) and 276 instances associated with them. Both concepts and instances are associated with encyclopedic articles.

3.6.1 Proportion Metaphor

We use proportion metaphor to map the number of encyclopedia articles or instances to the size of the wedge corresponding to a concept (an example shown in Figure 3.4 b). This proportion metaphor is quite familiar with most of end-users, and we believe it can help the users quickly get some awareness on the knowledge dataset. If the advantage users (for example, an expert on the domain) do not wish to define this kind of meaning with Memory Island, the children of a concept will equally share the available space(an example shown in Figure 3.4 a), that mean each sub-concept in the same hierarchical level have the same importance.

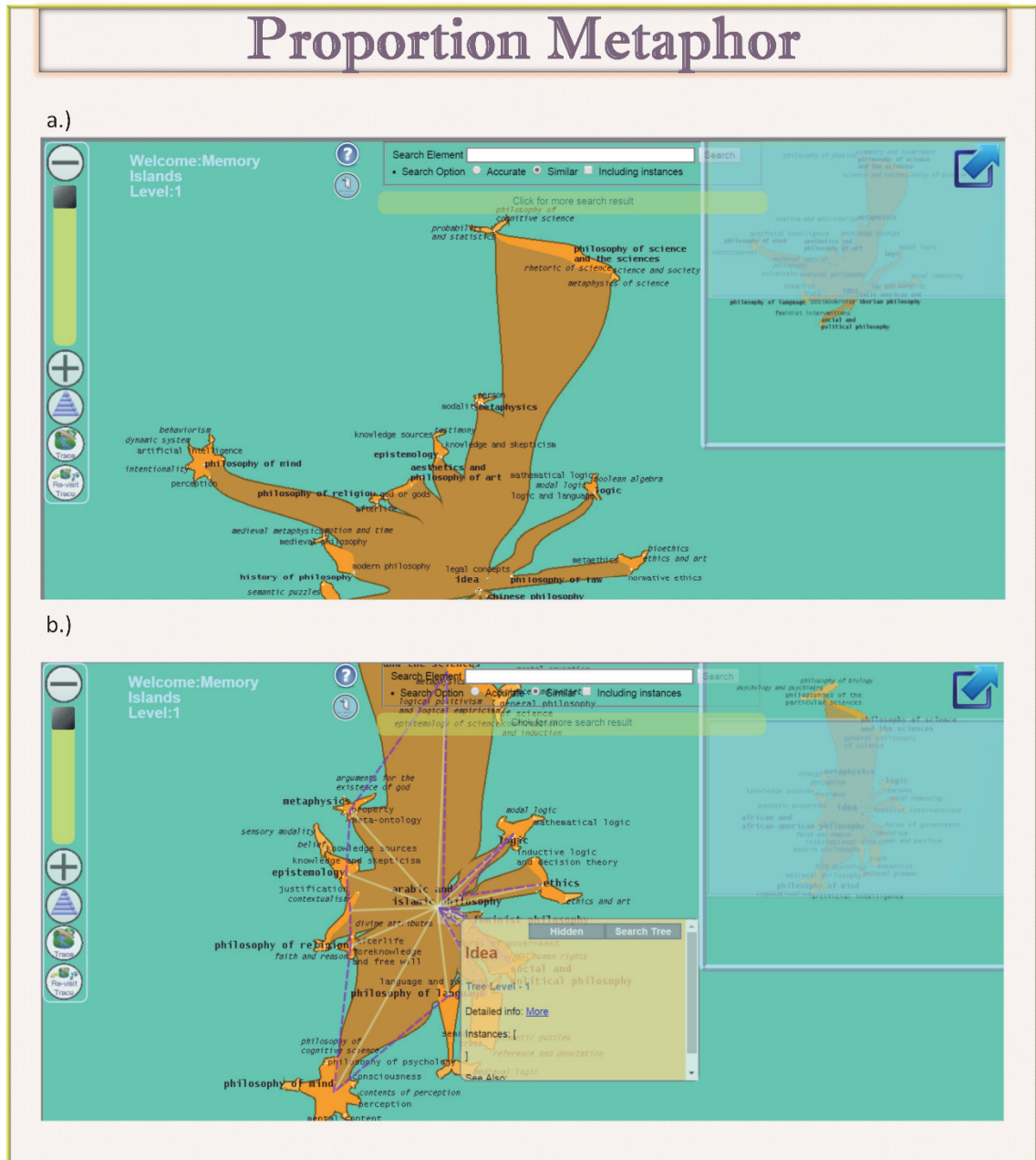


Figure 3.4 Two examples generated for the same ontology InPhO with different proportion metaphor with Memory Island.

3.6.2 Distance/centrality Metaphor

The center of an island or a sub-island is the most important concept in the hierarchical structure, which emphasize the hierarchical relations in the knowledge's skeleton. The distance from a sub-concept or sub-cluster to the parent concept maps to the number of instances and sub-concepts with a concept (e.g., Figure 3.5). The distance metaphor we introduced is similar likes the weight in Graph Theory in Computer science. For the experts in Information Science, they believe use this distance metaphor to map the similarity of a concept according to another

related ontology/database about the parent concept will be more reasonable. This is also a promising future research direction for Memory Island technique.

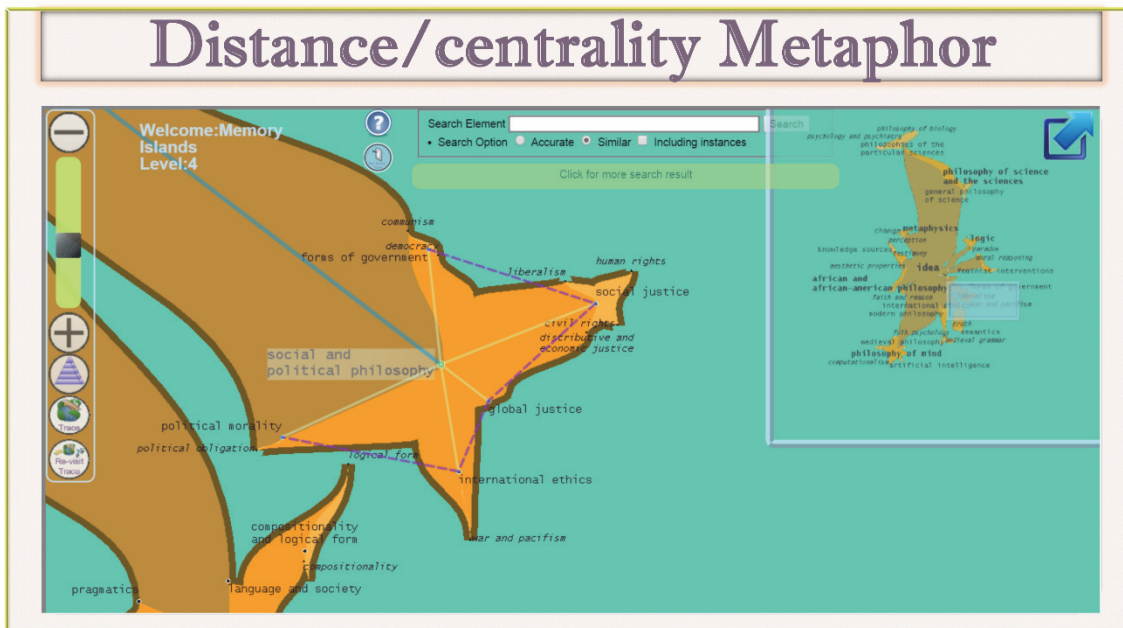


Figure 3.5 An example about the distance/centrality Metaphor, the high level is located in the center of sub-island, the distance from a location of concept to its parent concept's location is calculated with number of instances, and sub-concepts (weight) associated with this concept.

3.6.3 Font Attributes and Point Attributes

The label's font sizes map the importance of a concept, while different types of font map the group of concepts (e.g., The SEP Memory Island, we used the taxonomy level to group ideas, from general to specific ones). We used "Lucida Regular Bold" for the top two level of concepts, "Lucida Regular Plan" for the 3rd tree level concepts, and "Lucida Italic" for the rest concepts (Figure 3.6). We selected these fonts are selected with the help of an expert of typography.

We calculate the label's font size by both the label's hierarchical level and the zoom-level. When the users zoom-in on the map, the size of label need to be suitable increased, in order to make the map still readable for the end-users. The cartographic point's (location or places) type describes a concept's type in the dataset (e.g., the concept's type in ontology), if all the concepts in the Memory Island are the same type, we use it to emphasize the importance in hierarchy (Figure 3.7). We use the size of point to emphasize the hierarchical levels; it can be also decided by the number of instants associated with the concept. This is similar to the cartographic tradition of a map of cities, thus it is easy for the end-users to understand this mapping, even without specific description on it.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

Similar to our work, Brath and Banissi[78] also study the font attributes with knowledge maps, according to them, the end-users can well-understood and benefited from the font metaphor. They stated that the visualization with font attributes improve the performances of the end-users.

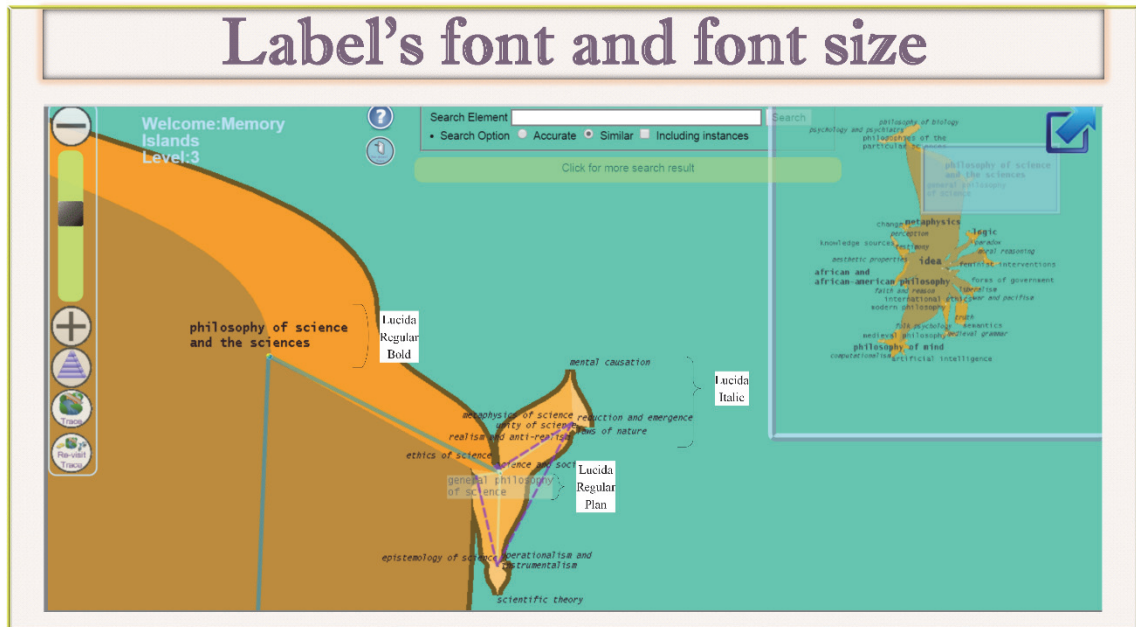


Figure 3.6 The fonts we used with Memory Islands.

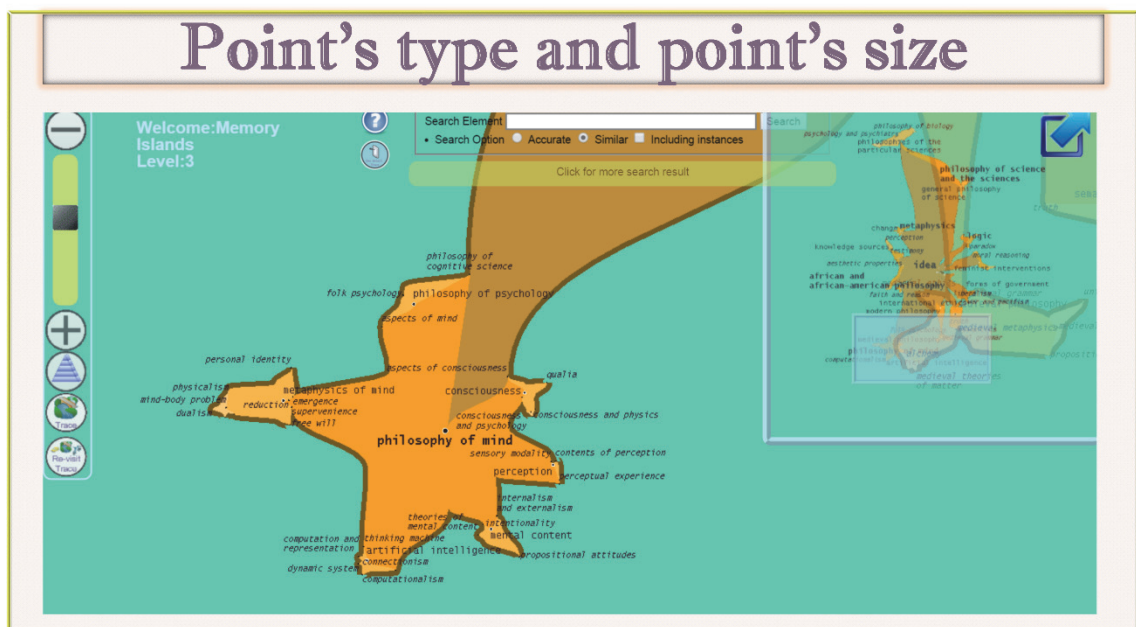


Figure 3.7 Points (locations) attributes in Memory Islands, points may have different types and different sizes.

3.6.4 Colors

Coloration strategy has been widely used in cartographic visualization; a well-designed coloration strategy could empower the visualization. Coloring can be done by applying a technique such as Temporal Trend [56]. In Memory Islands, we color the different hierarchic level clusters with the same group of colors shown in the schema of color gradients. We use an opposite color to simulate the sea of unknown. An area of a concept in the map can be colored by a special type if we want to emphasize its special inter-relation to other parts (Figure 3.8)



Figure 3.8 The color wheel shows our color selection strategy, and an example of Memory Island generated from table of contents of a paper; the main part of the island is colored by using the colors in-group A (in the wheel), the sub-island for chapter 4 (about technical details) is colored by using the colors of group B.

3.6.5 Paths and Gaps

We introduce the cartographic paths to express two different cartographic means. The first type of paths show the hierarchy (data lines), and their color is used to express the type of relation between concepts (Figure 3.9).

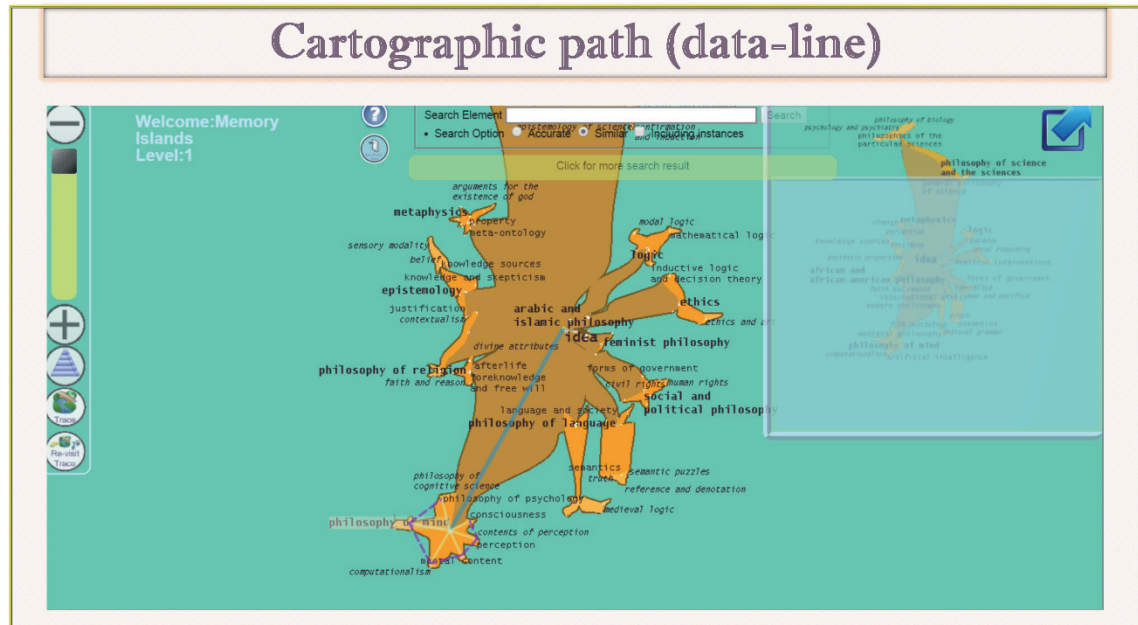


Figure 3.9 The blue and white data-lines describe the relations (subsumption) associated with the selected concept philosophy of mind.

We design the second type of paths to show the orders of elements in the knowledge. Similar concepts or clusters will be arranged more closely and the path in violet between them can be shown when demanded (Figure 3.10). The order between the nodes with the same high-level concept is measured by the semantic similarity. We use the following metaphor: a cartographic path (colored in violet) between two places means a low cost to move between them (e.g., there are airline between these locations), therefore for each concept, its neighbor(s) in these violet route is (are) its most related (similar) concept(s).

Meanwhile, we introduce the gaps (such as gulfs) between the sub-islands of clusters or concepts (Figure 3.9). Those are usually harder to emphasize in the map-like visualization by using clustering (SOMs) or space filling algorithms. And according to the evaluation study of Jianu et al.[79], with the map with disjointed areas, the user performs equally or better than with other maps.

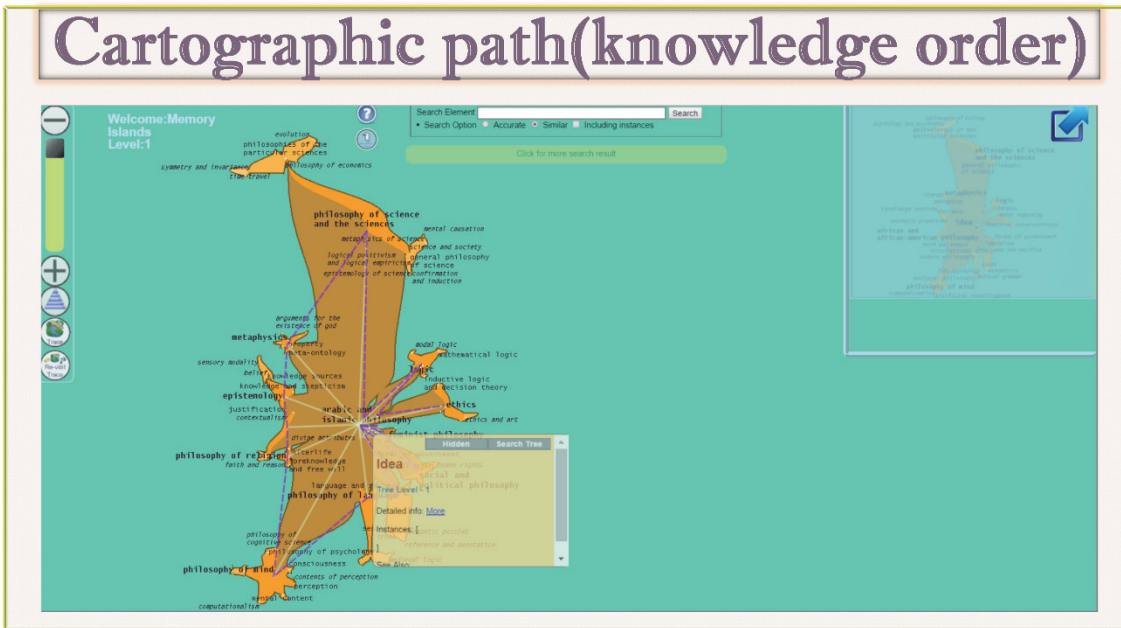


Figure 3.10 Cartographic paths (violet) between two concepts indicated the two concept is more closed.

3.7 Memory Island Prototype Algorithm

Based on the reflections mentioned above, Memory Island technique consists of: (1) extracting concepts, relationships (skeleton) and the other types of information from sources. (2) Automatically generating the visual representation corresponding to the given knowledge and displaying all the information and phenomenon found from the knowledge. (3) generating a user-interactive interface to help the user to navigate and memorize that knowledge. Our prototype algorithm for Memory Island technique is described in Algorithm 1.

Algorithm 1 Memory Island Prototype Algorithm

Memory Island Prototype Algorithm
Input: Hierarchical knowledge likes ontology
Output: An interactive Memory Island
Step 1 Extract skeleton: Extract the skeleton of knowledge (and parse knowledge (e.g., ontology)), build a Memory Island Tree to store all information found from the knowledge and the related resources.
Step 2 Hierarchical reorganization (Order of elements in knowledge): Reorganize the tree structure according to one knowledge order (e.g., semantic similarity).

Step 3 Island generation: Apply our Island algorithm to create an island representation of the knowledge.

Step 4 Labeling and map generation: Initialize the map size according to the zoom-level (start at 0), project the island to the map area, and then initialize the labels associated to each point according to the given configuration (random placement for first zoom-level), then apply labeling algorithm.

If (no overlap appear in map) then

Go to Step 5;

else

Increase zoom level. (Map size increase 4 times.) Save current label placement configuration for next zoom level and go to Step 4;

end-if

Step 5 Interactive Users Interface: Create images for each zoom level and store the information of each concept of each zoom level in a database. Generate a web-scale (HTML5+javascript) Memory Island interface with interactive functions as the output.

End algorithm

Based on the discussion in this chapter, we can find that our Memory Island algorithm can overcome the limitation of many the existing ontology visualizations, who only provide a visual picture of the taxonomy and do not display all its information contents. In the rest of Part II, we describe this technique with all its details for each step in this prototype, to explain how this technique can overcome the limitation of these techniques based on the Graph Drawing (difficult to display all concepts' labels, and a lot of useless empty spaces) and the Treemaps (less evidence of relations).

Chapter 4 HIERARCHICAL REORGANIZATION BY SEMANTIC SIMILARITY

CONTENTS

<i>4.1.1 Related works</i>	55
4.1.1.1 Semantic similarity measure	55
4.1.1.2 WordNet and its application for semantic similarity measure.....	56
<i>4.1.2 Hierarchical Reorganization by Semantic Similarity</i>	56

Reorganization is a worthwhile topic for design visualization, the organization of the other data type of visualization can empower the visualization technique. For instance, in 2013, Venturini et al. [80] worked on re-organization for visualizations on the online analytical processing (OLAP), and their findings in their user study indicated that this reorganization process is useful for the users.

For knowledge visualization, order of concepts is one of the key concept. As we have discussed, the similar or dissimilar between the concepts is the key process for the performances of information seeking tasks. Therefore, we introduce the hierarchical reorganization using a semantic similarity measure for Memory Island.

4.1.1 Related works

In this section, we briefly review some basic concepts with semantic similarity and related works with the WordNet Lexical Database, which we used in our reorganization process. The existing general approaches on semantic similarity measures using WordNet or other corpus are reviewed by Slimani[81] in 2013.

4.1.1.1 Semantic similarity measure

Semantic similarity is important to measure concepts with semantic information. Varelas et al.[82] stated, “Semantic Similarity relates to computing the similarity between concepts which are not lexicographically similar.” Lin has given three definitions of similarity in [83] as the following:

- “Intuition 1: The similarity between A and B is related to their commonality. The more commonality they share, the more similar they are.”
- “Intuition 2: The similarity between A and B is related to the differences between them. The more differences they have, the less similar they are.”
- “Intuition 3: The maximum similarity between A and B is reached when A and B are identical, no matter how much commonality they share.”

Some popular semantic similarity methods implemented and evaluated using WordNet as the underlying reference ontology (for general purpose).

4.1.1.2 WordNet and its application for semantic similarity measure

Princeton University have developed a English lexical Database called WordNet¹⁰[84, 85]. In this online database, it grouped the words (nouns, verbs, adjectives, and adverbs) into sets of cognitive synonyms (synsets). It is common in the domain of ontology engineering; the name of a concept is noun or a set of nouns. Therefore, in Memory Island we consider only the measurement of semantic similarity of noun or a set of noun from the labels of the terms in the visualized hieratical datasets (e.g., ontologies).

Several researchers proposed different methods based on WordNet for determining semantic similarity between terms. Petrakis et al. [86] have summarized the existing works on Semantic Similarity using WordNet in four categories:

- Information Content Methods: in these methods, WordNet is used as a statistical resource(corpus) for computing the probabilities of occurrence (of terms), they use these probabilities to calculate the difference in information content of the two terms (e.g., the work of Seco[87]).
- Edge Counting Methods: These edge-counting methods, such as the approaches proposed in [88-90], measure the similarity by considering the length of the path of linked-terms and on terms' positions in the taxonomy (e.g., level).
- Feature based Methods: Measure the similarity of two terms by using a function of one of their features, such as their properties or their relationships, to other similar terms in the taxonomy.
- The Hybrid methods: as its name, it combine the above three approaches.

4.1.2 Hierarchical Reorganization by Semantic Similarity

The order of elements (organization) is important for creating a spatial representation for the hierarchical elements, especially for the knowledge. We consider the knowledge (concepts) orders by using a measure of semantic similarity. An intuitive way of visualizing these orders is grouping together closely related terms and spacing further apart the less similar ones.

When we want to implement hierarchical reorganization for Memory Island, we need to evaluate each method by measured and compared their performance based on some sets of terms in which we can use human sense of understanding to evaluate approximately the similarity score and also on the computation time. As a result we found out from the works of Pirro and

¹⁰ <http://wordnet.princeton.edu/>

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

Euzenat [91] [92], their approaches called FaITH similarity measure has given a good result with a very good computation time.

Therefore we used WordNet[84, 85] for computing the similarity score with the FaITH similarity measure[92]. We take into consideration the sets of nodes with the same parent node in the tree structure and try to re-order their appearance from leaves to root. We use the label of each concept for calculating similarity score. We then compare each of every pair of concepts and calculate the score. This way we are developing a fully connected graph with a similarity score attached to it. In order to reorder the sub-concepts of one concept, we are solving this problem by reducing it to the traveling salesman problem [93]. We discuss more details about the implementation of this hierarchical reorganization in section 8.1.1.

Chapter 5 VISUALIZING HIERARCHICAL DATA AS ISLANDS

CONTENTS

5.1 ISLAND GENERATION.....	61
5.1.1 Polyle II Algorithm	61
5.1.2 Memory Island Algorithm.....	64
5.1.3 Discussion on Memory Island Algorithm.....	67
5.2 RESHAPING THE RESULTING ISLAND	68

In this chapter, we introduce our island generation algorithms based on the idea of drawing a tree (skeleton) and the contours of each nodes to make an island from given knowledge.

5.1 Island Generation

Memory Island algorithm can be said as an inverse process of the schematization techniques, it build an island (polygon) based on the tree-structure skeleton of the knowledge. We defined a "Crown" of a concept as a small sub-island, which is a polygon created by the concept and its sub-concepts. For the simplest case, a crown of a concept is a point itself or a specific contour created by its children's crowns, like a contour of star Glyphs (within a plane). We also discussed why we safely use this island representation (crown) instead of the classic node-link diagram.

5.1.1 Polyle II Algorithm

Based on the Memory Island idea discussed in Chapter 3, we firstly introduce a new island generation algorithm called Polyle II. The Polyle II algorithm aims to generate an island based on the skeleton of knowledge (weighted tree). It draws the tree-structure in a 2D plane to form the island representation of the knowledge. The detail of this algorithm is described in Algorithm 2.

Algorithm 2 Polyle II algorithm

Input: Memory Island Tree (weighted tree)
Parameter: Unit increment for decay rate UI and Threshold of decay rate T ($T < 1$).
Output: An island polygon
Step 1 Initialization: Create a polygon P (Centre point of P is associated by the root node) by weighted tree T's root and its children (the first level nodes). All these tree nodes marked as unfinished; Each node's decay rate set to 0; Initial current node as root node
Step 2: Spatialization: If current node is root, then Go to Step 4. Else

Create a crown (with its child-nodes) for current node according to its weight and its decay rate.

The distance between the current node and its parent is calculated by a function of its weight $f(\text{weight}) * (1 - \text{its decay rate})$.

End If-Then-Else

Step 3: Test Overlap

If replace the current node by its crown in the polygon P Then

If this crown is not overlap with the existing polygon P. Then

go to step 4;

Else If current node's decay Rate < T;

Decay rate += UI.

End If

Go to step 2;

Else

Cannot generate an island by Polyle II

Return a polygon generate by using radial layout algorithm.

End If-Then-Else.

Step 4: Create the midpoints

Construct the midpoints between each pair of 2 nodes in the created crown. These mid-points used to emphasize the proportion metaphor.

Step 5: Update current island polygon

Replace the current node in the polygon by its crown (with midpoints). Mark current node as finished.

Step 6: Verify the end condition:

If no point (node) in the island polygon has been marked as unfinished Then

Return the current Polygon as result;

Else

Set the current node to the next non- finished node in Polygon; Go to Step 2;

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

End If-Then Else

End algorithm

By using the Polyle II algorithm, we can generate many delightful maps for many different datasets, as shown in Figure 5.1 However it has a limitation concerning the max-displaying sizes. As a consequence of reducing the distance between a current node and its parent the available space will also decrease. It is thus possible that in some cases there will not be enough space for assigning a minimum-distance between all such pair of points (to avoid overlapping appearing in the island polygon). An example of this overlapping is shown in Figure 5.2). Normally many InfoVis approaches are limited to scaling; in addition to that, this algorithm is also unable to establish what this limit is. Moreover, this algorithm reduces the distance (size of created crown), thus violating the distance metaphor. To overcome this problem, we proposed a novel island generation algorithm (Memory Island algorithm), based on the idea of Polyle II algorithm.

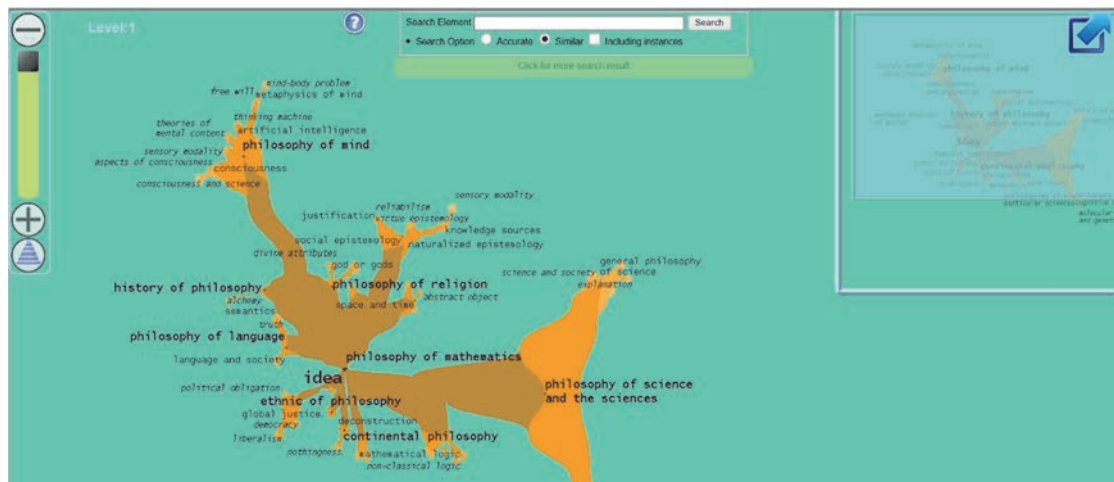


Figure 5.1 An example of Memory Island generated with Polyle II algorithm.

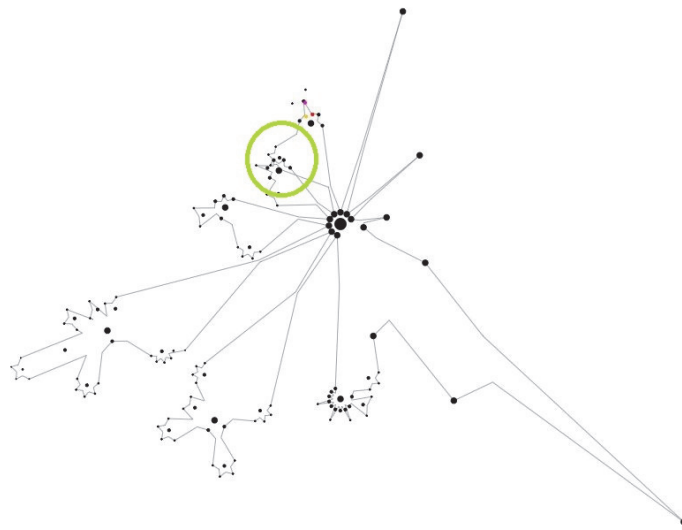


Figure 5.2 An example generated by Polyle II Algorithm, it has the overlapping between its sub-islands.

5.1.2 Memory Island Algorithm

Like some existing tree drawing algorithms, we design a recursive algorithm to build the island. Memory Island algorithm builds simple crowns(a point or a specific contour) for the low level concepts, and then uses the these crowns (small islands) to generate a larger island for a high level concept by assigning them to a designated position; we keep that high level concept in the center of that new crown (island). The following pseudo-code Algorithm 3 is an outline of our algorithm.

Algorithm 3 Memory Island Algorithm

Input: Memory Island tree Node : n

Output: A Crown polygon of the n

Begin Algorithm

If n.type == leaf Then

Return a crown with only one center point;

Else

childNode = n.leftMostChild;

While childNode != null Do

childNode.crown = Memory_Island_Algorithm(childNode);

```
childNode = childNode.rightSibling;  
childNode.crown = Memory Island Algorithm(childNode);  
End While  
Return generateCrown(n);  
End If-Then-Else  
End Algorithm
```

The pseudo-code Algorithm generateCrown is the function "generateCrown". The measure to calculate the wedge area of each child node is measured by our Proportion Metaphor, the parameter: "reservedWedge" is used to emphasize the cartographic gap between the first sub-concept and the last sub-concept (they are not closed) with a specified (concepts or knowledge) order. The functions "rotation" and "transition" assigned the crown polygon to a designated position. The "Test" function tests the crowns displays in their assigned position with restrictions. We test with the following restrictions: 1). "Is there overlapping between the assigned crowns?" 2). "Is a crown assigned to an invalid area (for example, is the center point of existing island inside of its crown)". Unlike the "hyperbolic tree" or radial tree layout algorithms, we do not constraints a node place all its descendants in its own wedge (a crown inside a wedge). It is also different from the approaches based on the "Ringed Circular Layout", from which the results are normally not in a plane [48]. The "FarFromParent" function will increase the distance of the children's crowns from the center point in considering the measure to calculate the distance, to retain the "distance metaphor". The size increase of each crown is noted. The "adjustCrownsSize" function adjusts the size of crowns (of the same level children's nodes) to keep the distances in different crowns comparable.

Algorithm 4 generateCrown

Input: Memory Island tree Node

Parameter: reservedWedge: the wedge for simulate the cartographic gap

Output: A Crown polygon of the n

Begin Algorithm

Assign n to the center position of the plane;

If n.type == root Then

```
    calculate each children's wedge (with the proportion Metaphor) with the whole
plane;
    Else
        calculate each children's wedge (with the proportion Metaphor) in the area (2 *
Pi - reservedWedge) ;
    End If-Then-Else
    child = n.leftMostChild;
    While child != null Do
        calculate its distance r to the center point by the metaphor distance;
        calculate its degree d with its assigned wedge (center of the wedge);
        child.crown = rotation(child.crown, d);
        child.crown = transition( child.crown, r, d );
        crownsList.add(child.crown);
    End While
    adjustCrownsSize(crownsList) ;
    While Test(crownsList) = not pass Do
        farFromCenter(crownsList);
    End While
    Return createCrown(crownsList);
End Algorithm
```

The function "createCrown" generates a crown polygon (island) for a node from its children nodes' crowns. We build a middle point between two children's crowns(between their wedges, for example, in Figure 5.3, points A, B, C) to emphasize the cartographic gap between the concepts, following the "Proportion metaphor".

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang- June 2015

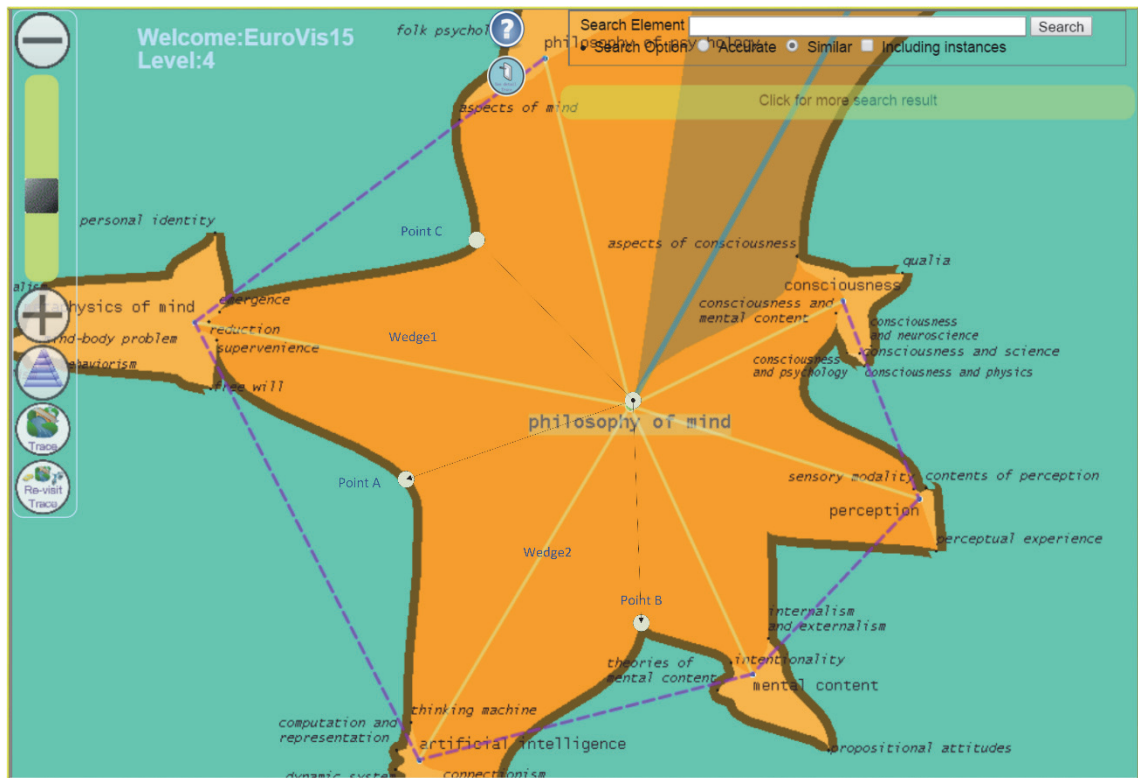


Figure 5.3 An illustration of the wedges of the concepts with the same parent.

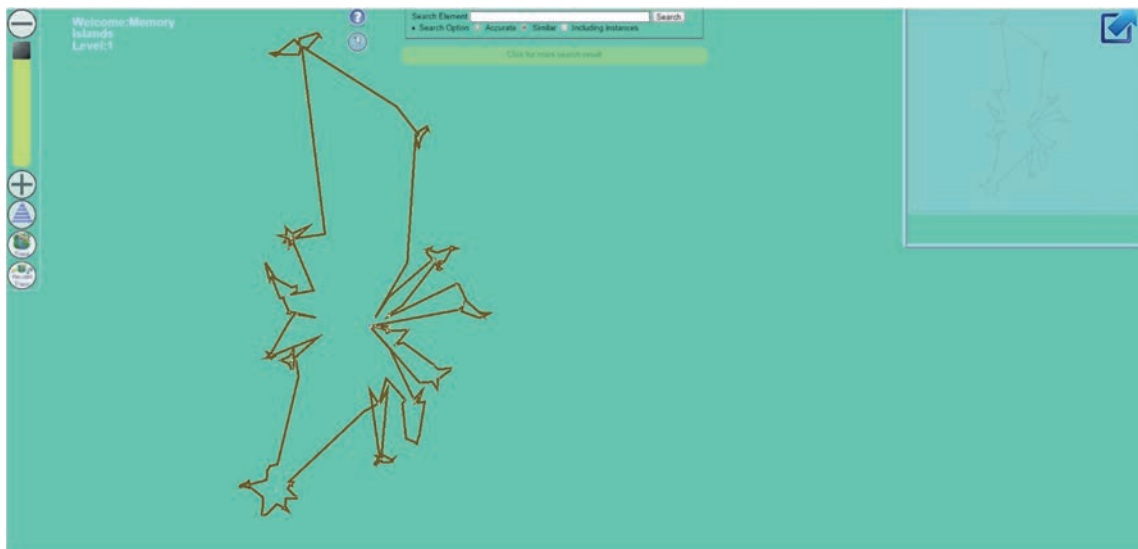


Figure 5.4 The island generated with Memory Island algorithm (without labeling and cartographic means).

5.1.3 Discussion on Memory Island Algorithm

To prove our island algorithm is an easier task. Similar to the layout idea of hyperbolic tree [94, 95], supposing we have a circle in a plane, when the size of each crown polygon is fixed, when we increase the distance of each child crown to the center point of the circle, the circumference and area of the circle grow exponentially, there will bring a lot of new room and

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang - June 2015

there will be a moment that all crowns are displayed in the plane without overlapping and satisfy the conditions in Test function. If we need to compel a node to place all its descendants' crowns in its own wedge, the algorithm will dump to a variants of On Balloon Drawings algorithms [96], this is the worst case for our algorithm to use the space. If we compare the island generated by this algorithm and the radial Memory Island (Figure 5.5) with same metaphors (if possible), we can find that our algorithm generates an island that can bring more sense than the radial one.

In Memory Island, the links between the nodes are less evident than the representation in the nodes graph (data lines), the question of Similarity Perception need to be considered. Recent research (from InfoVis 2014) on the influence of contour on Similarity Perception [97] states that for low number of dimension any glyph variation (data lines, contour and data lines + contour) can safely be used for data similarity judgment. Therefore, our Memory Island algorithm does not make the hieratical links less evident likes the tree mapping approaches.

Furthermore, with our algorithm, for each non-leaf node, the room of its crown can be used to create many interesting visualizations by simply applying an existing space-filling algorithm. All the spaces in the island can be well used as the map of earth. We will discuss it more detail in section 13.2.

When we have the spatialization of the knowledge, we need to consider how to improve this island representation, which can attract more end-users, and help them to improve their performance with Memory Island. Thus, we introduce a reshaping algorithm based on our resulting island polygon.

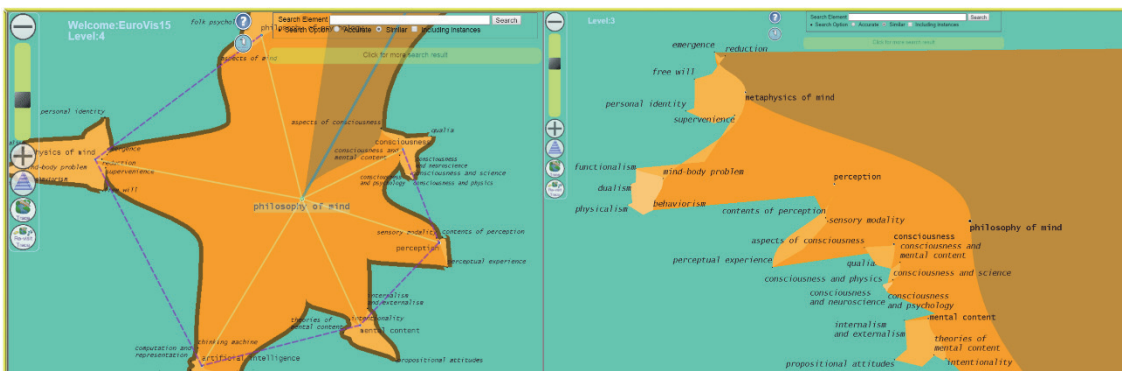


Figure 5.5 Left: Visualization SEP with InPhO Ontology with the metaphors we defined. Right: We generate a radial-layout Memory Island from the same dataset.

5.2 Reshaping the Resulting Island

The cartographic representations with more natural forms can bring advantages to users for knowledge navigation and memorization. According to feedbacks from our user studies, most of the users like the curved Memory Islands more than the list, node link diagram and the original Island. Recently scientific research supports our proposition: according to the exhibit "Beauty

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

and the Brain Revealed" at the 2014 AAAS Art Gallery, "humans have an affinity for curves and they claim that they have the scientific data[98] to prove it". Thus with memory Island technique, we introduce a proposition for re-shaping the island, using an algorithm based on Bezier curves. The principle of this method is (properly) replacing the straight lines (in Figure 5.6) by corresponding Bezier curves (e.g., it does not cause heavy overlapping inside of the island). The Bezier curves have already been well studied, the control point(s) used to generate the curves can easily be defined by a function of the two given points in the line. We discuss more details of this implementation in Chapter 7.

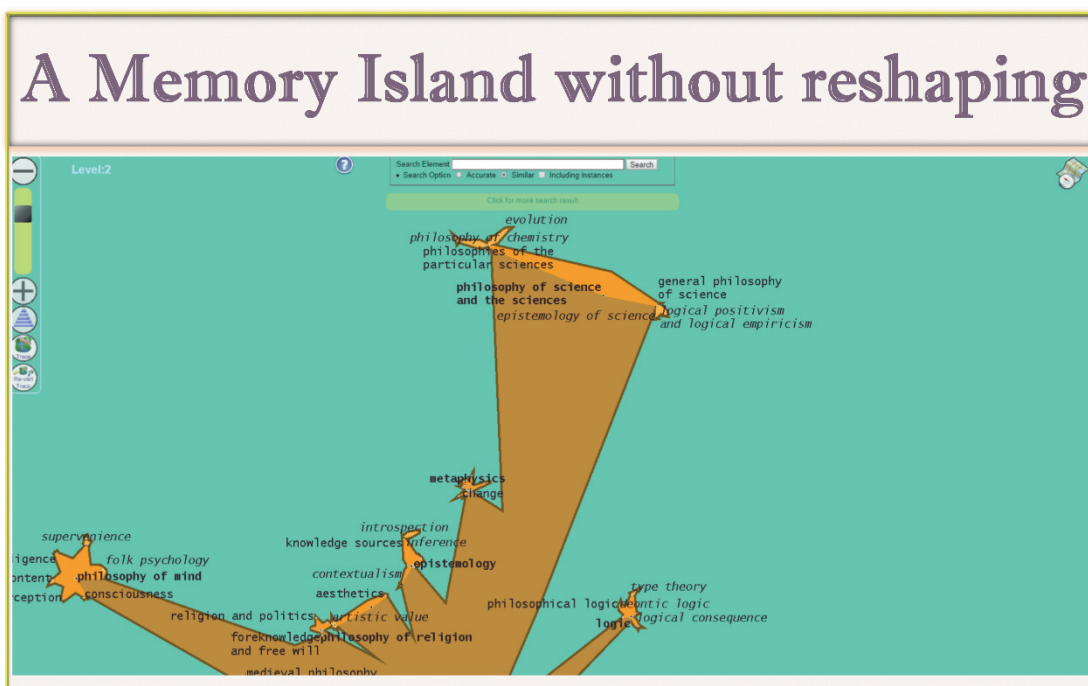


Figure 5.6 A Memory Island without reshaping process.

Chapter 6 LABELLING AND MAP GENERATION

CONTENTS

6.1 RELATED WORKS	73
6.1.1 PFLP Problem	74
6.1.2 PFLP Algorithms.....	75
6.1.2.1 Greedy algorithms.....	75
6.1.2.2 Based on the local search techniques	76
6.1.2.2.1 Discrete Gradient Descent.....	76
6.1.2.2.2 Tabu Search (TS) for labeling.....	76
6.1.2.3 Stochastic algorithms for labeling.....	77
6.1.2.3.1 Simulated Annealing (SA) for labeling	77
6.1.2.3.2 Genetic Algorithm (GA) for labeling.....	78
6.1.2.4 Dynamic Labeling for data visualization and labeling for dynamic maps.....	78
6.1.3 Google Map Mechanism	79
6.2 LABELING AND MAP GENERATION IN MEMORY ISLAND	80
6.3 DISCUSSION	82
6.3.1 Improvement of the Label placement algorithm.....	82
6.3.2 Apply Area-features label placement algorithm	83

Cartographic labeling is the most important concept in visualization for the use of humans. Despite labeling have been studied for long time, it is still one of the most challenging topic, both for the InfoVis researchers and cartographers. It is worth to note that naively placed labels caused excessive overlaps, which make the result map or graph difficult or even impossible to read, and cannot be used by the human end users. (e.g., the island showed in Figure 6.1, with this figure, the random labelling causes a lot of overlapping).

In this chapter, we start with the literature review on cartographic labeling problem, then we discuss the labeling with Memory Island and we propose our labeling and map generation mechanism for Memory Island.



Figure 6.1 A map with heavy label overlaps. In this example, we randomly placed the labels in this map. The human users cannot discovery through a map with heavy label overlaps.

6.1 Related Works

Cartographic label placement is refer to the label (text) insertion process in maps. Three independent researchers have shown that cartographic labeling is an N-P hard problem, and they have established the NP-completeness of the admissible-labeling problem [99-101].

Typically, the problem of label placement is usually divided into three different sub-types [102]: labeling of point features (such as cities, schools, hospital or mountain peaks) with horizontal labels, line features (such as rivers or roads) with tilted labels and area features (such as countries, seas or oceans) with curved labels. This basic logic for labeling was proposed by Yoeli in 1972.

There also exist a type of labels placement algorithms (e.g., the work of Wagner and Wolff [103] and Edmondson et al.[104]). It tries to solve the general map-labeling problem by considering all the features in the map as sites (the features), and then it consists in labeling these sites with a given set of candidates (e.g., rectangles, circles, ellipses, irregularly shaped labels).

As each concept is associated with a point feature in our resulting Memory Islands, therefore the labeling problem for the Memory Island technique is Point-Feature Label Placement (PFLP) problem. In the rest of this section, we give a brief resume on the PFLP with its existing algorithms. Some existing general map labeling algorithm was reviewed by Kern and Brewer in [105].

6.1.1 PFLP Problem

The PFLP is the problem of placing names (text labels) adjacent to point feature on a map or diagram (e.g., the node-link diagram in Figure 2.11). The labeling challenges are readable of the result map with all pertinent information displayed. In another word, the goal of PFLP is to choose positions for the labels that do not give rise to label overlaps and that minimize obscuration of features. Christensen et al. [106] stated that PFLP problem can be considered as a combinatorial optimization problem. Like the all-combinatorial optimization problems, it has two aspects: a **search space** and an **objective function**.

The **search space** for labeling is characterized by all the potential label positions. For example, a search space of a single label is its eight potential positions in Figure 6.2. The **objective function** can be decided with different labeling quality measures. In general, the **objective function** is to minimize the number of overlaps (the quality measure label visibility and readable of map). We can also consider the cartographic preferences conducted by Wu and Buttenfield [107] in calculating this objective function. According to the stat of art paper of Christensen et al. [106], most existing PFLP algorithms defined an **objective function** depends on the following factors:

- The amount of labels' overlapping appeared in the result map.
- A priori preferences among the set of the eight potential label positions. For example, the cartographic preferences show in Figure 6.2.
- The number of points without associated label.

Recently, more existing general label placement quality measures and existing labeling rules was reviewed by Van Dijk et al. [108, 109], in their work, the following 4 factors need to be considered: Aesthetics, Label visibility, Feature visibility and Association quality. Kern and Brewer used their proposal to evaluate the existing map labeling works in [105]. In the rest of this section, we review the ideas of the existing PFLP algorithms.

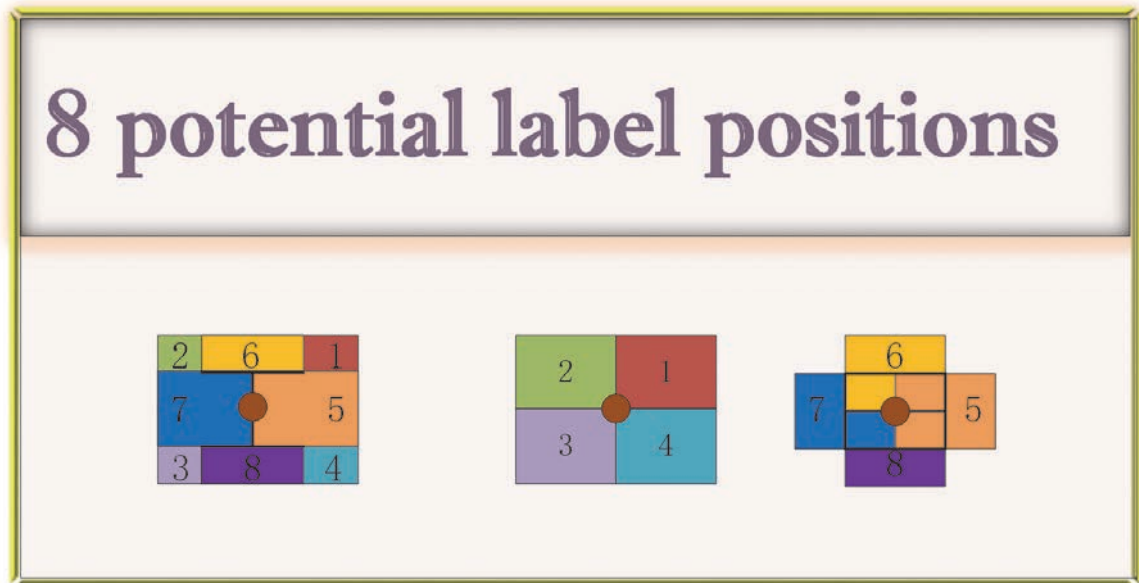


Figure 6.2 A set of potential label positions with their relative desirability. This set of potential label positions was proposed by Christensen et al. [110], based on the work of Yoeli [102]. The number of each position was used to describe the cartographic preferences for repositioning the label. (Label positions' preferences: 1>2>3>4>5>6>7>8) [107]

6.1.2 PFLP Algorithms

Obviously, we can place the labels by using the straw-man random-placement. It is the simplest way (if we do not consider the quality of result map) to solve the label placement problem with computer (Figure 6.1). Each label's position is random given from one of eight possible position shown in Figure 6.2. However, this algorithm serves as an effective lower bound, and often has a low value in the **objective function** according to some state of arts papers of label placements algorithms[105, 106]. Here we briefly review some important labelling algorithm for PFLP. The important researches in labeling published before 1995 have been compared by Christensen et al in their empirical study in [106].

6.1.2.1 Greedy algorithms

The greedy algorithms make the locally optimal choice at each stage as their name greedy. They want to find the global optimum, but for many problems, they do not have the ability to

jump out of the local-optima. The simplest greedy algorithm for PFLP places consecutive labels on the map with the positions that result in minimal extra overlap of labels. Its results are not satisfactory even for very simple map, but it is tremendously fast compared with stochastic approaches. Therefore, for the cartographic and graph drawing problems for the use of human users, when the speed of map generation is more important than the quality of the labeling, greedy algorithms are still a good choice.

For this reason, some researchers still work on the greedy algorithms for labeling, for example, in 2007, Cravo et al. [111] proposed a greedy randomized adaptive search procedure based on the conflict graph to solve the PFLP problem, they stated that their approach is generated better solutions than all other existing works (before 2007) in literature in reasonable times.

6.1.2.2 Based on the local search techniques

Local search is a metaheuristic method; the local search techniques have been well studied for many years with different optimization problems. In this section, we review two type of local search technique for labeling: Discrete Gradient Descent and Tabu Search.

6.1.2.2.1 *Discrete Gradient Descent*

The discrete gradient descent algorithm is a typical local search technique based on a discrete form of a gradient descent. An outline of this type of algorithm is given by Christensen et al. [106] as shown in the following:

1. For each feature, place its label randomly. (Random placement in the beginning)
2. When an improvement in the **objective function** is possible:
 - (a) For each feature, consider moving the label to each of the alternative positions.
 - (b) For each such potential move, calculate the change in the **objective function**.
 - (c) Implement the single label repositioning that result in the most improvement.

Although these algorithms often produce much better result than the greedy algorithms, they still do not have the ability to escape from local minima.

6.1.2.2.2 *Tabu Search (TS) for labeling*

In 1986, Glover proposed the Tabu (The word Tabu means forbidden) search method [112]. These Tabu search approaches improves the performance of local search, they accept the worse changes to who create a worse solution when there is no improving change. This move who violated a rule (e.g., worse change in the **objective function**) will be marked as “Tabu” (TS approaches usually provided a Tabu list with a given list-size).

In 2002, Yamamoto et al. [113] employed a Tabu Search (TS) algorithm to the PFLP problem. Their finding in their experiments indicated that their TS approach have a result even better than the Simulated Annealing (SA)'s and Genetic Algorithms (GA)'s. However, with this TS approach, the quality of the result depended on the given size of its Tabu list. Thus, even it have a better result, it is not suitable for solving the labeling with Memory Island, because we cannot predicate the best Tabu list size for each problem or ask our users have the ability to decide the size of Tabu list.

6.1.2.3 Stochastic algorithms for labeling

By incorporating a probabilistic or stochastic element into the search, stochastic methods can solve many difficult problems. Obviously, as their name, the stochastic approaches always have the ability to jump out of local minima, as they are stochastic methods. In this section, we review the two well-applied stochastic methods for labeling: the Simulated Annealing (SA) and Genetic Algorithms (GA).

6.1.2.3.1 Simulated Annealing (SA) for labeling

The simulated annealing was inspired by the annealing in metallurgy, the SA approaches for optimization were proposed independently by Kirkpatrick et al. [114] in 1983 and Cerny[115] in 1985. This stochastic gradient-descent method allows the worse movement in directions other than that of the gradient. Sometimes, SA method allowed getting worse rather than better, which make SA approaches have the ability to jump out of the local-minima.

The first algorithm based on simulated annealing for PFLP problem was introduced by Christensen et al. [110], the essential characteristics of these simulated-annealing algorithm for PFLP can be summarized in the following outline:

Given a threshold t , a temperature *temp* and an annealing schedule *AS*.

1. Randomly place all labels in the map. (Initial Configuration)
2. While the Temperature *temp* do not falls below the threshold t
 - (2.1) Decrease *temp* (according to schedule *AS*).
 - (2) Pick a label and move it to a new position (one of the 8 position in Figure 6.2).

Method 1: random choose a new position[110].

Method 2: the cartographic preferences [107](shown in Figure 6.2) was considered for re-positioning the label.

- (3) Compute ΔE , the change in the **objective function** caused by repositioning the label.

(4) If the new labeling is worse, undo the label repositioning with probability $P = 1.0 - e^{-\Delta E/T}$.

Then Edmondson et al extended this work [104], proposed an general SA labeling algorithm for general labeling problem (point, line and area), they introduce a scoring function to evaluate the quality of alternative individual label placements. Zoraster [116] applied this approach to the real work problem: the oil field based map. His work shown SA approaches have the abilities to solve the real-work problem, and can be wildly applied in many domains.

6.1.2.3.2 Genetic Algorithm (GA) for labeling

Genetic algorithm was firstly proposed and used by Holland [117] in 1975, nowadays, the genetic algorithms (GAs) is the most famous stochastic search methods in evolutionary computation (evolutionary algorithms, a family of algorithms inspired by the evaluation in Biology), and they applicable to a great variety of difficult problems, such as the Bayesian Network Structure Learning Problem. Many approaches based on the genetic algorithm was also proposed for the labeling, van Dijk has written a Ph.D. thesis[109] on this topic.

Like the GAs for other problems, each new solution generated by the application of selection, recombination (crossover), and mutation for generations must be evaluated (given a value according to the **objective function**). Unfortunately, an offspring solution of GAs usually contains much more than one changed element, therefore, when we want to compare each of its two solutions, a fast incremental calculation of the objective value (as in the SA approach, pre-calculate the score changes, detail can be found in [110]) is impossible.

6.1.2.4 Dynamic Labeling for data visualization and labeling for dynamic maps.

Many visualization tools need dynamic labeling algorithms for displaying labels. For example the Excentric labelling approach (Dynamic Neighborhood Labeling) proposed by Fekete and Plaisant [118, 119] deal with the display label overlapping, the labels of the objects located around the cursor, Recently with the increasing practical needs, Been et al. proposed an approach for generating the consistent map with dynamic labeling[120] have been proposed. In this model, the distracting actions (such as pop) are not allowed, then some researchers of computational geometry works on this topic to optimize its performances[121, 122].

However normally, for better performants, these approaches need the end-users have a powerful machine or have a well bandwidth for their Internet connections. We need to consider the effect/cost for applying a dynamic labeling algorithm.

6.1.3 Google Map Mechanism

Zoom and panning technique were widely used for cartography and cartographic/geographic visualization to display more detail information by increasing the displaying space. With the nowadays technique, this type of interactive function can help the map-like approaches to overcome its defects – display all the information contents in a limited space. Google Map¹¹ or other online map service, such as the Apple Map application for iOS and Baidu Map¹², have totally changed our life, and nearly everyone in the cities have been used one of them for at least one time in their daily life.

Google firstly introduced their Google Map mechanism to display all-important geographic information in their map service with zoom and pan function. It was supported by all-important browsers (Chrome, Mozilla, Opera etc.) and by nearly every smart phones and Tablets.

The basic idea behind the generation of image tiles in the Google Maps application is to divide the world into images of 256x256 resolutions. A small image can be easily loaded using even with a poor bandwidth connection. If the tiles image is very small and we have to load too many images at a time (therefore open too many TCP connections), the loading time would grow too much.

Therefore, for the first level of zoom, the map consists of a single image:



Figure 6.3 First zoom level Map in Google Map. Capture from the web service application Google Map.

For the second level of zoom, the map consists of four images:

¹¹ Google Map is, as its name suggests, an application which displays the map of the world proposed by Google. <https://maps.google.com/>

¹² An application of map of the world, proposed by Baidu. Widely used in China. map.baidu.com



Figure 6.4 The second zoom-level Map for the map of earth, Capture from Google Map application.

We obtain the following rules for each zoom level:

$$Number_{images} = 4^{zoomlevel},$$

$$Width_{map} = 256 * 2^{zoomlevel}$$

$$Height_{map} = 256 * 2^{zoomlevel}$$

6.2 Labeling and Map Generation in Memory Island

Labeling is the most important concept in visualization for the use of humans. A user cannot accomplish his task of discovering knowledge through the knowledge map without all its labels well displayed. If some labels were deleted from the map, the purpose of creating the knowledge map would be missed.

As the frequently questions in the domains of cartographic visualization, the labels have always has serious overlap which is even hard to deal with label placement. We need a method to see clearly all the labels that means a suitable size and configuration of the labels. Furthermore to avoid the situation to delete some labels (the case that there is no space for displaying all the labels), we have decide to make our visualization approach to be Zoom-able.

One simple way to display the label is focus+detail technique, when the users click a concept, another visualization about this point will be displayed for the users to see the labels. For example, the hyperbolic tree [94] simulates the distortion effect of fish-eye (Hollands et al

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

compared the fisheye and scrolling views in [123]) lens, and some variants of radial layout visualizations used a Focus + context technique to focus on a part of the contents [124]. However, it will make the users to change the content of their navigating through the map. Another method is classic zoom function for images and picture. With zoom-in, more space for displaying labels will, but it is hard to make sure there will be a solution with the dynamic labeling algorithm, and more difficult to make sure that the users can correctly use the zoom function to see all the information contents.

We would like propose a more natural method to show all the labels associated with each location (point), by simulating the mechanism of an interactive geographic map (e.g., the Google Map Mechanism).

We project the resulting island (polygon) to a map with size 512px*512px, then we randomly place the labels in one of its 8 possible position, then apply a point-feature label placement algorithms(PFLP)[106] based on Simulated Annealing approach, this algorithm will finish when we reach a threshold(temperature) or there is no overlap in the map. If there is at least one pair of labels overlapping, we display the labels well placed in the map, then we give priority to the high hierarchical level for the labels with overlapping. We increase the size of the map by four times, and appropriately increase labels' size, we re-apply the PFLP algorithm by using the last zoom-level configuration (position of labels) until, with a specific zoom-level, and there is no labels' overlapping. Using the last zoom-level configuration gives the already displayed labels more chances to maintain its position in next zoom-level. Then for each zoom-level, we color the island with the relationship between the concepts, and generate the image tiles of each-zoom level for the users' interface (Figure 6.5).

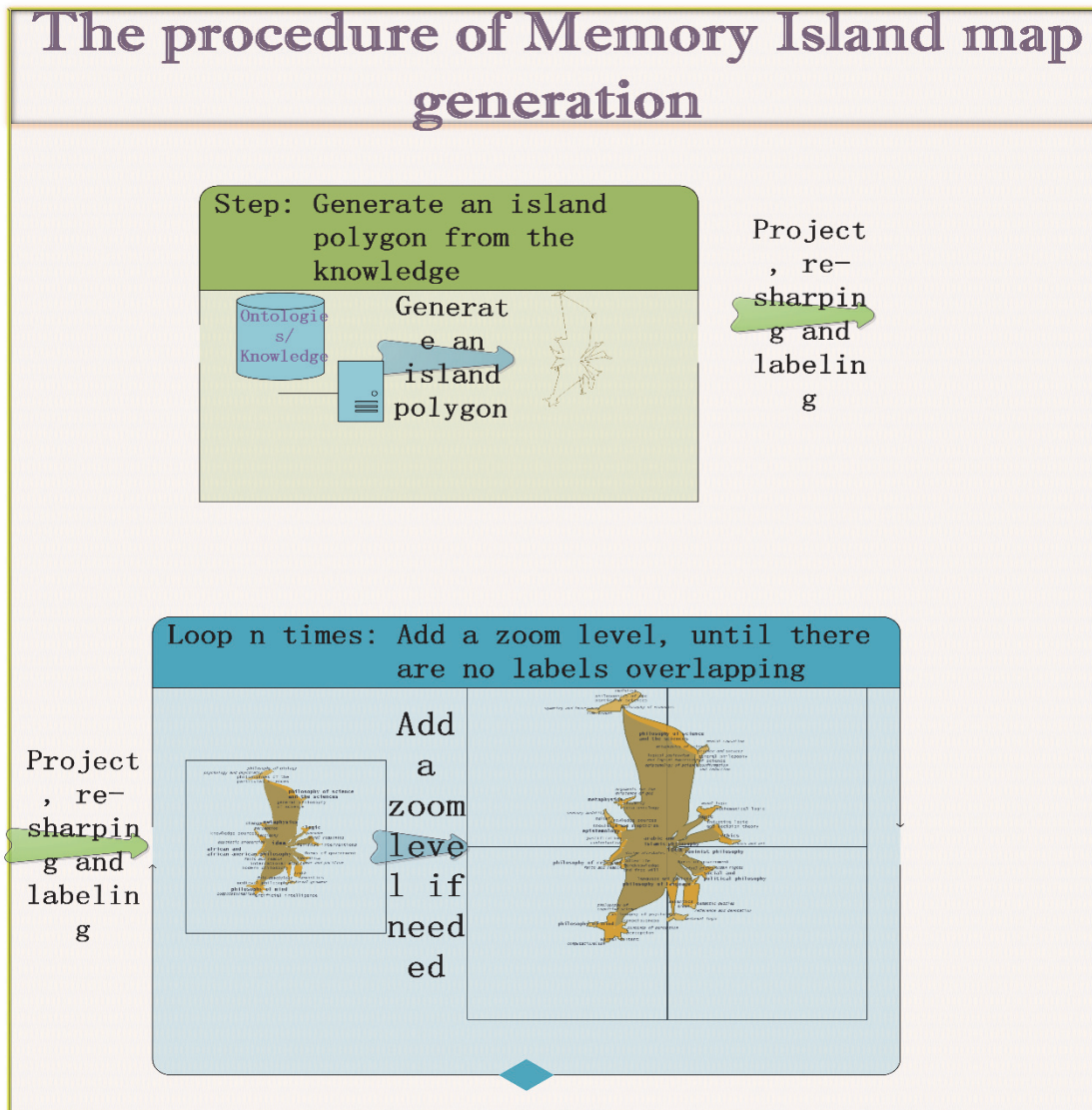


Figure 6.5 Memory Island's labeling and map generation mechanism.

6.3 Discussion

When we have done the evaluations, we asked the participants in these experiments for their advices to help us improve our Memory Island, some users mainly focus on the following points:

- Sometimes the user needs to zoom-in for many times.
- Few labels displayed without the label of its parent in some zoom level.

To end these, in this section, we propose two propositions for improving the Memory Island technique in the future.

6.3.1 Improvement of the Label placement algorithm

As we can find in our prototype algorithm (Algorithm 1), when we do not have enough space for displaying the labels, we need to add one zoom level. After some experiments on the

Memory Islands with different ontologies, if we compare the two labels' configurations in Figure 6.6 we may reduce some zoom levels by changing the orientation of the labels (texts), by extending the set of potential label positions (Figure 6.2), that is an operator mutation to change the orientation of a group of labels. Although this is abnormal in cartographic labeling in Computer Science, according to our discussion with Charles van den Heuvel (an expert on cartography), he believes we can add the orientation of labels to the PFLP problem. This will be a valuable research direction, to propose and apply a new PFLP labeling algorithm for Memory Island.

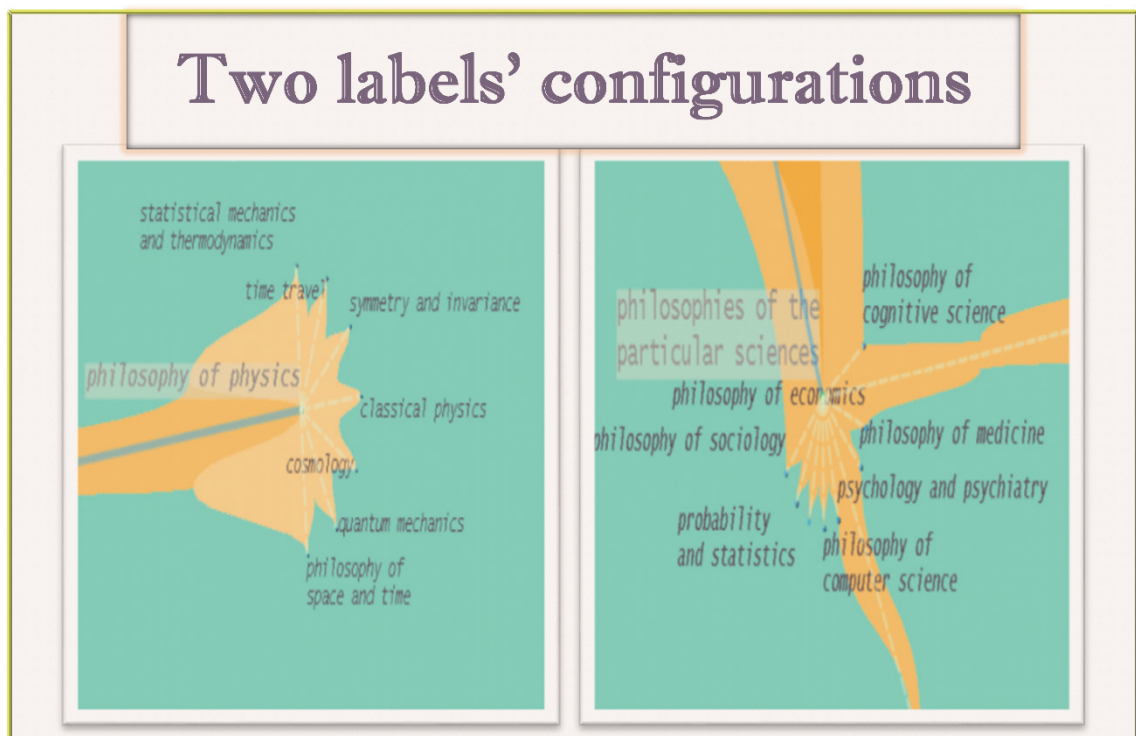


Figure 6.6 Two different labels' configurations found in Memory Island.

Meanwhile, Alvim and Taillard [125] proposed an heuristic algorithm for solving real work large labeling problem called POPMUSIC, they sated that the computation time of their algorithm increases almost linearly with the size of labels. By using their approach, we may reduce the time for labeling for the huge large problem. In the work of Been et al [121] [122] we mentioned, their proposed method for consistent dynamic map labeling, deals with problem of the labels display when the users zoom and pan. That also gives us a future direction of Memory Island technique: to apply an affordable dynamic labeling algorithm to reduce the time for the map-images generations.

6.3.2 Apply Area-features label placement algorithm

With Memory Island, each concept (non-leaf node) is not only associated to a point feature on the map, but it also has its own area in the map (except the leaf nodes). Although we believe

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

that these areas can be used with a space-filling algorithm to provide more information, we can also apply the area-feature label placement algorithm to generate the Memory Island. According to our discussion with Andre Skupin, these area-feature labeling algorithms can use the full space of the island, and reduce the max zoom-level. Therefore, less tiles images need to be generated and less zoom operators for the end-users during their information seeking processes for the same knowledge.

Chapter 7 MEMORY ISLAND INTERFACE

CONTENTS

7.1 OVERVIEW+DETAIL INTERACTIVE INTERFACE	87
7.1.1 <i>Map-like Focus + Context and Element highlighting technique:</i>	87
7.1.1.1 Pan, Zoom and Overview	87
7.1.1.2 Details on demand	88
7.1.1.3 Search.....	89
7.1.2 <i>Map interactive functions</i>	89
7.1.2.1 The visited trace function	89
7.1.2.2 Re-visit function.....	90
7.1.3 <i>Design of Memory Island Interface</i>	90
7.1.3.1 Interactive Knowledge Map.....	91
7.1.3.2 Overview panel	91
7.1.3.3 Search Area.....	91
7.1.3.4 Control Panel.....	92
7.1.3.5 Interactive Tree view.....	92
7.1.3.6 Other features	92
7.2 DISCUSSION	93

7.1 Overview+Detail interactive interface

To make interactive interface easily using is an important part of the visualization technique. With Memory Island, we would like to make our interactive functions more intuitive to the users (likes using a geographic map) and it brings some geographic metaphors (cartographic means) to end-users. A small database of each concepts' information found from the knowledge, will be automatically generated to support its interactive functions:

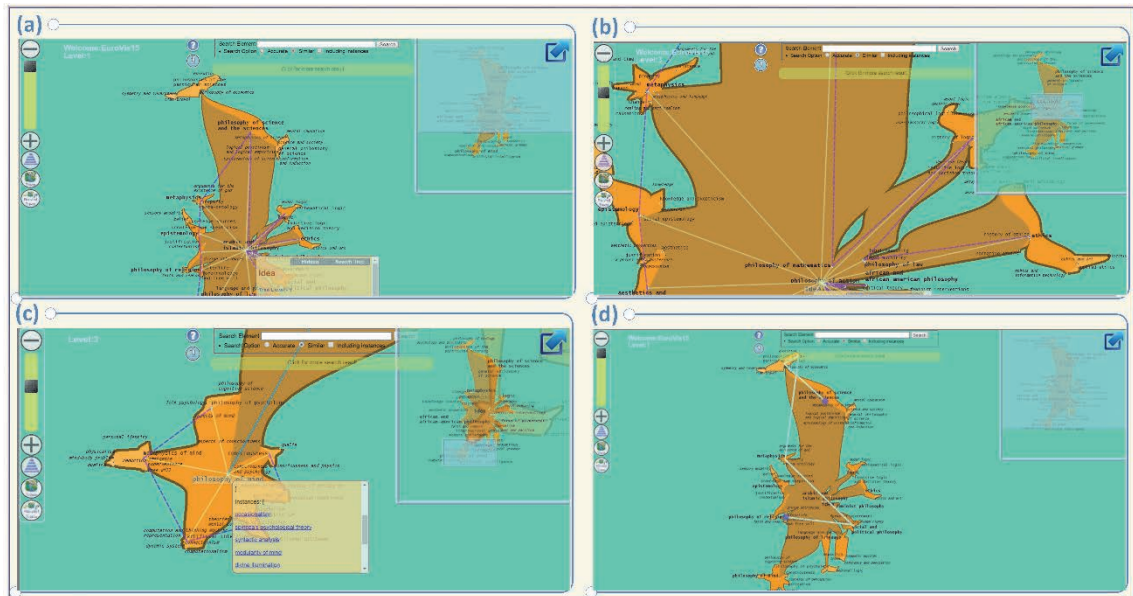


Figure 7.1 Navigating the SEP's Memory Island with interactive function: (a) (b): The cartographic gaps and violet paths are used to emphasize the order of concepts. For example, the philosophy of mathematics is close to the logic and philosophy of science and the science. (c): A Detail-on-demand window provides detailed information about the concept. Users can access to the SEP. (d): shows the visited trace in the map for a user named "EuroVis15".

7.1.1 Map-like Focus + Context and Element highlighting technique:

7.1.1.1 Pan, Zoom and Overview

- i. Pan and Overview: we provide an overview map to show the current position on the overview structure of the knowledge. With the pan function, users can easily navigate and explore the knowledge, like using a geographic map, such as the case show in Figure 7.2.
- ii. Zoom: it helps the users to focus on one part of information on the island without changing navigation context. For instance, in Figure 7.2, we zoom in to focus on the philosophy of mind, to see its full details. With the map-like zoom function, all the information contents can be displayed in the result map.

The overview panel of Memory Island Interface

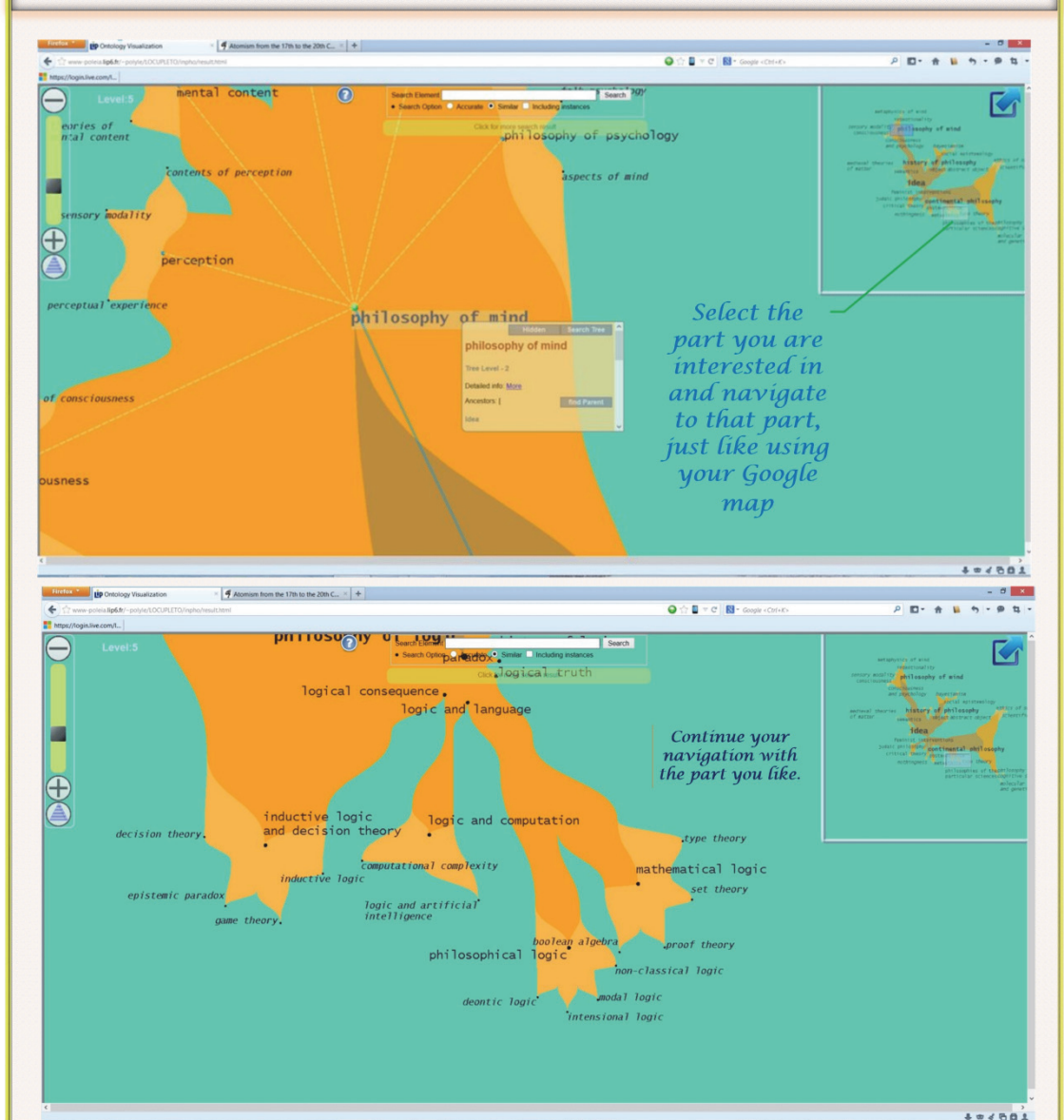


Figure 7.2 The overview panel shows the current position in the overview structure; the users can also use this panel to change the detail contexts to selected part of Memory Island. This function is friendly for most of users, because they are already familiar with the geographical information application, such as the Google Map.

7.1.1.2 Details on demand

By clicking on a label, a detail information window appears and user can access the source (e.g., encyclopedia) of the chosen concept. The instances associated with this concept will be shown in this window. The Data-lines can emphasize the connections between related points.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

We use the accessible paths between the concepts' points to emphasize the orders of concepts (knowledge order). (Figure 7.1 (c) (d)).

7.1.1.3 Search

Users can search from the concept and instances. For example, if a user gets a concept from a query, he may want to explore the area around that concept, he can quickly focus on that context and start his discovery by using the search function. The result will be highlighted with its data lines, and display in the center of the map (Figure 7.3).

The users can also use the interactive Tree-view panel provided in Memory Island, who the users click a concept in this concepts 'list, This concept result will be highlighted with its data lines, and display in the center of the map (with the zoom-level to display its label).

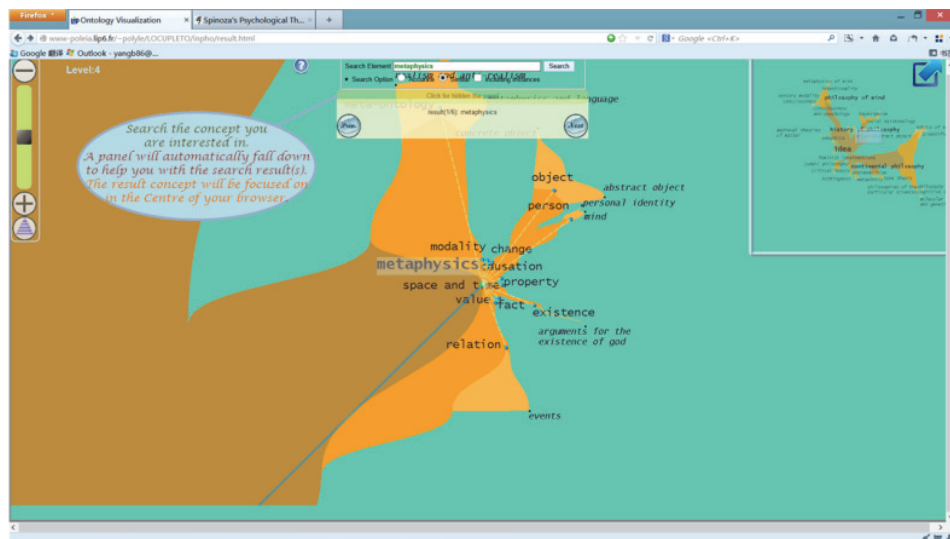


Figure 7.3 The search function in Memory Island Interface.

7.1.2 Map interactive functions

These functions allow users to use Memory Island as the geographic map; all the related researches associated with the knowledge, can be presented with Memory Island as geographic knowledge. Here we show two of them proposed within this thesis.

7.1.2.1 The visited trace function

Users can choose to save the trace of their visiting with their local browser, they can see their visited trace by click the Draw-trace button, and they can share their own experiences with Memory Island for collaboration. The visited trace draw on the map (Figure 7.1 (d) and Figure 7.4) give people (e.g., teachers or knowledge experts) a general overview and awareness of the navigation and information seeking process of a user.

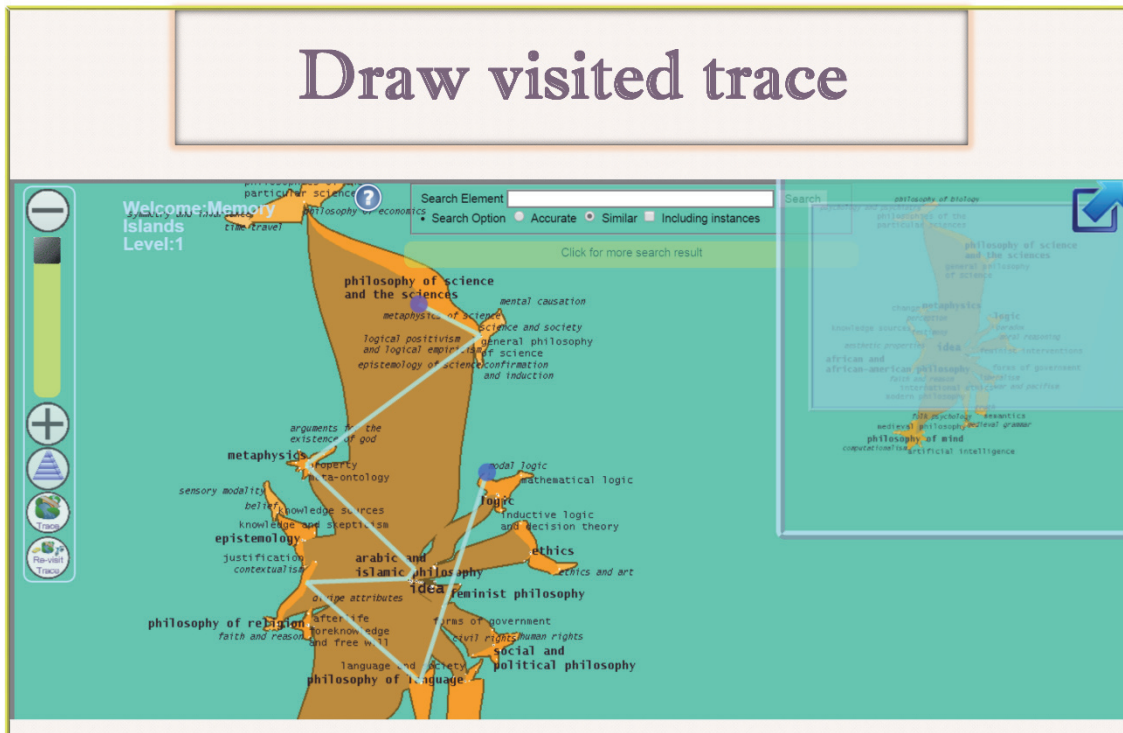


Figure 7.4 Draw the visited trace on the Memory Island.

7.1.2.2 Re-visit function

It allows users to revisit their visited concepts with the chronological order. The visited concepts will displayed in the center one by one in the chronological order (of visiting) by clicking the re-visiting button. With this function, the users can easily achieve the task of synergistic learning and interdisciplinary cooperation. For example, the teachers can share their students with their navigating experiences, and the students can use this function, to visit the concepts in the same order of their teachers. They can also share their experiences with other students.

7.1.3 Design of Memory Island Interface

Memory Islands are displayed in a HTML5 based web interface; it is able to use Memory Island not only on typical computer screen but also with tablets and smart phones. I design the interface of Memory Island as show in Figure 7.5. In this section, we discuss each parts in this interface in detail.

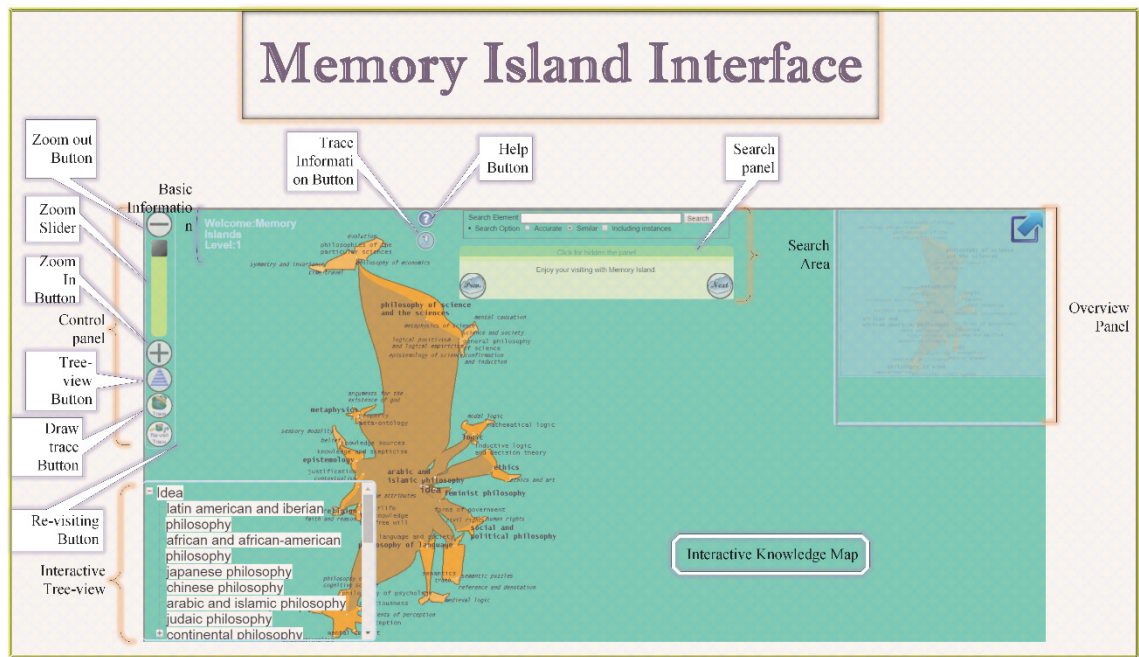


Figure 7.5 The Memory Island Interface with notations.

7.1.3.1 Interactive Knowledge Map

This part is the Memory Island itself with pan map function enabled user to move an island to any part of the page, and the zoom function can be applied by simply double clicking on the map, it shows/hides the detail information on the map. Another method to use zoom function is easily scroll mouse or uses your finger to spread and pinch on a touch screen devices.

7.1.3.2 Overview panel

Overview panel is placed on the top right of this interface, it shows an overview structure of the island. This function will become more useful when user have to explore an island in higher zoom level. They will not lose their positions on the island (context) and could be easily go back or forth from one position to another. This panel also help the user to pan the island by drag and drop a small square on it, it allow the users faster move over the island, like using a geographic map. If users think this panel is useless, they can hide it to gained more space for the interactive knowledge map (Figure 7.6).

7.1.3.3 Search Area

The search function helps the user to fast focus on some specific concepts or instances. This search function in Memory Island has two main choices for user to choose for either exact answer or just similar answer. In addition, optional choice is to include instances into result list or not. In case that there is more than one result, I give the priority to the more general concepts for displaying in the search panel. The result of search will be highlighting in two ways; first, we display its name in the search panel under the search form and second we display and highlight

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang - June 2015

(data-lines and paths) the result concept in the center of the island (Figure 7.6). The next and prev. buttons helps the users with the cases that more than one research results are available.

7.1.3.4 Control Panel

Control panel is design to support the interactive functions of Memory Island. The zoom-in, zoom-out buttons, and zoom slider similar to the online map services help the users with zoom function. Users can slide trough the slider to zoom in and out of the island with an indicator next to slider to say in which level is at that time. The tree-view button allow the users to display/hidden the interactive tree-view, and the Draw-trace button and re-visiting button allow the users to use our Map interactive functions by simple click on these buttons.

7.1.3.5 Interactive Tree view

Most of the users are familiar with a tree view (indented list). This represent come with an old style maximize and minimize of node enabled. Users can use this part to help improve their experiences when using Memory Island application somehow. When they click on a concept in this list, that concept will be displayed in the center of Map (with propos zoom-level).

7.1.3.6 Other features

The help button show an instruction panel to tell the users how to use Memory Island Interface. When users click on the trace information button, the detail information about their visited concepts (if they choose to save their visited traces) will be shown in a table. The Basic Information Area show the basic information of current navigation, including the user name and the current map zoom level.



Figure 7.6 The more general concepts will have the priority to display in the result panel, in this example, we hidden the overview map to have more displaying space for the knowledge map.

7.2 Discussion

Although our Memory Island interface has provided many interactive functions, some other advanced collaborative functions can be introduced in the future. For example, an interesting interactive function is to support some local changes for the users with the solid knowledge background (domain experts). Another interesting interactive function is allow the users to add the icons or landmark shapes (Figure 7.7) on the map; they can add their own understanding to the Memory Islands. Then we can re-build the knowledge from the different revised Memory Islands.

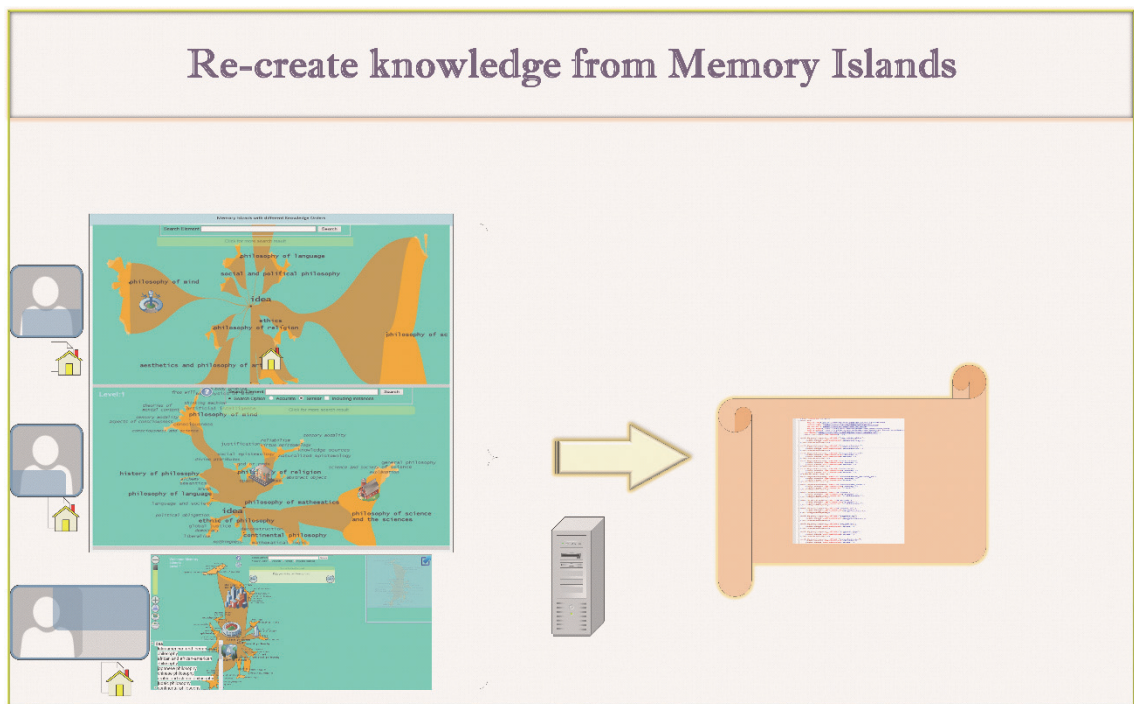


Figure 7.7 An example of the (local) modification function and the perspective on re-generate knowledge from different modified Memory Islands.

Part III.

Implementation and Applications

*“When we mean to build,
We first survey the plot, then draw the model;
And when we see the figure of the house,
Then we must rate the cost of the erection;
Which if we find outweighs ability,
What do we then but draw anew the model
In fewer offices, or at least desist
To build at all?”*

William Shakespeare [1598],

Henry iv, part 2

Chapter 8 MEMORY ISLAND IMPLEMENTATION

CONTENTS

8.1 MEMORY ISLAND APPLICATION'S SUB-SYSTEM COMPONENTS.....	99
8.1.1 Knowledge Extraction	100
8.1.2 Island Generation.....	102
8.1.3 Label Placement and Map Generation.....	103
8.1.4 Memory Island Interface Generation.....	104
8.2 THE RUN-TIME OF MEMORY ISLAND APPLICATION	104
8.3 DISCUSSION	105

In this chapter, we discuss the technique details for implementing our Memory Island technique. Based on our prototype algorithm as shown in Algorithm 1, in this chapter, we progress towards an architecture for generating cartographic representation from a given knowledge. To implement each step in our prototype algorithm, we design four subsystem components as shown in Figure 8.1: Knowledge Extraction, Island Generation, Labeling and Map Generation and Memory Island Interface Generation. The activity diagram of our Memory Island application is shown in Figure 8.2.

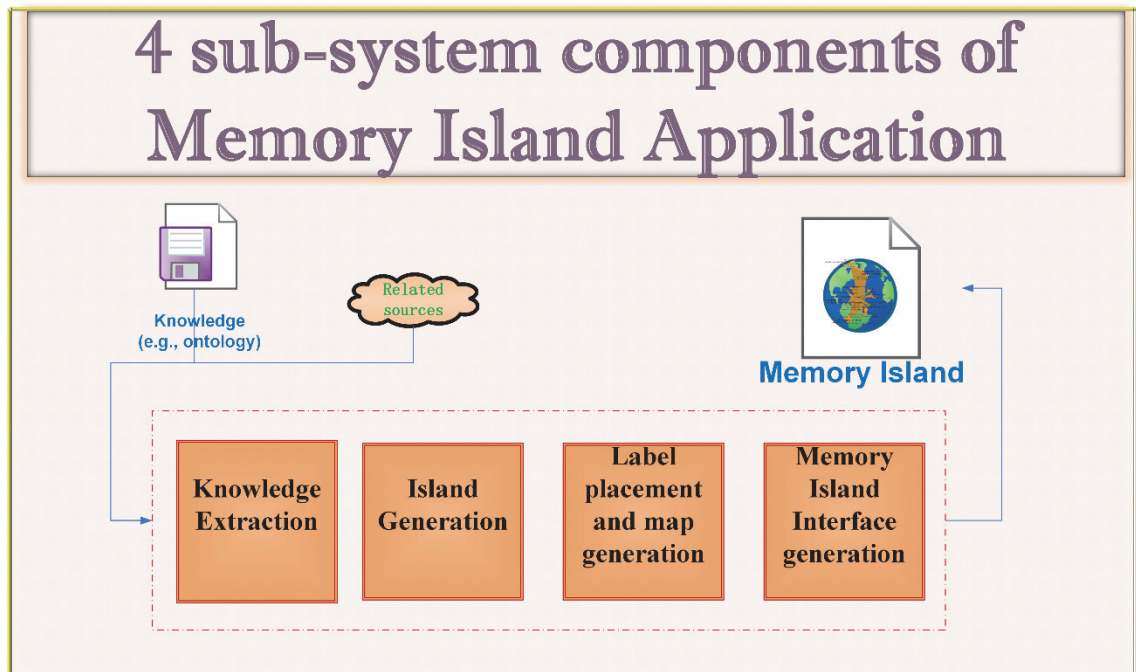


Figure 8.1 The 4 sub-system components of Memory Island Application

8.1 Memory Island Application's sub-system components

In this section, we discuss the technique details of each sub-system component in Figure 8.1.

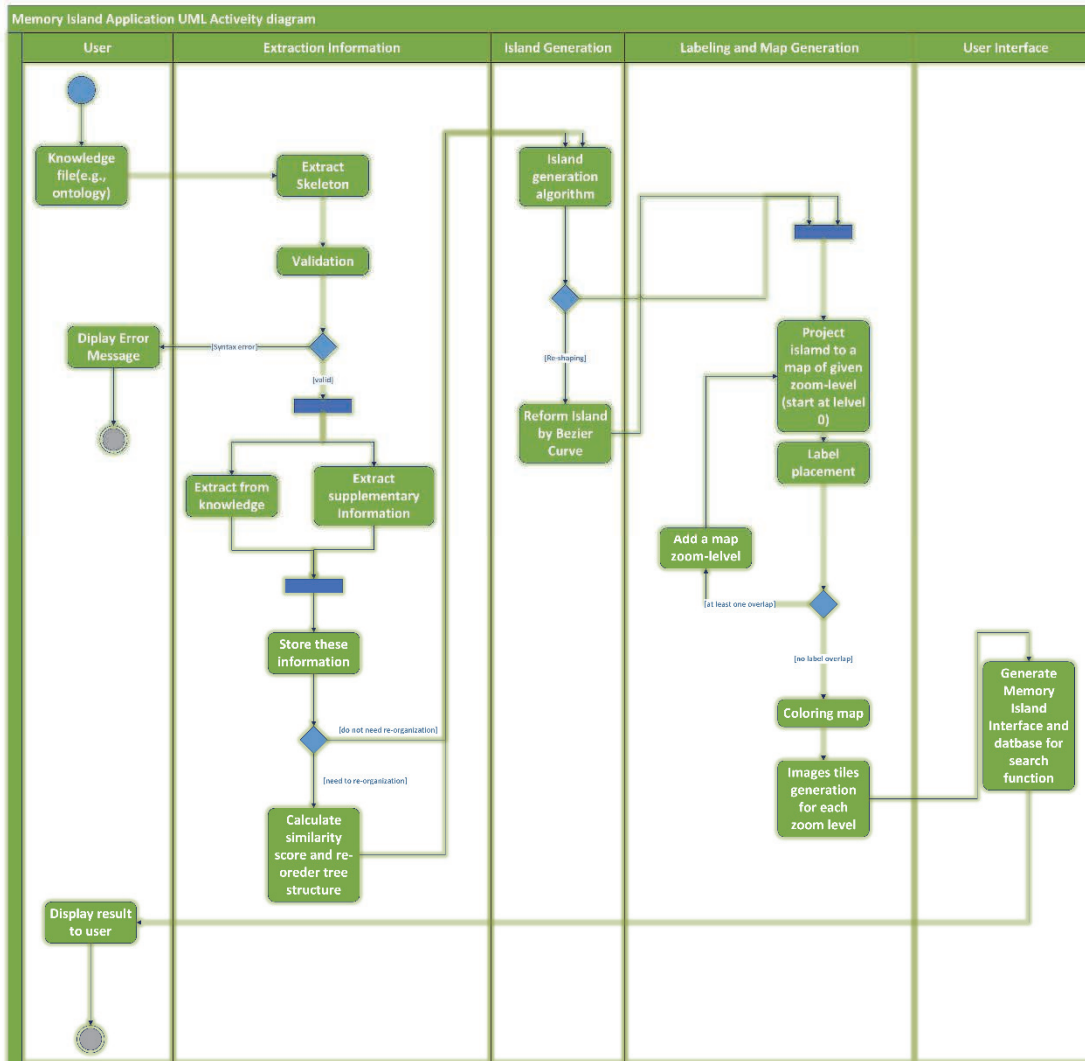


Figure 8.2 The activity diagram of our Memory Island Application

8.1.1 Knowledge Extraction

The sub-system component extracts information from the given knowledge dataset and the associated web-sources such as Wikipedia, dictionaries or encyclopedias. For example, with the InPhO Ontology, we used the InPhO API¹³ to get the supplementary information from the InPhO web site and the online Stanford Encyclopedia of Philosophy. We used a tree data structure to manager these information. Within this component, we also implemented the re-organization process for this tree structure method using the FaITH semantic similarity measure, to help us to present and express the knowledge order (of concepts).

¹³ InPhO API is available at <https://inpho.cogs.indiana.edu/docs/>.

As we mentioned in section 4.1.2, we use WordNet[84, 85] for computing the similarity score using the FaITH similarity measure[92]. We take into consideration the sets of nodes with the same parent node in the tree structure and try to re-order their appearance from leaves to root. In order to reorder the sub-concepts of one concept, we are solving this problem by reducing it to the traveling salesman problem [93]. Here I give an example.

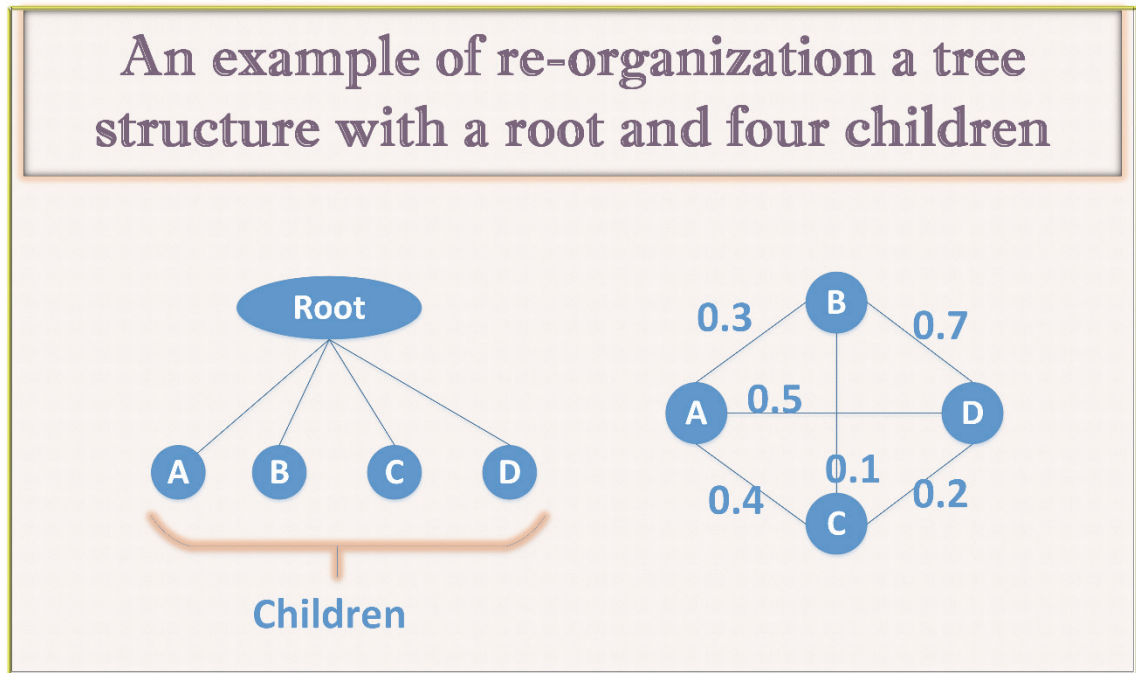


Figure 8.3 Left: an example of a tree dataset (with a root and four children). Right: The fully connect graph of children nodes based on their similarity score.

Figure 8.3 shows a simple tree structure with one root and four children, each child is associated with a specified label. As explained in section 4.1.2, Memory Island application starts by comparing each pair of terms from the nodes under the same parent and gives a score to the edge between nodes. We then built a fully connected graph (Figure 8.3, right) with similarity scores attached to it.

In order to reorder the sub-concepts of one concept, we solve this problem with heuristic traveling salesman algorithm[93] (e.g., the greedy algorithm). We then orderly sort the scores from maximum to minimum and then add the first pair to result list, then add the second highest score until we have all nodes in the result list. From the example in Figure 8.3, the result will be reordered from elements = [A; B; C; D] to elements = [B; D; A; C].

Beside of the reorganization method using Semantic Similarity, in the early time of our Memory Island technique, we also implemented a symmetric reorganizing for the elements in a tree structure with considering the volume of concept (weight). It tries to reorganize the tree structure to generate a more symmetric structure (Figure 8.4).

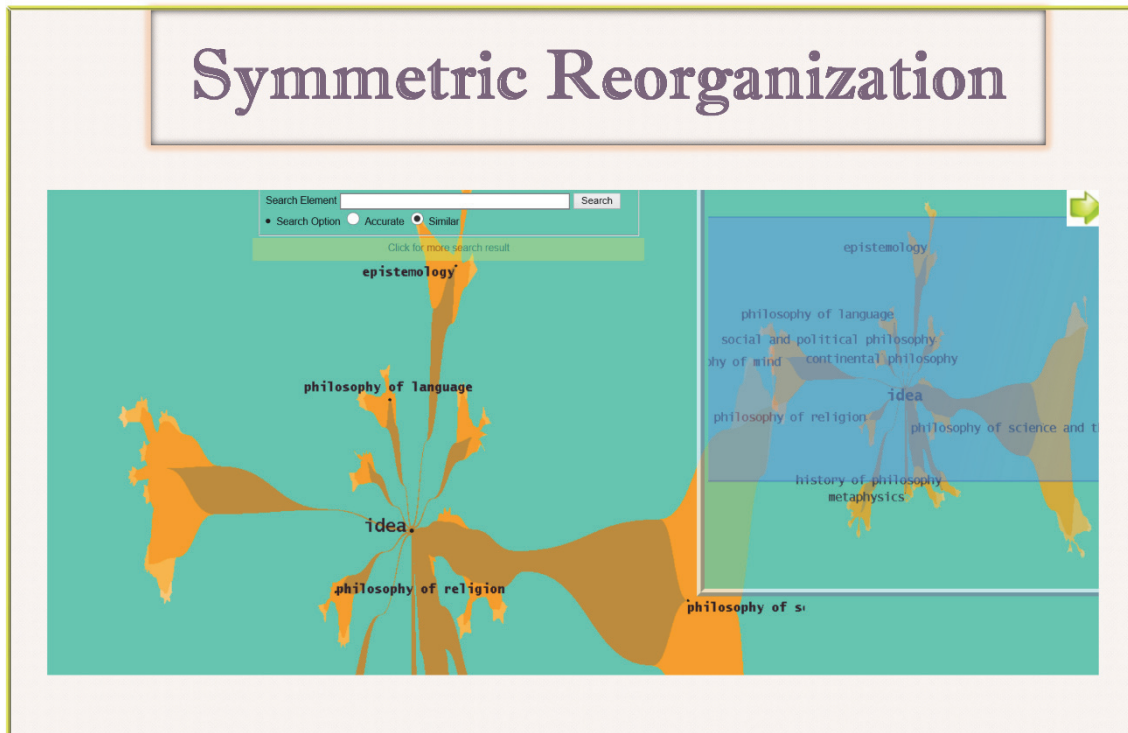


Figure 8.4 The symmetric reorganizing of Memory Island Application.

8.1.2 Island Generation

This part of the system deals with the generation of visual representation. It implemented the MI algorithm, and deal with the problem of island's re-shaping (e.g., using a curve to replace a line) and the mechanism for the coloration of the Memory Island. We already discussed our algorithms in full-detail in section 5.1. In this section, we give an example to express our island re-shaping method using Bezier Curve.

The idea of Bezier Curve was firstly proposed by Paul de Castel'jau in 1959. In 1962, in a widely known work of Pierre Bezier, he applied this method to design automobile bodies, which is the reason why this parametric curve was called Bezier Curve. A Bezier Curve can be created by given two auxiliary points P1, P2, shown in Figure 8.5.

In Memory Island application, we use the following formula to obtain the coordination of two auxiliary points P1 ($x_1; y_1$) and P2 ($x_2; y_2$). The start point of the curve is P0 ($x_0; y_0$) and it ends at P3 ($x_3; y_3$), these two points are in the island polygon created by the island generation algorithm. We then used Bezier curve to replaced straight line of the island polygon.

$$x_1 = x_2 = \frac{x_0 + x_3}{2};$$

$$y_1 = y_0;$$

$$y_2 = y_3;$$

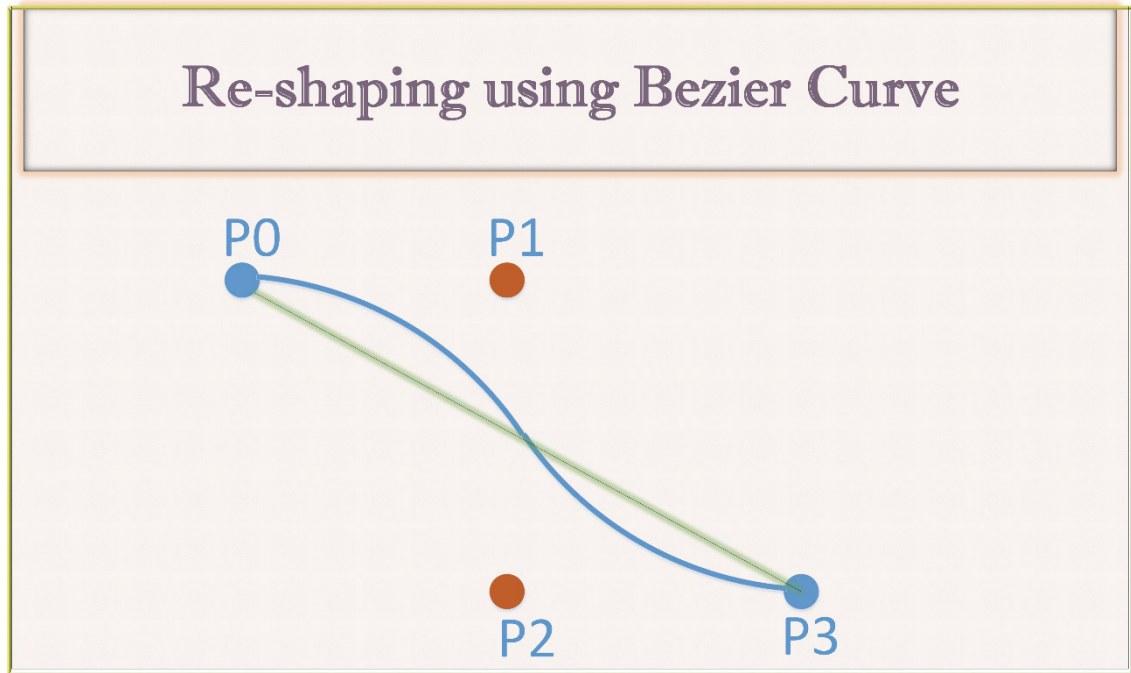


Figure 8.5 We Re-shape the polygon by using Bezier Curve; it replaces the straight line with the curve to make more a natural visual representation.

A Memory Island generation algorithm based on the radial layout (Figure 5.5, right) is also implemented in the Memory Island Application, because each layout will be particularly useful for some specific cases. The Memory Islands makers can also choose to use this radial Memory Island for visualizing the knowledge dataset.

8.1.3 Label Placement and Map Generation

This part of the system deals with how to place the labels and to generate the map for the end-users, it implements the idea (shown in Figure 6.5) we discussed in Chapter 6. In our implementation of Memory Island, we use the stochastic algorithms similar to Simulated Annealing (SA) [110] for labeling, as we discussed, using SA to resolve PFLP has the following advantages:

- Stochastic methods have the ability to jump out of local minima; PFLP using SA can have an equivalent, even better result to Tabu Search (TS) and Genetic Algorithms (GA) [106].
- No need to define the parameter(s) (e.g., the size of Tabu List in [113]) for each specific problem to obtain the best result.
- Normally it is faster than the classic GA algorithms[106].

In order to provide the users a better experience on navigation, we also consider the problem of labels' consistent. For each zoom level's map, we used the final configuration (label

placement) of the previous zoom-level as the initial state. This method provides a higher possibility for a label already displayed in the last zoom-level to hold its label position in the next level. As the SA approach accepts the worse re-positioning and therefore, it can jump out of the local optima. When we display the labels in the map of each zoom-level, we do not include the label of a concept that is obscured by a label of a more general concept, this mechanism give the general concepts more chances to be displayed in the map.

We presented our coloration mechanism in section 3.6.4, here we discuss the technique detail of implementation this mechanism in Memory Island. We define the coloring algorithm based on the idea of the BFS (Breadth-First Search); we color the root then its children then its grandsons, grandsons 'sons, etc. That ensures the correct color sequence: the whole island (crown of root node) in color for level 1, then the sub-parts of island (crowns of first tree level nodes) in color for level 2, then the sub-sub-islands (crowns of second tree level nodes) in color for level -3, etc. The detail colors we used for our application can be found in Appendix (Table 3).

8.1.4 Memory Island Interface Generation

For Memory Island application, we design a web-scale user interactive interface (in HTML 5) to display our result knowledge, we have described it in detail in Chapter 7. This web-scale interface allows the users to access Memory Island everywhere, from PCs or Smart Phones. It is also integrated with a small search engine, a small database about the information of concepts are generated with this interface.

8.2 The run-time of Memory Island Application

We also design an experiment to see how the Memory Island application evolves when a characteristic of knowledge dataset changes. For example, we increase the number of nodes in the ontology and fix other variables such as average number of children all over the tree, maximum number of children of nodes.

We set up an experiment to generate virtual ontologies in which we can control their characteristics. Then we measured the runtime of each module. We ran different ontologies generated with the same characteristics for at least three times and then computed the average value of the runtime. Figure 8.6 shows several pie charts of the time taken by the application when the number of concepts is 300, 500, and 700, respectively.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

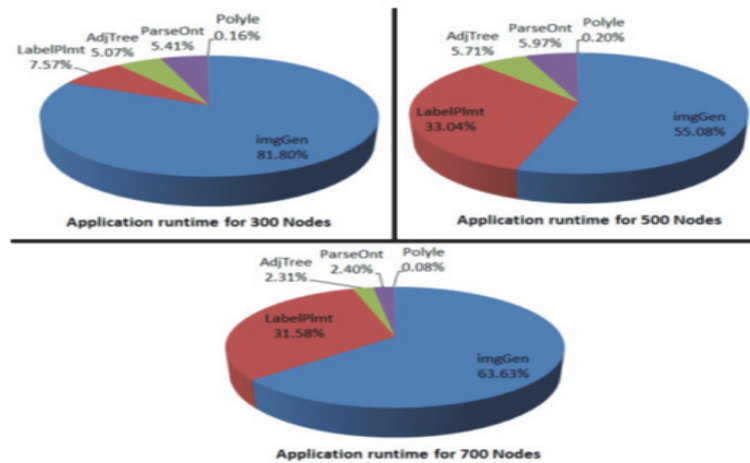


Figure 8.6 The runtime for our implementation of Memory Island Application¹⁴

8.3 Discussion

The results shown in Figure 8.6, indicate that when the number of nodes is low (e.g. at 300 nodes), the most usage time is spent on image tiles generation (map generation), while all other modules use the same proportion of the remaining time. Nevertheless, when the number of nodes was increased to 500 and 700, the label placement (PFLP is an NP-hard problem) module also increased significantly. The reason for this is the fact that we fixed the average number of children per parent, so that when the number of nodes increases, the probability of increasing the depth of skeleton of an ontology becomes higher. This could lead to a generation of an island in which we have less space to place labels in low zoom-levels, and the application relies on the deeper zoom levels for displaying all labels, which increase the time spending on images tiles generation. Furthermore, when we have a higher number of nodes, we have to place more labels.

Knowledge order (elements' order in knowledge or organization) is an important topic for creating a spatial representation. Some may like the symmetrical spatial representation. (Fig. 8.7 upper image), while others may like to organize the elements by their meanings – putting those that display related concepts close together and non-related concepts far apart(Fig. 8.7). This will be a valuable direction for future researches.

¹⁴ PolyLe is the name of Island generation Algorithm in our implementation. LabelPlmt: Label Placement; AdjTree: our method for re-ordering tree structure. imgGen: generation map and its image tiles. ParseOnt: Knowledge Parse.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang - June 2015

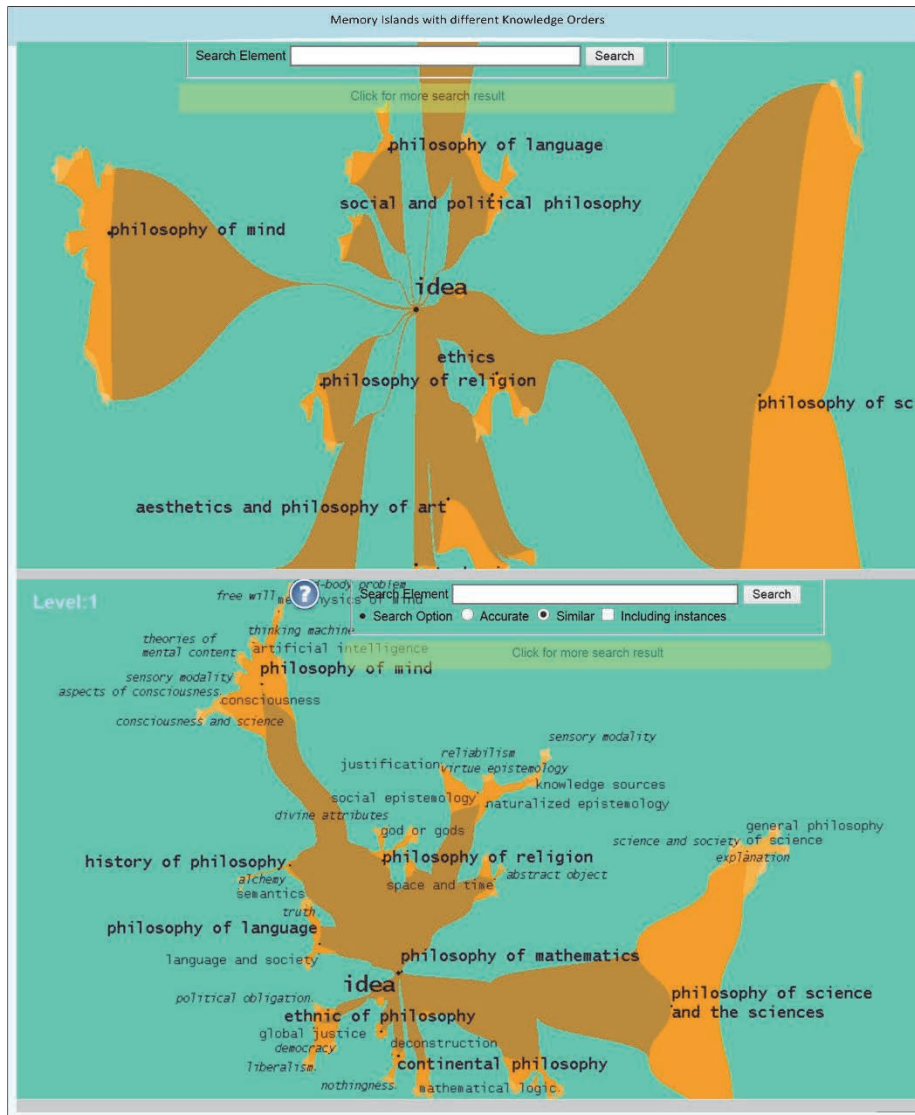


Fig. 8.7. Two maps generated for the InPhO Ontology with reorganization.

Additionally, we have introduced a small search engine in our visualization tool. Currently performing only simple searches over the taxonomies and the instances, we aim at developing it to take the advantages inside the ontology (e.g., it could give related results or close results for a query), in addition to the exact matches. We would like to discuss with the experts in IR (Information Retrieval) community about the search functions that we should embed in our tool. We think that feedbacks from the IR community will surely render our resulting knowledge maps more powerful and useful for both end-users, knowledge designers and engineers.

Chapter 9 CREATING KNOWLEDGE MAPS USING MEMORY ISLAND

CONTENTS

9.1.1 <i>Text and Document Data</i>	109
9.1.2 <i>Ontologies</i>	110
9.1.3 <i>Hierarchical Dataset</i>	112
9.1.3.1 <i>Table of contents of 20 books</i>	112
9.1.3.2 <i>Public debate Topic</i>	112
9.1.4 <i>Project OBVIL – Digital Humanities Dataset</i>	113
9.2 DISCUSSION	114

In this chapter, we discuss how to generate insightful knowledge maps using Memory Island technique. If we can find a tree-structure skeleton from an underlying knowledge, then this knowledge can be visualized by Memory Island technique. We give some classic cases studies of creating knowledge maps using Memory Island technique. Other examples can be found in our websites¹⁵.

9.1.1 Text and Document Data

With the development of Digital Humanities and Digital Libraries, there are needs of visualization tool for visual analysis the documents with the Text Mining techniques and other AI techniques. Therefore, visualization of text and document data becomes a popular research topic. For example, in 2008, Stasko et al. proposed a visualization tool Jigsaw [126] for visualizing text analysis. Recently, in end of 2014, Brehmer et al. [127] proposed a visual document mining tool for investigative journalists. They introduced the text mining techniques to their visualization tool. These text-mining techniques, such as Named Entities Recognition, can help its users to visual analysis text corpus.

In this thesis, we also give some preliminary examples we did with different Text corpus and Document dataset. For example, we use an original unsupervised approach for Named Entity Recognition and Disambiguation (UNERD) [128] with a French knowledge-base and a statistical contextual disambiguation technique that slightly outperformed Stanford's NER Classifier (when trained on a small portion of manually annotated data). It helps us visualize some children books for such entities as People, Locations, or Organizations. In project Locupleto, We try to build a knowledge structure based on the table of contents of book with the named-entities recognized in each chapter (Figure 1.2 & Figure 9.1). The zoom and detail-on-demand function provides more detailed information (e.g., access to the text in book); the paths are shown to improve navigation performance; the visited trace help them to share their reading and learning experience with their friends; search function allows the young users to search the concepts they think interested.

Currently we are experimenting our technique with the representation of extracurricular books and documents to help the students enhance their learning. The objective of our method is to visualize the contents to help the students improve their performance in learning and their memorability.

¹⁵ <http://www-poleia.lip6.fr/~polyle/>

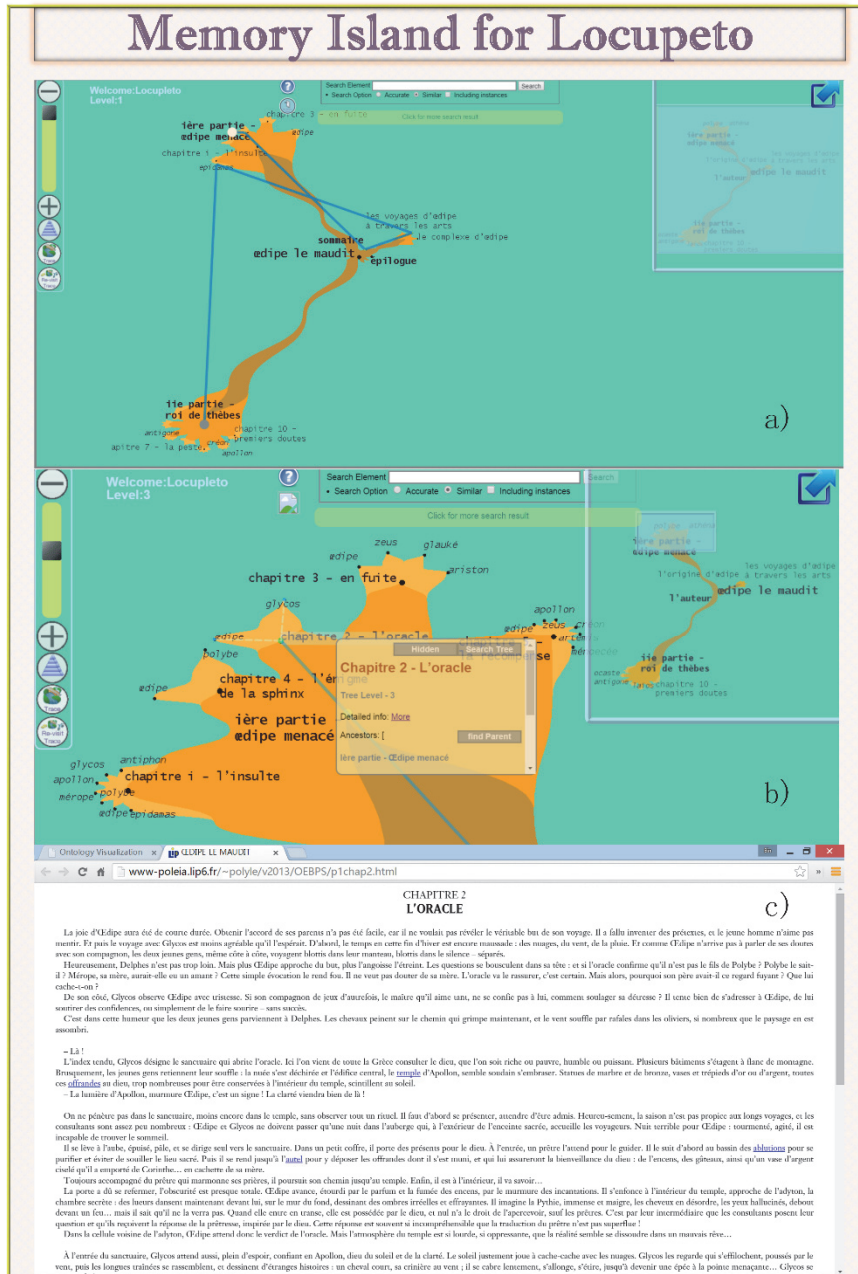


Figure 9.1 A children book “Oedipe le maudit” (copyright belongs to SEJER) visualized with the Memory Island technique. We identified People entities from the books, and displayed them with the corresponding chapter.

9.1.2 Ontologies

In this section, we show how to create insightful Memory Islands from ontologies. We illustrate these applications by illustrating one concrete case: Navigating through the encyclopedia of Philosophy with the InPhO. The Indiana Philosophy Ontology is a project on modeling the discipline of philosophy created by Indiana University of Bloomington. InPhO is dynamic ontology for the field of philosophy, and it is generated from the over 13 million word in the Stanford Encyclopedia of Philosophy (SEP) [129], and it is populated and extended by

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang- June 2015

combines statistical text processing, information extraction, human expert feedback, and logic programming. The SEP is an online and freely available encyclopedia of philosophy developed and managed by Stanford University since 1995. The InPhO ontology was integrated in the SEP (editorial workflow), it provided important metadata features and ontology driven (conceptual) navigation.

Although the InPhO is not large like the Gene ontology, but it is a real-world ontology with all-important features of an ontology: this ontology is not a simple tree structure (e.g., one concept in this ontology may have multiple parents). It has 265 ideas (277 nodes in its skeleton) and 276 instances associated with them. Both concepts and instances are associated with or related to some encyclopedic articles.

We used the InPhO API to help us for creating the cross-references to the SEP (e.g., Figure 9.2); benefited from the power of InPhO ontology, our Philosophy's Memory Island, realize the ontology driven conceptual navigation. As we discussed in Part II, we presented the phenomenon found from InPhO and the characteristics of SEP (from InPhO API) with the geographic metaphors: distance, proportion, colors, and another cartographic means. (See section 3.6 for more detail information). We have also applied Memory Island technique with many existing ontologies, such as the ontologies provided in Protégé or the ontologies mentioned in the user studies (in Part IV). Although each ontology has its special features, the basic principle to create and generate is the same as we did with InPhO. Many other ontologies' Memory Islands can be find in our website (<http://www-poleia.lip6.fr/~polyle/>).

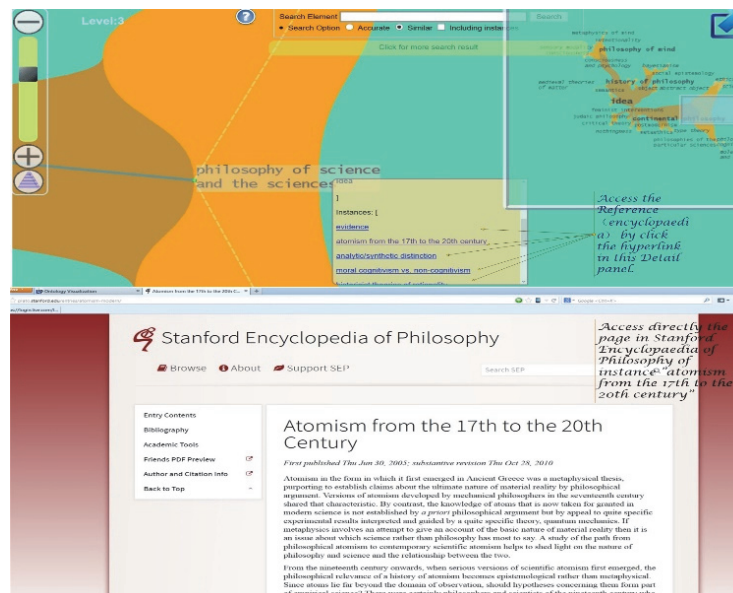


Figure 9.2 Users could navigate through the Stanford Encyclopedia of Philosophy with the help of Memory Island. In this example, we focus on the instance “atomism from the 17th to the 20th century” associated with concept “Philosophy of science and the sciences”

9.1.3 Hierarchical Dataset

In this section, we discuss how to apply Memory Island technique to the existing hierarchical datasets.

9.1.3.1 Table of contents of 20 books

This part of work was collaborated with Gilles Rouffineau of *École supérieure d'art et design Grenoble-Valence*. The dataset is a large table of contents of a collection of books (20 tables of contents). In these Memory Islands, the distances between chapters and its sup-chapter are measure by the number of subchapters of this chapter (the one used in Figure 9.3) or number of pages in this chapter. We can also generate the cross-references for accessing the contents of books. (It is not available in the public version of this Memory Island, due to the limitation of Copyrights). We are currently plan to integrated these Memory Islands with a visualization system of Digital Libraries (see Chapter 13).

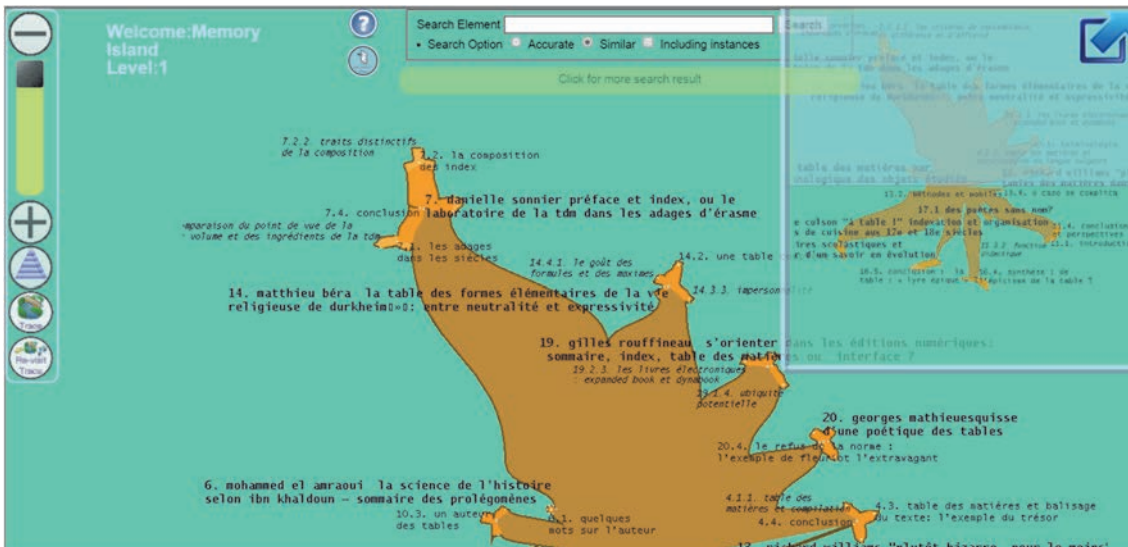


Figure 9.3 The Memory Island for a large Table of Contents created for *École supérieure d'art et design Grenoble-Valence*.

9.1.3.2 Public debate Topic

A second example is a representation of a public debate (Fig. 9.4) in the French Parliament - the National Assembly. Based on the documents of a public debate, we can link debate topics with different speakers as well as cluster sub-topics around the main topics. These Memory Islands are generated from the real-word topics in the public debates.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

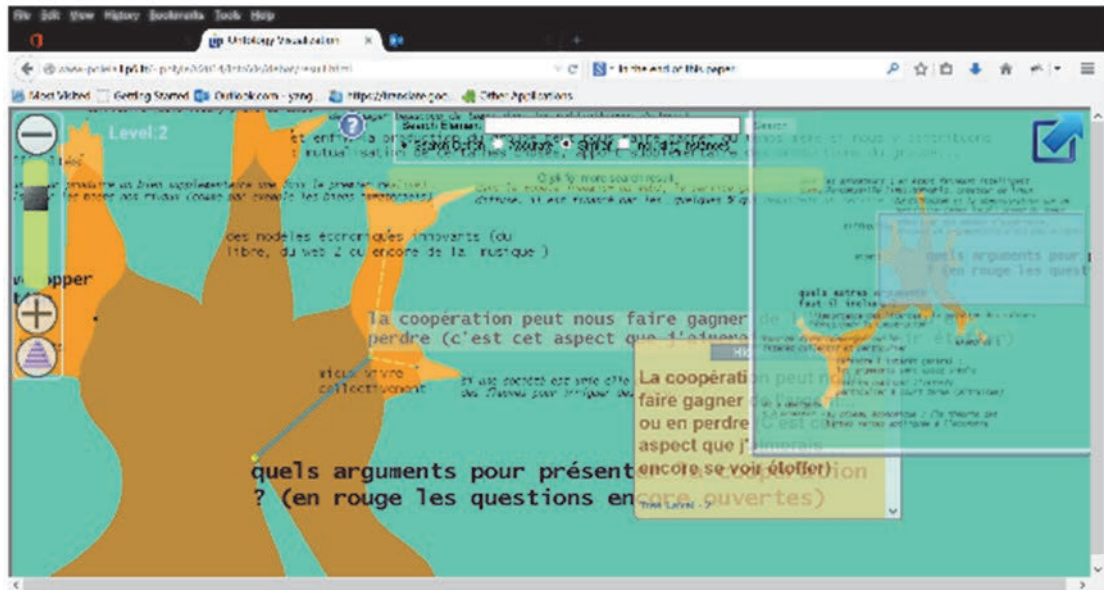


Fig. 9.4. A public debate (topics) in the French Parliament visualized with Memory Islands.

9.1.4 Project OBVIL – Digital Humanities Dataset

The project Labex OBVIL (l'Observatoire de la vie littéraire) intends to develop all the resources offered by computer applications to examine both the French literature of the past as more contemporary. It promotes scientific research in the field of Digital Humanities. Memory Island technique is applied in this project, to visualize the datasets of Digital Humanities.

For example, in Figure 9.5, we show some examples we did with the *Le Mercure Galant* journals, we build an ontology to represent and manage this corpus, and then we visualize this corpus using Memory Island technique. With these Memory Islands, the users can access the texts about the topic or the related journal issues.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

Bin Yang - June 2015

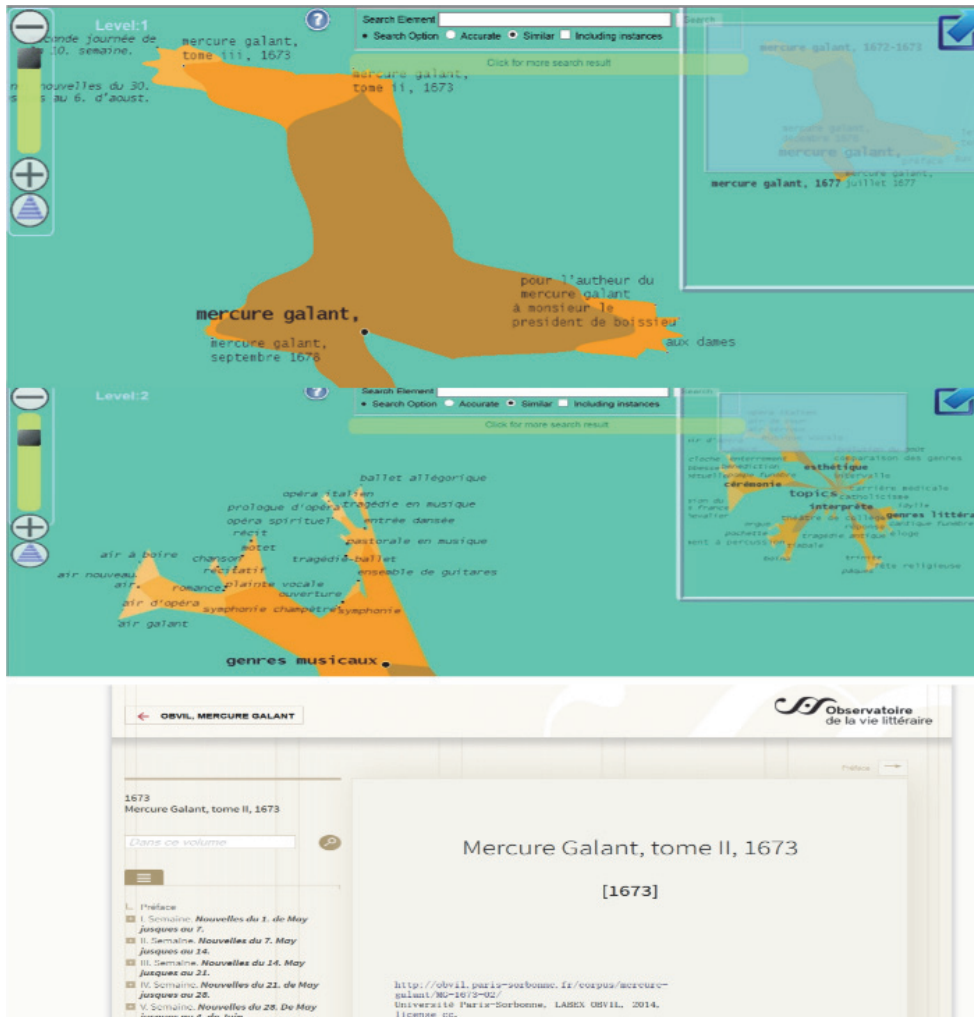


Figure 9.5 The Memory Islands for the *Le Mercure Galant* journals. The top image shows the islands for all the documents in the website of OBVIL and we can access the journal by this visualization. The second island is visualization of an ontology with the topics (nearly 1000) of *Le Mercure Galant* Journals.

9.2 Discussion

In this chapter, we show some of the knowledge maps that we created using Memory Island technique. Our technique takes advantage of the users' familiarity with a tool as common as a map to help them achieve complex visualization (information seeking) tasks. The metaphors we designed enable users to have in-depth insight into the given knowledge.

In the future, addition to test this technique with other real-world problems, it will be worth to discuss with information scientists about the quality of the mappings we could find in the knowledge, in order to improve this visualization technique.

Part IV.

Validation and Evaluation of Memory Island

“Firmness, usefulness, delight”

Marcus Vitruvius [22 BC]

From <De Architectura>

“People don’t want to buy a quarter-inch drill.

They want a quarter-inch hole.”

Theodore Levitt,

marketing professor, Harvard Business School

Chapter 10 VALIDATING MEMORY ISLAND

CONTENTS

10.1 VISUALIZATION MANTRAS	119
10.2 USER STUDY	119
10.2.1 <i>Methods and Visualization Set-up</i>	120
10.2.2 <i>Subjects</i>	120
10.2.3 <i>Procedure</i>	121
10.2.4 <i>Result</i>	122
10.3 DISCUSSION	123

In order to validate Memory Island technique, we organized a preliminary user study based on the InfoVis mantras, summarized as following:

10.1 Visualization mantras

The methodology we used to prepare tasks and questions are based on each point of the Ben Schneiderman's visualization mantras for visual data analysis ("overview first, zoom and filter, then details-on-demand" [130]). In the user study, we designed a number of tasks to complete on each given ontology representation.

- 1. Overview: users are asked to guess the domain of ontology or to guess which concept of the ontology (a node in the visualizations) contains the most number of sub-concepts;
- 2. Zoom: to check the zoom interactive function, we ask the users questions such as how many descendants of a given concept;
- 3. Filter: no questions ask for this type of task, as Memory Island do not provide a function of hiding parts of the island;
- 4. Details-on-demand: users have to search for a specific node and then ask the questions like what is its ancestor or its descendant node;
- 5. Relate: this task ask the participants to find the relationship among the items (concepts). Thus, we asked them to compare two nodes of the same ontology (easy questions, such as counting its children);
- 6. History: we ask the users what part of ontology they used for previous questions to check how well they can keep the information in their mind when exploring an ontology;
- 7. Extract: we do not check this mantra, because current Memory Island technique do not provide this function.

10.2 User Study

We designed a number of preliminary tasks in which participants respond to some basic questions on ontology visualizations. Sometimes, the Response Times (RT) was normalized using a log transformation before they were analyzed. However, with this experiment, we used the original RT to validate the Memory Island technique. We also ask the users for their preference for the use of knowledge navigation; the users need to choose a visualization tool for navigating through large information space (for example, navigating through InPhO ontology).

10.2.1 Methods and Visualization Set-up

The evaluation subjects are inspired by Shneiderman's InfoVis mantras [131]. We had in total 15 international participants with different levels of expertise for all the tasks (zero knowledge of ontology: four users; have background in the field of ontology: 11 users). Ages ranged from 22 through 42 years old, no experts in visualization took part in this experiment. Another two people who helped us to test the subjects only participated in S5 to give their preferences.

We used the classical node-link diagram, indented list, and Memory Island. We provide a basic version of Memory Island, which only allows the users to use the zoom and pan (with overview panel) and the detail on demand function to navigate through the map.

10.2.2 Subjects

Participants were asked questions with respect to four ontologies (InPhO [2] ontology, it has 265 classes, 277 nodes in its skeleton with 276 instances), Software Ontology¹⁶ (SWO), Material Ontology¹⁷ and ONTOderm¹⁸ ontology (size 50-200 nodes). We run these experiments with Memory Island, the classic node-link diagram and the indented list (see Figure 10.2). The following are the ontology retrieval tasks and subjective (preference) task used in this study. We summarized our full list of all designed subjects and detail questions in Appendix (Table 4):

- **S1. Overview(Guess domain and portion):** Participants are asked to guess the general domain of ontology or to guess by determining which portion of ontology contains the biggest number of concepts.
- **S2. Zoom(Count descendants):** This task aims to check the zoom function; we asked how many descendants of a given node can the users find. It is similar to the node based tasks in the existing works.
- **S3. Details-on-demand(Find ancestor):** Users have to search for a specific node and find its ancestor. This task can be called as the tree/network tasks in some studies.

¹⁶ Software Ontology (SWO) is a resource for describing software tools, it is available at: <http://theswo.sourceforge.net/>.

¹⁷ Material Ontology is an infrastructure for exchanging material information and knowledge. It is available at: <http://musigny.rds.toyo.ac.jp:8080/>.

¹⁸ ONTOderm is domain ontology for dermatology. It is available at: <http://www.ncbi.nlm.nih.gov/pubmed/18713597>.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

- **S4.Relate(Compare):** We asked users to compare two nodes with the same ontology and set some tasks such as counting its children in order to make comparisons.
- **S5.Preference:** In the end, we ask the users to choose the visualization they preferred for the ontology navigation task.

10.2.3 Procedure

This survey was set up via an online questionnaire, we measure time spent on each question and keep result into a small database. The participants start with some warm-up questions, and then answer the real ones (S1-S5). A DMKM¹⁹ student of Erasmus European Master Program help us to set-up this user study to make the result more just.

¹⁹ DMKM-Data Mining and Knowledge Management.

10.2.4 Result

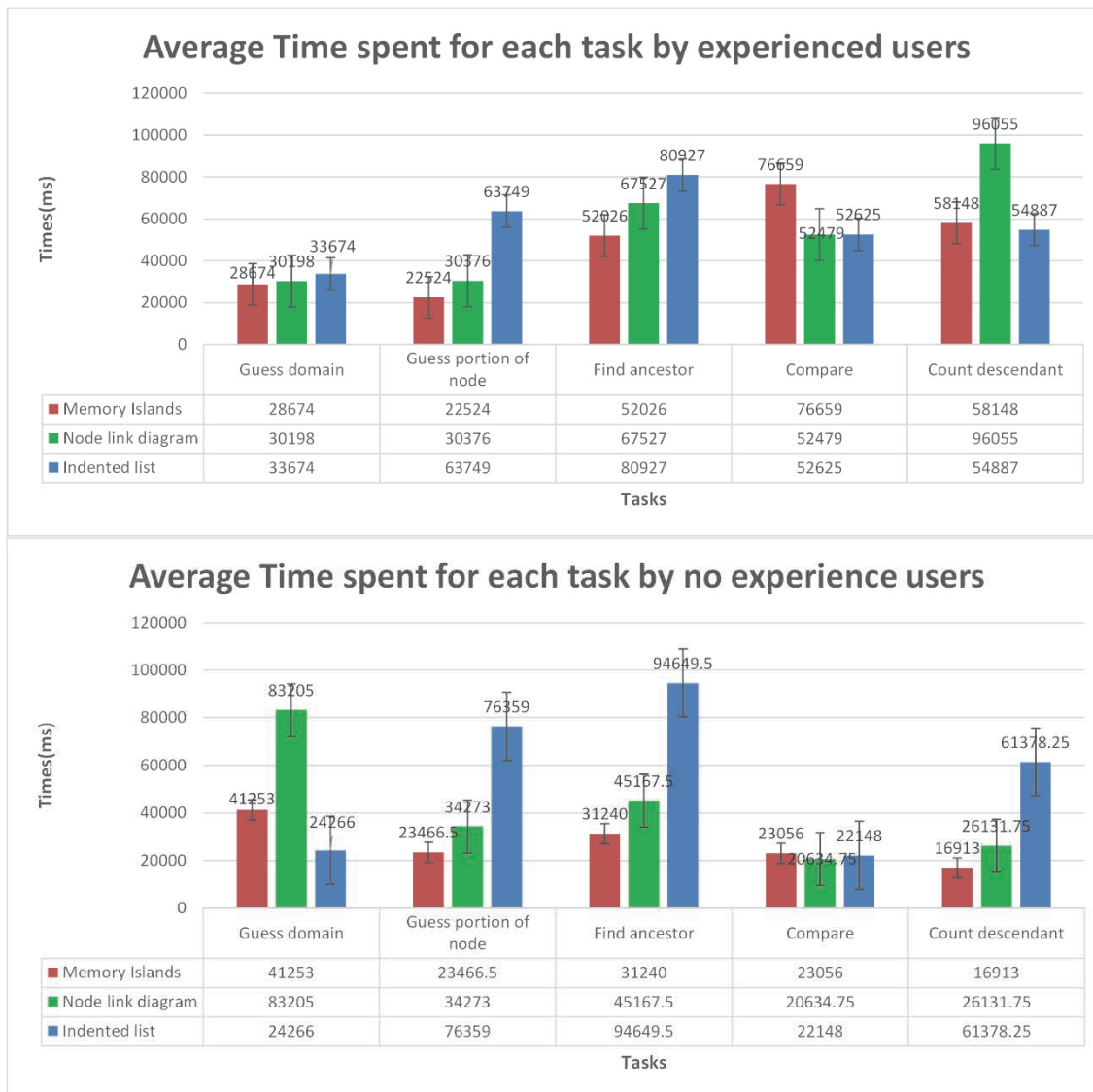


Figure 10.1 The mean response time of different groups of users, for different tasks.

We present the results of users' time spending in Figure 10.1 and their preferences are illustrated in Figure 10.2. With Figure 10.2, we find that most of users like to use Memory Island for their visual knowledge navigation. Figure 10.1 shows that the use of Memory Island provides advantages for non-experienced users tackling realistic browsing and visualization tasks. Meanwhile, the user with background knowledge in ontology or InfoVis can use Memory Island as effective as their most familiar tool, which is the indented list.

In this experiment, we are not interested in studying the accuracy, because the asked questions are basic and simple. For the future experiments with complex question, both task completion times) and accuracy need to be analyzed with statistical methods, such as the pairwise comparisons (e.g., Posthoc Tukey's HSD).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands

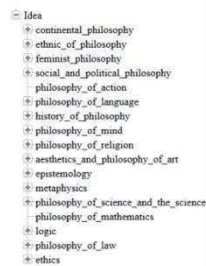
Bin Yang- June 2015

PART 3: Preference tasks

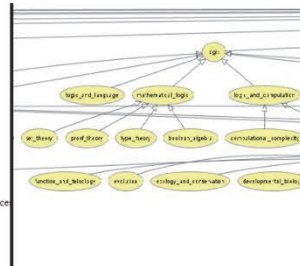
Question 26 - When you were asked to go through the ontology and find parent or child node, Which representation do you prefer?

Answer:

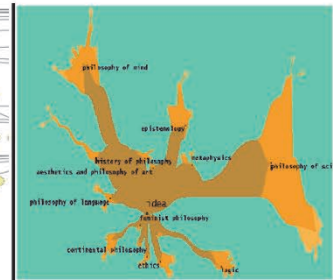
Why:



Class Browser



Node Link Diagram



Memory Island

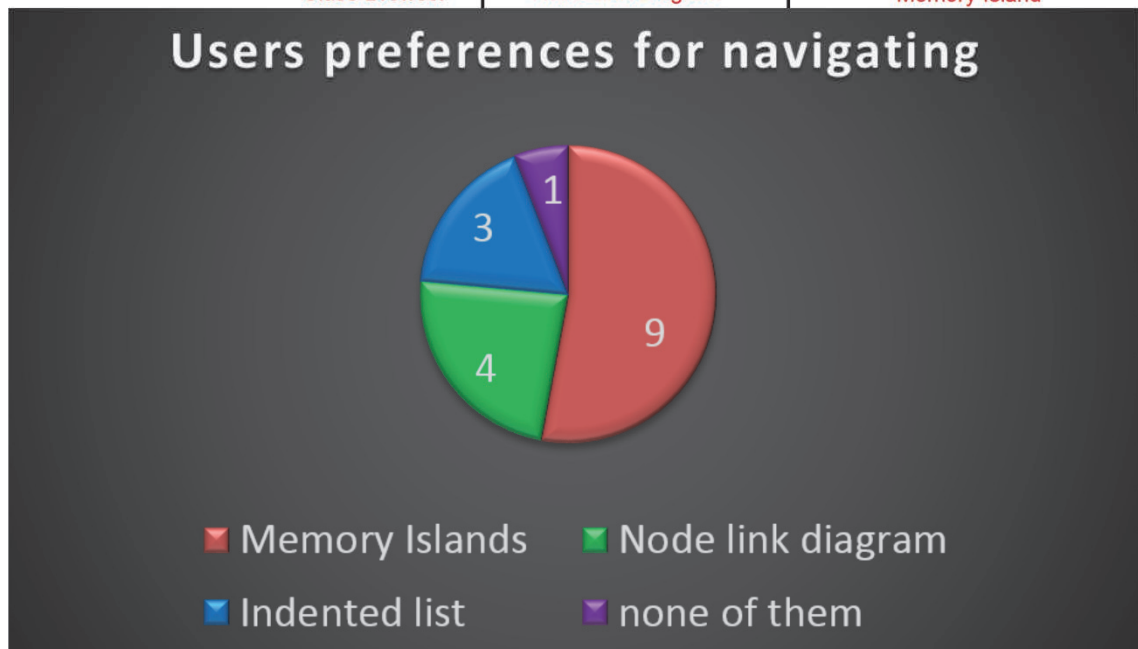


Figure 10.2 Users' preferred tool for navigating ontology.

10.3 Discussion

From the results shown in Figure 10.1, we can see that non-experienced users spent less time when using Memory Islands to complete the task and that the time required for completing the task increased when the Node link diagram and Indented list were used. The same amount of time was spent on the tasks, when users were asked to compare features of two nodes from the same ontology at the same time.

On other tasks, although users always spent less time when using Memory Islands as opposed to a Node Link diagram, the difference was not significant. When it comes to experienced users, the results show a similar distribution of time spent per task to those measured, but with a smaller difference between values recorded for each visualization tool. This may be explained by the fact that experienced users could easily understand the tree structure and the

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

ontology based on their previous knowledge, thus they could navigate through the ontology more efficiently.

For the task of the portion of a concept (the ‘overview’ task from visualization mantra), both groups needed less time to complete the task when using Memory Islands (Figure 10.1), this can be explained by the fact that users tend to be familiar with map representations. When groups are compared, the results show that Memory Island technique is more advantageous for the users without background knowledge. The experiments with the users without background knowledge will be interesting and useful, because their participants are like the real-world end-users.

Feedbacks received from experienced users indicated that even though they had worked on ontology before, they were not able to use the Memory Islands with its full advantages. For instance, few of them even did not realize that they could interact with an island. This is an interesting point we need to explore and to back up further with researches from the human computer interaction (HCI) field. The learning curve of users when they first interact with a new interface is a very important factor in determining the training process and for planning evaluation of visualization. Another limitation of this experiment is all participants respond to the same questions with identical order; it does not consider the learning effect of users, which needs to be, avoided in the future experiments.

Chapter 11 PSYCHOLOGICAL EVALUATION

CONTENTS

11.1 RELATED WORK	127
11.1.1 <i>Users' requirements analysis</i>	129
11.1.2 <i>Evaluations of Ontology Visualization tools</i>	129
11.1.3 <i>Evaluations of Tree Visualization tools</i>	131
11.1.4 <i>Evaluations of Map-like Visualization tools</i>	131
11.2 PSYCHOLOGICAL EXPERIMENTAL PROTOCOL	132
11.2.1 <i>Ontology Browsing Task</i>	132
11.2.2 <i>Ontology Understanding Task</i>	133
11.2.3 <i>Ontology Remembering Task</i>	133
11.2.4 <i>The Subjective Task</i>	134
11.2.5 <i>Suggested Evaluation Procedure</i>	135
11.3 A PRELIMINARY USERS STUDY.....	136
11.3.1 <i>Methods and Experiment Set-up</i>	136
11.3.2 <i>Subjects</i>	136
11.3.3 <i>Procedure</i>	136
11.3.4 <i>Result</i>	137
11.4 DISCUSSION WITH THE PAST EVALUATION EXPERIMENT	139

The objective of our evaluation experiment is to verify if Memory Island can help its users with their information tasks and knowledge memorization. In this chapter, we review the existing works for evaluating the visualizations of hierarchical knowledge: ontology or taxonomies (trees) visualizations, and the network visualization techniques. Based on these works, we introduce our evaluation protocol, and then we present a preliminary user experiment based on it.

11.1 Related work

As we described in Chapter 1, trees and ontologies have proven to be useful tools and were widely applied in several domains. There are growing needs for effective ontology and tree visualization. Various researchers have been trying to compare ontology and tree visualizations tools and evaluate their efficacy and performances.

In this section, we review previous related papers (see Table 1) on evaluation or comparisons of ontology and tree visualizations. Some Map-like visualization tools are generated based on Node-link diagram, and they have been already used to visualize the hierarchical data. For this reason, we also review the recent evaluation work with Map-like Visualizations.

Table 1 the existing works on evaluation and users' requirements analysis

Year	Online/Local or Controlled	Authors	Visualization tools (Purpose)	Main Tasks	Number of participants
2014	Online	Jianu et al. [79]	Map-like Visualization: Colored node-link diagram, LineSet[132], GMap[133] and Bubble-Sets[49].	10 types of tasks (network).	30-70
2014	Controlled	Saket et al. [134]	Map-like Visualization: Node, node-link, and Node-link-Group Diagrams	3 types of tasks.(node, network and group)	36
2010	Local	Song et al. [135]	Tree-visualizations: node-link diagram	Browsing, re-visit and	18

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

			visualization with two extensions of it	topology understanding	
2009	Online/Local	Kriglstein [136]	Users' requirements: for ontology visualizations	interviews	16
2007	Online	Flaconer et al. [137]	ontology mapping tools	Context + tool + process	28
2006	-	Flaconer et al. [138]	Users' requirements: for ontology (mapping) visualizations.	13 tasks must be supported (during the mapping process)	-
2006	Local	Katifori et al. [39]	Ontology visualizations: InfoVis plugins with Protégé.	7 ontology information retrieval tasks	13
2005	Local	Bosca et al. [139]	3D ontology visualizations: OntoSphere	Simple ontology browsing, "conceptual consistency" checking, and ontology development	8 users for ontology browsing
2005	-	Tu et al. [58]	Review a ontology visualization tool	3 tasks designed for review their user interface	-
2001	Local	Barlow and Neville [140]	Tree visualizations: Normal Tree, Tree ring, Icicle Plot and Treemap	Tasks about the topology of the tree and comparisons of node size	15
2000	Local	Risden et al. [141]	Tree visualizations: 3D hyperbolic interface with	2 types of tasks with different levels of complexity	16

			two 2D browsers		
--	--	--	--------------------	--	--

11.1.1 Users' requirements analysis

In 2006, Flaconer et al. [138] analyzed requirements and end-user tasks in ontology mapping tools for cognitive support. They provided a list of 13 tasks that must be supported during the mapping process (1.Navigation of ontologies. 2.Incremental navigation. 3.Identification of “candidate-heavy” ontology regions. 4.Browsable list of candidate mappings. 5.Information about the reasons a mapping was suggested. 6.Context for mapping terms. 7.Definitions for mapping terms. 8.Conflict resolution and inconsistency detection. 9.Ability to save the verification state; 10.Verification of mappings through execution. 11.Direct creation and manipulation of the mappings. 12. Navigation of verified and manually specified mappings. 13. Progress feedback.). They also proposed a plugin architecture named PROMPT for ontology management, PROMPT helps users to assemble a comprehensive ontology-mapping tool.

In 2009, Kriglstein [136] performed a user requirement study by analyzing the users' requirements on ontology visualization. He conducted both an online and face-to-face survey, to find out user expectations and attitudes about ontology visualizations.

He designed his interview contained open-ended, multiple choice and combined questions. He selected 16 participants, 12 of them were interviewed online and 4 of them were interviewed face to face. 8 of them were experts (solid knowledge about ontologies) while 8 were semi experts (who have basic knowledge about ontologies). Their findings indicate that providing overview-detailed views, browsing and updating, easy to learn and understand are main user expectations.

11.1.2 Evaluations of Ontology Visualization tools

In 2005, Bosca et al. [139] proposed the OntoSphere technique, who aim to visualize ontology on a 3D vie-port. They tested their visualization tool through a user study with three type of user tasks: simple ontology browsing, “conceptual consistency” checking (applied to real world cases), and ontology development. They have only 8 participants for the ontology-browsing test. The result of their experiment indicated that the OntoSphere is promising, and could be extended as a user-friendly ontology editor tool. Although the number of participants seems to be few (8 users), the three tasks they designed are interesting when we want to design a visualization technique.

In the work of Tu et al. [58] (2005), in their ontology visualization tool, they assessed the importance of each class and during visualization, labels only most important classes on the

screen. In addition, to show relationship between various classes, they used data-lines to connect them. In order to evaluate these designs, they reviewed their interface under different tasks, i.e. ontology navigation, ontology retrieval, and ontology instance analysis. Although their work does not perform a user study / experiment, it proposes a nice idea for evaluating ontology visualization tools.

In 2006, Katifori et al. [39] studied the InfoVis plug-ins with Protégé software: Protégé Class Browser, Jambalaya²⁰, TGVizTab²¹ [142], and OntoViz²² were the 4 studied visualization methods. They used their “University” ontology (contained 205 classes) in their study, this ontology described the University of Athens. They selected a group of 13 users (5 male 8 female) in their experiment, these users were the students from history-related departments (8 students) and researchers (5 computer experts²³).

They created a list of ontology information retrieval tasks (7 tasks) and asked participants to answer a set of questions. They measured the response time (RT) and correct answers percentage. Moreover, they had a questionnaire for collecting users’ opinion on various characteristics, the perceived ease of use and usefulness of each visualization methods. Their concluded that “Class Browser (in Protégé) is the promising tool in measured times (RT), correct answer percentage, and questionnaires” [39].

In 2007, Flaconer et al. [137] conducted a user survey on ontology mapping tools (Chimaera[143]²⁴, COMA++[144], FOAM, MoA Shell, OLA (OWL Lite Alignment), PROMPT[145] and QOM[146]). Their survey consisted of multiple choice and open-ended questions (totally 17 questions). They have 28 participants in this on-line survey. Their findings are “types of problems users are experiencing, features they’d like to see improved, some insight into their mapping and team process, and which tools are being used by the community” [137].

²⁰Available at <http://thechiselgroup.org/2004/07/06/jambalaya/>

²¹Available at <http://users.ecs.soton.ac.uk/ha/TGVizTab/>

²² Available at <http://protegewiki.stanford.edu/wiki/OntoViz>

²³The researchers in the Department of Informatics and Telecommunications of the University of Athens.

²⁴ Available at <http://www.cgl.ucsf.edu/chimera/>

11.1.3 Evaluations of Tree Visualization tools

In 2000, Ridsen et al. [141] compared a 3D hyperbolic interface with two conventional 2D browsers using the snap.com hierarchy contents. Participants in their study were 16 males; they are all web engineers (work for large web sites or portals). They have 2 types of tasks with different levels of complexity: 1) the verification of an concept is an existing category or new category, and 2) the task about the hierarchy, for example, if it's a new category, they asked the users find the best place in the hierarchy for it?

They stated “the study demonstrated the strengths and weaknesses of those three interfaces, and there were no significant differences across them in overall user satisfaction”.

In 2001, Barlow and Neville [140] compared four different hierarchical visualizations: Normal Tree (Organization Chart), Tree ring (radial space-filling), Icicle Plot(space-filling with empty space, similar to the concept of castles[147]) and Treemap. In their study, the visualizations are evaluated in the context of decision tree analyses prevalent in data mining applications. They selected fifteen coworkers for their experiment (7 male, 8 female). They found that Treemap-style was the slowest for most tasks. In addition, they suggested “either the tree ring or icicle plot is equivalent to the organization chart”.

In 2010, Song et al. [135] performed a comparative study on three visualization methods for hierarchical structures. They compared the conventional node-link diagram visualization with two extensions of it: a list view with a scrollbar and a multicolumn interface. Their user study included three important tasks: browsing, revisit, and topology understanding. They selected 18 participants (9 males and 9 females) in their experiment. They collected the responds time (total task time), correctness of the answer, and the total length of panning and performed statistical analysis on the results to compare three visualizations.

Based on this experiment, their conclusion indicates that, users are able to browse and understand the tree structure faster with the multi-column interface than the other two interfaces. In addition, they showed that users liked the multi-column more than the two other tree visualizations. Beside this conclusion, the strength of this work is the procedure of their experiments.

11.1.4 Evaluations of Map-like Visualization tools

In 2014, Jianu et al. [79], compared the variance Node-Link Diagrams used to display group information, they test 4 different tools: Colored node-link diagram(e.g., Figure 2.8), LineSet[132], GMap (e.g., the base map of Figure 2.15) and Bubble-Sets(with colored contiguous contours around nodes)[49]. They performed an online comparative with about 800 subjects of 10 types of tasks: Task 1. Intuitive group membership; Task 2. Deliberative group

membership; Task 3. Number of sets; Task 4. Relative group size; Task 5. Node degree estimation; Task 6. Path tracing; Task 7. Neighbors' selection; Task 8. Highest degree node; Task 9: Tracing paths over groups; Task 10: Memory.

Their users' experiments have at least 30 users for each task (the numbers of participants for each task is about 30-70 people). Number of users for each task is indicated after the task name. Their findings indicate that the map-like approach "Bubble-sets perform better than others for the tasks involving group membership assessment," and the GMap can help its users to improve their ability of memorization (for short time) they called this memorability. Another remarkable finding in their work is the map, such as GMap with disjointed areas, performs equally or better than other approaches.

Also in 2014, Saket et al. [134] also conducted a user's experiments with the different types of node-link diagrams: Node Diagrams(N diagrams), node-link Diagrams(NL diagrams), and Node-link-Group Diagrams(NLG diagrams). They stated that the tools used in the works of Jianu et al. [79] are NLG diagrams. In this work, they evaluated the different diagrams with three types of tasks: Node-Based Tasks (e.g., ask the users what is the color in diagrams of a given concept), Network-Based Tasks (tasks about the nodes and its links) and Group-based Task (tasks about the nodes, its links and its groups, e.g., find the neighboring groups of a given group). They recruited 36 participants (23 male, 13 female), with aged from 21 to 32 years. Before their controlled experiment, they informed their participants about the purpose of the study, the data, and the used technique. Their finding indicated that adding link and group representations does not negatively affect performance (time and accuracy) of node-based tasks and network-based tasks, but improve the users' performances with group-based tasks.

11.2 Psychological experimental protocol

Inspired by the existing evaluation works, we propose a psychological experimental protocol, and it contains three tasks: ontology browsing, ontology understanding, and ontology remembering. This experimental protocol aims to evaluate the psychological advantage of Memory Island technique: We designed the psychological sub-tasks to verify whether Memory Island can help its users with their information seeking tasks and knowledge memorization.

11.2.1 Ontology Browsing Task

Navigating through the content of ontology is the main purpose of this task. During the task, we would like to investigate ontology browsing with different visualization interface by asking users to answer a set of questions. The answers we obtain from various participants during the user study indicate how well they are able to browse and navigate ontology (taxonomy).

- During the task, the users need to answer three sub-type questions:
 - **T1.1. Path tracing.** The participant need to follow a path from root to a selected leaf node, select the node (concept) with the highest number of children's nodes.
 - **T1.2. Find parent.** Select the parent node P of a given concept X.
 - **T1.3. Find children.** Select the sub-concept(s) of a given concept Z.

11.2.2 Ontology Understanding Task

This task tests how well participants could understand the structure of an ontology (typology) using the different visualization techniques. The percentages of correct answers help us compare the visualization approaches. The task, such as find a specific class, can be used to verify whether the users have really understood the structure of ontology (taxonomy).

- Two subjects was designed to test if the users can understand the structure of ontology.
 - **T2.1. Find a specific class(with a special charaterisitics).** Navigate throught the ontology with the visualization, and selected the node whois the child of given concept X, and it has n children.
 - **T2.2. Find a specific class(Complex).** Select the node who has N children, which is the offspring of given concept X, and is in level L of the taxonomy of ontology. The users need to browse X, then to find the answer node.

Question sample: Browse the class “X” and go through its subclasses (path) until you find a specific class, which has a special characteristic. The participants need to provide the name of that class or an instance associated with it.

11.2.3 Ontology Remembering Task

The main purpose of the third task is to test how well the different visualizations could help users remember the positions of the classes on a previously visited path. In this task, we need to ask participants to revisit the previously visited classes after performing the browsing and the understanding task. If participants remember the approximate positions of the previously visited nodes, they are able to finish this task quicker and more accurately. It verifies whether Memory Island can help its users with their knowledge memorization and facilitate their remembering of the ontology. This task is called the memory (short-term) task or the memorability task.

- Four sub-tasks are designed for this memory task.
 - **T3.1. Recall location.** Select the location of a given concept in an image (see Figure 11.1). This given node is the answer for T1.2. (node P).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

- **T3.2. Recall hierarchy(simple).** Choose the node who has the given children X and Y. The answer of this question is the given concept Z in T1.3.
- **T3.3. Recall hierarchy.** Given a list of nodes, the participants need to select the node who has a different parent than the others. Except the answer node, all the others are the children of the answer of T2.2.
- **T3.4. Recall the path.** Select the the correct path from given concept X to concept Y. The answer of this subject is the path given in T1.1.

Question sample: Which one is the parent of the classes “X” and “Y”?

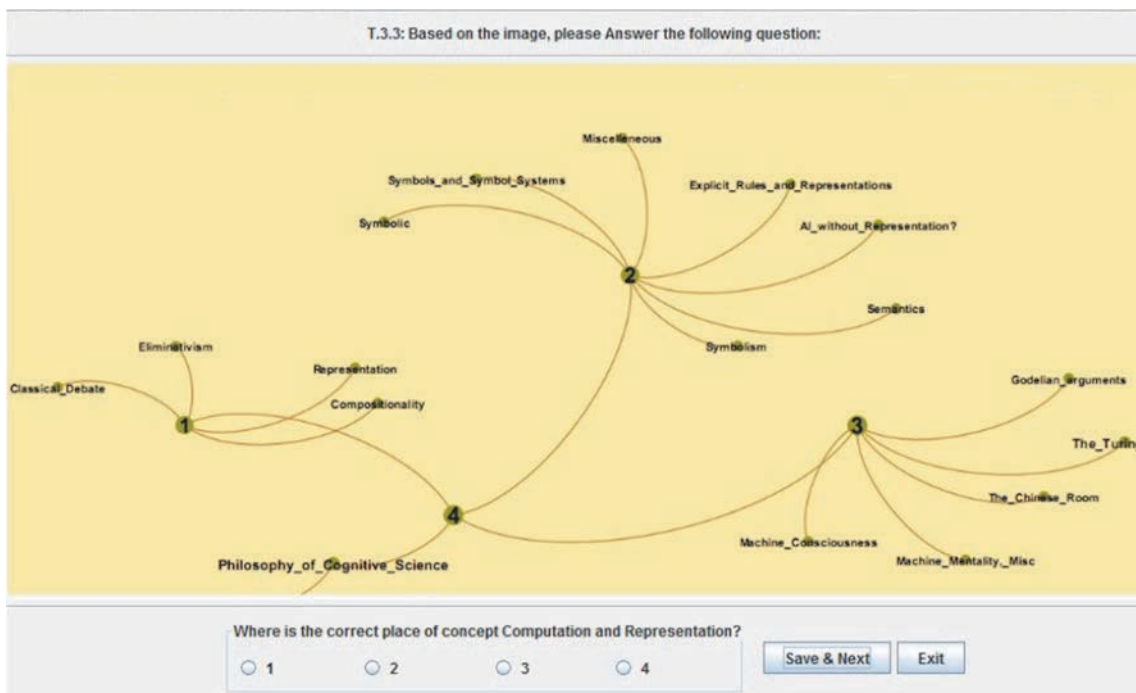


Figure 11.1 An example of the task T3.1 in our preliminary implementation of user study software.

11.2.4 The Subjective Task

After each session of experiments, the participants need to fill some subjective questions. They replied on a 5-point Likert scale (5= completely satisfied and 1= completely unsatisfied) to show how much they like the visualization.

- **T4.1.** This visualization is easy to learn.
- **T4.2.** This visualization is easy to use.
- **T4.3.** This visualization is fun.
- **T4.4.** It is easy to browse the ontology with this visualization.
- **T4.5.** It is easy to memorize the ontology with this visualization.

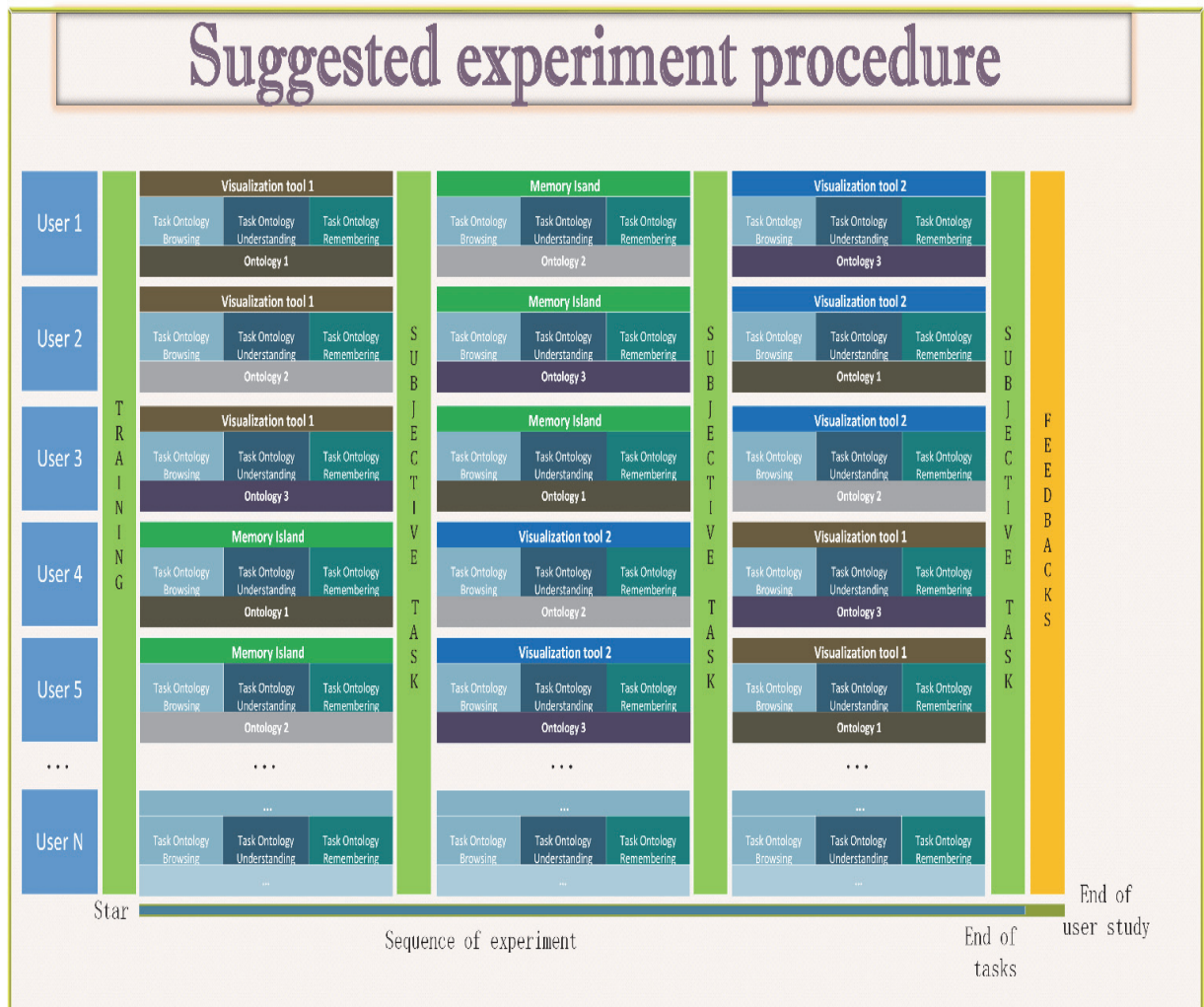
Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

- **T4.6.** It is easy to remember the ontology with this visualization.
- **T4.7.** I like this visualization in general.
- **T4.8.** I like to use this visualization in future.

11.2.5 Suggested Evaluation Procedure

With this evaluation protocol, each participant need to perform all three types of tasks, using different visualization tools, with different ontologies. Similar to the work of Song et al. [135], we need to counterbalance the order of visualizations and ontologies to avoid the learning effect. Our suggested experiment procedure using the Latin square design is shown in Table 2. Preferably, the whole time of the experiment should not exceed 30 minute.

Table 2 Suggested experiment procedure for our psychological experimental protocol



11.3 A Preliminary Users Study

In order to make its result impartial (to avoid the case that the authors or experts of the visualization technique may subconsciously choose to design the questions favorable or unfavorable to their preferred tool), this preliminary users study is organized by a master student (Laxenaire) as her independent master thesis on evaluation different ontology visualizations[10].

11.3.1 Methods and Experiment Set-up

This experiment followed our suggested protocol and suggested procedure show in Table 2. This experiment was set-up in a local machine and was controlled by the master student.

For this user study, the master student selected 15 participants with different levels of expertise, from ages 25 to 54. Participants were asked questions with respect to three ontologies Software Ontology (SWO), Material Ontology and ONTOderm ontology (size 50-200 nodes) with Memory Island (without search function), Gephi, and the indented list.

11.3.2 Subjects

In this preliminary empirical experiment, Laxenaire defined the subjects based on the three tasks in our psychological experimental protocol, and the detail questions used in this experiment were designed by the student, based on our psychological experimental protocol. She has defined one question for each sub-task we proposed in the evaluation protocol. See [10] for detail information.

It worth to noticed that during her experiment, she designed the subjective task

11.3.3 Procedure

The participants asked to perform all the subjects of the three tasks: browsing, understanding, and remembering. After each session with a visualization tool, she asked participants to fill out questionnaires for subjective evaluation. The same procedure is repeated with the other visualizations.

The users may have a different order of ontologies used with different tools, as we described in our suggested procedure (in Table 2). Laxenaire decided to use the following rule based on their given user's ID.

- For the users with $ID \bmod 3 = 0$: the order of ontologies is Software Ontology (SWO), Material Ontology, and ONTOderm ontology.
- For the users with $ID \bmod 3 = 1$: the order of ontologies is Material Ontology, ONTOderm ontology and Software Ontology (SWO).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

- For the users with ID mod 3 = 2: the order of ontologies is: ONTOderm ontology, Software Ontology (SWO) and Material Ontology,

Due to the limitation of numbers of participants in the experiment and the limited time to set-up the experiment for the intern, Laxenaire did not change the order of visualization tools in her experimental application. The order of visualizations tools for all the users are the same: Indented List, Memory Island, and Gephi.

Laxenaire give the participants a tutoring on ontology and the 3 visualization tools before starting this application. The results of this experiment are stored in a server and all the participants are controlled during their experiments.

11.3.4 Result

The respond times for browsing and remembering tasks are considered, the answers of the each question for each participants are also stored. With this experiment, Laxenaire first compared the time spending for browsing and remembering, and then she studied the correct rate for the tasks.

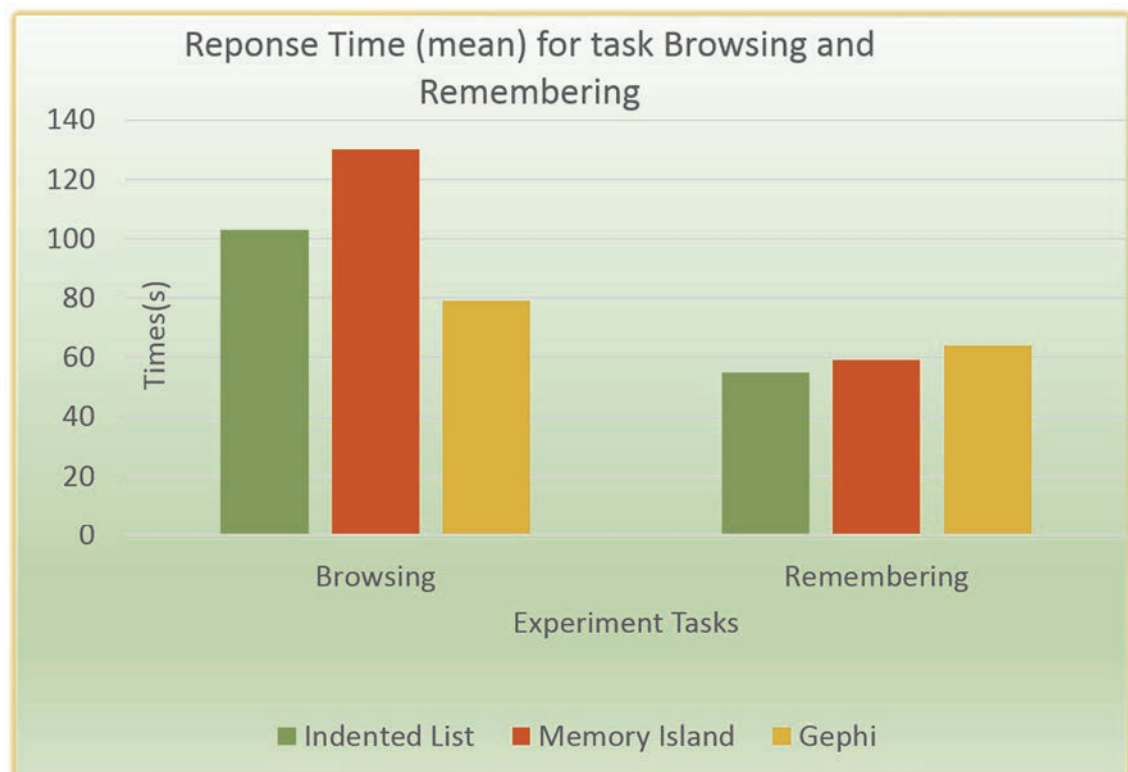


Figure 11.2 The responds time (mean) for task browsing and remembering, data from [10].

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

The responds time (mean) for browsing task and remembering is shown in Figure 11.2 and Figure 11.3. The overall correct rate for each visualization tools of the three tasks is shown in Figure 11.4. Detail about these results is reported in [10].

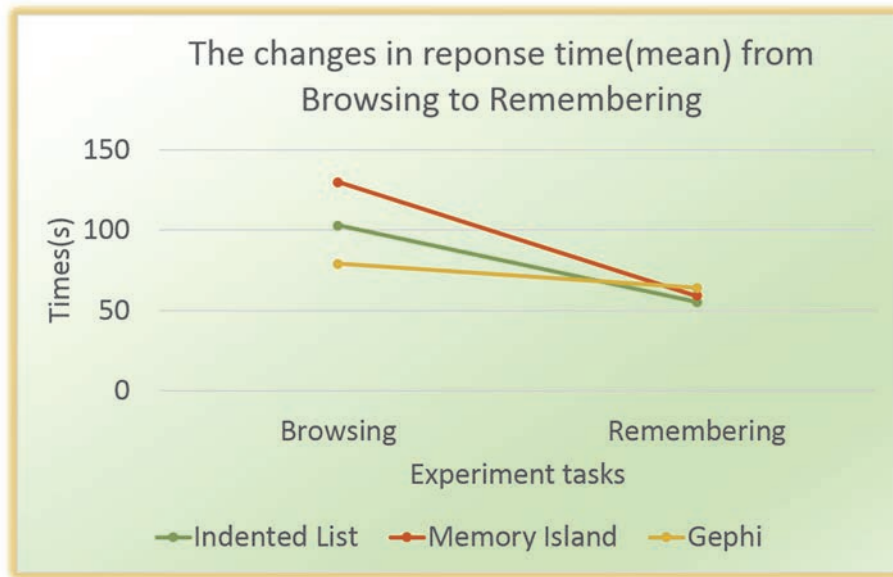


Figure 11.3 The changes in RT from Browsing to Remembering, data from [10].

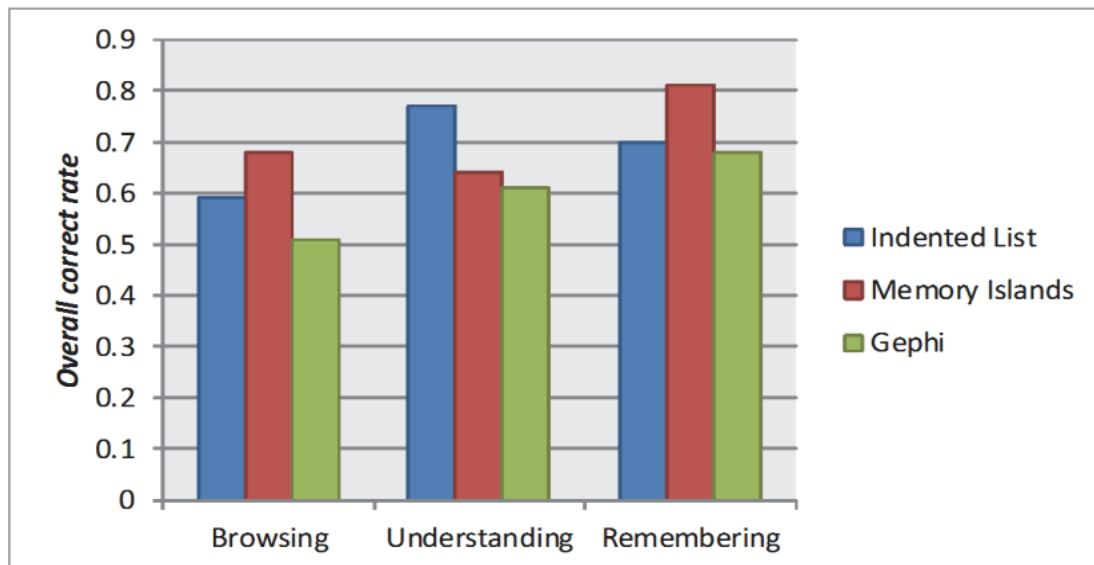


Figure 11.4 Overall correct rate for each visualization tool on the three tasks (data from [10]).

In the end, she studied the users' preference with their answers in subjective questionnaire. In general, none of these three tools is particularly preferred by the users. The preferences rates of users are: Memory Island: 2.71, Gephi: 2.75, and Indented List: 2.75. (A different Likert scale is used based on the method that "the lower is better").

11.4 Discussion with the past evaluation experiment

For the part of the psychological experimental protocol, the results in Figure 11.4 indicate that Memory Islands offers an advantage with the ontology-browsing task and the ontology-remembering task. Regarding the ontology-understanding task, the experiment was not able to provide any conclusive evidence since users are more familiar with indented list than with Memory Islands and Gephi. The users spend more time for the tasks of browsing with Memory Islands. When they go to the remembering tasks, the users can effectively use Memory Islands to correctly response these questions.

Unlike our validation experiments, after the short training section (on basic knowledge of ontology and three visualization tools), there is not a warm up section (sample questions) for the participants to learn how to perform the tasks, before the real-experiments. That may make the users spend more time with Memory Islands, as they may never use this type of visualization tools.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

Part VI. Conclusion and Perspectives

“The Most Beautiful House in the World

(is the one you build yourself)”

Witold Rybczynski [1989]

Chapter 12 CONCLUSIONS

Our contribution proposes an original technique inspired by the "Art of Memory" technique for visualizing hierarchical knowledge as insightful islands. With the power of knowledge, such as that of an ontology, we can generate a Memory Island, which simulates how great scientific figures used the "Art of Memory" for their knowledge. Besides ontologies, we also tested our technique with text and documents data, and employed a named-entity recognition approach to help us with visualizations of children's books. For example, we generate a Memory Island based on the table of contents with the named entities recognized in each chapter.

From the results of the user experiments, we find that Memory Island provides some advantages, especially for non-experienced users tackling realistic browsing. The geographic metaphors and cartographic means we designed for this technique can aid the users for their visual knowledge discovery. Moreover, from the users' preferences, we conclude that most of them appreciate the use of Memory Island for navigating large ontologies, finding it more interesting to navigate knowledge through an interactive island. From the evaluation experiment comparing it to different other visualization tools for ontologies (indented list and node-link diagram visualization tool), Memory Island helps with navigating and memorizing knowledge for most users. These findings indicated that the Memory Island technique well solved the research problem described in the beginning of this thesis: the meaningful and insightful Memory Islands generated by our proposed technique, help the human users with their information seeking tasks and improve their memorability.

Furthermore, our technique also opens a door for other researchers to extend the geographic visualization technique to visual knowledge analysis. The many valuable future researches, which could be established based on the Memory Island technique, are presented in **Chapter 13**. We hope that the other researchers will find our work useful, and it will be a great pleasure for us to see visualizations based on Memory Island released in the future.

Chapter 13 FUTURE WORKS

In this chapter, we present the future research opportunities for the Memory Island technique. Some (short-term) perspectives and improvement directions of the technique are discussed in the end of the related chapters.

13.1 Develop the Visualization Tools for Digital Humanities

As we mentioned, the island generated by Memory Island technique can be used as a base map to create various useful visualizations. For example, we are currently working on overlaying the result Memory Islands generated using the NER technique, with a heat map of named-entities' frequency (occurrence in the corpus). The interest is to build a visualization tool for visually analyzing text and document data. Then, we would like to introduce the time dimension to visualize the changing of named-entities; users can see the change of the frequency of named-entities, on the basic map of a specific relation between the entities (for example, friendship, or co-working).

Many other techniques for visualization of geographic information can be applied to visual knowledge analysis and visual knowledge discovery by using the Memory Islands generated by our technique.

13.2 Improve the Usage of space

As we discussed with our Island generation algorithm, we would like to use the empty room of each concept in the island with a space-filling algorithm to present more information found from or associated with the knowledge.

Then we can propose a novel cartographic space-filling algorithm to add more detailed cartographic features (e.g., cities, forest, river, buildings in a city, etc.) in order to visualize more detailed information on Memory Island. For example, with Google Map, in Figure 13.1 when

we focus on a specific location (city of Paris, with zoom-in function), detailed information about the city will appear in deeper zoom levels.



Figure 13.1 an example of zoom function of Google (when we zoom-in to deeper levels, we can see more detailed geographic information), screen shot from the Google Map web application.

13.3 Evaluation of the long-term memorization

Although maps have been proved as useful tools for reviewing the knowledge for learning and studying, we still need to evaluate the long-term memorization using the Memory Island. We are planning an evaluation on the topic of Long-Term Knowledge Memorization with different visualizations (e.g., the map-based approaches). We plan to conduct this user experiment within the COST Action TD1210: people in different age groups, from different nations and with different backgrounds.

We need to ask the participants to visit and revisit (review) the Memory Islands for a period, and we need to control the time spent on this learning process. The participants would need to check the visualization the next day, the third day, the fifth day, in a week, 2 weeks, and do some tasks for each period of a month.

We plan to design the subjects with help of the experts in cognitive science. These questions will test whether the Memory Islands allows better retention of ontologies (memorization) by counting how many labels they can still remember after learning a period. An example of our first attempt is shown in Figure 13.2.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands Bin Yang- June 2015

Another interesting topic would be the setting of a user study with the students of PRES Sorbonne Universities (within the Labex OBVIL). The students would use the Memory Islands of some knowledge dataset during their course (for example a course on InfoVis or Digital Humanities), and they will be tested in the end of this course with the questions about these knowledge.

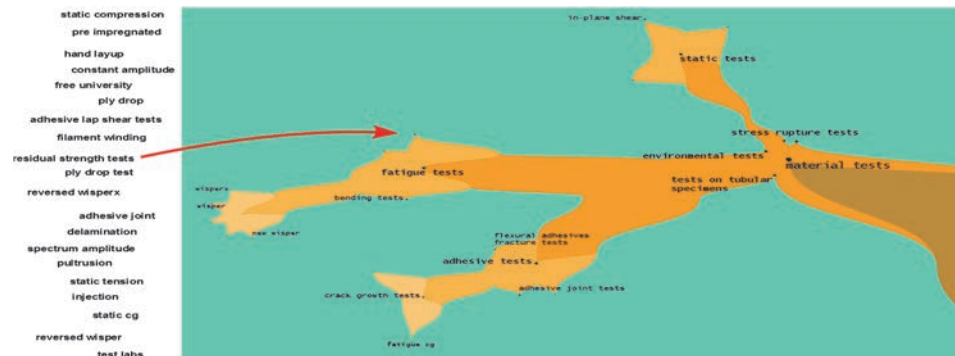


Figure 13.2 Evaluate the long-term memorization effect of the Memory Island. In this figure, we ask the participants to put some name of concepts to its locations.

13.4 User Interface and interactive functions

We have introduced some features for the Project Locupleto to help adolescents and adults with reading and learning from books. Such features include the visited traces of a reader or a group of readers.

In the future, we want to improve this interface to support more interactive functions (e.g., modify), and work on how to generate a more beautiful island with some aesthetics experts. The future Memory Island interface could become more powerful if we can collaborate with the experts of the domain of Human Learning.

Another research direction is to design how to use the third dimension to present additional information with the generated 2D Memory Island (e.g., a sub-island of philosophy of science with the associated thinkers, in a 2.5D design, as shown in Figure 13.3).

As we discussed in section 7.2, the future interface of Memory Island can further develop with more interactive functions, such as modify and re-generate knowledge. We could create a visual knowledge Editing Software based on the future Memory Island Interface. Another interesting point is that besides the function, which allows the users to add the icons or landmark shapes (Figure 7.7) on the map, we can design a method to automatically add related icons on the map (Figure 13.4).

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

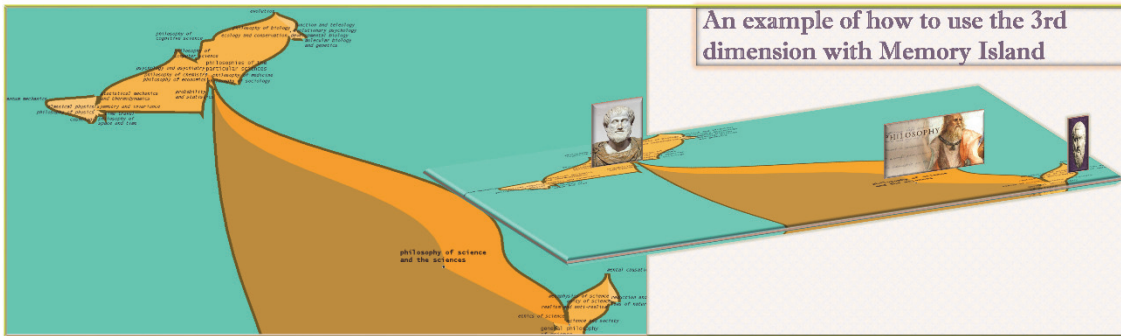


Figure 13.3 An example for using the third dimension (2.5D design) with Memory Island.

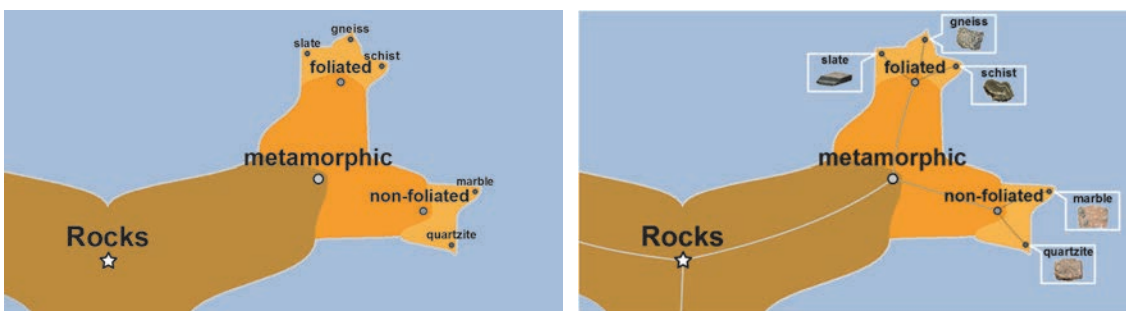


Figure 13.4 Imaginary renderings for the future application interface. With icons associated with each concepts.

13.5 Integrate Memory Islands to platforms of DH and DL

Our generated Memory Islands can be integrated with many platforms and tools in Digital Humanities and Digital Libraries. One of them is the Prévu project²⁵ on Digital Libraries. Developed in collaboration with Labex OBVIL, it aims to develop a platform for accessing and visualizing a library data corpus. It provides an intuitive interface (an example is shown in Figure 13.5) based on our cognitive abilities to locate and store an object (the book) in an organized space and it tries to help its users to find the books they are interested in.

²⁵ Website of Prévu project is www.prevu.fr, the participants of this project are PARIS8, EnsadLab/EN-ER, University of Michigan, Bibliothèque Universitaire de Paris 8 and Campus Condorcet.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

It would be very interesting if we could integrate Memory Islands with the topics of books (classifications) as an interface to this Digital Library. These Memory Islands could also interact with the elements in the 3D platform, to help the users find the geographic locations they want to access.

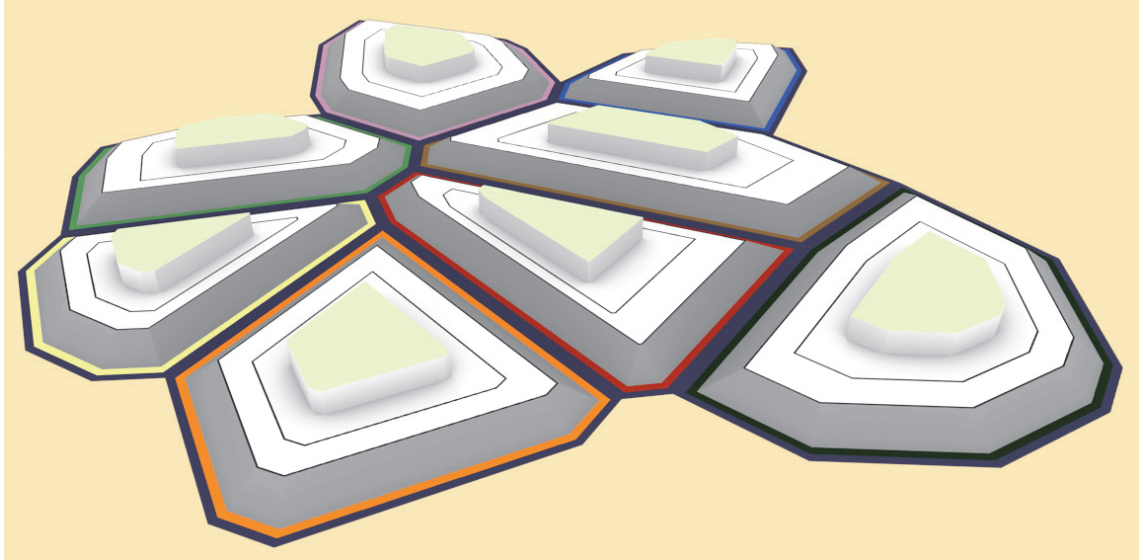


Figure 13.5 An example of the Prévü 3D platform. Image created by Donatien Aubert of EnsadLab EN-ER, and reproduced in this thesis with permission.

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

References

- 1 Hascoët, M., and Beaudouin-Lafon, M.: 'Visualisation interactive d'information', Information, Interaction, Intelligence, 2001, 1
- 2 Gleiser, M.: 'The Island of Knowledge: The Limits of Science and the Search for Meaning' (Basic Books, 2014. 2014)
- 3 Johnson, B., and Shneiderman, B.: 'Treemaps: a space-filling approach to the visualization of hierarchical information structures', in Editor (Ed.)^(Eds.): 'Book Treemaps: a space-filling approach to the visualization of hierarchical information structures' (IEEE, 1991, edn.), pp. 284-291
- 4 Ware, C.: 'Information Visualization: Perception for Design (3rd edition)' (Morgan Kaufmann, 2012. 2012)
- 5 Auber, D.: 'Tulip-a huge graph visualization framework', in Junger, M., and Mutzel, P. (Eds.): 'Graph Drawing Software (Mathematics and Visualization)' (Springer, 2003), pp. 105-126
- 6 Lin, X., and Ahn, J.-W.: 'Challenges of knowledge structure visualization', in Editor (Ed.)^(Eds.): 'Book Challenges of knowledge structure visualization' (Ergon Verlag, 2013, edn.), pp. 73-87
- 7 Dzbor, M., Motta, E., Aranda, C.B., Gomez-Perez, J.M., and Lewen, O.G.a.H.: 'Developing ontologies in OWL: An observational study', in Editor (Ed.)^(Eds.): 'Book Developing ontologies in OWL: An observational study' (2006, edn.), pp.
- 8 Torrance, E.P.: 'Dyadic interaction as a facilitator of gifted performance. ', Gifted Child Quarterly, 1970, 14, (3), pp. 139-143
- 9 Yates, F.A.: 'The art of memory' (University Of Chicago Press, 1968. 1968)
- 10 Laxenaire, M.: 'Evaluation comparative de techniques de visualisation d'ontologies', University Pierre and Marie Curie, 2012
- 11 Card, S.K., Mackinlay, J.D., and Shneiderman, B.: 'Readings in information visualization: using vision to think' (Morgan Kaufmann Publishers Inc, 1999), pp. 1-34

- 12 <http://datavis.ca/milestones/>
- 13 Keller, T., and Tergan, S.-O.: 'Visualizing Knowledge and Information: An Introduction. Visualizing knowledge and information', LNCS 3426 - Knowledge and Information Visualization Knowledge and Information Visualization 2005, pp. 1-23
- 14 Chen, M., Ebert, D., Hagen, H., Laramee, R.S., Liere, R.v., Ma, K.-L., Ribarsky, W., Scheuermann, G., and Silver, D.: 'Data, information, and knowledge in visualization', IEEE Computer Graphics and Applications, 2009, 29, (Issue No.01 - January/February), pp. 12-19
- 15 Schank, R.C., and Abelson, R.P.: 'Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale', in Editor (Ed.)^(Eds.): 'Book Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale' (Hillsdale, N.J. : Lawrence Erlbaum Associates, 1977, edn.), pp.
- 16 Biljon, J.A.V.: 'Knowledge visualization in academic reporting: a human-computer-interaction perspective', in Editor (Ed.)^(Eds.): 'Book Knowledge visualization in academic reporting: a human-computer-interaction perspective' (2012, edn.), pp.
- 17 Fekete, J.-D.: 'The InfoVis Toolkit', in Editor (Ed.)^(Eds.): 'Book The InfoVis Toolkit' (IEEE Press, 2004, edn.), pp. 167-174
- 18 Noy, N.F., Ferguson, R.W., and Musen, M.A.: 'The knowledge model of Protege-2000: Combining interoperability and flexibility', in Editor (Ed.)^(Eds.): 'Book The knowledge model of Protege-2000: Combining interoperability and flexibility' (2000, edn.), pp.
- 19 Francisco, A.P., C. Vaz, P.T.M., Melo-Cristino, J., and M. Ramirez, J.A.C., BMC Bioinformatics, 2012, 13, (1)
- 20 Bastian, M., Heymann, S., and Jacomy, M.: 'Gephi: an open source software for exploring and manipulating networks', in Editor (Ed.)^(Eds.): 'Book Gephi: an open source software for exploring and manipulating networks' (The AAAI Press, 2009, edn.), pp.
- 21 Munzner, T.: 'A Nested Model for Visualization Design and Validation', IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2009), 2009, 15, (6), pp. 921-928
- 22 Meyer, M., Sedlmair, M., and Munzner, T.: 'The Four-Level Nested Model Revisited: Blocks and Guidelines', in Editor (Ed.)^(Eds.): 'Book The Four-Level Nested Model Revisited: Blocks and Guidelines' (2012, edn.), pp.

- 23 Eisner, J.: 'State-of-the-art algorithms for minimum spanning trees: A tutorial discussion', in Editor (Ed.)^(Eds.): 'Book State-of-the-art algorithms for minimum spanning trees: A tutorial discussion' (1997, edn.), pp. 77
- 24 Cormen, T.H., Leiserson, C.E., Rivest, R.L., and Stein, C.: 'Minimum Spanning Trees': 'Introduction to Algorithms,' (MIT Press and McGraw-Hill, 2009), pp. 624-631
- 25 Newman, M.E.J.: 'Fast algorithm for detecting community structure in networks', PHYSICAL REVIEW 2004, 69
- 26 Chimera, R., and Shneiderman, B.: 'An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents', ACM Transactions on Information Systems, 1994, 12, (4), pp. 383-406
- 27 Egan, D., Remde, J.R., Gomez, L.M., Landauer, T.K., Eberhardt, J., and Lochbum, C.C.: 'Formative design-evaluation of SuperBook', ACM Transactions on Information Systems 1989, 7, (1)
- 28 Kumar, H.P., Plaisant, C., and Shneiderman, B.: 'Browsing hierarchical data with multilevel dynamic queries and pruning', International Journal of Human-Computer Studies, 1997, 46, (1), pp. 103-124
- 29 Noy, N.F., and McGuinness, D.L.: 'Ontology Development 101: A Guide to Creating Your First Ontology', in Editor (Ed.)^(Eds.): 'Book Ontology Development 101: A Guide to Creating Your First Ontology' (2001, edn.), pp. 25
- 30 Gruber, T.R.: 'A translation approach to portable ontology specifications', in Editor (Ed.)^(Eds.): 'Book A translation approach to portable ontology specifications' (Stanford University, 1993, edn.), pp. 27
- 31 Gruber, T.R.: 'a translation approach to portable ontology specifications', Knowledge Acquisition, 1993, 5, (2), pp. 199-220
- 32 Sowa, J.F.: 'Ontology, Metadata, and Semiotics', in Editor (Ed.)^(Eds.): 'Book Ontology, Metadata, and Semiotics' (Springer-Verlag, 2000, edn.), pp. 55-81
- 33 Amann, B., and Fundulaki, I.: 'Integrating Ontologies and Thesauri to Build RDF Schemas', in Editor (Ed.)^(Eds.): 'Book Integrating Ontologies and Thesauri to Build RDF Schemas' (Springer, 1999, edn.), pp. 234-253

- 34 Buckner, C., Niepert, M., and Allen, C.: 'InPhO: The Indiana philosophy ontology', *APA Newsletter on Philosophy and Computers*, 2007, 7, (1), pp. C26-C28
- 35 Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E.: 'Ontology Visualization Methods –A Survey', *ACM COMPUTING SURVEYS (CSUR)*, 2007, 39, (4)
- 36 Rivadeneira, W.: 'A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces', in Editor (Ed.)^(Eds.): 'Book A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces' (University of Maryland HCIL, 2003, edn.), pp.
- 37 Souza, K.X.S., Santos, A.D.D., and Evangeista, S.R.M.: 'Visualization of Ontologies through Hypertrees', in Editor (Ed.)^(Eds.): 'Book Visualization of Ontologies through Hypertrees' (2003, edn.), pp. 251-255
- 38 Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., and Dantzich, M.v.: 'Data Mountain: Using Spatial Memory for Document Management', in Editor (Ed.)^(Eds.): 'Book Data Mountain: Using Spatial Memory for Document Management' (ACM, 1998, edn.), pp. 153-162
- 39 Katifori, A., Torou, E., Halatsis, C., Lepouras, G., and Vassilakis, C.: 'A comparative study of four ontology visualization techniques in Protege: experiment setup and preliminary results', in Editor (Ed.)^(Eds.): 'Book A comparative study of four ontology visualization techniques in Protege: experiment setup and preliminary results' (IEEE, 2006, edn.), pp. 419-423
- 40 Rosvall, M., and Bergstrom, C.T.: 'Maps of random walks on complex networks reveal community structure', *PNAS(Proceedings of the National Academy of Sciences)*, 2008, 105, pp. 1118-1123
- 41 Eide, Ø.: 'An inquiry into the relationship between verbal and map-based expressions of geographical information', King's College London, 2012
- 42 <http://eagereyes.org/blog/2010/the-difference-between-infographics-and-visualization>
- 43 Okada, A., Shum, S.B., and Sherborne, T.: 'Knowledge Cartography: Software Tools and Mapping Techniques (Advanced Information and Knowledge Processing)' (Springer, 2008. 2008)

- 44 Quaggiotto, M.: 'Knowledge cartographies: Tools for the social structures of knowledge', in Editor (Ed.)^(Eds.): 'Book Knowledge cartographies: Tools for the social structures of knowledge' (2008, edn.), pp.
- 45 Quaggiotto, M.: 'Knowledge Atlas: a cartographic approach to the social structures of knowledge ', in Editor (Ed.)^(Eds.): 'Book Knowledge Atlas: a cartographic approach to the social structures of knowledge ' (2008, edn.), pp.
- 46 Auber, D., Huet, C., Lambert, A., Renoust, B., Sallaberry, A., and Saulnier, A.: 'GosperMap: Using a Gosper Curve for Laying out Hierarchical Data', IEEE Transactions on Visualization and Computer Graphics (TVCG), 2013, 19, (11), pp. 1820-1832
- 47 Duarte, F.S.L.G., Sikansi, F., Fatore, F.M., Fadel, S.G., and Paulovich, F.V.: 'Nmap: A Novel Neighborhood Preservation Space-filling Algorithm', IEEE Trans. Visualization & Comput. Graphics (Proc. InfoVis 2014.), 2014, 20, (12), pp. 2063 - 2071
- 48 Rusu, A.: 'Tree Drawing Algorithms', in Tamassia, R. (Ed.): 'Handbook of Graph Drawing and Visualization' (CRC Press, 2013)
- 49 Collins, C., Penn, G., and Carpendale, S.: 'Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations', IEEE Transactions on Visualization and Computer Graphics, 2009, 15, (6), pp. 1009 - 1016
- 50 Hong, L., Meng, F., and Cai, J.: 'Research on Layout Algorithms for Better Data Visualization', in Editor (Ed.)^(Eds.): 'Book Research on Layout Algorithms for Better Data Visualization' (2009, edn.), pp. 369-372
- 51 Skupin, A., and Fabrikant, S.I.: 'Spatialization methods: A cartographic research agenda for non-geographic information visualization', Cartography and Geographic Information Science, 2003, 30, (2), pp. 99-120
- 52 Gansner, E., Hu, Y., and Kobourov, S.: 'GMap: Visualizing Graphs and Clusters as Maps', in Editor (Ed.)^(Eds.): 'Book GMap: Visualizing Graphs and Clusters as Maps' (IEEE, 2010, edn.), pp. 201-208
- 53 Kohonen, T.: 'Self-Organizing Maps (Third Extended Edition)' (Springer Series in Information Sciences, 2001. 2001)
- 54 Fried, D., and Kobourov, S.G.: 'Maps of Computer Science', in Editor (Ed.)^(Eds.): 'Book Maps of Computer Science' (IEEE, 2014, edn.), pp. 113-120

- 55 Skupin, A.: 'From Metaphor to Method: Cartographic Perspectives on Information Visualization', in Editor (Ed.)^(Eds.): 'Book From Metaphor to Method: Cartographic Perspectives on Information Visualization' (IEEE Computer Society, 2000, edn.), pp. 91-97
- 56 Skupin, A.: 'The World of Geography: Visualizing a Knowledge Domain with Cartographic Means', in Editor (Ed.)^(Eds.): 'Book The World of Geography: Visualizing a Knowledge Domain with Cartographic Means' (National Academy of Sciences, 2004, edn.), pp. 5274-5278
- 57 Pampalk, E., and Dezember, W.: 'Islands of Music - Analysis, Organization, and Visualization of Music Archives', in Editor (Ed.)^(Eds.): 'Book Islands of Music - Analysis, Organization, and Visualization of Music Archives' (Vienna University of Technology 2001, edn.), pp.
- 58 Tu, K., Xiong, M., Zhang, L., Zhu, H., Zhang, J., and Yu, Y.: 'Towards Imaging Large-Scale Ontologies for Quick Understanding and Analysis', in Editor (Ed.)^(Eds.): 'Book Towards Imaging Large-Scale Ontologies for Quick Understanding and Analysis' (Springer Berlin Heidelberg, 2005, edn.), pp. 702-715
- 59 Gansner, E., Hu, Y., and Kobourov, S.: 'Viewing Abstract Data as Maps', in Huang, W. (Ed.): 'Handbook of Human Centric Visualization' (Springer, 2013)
- 60 Everitt, B.S., Landau, S., and Leese, M.: 'Cluster Analysis' (Oxford University Press, 2001, 4th Edition edn. 2001)
- 61 Balzer, M., Deussen, O., and Lewerentz, C.: 'Voronoi treemaps for the visualization of software metrics', in Editor (Ed.)^(Eds.): 'Book Voronoi treemaps for the visualization of software metrics' (ACM, 2005, edn.), pp. 165-172
- 62 Goethem, A.v., Reimer, A., Speckmann, B., and Wood, J.: 'Stenomaps: Shorthand for Shapes', IEEE TVCG (InfoVis'14), 2014, 20, (12), pp. 2053-2062
- 63 Thorndyke, P.W., Hayes-Roth, B., and Stasz, C.: 'Performance models for spatial and locational cognition', in Editor (Ed.)^(Eds.): 'Book Performance models for spatial and locational cognition' (the RAND Corporation report series, 1980, edn.), pp.
- 64 Thorndyke, P.W.: 'Distance estimation from cognitive maps', Cognitive Psychology, 1981, 13, pp. 526-550

- 65 Thorndyke, P.W., and Hayes-Roth, B.: 'Differences in Spatial Knowledge Acquired from Maps and Navigation', *Cognitive Psychology*, 1982, 14, pp. 560-589
- 66 Darken, R.P., and Sibert, J.L.: 'Navigating large virtual spaces', *Int. J. Hum.-Comput. Interact.*, 1996, 8, (1), pp. 49-71
- 67 Shneiderman, B.: 'Research Agenda: Visual Overviews for Exploratory Search', *Information Seeking Support Systems*, 2008, 11
- 68 Tobler, W.: 'A Computer Model Simulating Urban Growth in the Detroit Region.', *Economic Geography*, 1970, 46, (2), pp. 234-240
- 69 Ganascia, J.-G.: 'AC3 - Automatic cartography of cultural contents', in Editor (Ed.)^(Eds.): 'Book AC3 - Automatic cartography of cultural contents' (Springer-Verlag, 2007, edn.), pp. 253-263
- 70 Ganascia, J.-G., Madeira, C., and Fouladi, K.: 'An Adaptive Cartography of DTV Programs', in Editor (Ed.)^(Eds.): 'Book An Adaptive Cartography of DTV Programs' (Springer Berlin Heidelberg, 2008, edn.), pp. 253-262
- 71 Skupin, A.: 'On Managing Domain Knowledge: Ontology, Visualization, and Beyond', in Editor (Ed.)^(Eds.): 'Book On Managing Domain Knowledge: Ontology, Visualization, and Beyond' (2014, edn.), pp.
- 72 Cairo, A.: 'The Island of Knowledge and the Shorelines of Wonder. Visualization and Infographics for an Enlightned World', in Editor (Ed.)^(Eds.): 'Book The Island of Knowledge and the Shorelines of Wonder. Visualization and Infographics for an Enlightned World' (2014, edn.), pp.
- 73 Dobrowolski, T., Huntington, P., Nicholas, D., and Williams, P.: 'Reappraising Information Seeking Behaviour in a Digital Environment: bouncers, checkers, returners, and the like', *Journal of Documentation*, 2004, 60, (1), pp. 24-43
- 74 Westerman, S.J., Collins, J., and Cribbin, T.: 'Browsing a Document Collection Represented in Two and Three dimensional Virtual Information Space', *International Journal of Human Computer Studies*, 2005, 62, (6), pp. 713-736
- 75 Miller, N.E., Wong, P.C., Brewster, M., and Foote, H.: 'TOPIC ISLANDS-a wavelet-based text visualization system', in Editor (Ed.)^(Eds.): 'Book TOPIC ISLANDS-a wavelet-based text visualization system' (IEEE, 1998, edn.), pp. 189-196,

- 76 Cockburn, A.: 'Revisiting 2D vs 3D implications on spatial memory', in Editor (Ed.)^(Eds.): 'Book Revisiting 2D vs 3D implications on spatial memory' (Australian Computer Society, 2004, edn.), pp. 25-31
- 77 Cockburn, A., and McKenzie, B.: 'Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments', in Editor (Ed.)^(Eds.): 'Book Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments' (ACM, 2002, edn.), pp. 203-210
- 78 Brath, R., and Banissi, E.: 'Using Font Attributes in Knowledge Maps and Information Retrieval', in Editor (Ed.)^(Eds.): 'Book Using Font Attributes in Knowledge Maps and Information Retrieval' (2014, edn.), pp. 23-30
- 79 Jianu, R., Rusu, A., Hu, Y., and Taggart, D.: 'How to Display Group Information on Node-Link Diagrams: an Evaluation', IEEE Transactions on Visualization and Computer Graphics (InfoVis 2014), 2014, 20, (12), pp. 1530 - 1541
- 80 Lafon, S., Bouali, F., Guinot, C., and Venturini, G.: 'Hierarchical Reorganization of Dimensions in OLAP Visualizations.', IEEE transactions on visualization and computer graphics 2013, 19, (11), pp. 1833-1845
- 81 Slimani, T.: 'Description and Evaluation of Semantic similarity Measures Approaches', International Journal of Computer Applications 2013, 80, (10)
- 82 Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E.G.M., and Milios, E.E.: 'Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web', in Editor (Ed.)^(Eds.): 'Book Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web' (ACM, 2005, edn.), pp. 10-16
- 83 Lin, D.: 'An information-theoretic definition of similarity.', in Editor (Ed.)^(Eds.): 'Book An information-theoretic definition of similarity.' (1998, edn.), pp. 296-304
- 84 Miller, G.A.: 'WordNet: A Lexical Database for English. ', Communications of the ACM, 1995, 38, (11), pp. 39-41
- 85 Fellbaum, C.: 'WordNet: An Electronic Lexical Database' (1998. 1998)

- 86 Petrakis, E.G.M., Varelas, G., Hliaoutakis, A., and Raftopoulou, P.: 'X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies', *Digital Information Management (JDIM)*, 2006, 4
- 87 Seco, N., Veale, T., and Hayes, J.: 'An intrinsic information content metric for semantic similarity in wordnet', *Proc. of ECAI*, 2004, 4, pp. 1089-1090
- 88 Rada, R., Mili, H., Bicknell, E., and Blettner, M.: 'Development and application of a metric on semantic nets', *IEEE Transactions on Systems, Man and Cybernetics (Volume:19 , Issue: 1)* 1989, 19, (1), pp. 17-30
- 89 Leacock, C., and Chodorow, M.: 'Combining local context and WordNet similarity for word sense identification' (1998), pp. 265 - 283
- 90 Li, Y., Bandar, Z.A., and McLean, D.: 'An approach for measuring semantic similarity between words using multiple information sources', *IEEE Transactions on Knowledge and Data Engineering*, 2003, 15, (4), pp. 871-882
- 91 Pirro, G.: 'A semantic similarity metric combining features and intrinsic information content', *Data & Knowledge Engineering*, 2009, 68, (11), pp. 1289-1308
- 92 Pirro, G., and Euzenat, J.: 'A feature and information theoretic framework for semantic similarity and relatedness', in Editor (Ed.)^(Eds.): 'Book A feature and information theoretic framework for semantic similarity and relatedness' (2010, edn.), pp. 615-630
- 93 Nilsson, C.: 'Heuristics for the traveling salesman problem', in Editor (Ed.)^(Eds.): 'Book Heuristics for the traveling salesman problem' (Linkoping University, 2003, edn.), pp.
- 94 Lamping, J., Rao, R., and Pirolli, P.: 'A focus+context technique based on hyperbolic geometry for visualizing large hierarchies', in Editor (Ed.)^(Eds.): 'Book A focus+context technique based on hyperbolic geometry for visualizing large hierarchies' (1995, edn.), pp. 401-408
- 95 Lamping, J., and Rao, R.: 'The Hyperbolic Browser : A Focus+Context Technique for Visualizing Large Hierarchies', *Journal of Visual Languages and Computing* 1996, 7, pp. 33-55

- 96 Lin, C.-C., and Yen, H.-C.: 'On balloon drawings of rooted trees', *Graphics Algorithms and Applications (JGAA)*, 2007, 11, (2), pp. 431-452
- 97 Fuchs, J., Isenberg, P., Bezerianos, A., Fisher, F., and Bertini, E.: 'The Influence of Contour on Similarity Perception of Star Glyphs', *IEEE Transactions on Visualization and Computer Graphics (TVCG - InfoVis 2014)*, 2014, 20, (12), pp. 2251-2260
- 98 Vartaniana, O., Navarreteb, G., Chatterjeed, A., Fiche, L.-B., Lederf, H., Modronog, C., Nadalf, M., Rostruph, N., and Skovi, M.: 'Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture', in Editor (Ed.)^(Eds.): 'Book Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture' (2013, edn.), pp.
- 99 Kato, T., and Imai, H.: 'The NP-completeness of the character placement problem of 2 or 3 degrees of freedom', in Editor (Ed.)^(Eds.): 'Book The NP-completeness of the character placement problem of 2 or 3 degrees of freedom' (1988, edn.), pp. 11-18
- 100 Marks, J., and Shieber, S.: 'The computational complexity of cartographic label placement', in Editor (Ed.)^(Eds.): 'Book The computational complexity of cartographic label placement' (Harvard University, 1991, edn.), pp.
- 101 Formann, M., and Wagner, F.: 'A packing problem with applications to lettering of maps', in Editor (Ed.)^(Eds.): 'Book A packing problem with applications to lettering of maps' (ACM, 1991, edn.), pp. 281-288
- 102 Yoeli, P.: 'The logic of automated map lettering', *The Cartographic Journal*, 1972, 9, (2), pp. 99-108
- 103 Wagner, F., and Wolff, A.: 'A combinatorial framework for map labeling', in Editor (Ed.)^(Eds.): 'Book A combinatorial framework for map labeling' (Springer-Verlag, 1998, edn.), pp. 316-331
- 104 Edmondson, S., Christensen, J., Marks, J., and Shieber, S.: 'A general cartographic labeling algorithm', *Cartographica*, 1997, 33, (4), pp. 13-23
- 105 Kern, J.P., and Brewer, C.A.: 'Automation and the Map Label Placement Problem: A Comparison of Two GIS Implementations of Label Placement', *Cartographic Perspectives*, 2008, 60

- 106 Christensen, J., Marks, J., and Shieber, S.: 'An empirical study of algorithms for point-feature label placement', *ACM Transactions on Graphics (TOG)*, 1995, 14, (3), pp. 203-232
- 107 Wu, C.V., and Battenfield, B.P.: 'Reconsidering rules for point-feature name placement', *Cartographica*, 1991, 28, (1), pp. 10-27
- 108 Dijk, S.v., Kreveld, M.v., Strijk, T., and Wolff, A.: 'Towards an evaluation of quality for label placement methods', in Editor (Ed.)^(Eds.): 'Book Towards an evaluation of quality for label placement methods' (Internat. Cartographic Association, 1999, edn.), pp. 905-913
- 109 Dijk, S.v.: 'Genetic Algorithms for Map Labeling', Utrecht University, 2001
- 110 Christensen, J., Marks, J., and Shieber, S.: 'Placing Text Labels on Maps and Diagrams', *Graphics gems IV*, 1994, pp. 497-504
- 111 Cravo, G.L., Ribeiro, G.M., and Lorena, L.A.N.: 'A greedy randomized adaptive search procedure for the point-feature cartographic label placement', *Computers & Geosciences*, 2007, 34, (4), pp. 373-386
- 112 Glover, F.: 'Future Paths for Integer Programming and Links to Artificial Intelligence', *Computers and Operations Research* 1986, 13, (5), pp. 533-549
- 113 Yamamoto, M., Camara, G., and Lorena, L.A.N.: 'Tabu search heuristic for point-feature cartographic label placement', *GeoInformatica*, 2002, 6, (1), pp. 77-90
- 114 Kirkpatrick, S., Jr., C.D.G., and Vecchi, M.P.: 'Optimization by Simulated Annealing', *Science*, 1983, 220, (4598), pp. 671-680
- 115 Cerny, V.: 'Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm', *Journal of Optimization Theory and Applications* 1985, 45, pp. 41-51
- 116 Zoraster, S.: 'Practical Results Using Simulated Annealing for Point Feature Label Placement', *Cartography and Geographic Information Systems* 1997, 24, (4), pp. 228-238
- 117 Holland, J.H.: 'Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence' (MIT Press, 1992, 2nd edition edn. 1992)
- 118 Fekete, J.-D., and Plaisant, C.: 'Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization', in Editor (Ed.)^(Eds.): 'Book Excentric Labeling: Dynamic Neighborhood Labeling for Data

Visualization' (Department of Computer Science, University of Maryland, 1998, edn.), pp.

119 Fekete, J.-D., and Plaisant, C.: 'Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization', in Editor (Ed.)^(Eds.): 'Book Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization' (ACM New York, 1999, edn.), pp. 512-519

120 Been, K., Daiches, E., and Yap, C.: 'Dynamic Map Labeling', IEEE Trans. Visualization & Comput. Graphics, 2006, 12, (5), pp. 773-780

121 Been, K., Nöllenburg, M., Poon, S.-H., and Wolff, A.: 'Optimizing Active Ranges for Consistent Dynamic Map Labeling', Proceedings of Annual Symposium on Computational Geometry (SoCG'08), 2008

122 Been, K., Nöllenburg, M., Poon, S.-H., and Wolff, A.: 'Optimizing active ranges for consistent dynamic map labeling', Computational Geometry (Special Issue on 24th Annual Symposium on Computational Geometry (SoCG'08)), 2010, 43, (3), pp. 312-328

123 Hollands, J.G., Carey, T.T., Matthews, M.L., and McCan, C.A.: 'Presenting a graphical network: a comparison of performance using fisheye and scrolling views', in Editor (Ed.)^(Eds.): 'Book Presenting a graphical network: a comparison of performance using fisheye and scrolling views' (1989, edn.), pp. 313-320

124 Stasko, J., and Zhang, E.: 'Focus+Context Display and Navigation Techniques for Enhancing Radial', in Editor (Ed.)^(Eds.): 'Book Focus+Context Display and Navigation Techniques for Enhancing Radial' (2000, edn.), pp.

125 Alvim, A.C.F., and Taillard, É.D.: 'POPMUSIC for the point feature label placement problem', European Journal of Operational Research, 2009, 192, (2), pp. 396-413

126 Stasko, J., Görg, C., and Liu, Z.: 'Jigsaw: supporting investigative analysis through interactive visualization', Information Visualization (2008), 2008, 7, (2), pp. 118-132

127 Brehmer, M., Ingram, S., Stray, J., and Munzner, T.: 'Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists', IEEE Trans. Visualization & Comput. Graphics (Proc. InfoVis 2014.), 2014, 20, (12), pp. 2271-2280.

128 Mosallem, Y., Abi-Haidar, A., and Ganascia, J.-G.: 'Unsupervised Named Entity Recognition and Disambiguation: An Application to Old French Journals', in Editor (Ed.)^(Eds.): 'Book Unsupervised Named

Entity Recognition and Disambiguation: An Application to Old French Journals' (Springer, 2014, edn.), pp. 12-23

129 <http://plato.stanford.edu/>

130 Shneiderman, B.: 'The eyes have it: A task by data type taxonomy for information visualization', in Editor (Ed.)^(Eds.): 'Book The eyes have it: A task by data type taxonomy for information visualization' (1996, edn.), pp. 336-343

131 Asahi, T., Turo, D., and Shneiderman, B.: 'Using treemaps to visualize the analytic hierarchy process', Information Systems Research, 1995, 6, (4), pp. 357-375

132 Alper, B., Riche, N., Ramos, G., and Czerwinski, M.: 'Design Study of LineSets, a Novel Set Visualization Technique', IEEE Transactions on Visualization and Computer Graphics, 2011, 17, (12), pp. 2259-2267

133 Gansner, E., Hu, Y., and Kobourov, S.: 'GMap: Visualizing Graphs and Clusters as Maps', Computer Graphics and Applications, 2010, 30, (6), pp. 54-66

134 Saket, B., Simonetto, P., Kobourov, S., and Borner, K.: 'Node, Node-Link, and Node-Link-Group Diagrams: An Evaluation', IEEE Trans. Visualization & Comput. Graphics (Proc. InfoVis 2014.), 2014, 20, (12), pp. 2231 - 2240

135 Song, H., Kim, B., Lee, B., and Seo, J.: 'A Comparative Evaluation on Tree Visualization Methods for Hierarchical Structures with Large Fan-outs', in Editor (Ed.)^(Eds.): 'Book A Comparative Evaluation on Tree Visualization Methods for Hierarchical Structures with Large Fan-outs' (ACM, 2010, edn.), pp. 223-232

136 Kriglstein, S.: 'User Requirements Analysis on Ontology Visualization', in Editor (Ed.)^(Eds.): 'Book User Requirements Analysis on Ontology Visualization' (IEEE, 2009, edn.), pp. 694 - 699

137 Falconer, S.M., Noy, N.F., and Storey, M.-A.: 'Ontology Mapping - A User Survey', in Editor (Ed.)^(Eds.): 'Book Ontology Mapping - A User Survey' (2007, edn.), pp. 49- 60

138 Falconer, S.M., Noy, N.F., and Storey, M.-A.: 'Towards understanding the needs of cognitive support for ontology mapping', in Editor (Ed.)^(Eds.): 'Book Towards understanding the needs of cognitive support for ontology mapping' (2006, edn.), pp. 49-60

139 Bosca, A., Bonino, D., and Pellegrino, P.: 'Ontosphere: more than a 3d ontology visualization tool', in Editor (Ed.)^(Eds.): 'Book

Ontosphere: more than a 3d ontology visualization tool ' (CEUR Workshop Proceedings. , 2005, edn.), pp.

140 Barlow, T., and Neville, P.: 'A comparison of 2-D visualizations of hierarchies', in Editor (Ed.)^(Eds.): 'Book A comparison of 2-D visualizations of hierarchies' (IEEE, 2001, edn.), pp.

141 Ridsen, K., Czerwinski, M.P., Munzner, T., and Cook, D.B.: 'An initial examination of ease of use for 2D and 3D information visualizations of web content', Internal Journal of Human Computer Studies, 2000, 53, (5), pp. 695-714

142 Alani, H.: 'TGVizTab: An Ontology Visualisation Extension for Protégé', in Editor (Ed.)^(Eds.): 'Book TGVizTab: An Ontology Visualisation Extension for Protégé' (2003, edn.), pp.

143 McGuinness, D.L., Fikes, R., Rice, J., and Wilder, S.: 'The chimaera ontology environment', in Editor (Ed.)^(Eds.): 'Book The chimaera ontology environment' (2000, edn.), pp. 1123-1124

144 Do, H.-H.: 'Schema Matching and Mapping-based Data Integration', Universit"at Leipzig, 2006

145 Noy, N.F., and Musen., M.A.: 'The PROMPT suite: Interactive tools for ontology merging and mapping.', International Journal of Human-Computer Studies, 2003, 59, (6), pp. 983-1024

146 Ehrig, M., and Staab, S.: 'QOM - quick ontology mapping', in Editor (Ed.)^(Eds.): 'Book QOM - quick ontology mapping' (2004, edn.), pp.

147 Kleiner, B.K., and Hartigan, J.A.: 'Representing Points in Many Dimensions by Trees and Castles', Journal of the American Statistical Association, American Statistical Association, 1981, pp. 260-272

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

APPENDIX

CONTENTS

RELATED PUBLICATIONS AND COMMUNICATIONS	169
COLOR SCHEMES	171
QUESTIONS FOR VALIDATION EXPERIMENTS	173

Related Publications and Communications

Publications:

- Bin Yang; Jean-Gabriel Ganascia, Memory Islands: an approach to cartographic visualization. In: /Classification & visualization: interfaces to knowledge: proceedings of the International UDC Seminar 24-25 October 2013, The Hague, The Netherlands/. Edited by Aida Slavic, Almila Akdag Salah, Sylvie Davies. Wurzburg: Ergon Verlag, 2013, pp 137-152
- Bin Yang; Jean-Gabriel Ganascia,: “Creating knowledge maps using Memory Islands”, First Workshop on Knowledge Maps and Information Retrieval co-located with International Conference on Digital Libraries 2014 - ACM/IEEE Joint Conference on Digital Libraries (JCDL 2014) and International Conference on Theory and Practice of Digital Libraries, vol. 1311, London, United Kingdom, pp. 15-22 (2014)
- Bin Yang; Jean-Gabriel Ganascia,: Cartographie des connaissances dans les humanités numériques par l’Îles de mémoires -- une démonstration In: Artlier Visualisation d’information, fouille visuelle de données et nouveaux challenges en Big data et Humanités numériques at IHM 14, Lille, France (2014)
- Bin Yang; Jean-Gabriel Ganascia,: How to Visualize Information by using the notion of Memory Islands. In: / Atelier Visualisation d'informations, interaction et fouille de données at EGC 14, Rennes, 2014, pp.1-2.

National Workgroup meeting (Journée(s) de Groupe de Travail)

- Bin Yang; Jean-Gabriel Ganascia, An overview of Memory Islands – an approach for cartographic visualization, in: Journées Big Data & Visualisation Fouille et Visualisation de Données Massives Big Data Mining and Visualization (FDCFGGVIF13) 24-25 June, 2013, Paris, France.
- Bin Yang; Jean-Gabriel Ganascia, Visualisation des connaissances et de l’information en utilisant la technique des îles de mémoires, in: the Journées des groupes de travail EGC avec une session spéciale Big Data et Santé, co-organisée avec le thème e-Santé du GDR STIC-Santé.23-24/06/2014, Lille, France
- Bin Yang; Jean-Gabriel Ganascia - "Memory Islands: A technique for visualizing tree-structure Knowledge and Clusters with cartographic means" in: the journée du GT Visualisation, du GDR IGRV. 07/11/2014 Paris, France.

In submission

- Extended version of our paper “Creating knowledge maps using Memory Islands” for International Journal on Digital Libraries, Special Issue on “Knowledge Maps and Information Retrieval (KMIR)” (2015)

Papers accepted

- Bin Yang; Jean-Gabriel Ganascia, Navigating through Memory Island of Stanford Encyclopedia of Philosophy – a demonstration of the Memory Island Technique In 2015 CSDH/SCHN & ACH joint conference.(2015)
- Alaa Abi Haidar; Bin Yang; Jean-Gabriel Ganascia, Extracting and Visualizing Named Entities using Interactive Streamgraphs – A Case Study on First World War Data. In In 2015 CSDH/SCHN & ACH joint conference.(2015)

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

Seminar Talk

- Bin Yang & Jean-Gabriel Ganascia, Creating knowledge maps using Memory Islands, In “Use of ontologies in systems biology and health” Seminar, 28/10/2014, Paris, France.

Invited talk

- Bin Yang & Jean-Gabriel Ganascia: “How to Visualize Information by using the notion of Memory Islands”, Ways of SEEING MULTIDISCIPLINARY WORKSHOP ON VISUALISATION AND ANALYSIS, Galway, Ireland (2014)

Color Schemes

The colors we used for Memory Island application is shown in the following Table 3.

Table 3 Selected color schemes for Memory Island application

Color schemes	Orange	Green	Yellow
level 1 (In taxonomy)	#BF8930	#6F9C00	#A68800
level2 (In taxonomy)	#FF9F00	#8DB42D	#BFA530
level 3 (In taxonomy)	#FFB740	#ABF000	#FFD100
level 4+ (In taxonomy)	#FFCA73	#C2F83E	#FFDC40
Sea	#5ED2B8	#A767D5	#4575D4
Label (Text) color	#2C1721	#F60018	#5D2680

Questions for Validation Experiments

In the annex, we show the questions we designed for validation Memory Island. To achieve this appendices have their own heading style, which is visually similar to the main headings but are functionally different. We summarized the complete subjects we designed with Memory Island validation experiment in Table 4.

PART 1: Preliminaries questions

Question 04 - How many nodes can you see in this ontology representation? (State a number as the answer).

Answer: [Show solution](#) [NEXT](#)



Figure 0.1 An example of our on-line users' survey program.

Table 4 the complete list of subjects for validation of Memory Island.

Type of Tasks	Description	Sub-tasks	Questions
Preliminary task	This part of questions asked users for basic knowledge of ontology to check background of user in order to determine their level of experience and few simple questions for users to get familiar with the survey interface.	Background on ontology	Do you have any experience using or knowing about "Ontology" before?
		Background on Visualization	How do you evaluate yourself about the knowledge in data visualization interpretation?
		Warm-up questions	How many nodes can you see in this ontology representation?

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

			What is the depth of this ontology? (State a number as the answer).
Ontology retrieval task	There will be many kinds of questions for users to answer by searching for the answer from given ontology representation to test each point from visualization mantra.	Overview task	From the image of an ontology visualization, guess what is the domain of this ontology from your point of view? Guess which part of ontology do you think it contains least number of nodes?
		Zoom task	How many descendant that node "Fatigue Tests" have?
		Details-on-demand task:	What is the parent of parent of "theories of mental content" node? What is the depth of node "dynamic system"? What is the parent of parent of "constant amplitude" node?
		Relate task:	Between node "type" and node

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang- June 2015

			<p>"computable construct" which one has more number of direct children?</p> <p>Between node "Static Tests" and node "Polymer" which one has more number of direct children?</p> <p>Between node "NOT USER NAMED" and node "SIMPLE EXPRESSION" which one has more number of direct children?</p>
		History task	What part of the island did you used for answer previous question?
		Node ordering task	<p>If the representation you see in this question showed nodes based on their meaning. Which mean the closer meaning node will appear nearer to each other. Which part do you think node name "philosophy of</p>

Memory Island: Visualizing Hierarchical Knowledge as Insightful Islands
Bin Yang - June 2015

			religion" should appear?
Preference tasks	We ask users to compare different ontology representation and choose one that they like most for each of specific task given in each question.	Preference on guess the general domain	When you were asked to guess the general domain of ontology from representation, Which representation do you prefer?
		Preference on Navigating	When we ask you to go through the ontology and find parent or child node, Which representation (tool) do you prefer?