HAL open science

Automatic methods for assisted recruitment

Luis Adrian Cabrera Diego

To cite this version:

Luis Adrian Cabrera Diego. Automatic methods for assisted recruitment. Automatic Control Engineering. Université d'Avignon, 2015. English. NNT : 2015AVIG0205 . tel-01306282

HAL Id: tel-01306282

https://theses.hal.science/tel-01306282v1

Submitted on 25 Apr 2016

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.

ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

Thèse

Présentée pour obtenir le grade de Docteur en Sciences
de l'Université d'Avignon et des Pays de Vaucluse

SPÉCIALITÉ : Informatique

École Doctorale 536 « Agrosciences et Sciences »
Laboratoire d'Informatique (EA 4128)

Automatic Methods for Assisted Recruitment

par

Luis Adrián CABRERA DIEGO

Soutenue publiquement le 9 décembre 2015 devant un jury composé de :

M.	Mathieu ROCHE	Chercheur HDR, TETIS, Montpellier	Rapporteur
M.	Djamel A. ZIGHED	Professeur, ISH, Lyon	Rapporteur
Mme	Rocío ABASCAL-MENA	Maître de Conférences, UAM, Mexico	Examinatrice
Mme	Catherine BERRUT	Professeur, LIG, Grenoble	Présidente
M.	Juan-Manuel TORRES-MORENO	Maître de Conférences HDR, LIA, Avignon	Directeur
M.	Marc EL-BÈZE	Professeur, LIA, Avignon	Co-directeur

Ce travail a également été encadré par :

M.	Barthélémy DURETTE	Chercheur, Adoc Talent Management, Paris	Co-encadrant

Laboratoire Informatique d'Avignon

Association Nationale de la Recherche et de la Technologie

Consejo Nacional de Ciencia y Tecnología

Abstract

The massification of the Internet and computers has changed several aspects of our daily life and the way we apply to a job is not the exception. Nowadays, the recruitment and selection of applicants for a job is done through the use of information technologies, creating what it is known as e-Recruitment. Since the last 15 years, the researchers in Natural Language Processing have been studying how to improve the performance of recruiters with the help of the e-Recruitment. Several systems have been developed in this field, from the job and applicants search engines to the automatic ranking of applicants. In the last case, most of the developed systems consist in the comparison between the résumés of applicants and a job offer. Only one system makes use of résumés from past selection processes to rank newer applicants.

In this thesis we study whether and how we can use the résumés, without having to use past selection processes, to develop new methods for e-Recruitment systems. More specifically, we start with the automatic treatment of a large set of résumés used during real recruitment and selection processes. Then, we analyze and apply different proximity measures to know which are the most adequate to study the résumés of applicants. We introduce, after, an innovative method which consists on the Relevance Feedback and the use of proximity measures over uniquely the résumés to rank applicants. Finally, we present the study and application of a statistical measure which allows us to compare, at the same time, the job offer, one specific applicant and the rest of applicants, in order to rank all the job candidates.

Along this thesis we show that résumés have enough information about the selection processes, in order to rank the applicants. Nonetheless, it is important to choose correctly the proximity measure to use. As well, we present interesting outcomes from the triple comparison between résumés and job offers.

The results obtained in this thesis are the basis for a new prototype of an e-Recruitment system and hopefully, the beginning of a new way to create these.

Keywords

e-Recruitment, Information extraction, Proximity measures, Analysis of résumés, Natural language processing.

Résumé

L'utilisation massive de l'Internet et les ordinateurs ont changé plusieurs aspects de notre vie quotidienne et la façon que nous postulons pour un travail n'y fait pas exception. Aujourd'hui, le recrutement et sélection des candidats pour un poste se font en utilisant les technologies de l'information. Ceci a créé le recrutement électronique. Depuis les 15 dernières années, les chercheurs du Traitement de la Langue Naturelle ont étudié comment améliorer les performances des recruteurs avec l'aide du recrutement électronique. Beaucoup de systèmes ont été développés dans ce domaine, depuis les moteurs de recherche de candidats ou de postes jusqu'au classement automatique de candidats. Dans ce dernier cas, les systèmes développés font, pour la plupart, la comparaison entre les CV des candidats et les offres d'emploi. Seul un système utilise les CV de processus de sélection relevant du passé pour classer les candidats à un nouveau poste.

Dans le cadre de cette thèse, nous avons étudié la possibilité et la façon d'utiliser les CV, sans avoir à exploiter aucun processus de sélection précédent, pour développer nouvelles méthodes applicables aux systèmes de recrutement électronique. Plus spécifiquement, nous commençons par le traitement automatique d'un grand ensemble de CV utilisés pendant des processus réels de recrutement et sélection. Ensuite, nous analysons et appliquons différentes mesures de proximité pour savoir lesquelles sont les plus appropriées pour étudier les CV des candidats. Après, nous introduisons une méthode innovante qui repose sur le Relevance Feedback et l'utilisation de mesures de proximité seulement sur les CV pour pouvoir classer les candidats d'un poste. Finalement, nous présentons l'étude et l'application d'une mesure statistique permettant de comparer, en même temps, l'offre d'emploi, un certain candidat et les autres candidats ; le but est de pouvoir classer tous les candidats d'un poste.

Dans cette thèse, nous montrons que les CV contiennent assez d'information sur le processus de sélection pour pouvoir classer les candidats. Néanmoins, il est important de choisir correctement les mesures de proximité à utiliser. D'ailleurs, nous présentons des résultats intéressants de la triple comparaison entre les CV et les offres d'emploi.

Les résultats obtenus dans cette thèse forment une base pour la conception de nouveaux prototypes de systèmes de recrutement électronique et possiblement le début d'une nouvelle façon pour les développer.

Mots-clés

Recrutment électronique, Extraction d'information, Mesures de proximité, Analyse de curriculum vitæ, Traitement de la langue naturelle.

Acknowledgments

I would like to express my sincere gratitude to my Ph.D. advisors Dr. Juan-Manuel Torres-Moreno and Dr. Marc El-Bèze, for their continuous support, motivation, patience and immense knowledge. For the last three and one half years, your advices, ideas and guidance helped me in order to finish my research and this thesis. Moreover, your support, especially during the hard moments of my Ph.D., is something for which I will be eternally grateful.

I would like to extend my deepest gratitude to Dr. Barthélémy Durette, who was my supervisor in Adoc Talent Management. Without his support, motivation, advices and guidance, my research project and thesis would not have been the same.

Thanks to my examining committee, especially to Dr. Djamel Zighed and Dr. Matthieu Roche, not only for the time committed to reading my thesis but for presenting me with new ideas and points of view.

Gracias is the only word I can say to my parents, Ismael Cabrera Mancilla and Gabriela Diego Casimiro, and to my sister *Gaby*. You had always time to listen to me, to support me and to guide me. Without you, my Ph.D. would not have been the same.

I want to thank to my best friends, Juan Miguel Rolland and Miguel Mazkiarán, who supported me every time and were always available to talk in spite of the distance and time difference.

As well, I want to give my special gratitude to: Thomas Dufour, your support, friendship and words of encouragement made me have the best months in Paris and helped me to finish this thesis, *merci beaucoup !* Pamela Torres, without your friendship, my stay in Avignon would not have been the same or with the same amount of laughs.

I cannot forget to thanks to my friends: Karla Hinojosa, Alejandro Molina, Mariana Signoret, Laura Torres, Miriam Colin, Alejandra Carrillo, Cristiane Morgado and more I would like to list, you were like my family here in France. I loved the good times and *soirées* we passed together.

I want to express my sincere gratitude to Dr. Amandine Bugnicourt and Dr. Matthieu Lafon who gave me the opportunity to work with Adoc Talent Management. To work with you and all the team was something incredible that I will not forget.

Last but not least, I would like to thank to the *Consejo Nacional de Ciencia y Tecnología (CONACyT)* from Mexico who supported my Ph.D project from september 2014 to december 2015.

Contents

Contents vii

List of Figures ix

List of Tables xi

1 Introduction 1
- 1.1 Recruitment and Selection . . . 1
  - 1.1.1 Résumés . . . 2
- 1.2 The e-Recruitment . . . 3
- 1.3 Objective . . . 4
- 1.4 Background and Scope . . . 4
- 1.5 Structure of the Thesis . . . 5

2 State of the Art 7
- 2.1 The e-Recruitment Systems . . . 7
  - 2.1.1 Documents Searchers . . . 8
  - 2.1.2 Documents Segmenters and Annotators . . . 9
  - 2.1.3 Résumés Rankers . . . 11
- 2.2 Analysis of Résumés . . . 13
- 2.3 Conclusions . . . 15

3 Processing Automatically a Human Resources Corpus 19
- 3.1 Objectives . . . 22
- 3.2 Methodology . . . 22
  - 3.2.1 Document Conversion . . . 22
  - 3.2.2 Language Recognition . . . 23
  - 3.2.3 Résumé Detection . . . 23
  - 3.2.4 Résumé Uniqueness . . . 27
  - 3.2.5 Data Representation . . . 28
- 3.3 Results and Discussion . . . 29
- 3.4 Conclusions . . . 30

4 Textual Proximity Measures Applied to Résumés 33
- 4.1 Objectives . . . 34
- 4.2 Methodology . . . 35

4.2.1 Model . . . 35
4.2.2 Calculation of the Résumé Proximity . . . 37
4.2.3 Statistical Test . . . 39
4.2.4 Data . . . 41
4.3 Results . . . 42
4.4 Discussion . . . 46
4.5 Conclusions . . . 48

5 Automatic Assistance for Processing Large Quantities of Résumés 51
5.1 Objective . . . 52
5.1.1 Hypothesis . . . 53
5.2 Methodology . . . 53
5.2.1 Proximity Measure . . . 53
5.2.2 Relevance Feedback . . . 55
5.2.3 Evaluation Methodology . . . 58
5.2.4 Experiments . . . 60
5.2.5 Data . . . 61
5.3 Results . . . 61
5.3.1 AIRP versus Baselines . . . 61
5.3.2 AIRP with Implicit Relevance Feedback . . . 63
5.3.3 AIRP with Explicit Relevance Feedback . . . 65
5.3.4 AIRP Using the Best Parameters . . . 66
5.4 Discussion . . . 68
5.5 Conclusions . . . 72

6 Using the Proximity Between a Résumé, a Job Offer and others Résumés 75
6.1 Objectives . . . 76
6.2 Methodology . . . 76
6.2.1 Kullback-Leibler Divergence . . . 77
6.2.2 Trivergence . . . 78
6.2.3 Trivergence Applied to HRM . . . 79
6.2.4 Experiments . . . 80
6.2.5 Data . . . 81
6.3 Results . . . 81
6.3.1 Most Adequate Trivergence in a HRM Context . . . 81
6.3.2 Application of the Trivergence into a HRM Context . . . 83
6.4 Discussion . . . 84
6.5 Conclusion . . . 88

7 Conclusion and Perspectives 91
7.1 Perspectives . . . 93

A Box plots 97

B Supplement Results of the Implicit Relevance Feedback 99

Bibliography 101

List of Figures

3.1 Percentage of Selected candidates by the number of job offers. 21

4.1 Example of the possibles proximity calculation between 3 Selected résumés and 3 Rejected ones. 36
4.2 Percentage of French Selected candidates by the number of job offers. 41
4.3 ANOVAs' Effect Size for each measure in natural logarithm scale. Only the jobs with statistically different groups were taken into account. 43

5.2 Results, in terms of the Mean Average Precision, of AIRP and the two baselines: résumés sorted randomly by applicants' arrival date. 62
5.3 Results of the values for AIRP with Implicit Relevance Feedback for percentages from 10% to 50%. 63
5.4 Results of the values for AIRP with Implicit Relevance Feedback for percentages from 60% to 100%. 64
5.5 Results of the values for AIRP with Explicit Relevance Feedback. 65
5.6 Results of the values for AIRP with iterative Explicit Relevance Feedback using blocks of résumés. 68

6.1 Boxplots of each possible combination of the trivergence's elements. 82
6.2 Results of the analysis of 49 jobs with AIRP, with random sorting, with the proximity between résumés and a job offer, and with the most and the less adequate combination of trivergence. 84
6.3 The trivergence $\tau(X,Y,Z) = KL(X,Y) \cdot KL(Z,Y) \cdot KL(X,Z)$ presented as a diagram, where X is the job offer, Y the résumé of the hired applicant and Z the rest of résumés sent to a job. 86
6.4 The trivergence $\tau(X,Y,Z) = KL(X,Y) \cdot KL(Y,Z) \cdot KL(X,Z)$ presented as a diagram, where X is the job offer, Y the résumé of the hired applicant and Z the rest of résumés sent to a job. 87

A.1 Two examples of boxplots. 98

B.1 Results of the values for the AIRP with Implicit Relevance Feedback for percentages from 10% to 50%. 99
B.2 Results of the values for the AIRP with Implicit Relevance Feedback for percentages from 60% to 100%. 100

List of Tables

3.1 Number of job offers, job applications and files in the Recruitment Corpus. . 21
3.2 Evaluation, in terms of precision, recall and F-score, of the linear and radial SVM through a five-fold cross-validation 25
3.3 Confusion matrix of the Résumé Detector evaluation according the evaluation corpus annotated by Recruiter 1. 27
3.4 Confusion matrix of the Résumé Detector evaluation according the evaluation corpus annotated by Recruiter 2. 27
3.5 Evaluation of the Résumé Detector in terms of precision, recall and F-score. . 27
3.6 Number of résumés used during the processing of the Recruitment Corpus . . 30

4.1 Results of the Analysis of Variance (ANOVA) for each proximity measure. . . 42
4.2 Number of jobs supporting and refuting the experimental hypothesis about the proximity of selected and reject résumés. 44
4.3 Significance rate (SR), Testing rate (TR), Inference rate (IR) and Global score (GS) for each proximity measure. 46

5.1 Example of the calculation of the Precision (P), the Average Precision (AP) and the Mean Average Precision (MAP) for two priority lists. 59
5.2 Results of the Tukey's HSD test for AIRP versus the baselines. 62
5.3 Results of the Tukey's HSD test for AIRP with Explicit Relevance Feedback when 10 résumés are used. 66

6.1 Results of the Tukey's HSD test from the analysis of the application of the trivergence and its comparison with a set of baselines. 85

Chapter 1

Introduction

The *Human Resource Management (HRM)* is defined as the set of interrelated and strategic activities associated to the management of the organizations' employees (Armstrong and Taylor, 2014, page 5). The aim of these activities is to enhance employees well-being and performance but also to assure the organization future (Johnson, 2009; Watson, 2010;Boxall and Purcell, 2003, page 1).

According to Armstrong and Taylor (2014, page 4), among the activities that the HRM covers, we can find: human capital management, organization development, employee relations, knowledge management and people resourcing. This last activity is composed of several tasks, like talent management, workforce planning and, recruitment and selection of personnel. This last couple of tasks, recruitment and selection of personnel, is the one over we will focus in this thesis.

1.1 Recruitment and Selection

The *recruitment* and the *selection* processes are two activities belonging to HRM which offer the opportunity to organizations to change their staff (Schneider, 1987). They are frequently merged into the term *recruitment* although they are different (Searle, 2009).

The *recruitment* process are "those practices and activities carried out by an organization with the primary purpose of identifying and attracting potential employees" (Barber, 1998). The *selection* process consists in "choosing one candidate from a list of candidates for the desired role and level in the organization" (Pande and Basak, 2012). In other words, the recruitment process consists in the activities related with the invitation of job seekers to apply for a job, while the selection one corresponds to the tasks done for choosing the applicants following the organization's needs.

The recruitment and selection processes can be divided in 10 stages (Armstrong and Taylor, 2014, page 226):

1. Defining requirements
2. Attracting candidates
3. Screening applications
4. Interviewing
5. Testing
6. Assessing candidates
7. Obtaining references
8. Checking applications
9. Offering employment
10. Following up

While the first two stages belong to the recruitment process, the last 8 are part of the selection one.

Despite the fact that the selection process starts theoretically at the stage of *screening applications*, according to Dipboye and Jackson (1999), the selection process starts with the documents called *résumés*. A résumé is a document sent by a candidate after a formal application to a job offer (Armstrong and Taylor, 2014, page 234); this kind of documents will be presented in the next section.

1.1.1 Résumés

A *résumé*[1], from the past participle of the French verb *résumer* (to summarize), is a brief document, submitted as part of a job application, which can contain a wide range of personal information (American Heritage Dictionary of the English Language, 2014; Forbes, 2003). The personal information consists mainly in academic qualifications and achievements, work experience and extracurricular activities (Cole et al., 2003, 2007, 2009; Tsai et al., 2011). The objective of résumés is to make the recruiters to be interested into the job applicant and to create a connection between both (Huggins, 1977).

The résumés are an important component in the selection process (Knouse, 1994) as they are one of most important sources of personal data (Nemanick Jr and Clark, 2002). This is also because they have the capacity, to a certain extent, to predict which will be the applicant's behavior (Mael, 1991; Owens and Schoenfeldt, 1979) and performance (Hunter and Hunter, 1984; Owens, 1976). As well, the résumés can create strong first impressions to recruiters (Knouse, 1989; Wyant and Vise, 1979). In consequence, recruiters can change their

[1]Résumé is not the only term used in the literature. Another frequent term is the Latin locution *curriculum vitæ* which means course of life. For Sarada (2005) and Thompson (2000) both terms are synonyms. Therefore, in this thesis we will use the word résumé as common term.

perception about the applicant (Thoms et al., 1999), their consideration to do an interview (Cole et al., 2005, 2007) and the possibility to give a positive hiring recommendation (Tsai et al., 2011).

1.2 The e-Recruitment

In 1995, it was predicted that the future of the recruitment would be on the Internet (Edgley, 1995). Nowadays, we see that during the last 15 years, the massification of computers and the Internet have had an impact on the way people search for jobs and new employees (Rafter et al., 2000a; De Meo et al., 2007; Kessler et al., 2012; Parry and Wilson, 2009). In fact, Internet has become the main medium to recruit applicants (Bizer et al., 2005). Furthermore, now the recruitment and selection processes, with the help of the Internet and the information technologies, have become the *e-Recruitment* (Bizer et al., 2005; Radevski and Trichet, 2006; Yahiaoui et al., 2006; Kerrin and Kettley, 2003).

Most of the benefits from the use of the e-Recruitment have gone to companies and HRM. Today, the job offers can reach wider audiences: local, national or international (Montuschi et al., 2014). The operational costs of HRM have been reduced (Chapman and Webster, 2003). Moreover, according to Cappelli (2001), to do the recruitment and selection processes through the Internet costs one-twentieth as much of the original processes. In addition, the recruitment process has been reduced in time. For example, it is faster to post on-line job offers than using the traditional methods, like newspapers, college boards and journals (Elkington, 2005) but, it is also easier to spread job offers through specialized communities (Arthur, 2001, page 126). Finally, now it is possible to create interconnected databases, providing access to the data about previous applicants, current workforce or recently hired persons (Kim and O'Connor, 2009).

Regarding job seekers, the obtained benefits have consisted in time and easiness. Now, job seekers can search available job offers through the Internet (Looser et al., 2013) and apply to them faster, through an e-mail or a web form and an electronic résumé (Elkington, 2005).

At the end, the Internet has created an interactive network between job seekers, employers and recruiters which is available around the world 24/7 (Barber, 2006, page 1). Moreover, nowadays it is possible to find social networks related to the professional activities, where HRM, job seekers and others are in contact and interchanging information. Some of these social networks are *LinkedIn*[2], *Research Gate*[3], *Viadeo*[4] and *XING*[5].

[2]https://www.linkedin.com/
[3]http://www.researchgate.net/
[4]http://www.viadeo.com
[5]https://www.xing.com/

Although the e-Recruitment has brought the aforementioned benefits, some undesirable consequences have arisen for HRM, specially when high volumes of résumés are received (Barber, 2006, page 11). In first place, since the use of the e-Recruitment, there have been an important increment of the unqualified applications (Noelle, 2005; Faliagka et al., 2011). In second place, recruiters have difficult to manage correctly and rapidly the great amount of received data (Rafter et al., 2000b; Trichet et al., 2004).

In order to decrease the effect of these undesirable consequences over the HRM, the scientific community with the help of HRM have developed the *e-Recruitment systems*. An e-Recruitment system is a set of tools that utilizes different techniques from the *Natural Language Processing (NLP)* and the *artificial intelligence* in order to treat automatically the information located in the résumés, job offers and related documents.

The appearance of the e-Recruitment systems started 15 year ago, however, there are still problems to solve. For example, the detection of the résumés' sections, the use of knowledge resources, which are hard to build, to the résumés' lexical variety or the ranking of résumés according to their relevance for a job. Thus, in this thesis we present a set of experiments and innovative methods in order to assist the recruiters in the task of retrieving the most relevant applicants.

1.3 Objective

The main objective of this thesis is to develop new methods to assist HRM during the selection processes. In other words, to create different ways to help the recruiters during the reading of résumés. The reason to focus on this stage is that the résumés *bottleneck* is located at this part of the selection process. As well, because there are several windows of opportunity and we can develop and study new methodologies.

We decided that the developed methods should not substitute the recruiters neither decide by themselves who must be hired. This is because we do not consider that computers are ready nor prepared to be used in a so delicate task. Instead of it, the methods must make more efficient the recruiters by suggesting them which applicants should be analyzed in first place. In other words, the aim is to give a priority to the résumés with a quality while reducing the quantity of information to be analyzed and the time necessary to do it.

1.4 Background and Scope

This thesis is carried out in the scope of a *CIFRE (Conventions Industrielles de Formation par la REcherche)* project between the *Université d'Avignon et des Pays de Vaucluse* and *Adoc Talent*

Management.

Adoc Talent Management is a Human Resources company, settle in Paris, dedicated mainly into the recruitment and selection of specialized and doctorate professionals in several fields, like health and engineering.

All the data used in this thesis came from Adoc Talent Management, while the methods here presented were developed in collaboration with the Université d'Avignon et des Pays de Vaucluse.

This thesis does not present the development of an e-Recruitment system. Nonetheless, it sets the basis to create one prototype with innovative methods that were developed and tested over a large set of real data.

1.5 Structure of the Thesis

This thesis is composed of 6 more chapters. In Chapter 2, we introduce the state of the art related to the e-Recruitment systems and the analysis of résumés. Then, we present the automatic processing of a Human Resources corpus in Chapter 3. At Chapter 4, we make an analysis of different proximity measures in order to determine whether the résumés contain information about the selection process. In Chapter 5 we present the development of a methodology that which ranks the applicant through their average proximity. We present an innovative methodology for an e-Recruitment system, based on the dissimilarity of résumés and job offers, in Chapter 6. Finally, the conclusions and perspectives of this thesis are presented in Chapter 7.

Chapter 2

State of the Art

The e-Recruitment is the use of the Internet and other information technologies to perform the recruitment and selection processes (Bizer et al., 2005; Radevski and Trichet, 2006; Yahiaoui et al., 2006; Kerrin and Kettley, 2003). It was born as result of the massification of computers and the Internet around the world (Bizer et al., 2005) and it succeeded due to the benefits brought to HRM and job seekers. For example, job offers are available around the world (Montuschi et al., 2014) and can be posted in specialized communities within some clicks (Arthur, 2001, page 126); job seekers can apply in seconds through an e-mail with an electronic résumé attached (Elkington, 2005).

Nonetheless, as the e-Recruitment become more and more popular, some undesirable consequences appear to HRM, specially when several jobs seekers apply (Barber, 2006, page 11): the recruiters cannot manage correctly and rapidly the applications (Noelle, 2005; Faliagka et al., 2011) and a great number of them belong to unqualified candidates (Rafter et al., 2000b; Trichet et al., 2004). To solve these problems, the scientific community with the HRM, started 15 years ago to develop the e-Recruitment systems.

2.1 The e-Recruitment Systems

An e-Recruitment system is a set of computational tools with the goal of assisting HRM and, in lesser extent, job seekers. These systems can provide support from the search of résumés or job offers to the pertinence analysis of résumés respecting a job. In this section, we present the most representative e-Recruitment systems developed in the last 15 years; these are divided into 3 types and will be described in detail as follows.

2.1.1 Documents Searchers

In the literature, the first kind of e-Recruitment systems that we find are those related to the search of information. To be more specific, they are systems with tools developed to explore résumés, applicants profiles and/or job offers located in databases.

An excellent example of a document searcher is the website *Web del empleo* developed by García-Sánchez et al. (2006). The goal of this e-Recruitment system was to facilitate the search for a job or job seekers. At the website, the job seekers could create an on-line profile, where it had to be provided their work experience, education, skills, personal data, among other information; this profile worked like an on-line résumé. The employers or HRM were able to create on-line the job offers. This last document had to indicate the researched features, like the professional experience, the technical skills, the salary and the job location. In both cases, for almost all the fields the vocabulary was restricted thanks to an ontology[1]. In other words, the information that could be used by the candidates, employers and HRM for their profiles or job offers, had to follow certain rules. For example, there were a predefined list of nationalities, job sectors and country locations. The reason to restrict the vocabulary was to facilitate the search of job offers and résumés in the website without having to use complex search systems.

The system presented by Yahiaoui et al. (2006) follows the same philosophy of *Web del Empleo*. However, they differ in their methodology. While in the first one, the users had to fill on-line forms to create résumés and job offers, in the work of Yahiaoui et al. (2006), the documents had to be uploaded directly to their server. Once the résumés or job offers were on the server, the users were asked to annotate their own document using an on-line interface. The annotation consisted in marking the words or concepts related to the personal information (e.g. age, gender) or to the fields of computer science and telecommunications. The annotated documents were linked to a set of ontologies[2], the goal was to allow the enrichment of the documents' information. Users were allowed to search for résumés inside the database according to a job offer and vice-versa. The documents were retrieved from the system's data through a semantic matching method. For the employers and HRM, this approach consisted in finding the résumés that contained, semantically speaking, similar expressions to their job offer. For the job seekers, the method consisted in matching the job offers that contained expressions which were semantically similar to the ones presented in their résumé.

Another e-Recruitment searcher is *Screener* (Sen et al., 2012). In this system the job seekers had to upload their résumés to a server in order to be segmented and annotated automatically into different sections, like professional experience, education and skills. The recruiters and/or employers could search for the résumés through queries. The results were given by Screener according to the matching score calculated between the query and the

[1]An ontology is a structure that represents a portion of the human knowledge in a computational format; these will be broached in Section 2.1.2.

[2]Idem.

different sections of the résumés.

As it could be seen, to use sections or parts of a résumé is a common denominator in the search systems. Thus, to comprehend better the segmentation and annotation of the résumés, we will discuss, some tools developed in this field, as follows.

2.1.2 Documents Segmenters and Annotators

To being able to process correctly the documents in an e-Recruitment system it is necessary, in some occasions, to separate and annotate the different kinds of information located in each file. Some systems, make use of on-line forms to keep the data with a specific format and vocabulary (García-Sánchez et al., 2006). Some other systems ask the users to separate and mark themselves the information located in their own documents (Faliagka et al., 2013).

Nonetheless, there are occasions that these methods are not performing enough or can be considered as limited. For example, even if a job seeker knows better than anyone their résumé, this does not mean that it will separate and annotate the information as would have expected a recruiter. As well, most of the job seekers have nowadays an electronic résumé and to be forced to fill manually a form on-line may be considered as undesirable. Therefore, we found in the literature e-Recruitment systems that have been developed with tools to segment and/or annotate the documents.

One document segmenter we can find in the literature belong to the e-Recruitment system *E-Gen* (Kessler et al., 2008b). It consisted in separating a document into fractions, for example, to divide in 3 equal parts the résumés and motivation letters. The basis of this method derived from the observation that not all the relevant information is located at the same part of a document. For instance, the last fraction of a document has more frequently the information which is not relevant for the selection process. In the case of résumés, it may be occupied by the hobbies, while in the motivation letters it can be found the letter closing.

To annotate whether the résumés were from an executive person or not, Clech and Zighed (2003) developed a tool based on the terminological analysis. To be more precise, the authors made use of a terminological extractor to obtain the terms and concepts related, for instance, to the competences, skills and education. The extracted terms and concepts were categorized through the *C4.5* decision tree algorithm (Quinlan, 1993) and a discriminant analysis. With the concepts and terms classified, a typological annotator was built.

An example of a document segmenter and annotator is the tool of the e-Recruitment system called *Screener* (Sen et al., 2012). The tool consisted in finding and marking the principal sections of résumés, like education, skills, competences and professional experience. To do

that, Screener made use of rules based on the probable headings for each sections and punctuations marks.

A similar document segmenter and annotor is *SegCV* (Cabrera-Diego et al., 2013). SegCV was a tool created to find, regroup and annotate the information related to the basic sections of French résumés: personal data, professional experience, education, skills and competences, and other information. It employed human-created rules which were based on the most common headers and subheaders of the sections. The résumés were presented in an XML format with the data grouped into the different sections found.

One problem that most of the e-Recruitment systems have to make frequently face is the existence of different ways to express one same concept. Although for human beings is easy to know that two concept are similar or equivalent, for computers it is not the case. Thus, to reduce the consequences of the vocabulary variation, some systems have implemented the use of external knowledge resources, like ontologies and taxonomies. An *ontology*[3] is a set of concepts and relations organized by type and subtype (Lacasta et al., 2010). These concepts and relations belong to a specific domain and were selected following the *savoir-faire* of a field expert (Sowa, 2000). A taxonomy is a vocabulary set designed to classify other resources by generalization and specialization (Guarino et al., 2009; Lacasta et al., 2010).

The *E-Rec-Sys* (Karaa and Mhimdi, 2011) is an example of a tool that makes use of an ontology to annotate the different kinds of information that were located in a résumé, like names, diplomas and languages. The process done by E-Rec-Sys consisted, in first place, to split the documents into words. Then, to analyze those words with a morphological dictionary and to delete those considered as stop words. Finally, to use the ontology called *ERECO (ERECruitment Ontology)* to annotate automatically the data. The ontology ERECO was a knowledge base created from the European standard of résumés, *Europass CV*[4]. This standard addresses the most common information located in the résumés, like personal data, work experience, education, skills and competences; it was created for all the languages of the European Union. The annotated résumés were presented in a XML format, following an HR-XML[5] schema which was designed to be used with Europass CV.

Following the path of Karaa and Mhimdi (2011), we can find the tool *Ontology-based Résumé Parser (ORC)* (Çelik et al., 2013). The goal and methodology was similar to the one used by E-Rec-Sys, however, it added new characteristics to increase the performance of the data annotation. One of the problems found at the E-Rec-Sys was the incapacity to annotate information that was not located at the ontology. To solve this problem, the ORC tool made use of semantic rules to infer new concepts and facts that could be added to the ontology automatically. The ORC tool had, as well, an acronym extender, i.e. a dictionary of acronyms

[3]The term ontology has several acceptations. In this case we make reference to the sense of *computer ontology*. For a deeper analysis of the concept *ontology* see Giaretta and Guarino (1995).

[4]https://europass.cedefop.europa.eu/fr/documents/curriculum-vitae

[5]http://www.hropenstandards.org/

to retrieve the original word(s), and an orthographic corrector. This last characteristic was included to avoid the problem that applicants and/or job offers could not be found due to orthographic mistakes.

Most tools make only use of ontologies to annotate the résumés or job offers, however Senthil Kumaran and Sankar (2013) utilized them as a way to represent documents. The tool, belonging to a more complex e-Recruitment system called EXPERT, generated the ontologies by extracting the concepts of résumés and job offers, and finding the relations between them. To achieve this, the tool made use of the parts of speech (POS) tags, named entities and semantic parsing.

The literature presents, as well, some other tools that instead of segmenting the résumés by sections, they do it according to the relevance or informativeness of the text structures (phrases, paragraphs, etc.). For example, Kessler et al. (2012) used *Cortex* (Torres-Moreno et al., 2001), an automatic summarization tool, to delete the information that could be considered as noise. Although the résumés are a summary of the life, for the authors not all the data presented at the résumés is of relevance to the recruiters.

In the case of Di Meglio et al. (2007), they asked to recruiters to analyze résumés of previous applicants in order to know which parts of the résumés were the relevant ones. The goal was to make learn an e-Recruitment system that could detect, in recent résumés, the parts that would be of interest to recruiters. A similar tool, is the one presented by Singh et al. (2010), where they make use of *Conditional Random Fields* to create a model with the capacity to extract relevant features of résumés.

2.1.3 Résumés Rankers

We can find in the literature more complex e-Recruitment systems, where the goal is not only to analyze the résumés and/or job offers but to give, as well, a specific ranking of the applicants. In other words, to assist the recruiters letting them know which applicants are the most relevant for a job.

The first example of a résumé ranker is *E-Gen* (Kessler et al., 2007, 2008,b, 2009, 2012). It was an e-Recruitment system, developed for French, that dealt from the reception of applications to the ranking of résumés. The processes done by E-gen can be divided into two parts. The first one consisted in the recovering of the documents from the applicants' e-mails and to determine which files were résumés and which motivation letters. This last task was done through a not supervised machine learning algorithm. At this part of the system, the résumés and motivation letters were also segmented following two methods: documents fractions and automatic summarization. The second part comprehended the proximity of the résumés and/or motivation letters with respect a job offer, i.e. how similar or dissimilar were these two kinds of documents. The authors made use of the proximity measures of

Cosine Similarity, Minkowski Distance (Manhattan and Euclidean) and Okabi. To rank the applicants, they sorted the résumés according to the proximity results. In order to increase the performance of E-Gen, the authors implemented a relevance feedback algorithm[6], which allowed to the system to reproduce better the judgment of recruiters.

Another system that we can find in the literature is *PROSPECT* a project developed by Singh et al. (2010). The objectives of the system were to allow the recruiters to search and screen résumés located in a database but also to give a ranking of the applicants given a job offer. In contrast to the system E-Gen, that used a document fractions or automatic summarization algorithms to segment the data, the PROSPECT system used a text-mining method. The goal was to obtain the information related to skills, education and professional experience, to name some of them. Being more detailed, PROSPECT used a combination of Conditional Random Fields (CRF), a lexicon, a named-entity recognizer, a data normalizer and a table extractor to obtain the relevant data. The extracted information of the résumés was saved in databases to allow the recruiters to do queries and find candidates for a job offer. To rank the applicants according to a job offer, PROSPECT made use of *Okapi BM25*, the *Kullback-Leibler divergence* and the *Lucene Scoring*[7].

As indicated in the Section 2.1.2, the use of ontologies and/or taxonomies have an important role in the e-Recruitment systems. Therefore, it is not odd to find résumé rankers that make use of the knowledge presented in ontologies or taxonomies to do their task. As follows, we present some example of this kind of e-Recruitment systems.

LO-MATCH (Montuschi et al., 2014), is platform developed for the European project MATCH1. Its objective was to correspond résumés and job offers through the analysis of professional competences. The LO-MATCH system was based on the use of ontologies to enrich the information located in résumés and job offers. The purpose was to make easier the process of searching, for job seekers and recruiters, the most adequate job and the best candidate, respectively. To be more specific, LO-MATCH utilized an ontology created by domain-experts and the lexical database WordNet[8]. To rank the job offers and résumés the system made use of a semantic similarity, i.e. LO-MATCH measured how similar, in terms of the semantic, were the expressions located in the documents. Among the characteristics that make remarkable this system is the possibility to hint the applicants for missing or low developed competences. As well, the recruiters had the possibility to examine alternative competences for their job offer and the weakest candidates' points.

Another example of a system using ontologies is *EXPERT* (Senthil Kumaran and Sankar, 2013). It is a résumé screening tool that ranked the applicants according to the job offer requirements. Despite the fact that EXPERT uses ontologies like the system LO-MATCH, the

[6]The relevance feedback is an error input sent by a user to a system with the objective to ameliorate the results. It will be discussed in Chapter 5.

[7]https://lucene.apache.org/core/

[8]https://wordnet.princeton.edu/

approach utilized by the former is different. The basis of the system EXPERT consisted in extracting the meaningful concepts from the résumés and job offers. Then, to use them to automatically build two types of ontologies, one related to the applicants (one per résumé) and one other to the job offer. To achieve this, the system utilized several tools, e.g. POS tagging, named-entity recognition, and semantic parsing. In order to select and rank the candidates for a job offer, EXPERT applied an ontology mapping. This means, a method that measures how similar were the candidate and job offer ontologies. The candidates with the closest ontologies to the one of a job offer were the ones positioned as the best for the job.

Since the apparition of the social networks over the Internet, specially those related to the professional activities (e.g. LinkedIn and Research Gate), some researches have connected them to the e-Recruitment systems. An excellent example of this is the work of Faliagka et al. (2011, 2013). The authors created an e-Recruitment system relied on the use of LinkedIn profiles and blogs to find the best candidates for a job offer. To be more specific, the system instead of using résumés, utilized and extracted the information located on the LinkedIn profiles of each job seeker. Then, the data related to the skills, education and professional experience was compared with a job offer, posted in their system, through semantic matching. The semantic matching was done using by an Information Technologies taxonomy. At the same time, the system offered the possibility to analyze linguistically the blog of each job seeker to determine the candidate's personality. They used the tool LIWC (Pennebaker et al., 2001), to analyze the blogs and to obtain an extraversion score. This system implemented, as well, an applicants ranking. To do that, the authors asked to some recruiters to classify, by relevance with a job offer, a set of résumés. From the manual classification they created a model, through automatic learning, that would predict the recruiters' judgment for any job offer.

As it could be seen along these sections, the résumés are a basic component of the e-Recruitment systems. Therefore, we present in the next one, the researches done to comprehend the résumés in a HRM context.

2.2 Analysis of Résumés

Despite the fact that résumés contain the most important information of our life, many researchers have asked themselves whether some data is more influent to recruiters than other. Because, as it indicated by NACE (1995), the résumés are evaluated according the likes and dislikes of the recruiters. Thus, in the literature we can find some researches trying to understand which is the data can change the recruiters' opinion about an applicant.

One of the first researches in this field was the work of Hutchinson (1984). In his paper he presented a study with the purpose of comprehend better, but empirically, how the structure of résumés changed the recruiters' advice. To achieve this, he sent to 500 personal

administrators a brief questionnaire about what kind of information they were inclined to see in a résumé. More specifically, the author designed a survey to learn about the preferred content in résumés of college students entering to the job market. The results, obtained from the 200 returned questionnaires, shown that the most relevant information of a résumé was the one related to the education qualifications and work experience. While the less important one was the data related to the social data, e.g. salary requirements, references required and photograph.

The research carried out by Knouse (1994) consisted in analyze how certain sections or information of the résumés affect the possibility of a person to be hired. To realize this research, the author asked to 89 members of the chamber of commerce to read and analyze several résumés. Those documents had only 5 sections: job objective, education, work experience, references and, activities and interests. Each recruiter, after reading a résumé, was asked to fill a questionnaire. This survey had for objective to evaluate the recruiter's impressions about the applicant. The results shown that a relevant education and job experience produced positive perceptions to the recruiters. In consequence, these increased the probability to receive a positive recommendation. Another point analyzed by the research was the *impression management*, i.e. the control of the information to keep a positive image of oneself. In this case, the author found that some phrases, in certain sections, could impress positively to the recruiters and to increase the possibility to get a positive recommendation.

Another research about college students' résumés is the one done by Thoms et al. (1999). In their paper they tried to find which information influenced the recruiters in order to interview a college job seeker. The difference between this work and the one of Hutchinson (1984) is their method. Instead of sending a survey to some recruiters, the authors sent to HRM a set of false résumés with the characteristics of college students entering to the job market. A fake candidate was considered positively recommended when it was called for an interview. The experimental résumés varied in length, academical qualifications (coursework, GPA[9]), work experience and accomplishments. The results of this work shown that the length of résumés, objective statements, good GPA, relevant courseworks and accomplishments lists had a positive influence in the decision of a recruiter to do an interview.

Nemanick Jr and Clark (2002) evaluated how the extracurricular activities had an impact in the résumé evaluation. Their investigation consisted, more specifically, to measure how aspects, like leadership and number of extracurricular activities impacted the recruiters. One peculiarity of this research is that the recruiters were university students, from psychology and statistics, instead of HR experts. False résumés of senior accounting and marketing majors were used in the project to achieve the objective. The methodology used to evaluate the résumés were that each student had to select the top 4 documents from a set of 12. All the résumés were pre-screened by professional recruiters. As results, the authors indicated that the recruiters used the extracurricular activities to draw several ideas about the candidates, specially inferences between the job activities and applicants' attributes.

[9]Grade Point Average.

In the literature we can find, as well, the work of Cole et al. (2004). It studied how the recruiters draw personality inferences about applicants only through the reading of résumés. Put differently, to know whether the recruiters develop an inference, positive or negative, about candidates after reading the résumés. Moreover, whether the inferences could affect the following stages of the selection processes. To achieve this, the authors asked to 244 recruiters to analyze 122 résumés from business seniors enrolled in last semester courses at large universities. The methodology of the project consisted in a questionnaire, applied to all the recruiters, with subjects like position-hiring scenarios[10], candidates' personality[11] and hiring recommendation. Among the conclusions arrived by the authors that we can highlight, is the fact that for the recruiters the information about the academical major and the career objectives were important. As well, that the recruiters' hiring recommendations were based on two things. First, the job characteristics had to be fulfilled by the candidates. Second, the personality traits of an applicant, inferred by a recruiter, should be of aid for the job performance.

One last example of this kind of studies is the work of Chen et al. (2011). Their analysis studied the aesthetics of the documents, the recruiters' moods, the job-related knowledge and the physical attraction. The experiment consisted in the analysis of 277 Taiwanese résumés by 62 recruiters of 30 different firms. The authors found that a positive mood was significantly related to the hiring recommendations but not the physical attraction. There was no significant relation between academic qualifications and the job-related knowledge. In contrast, the GPA had a relation with the academic qualification. The perception of job-related knowledge, and consequently the hiring recommendations, were increased when résumés had an extensive work experience.

2.3 Conclusions

In the first part of the chapter we presented the different types of e-Recruitment systems developed along the last 15 years, from the searchers of documents to the rankers of résumés.

We saw in this part, that no matter the type of system, the most common approach consists in the measure of how similar or dissimilar is the available information but also in the use of knowledge resources. Furthermore, we noticed that in the last years more complex e-Recruitment systems have been developed, merging the different types of them into only one. The goal is to offer better and unique solutions to HRM and job seekers.

[10]These scenarios consisted in setting, hypothetically, an applicant in a specific job position from an organization with different jobs available.

[11]They measured two kinds of personality: extraversion and conscientiousness.

As said previously, the use of knowledge resources, like ontologies or taxonomies, has become a common factor in several e-Recruitment systems. In our opinion their application in the real life is complicated. The reason is that they tend to be domain-related, i.e. they only contain information of a small set of knowledge fields, due to the time necessary to build them. However, it should be noted that there have been developed methods to enrich faster e-Recruitment ontologies (Roche and Kodratoff, 2006; Looser et al., 2013).

It should be considered that, despite the existence of different methods to separate and annotate the résumés, job offers and similar documents, it is not a task that is completely finished. In several cases the task is linked to the use of ontologies or rules that are human-created and that, in consequence, cannot to deal totally to the varied, and growing, vocabulary used in the documents.

We could observe, as well, that most of the e-Recruitment systems that rank documents need the résumés and the job offers to do the comparison and make a match. From the analyzed systems, we found only one that made use of résumés previously classified, to rank newer ones. Nonetheless, in our opinion the inconvenient of the method used, that was based on automatic learning, is the consideration that recruiters judge equally the requirements of all the jobs, no matter how dissimilar could be. To our knowledge, even with similar jobs, the recruiters behave different as the needs of the employers can vary from one post to another. Thus, to learn from past applications may not be the most correct or performing method to rank résumés.

In the second part of this chapter, we introduced the studies that analyzed the effects of the résumés over the recruiters. In other words, how and what information from a résumé trigger a positive or negative reaction in the recruiters' decisions. From the presented studies, we can discern that information coming, for example, from professional experience and education, have an impact on the decisions taken by recruiters. Furthermore, we can highlight that recruiters make inferences from the applicants using the résumés information. Nevertheless, it may be taken into account that most of these studies were done over the résumés of (under)graduated students. Therefore, the results may change in résumés of more experienced job seekers.

Returning to the e-Recruitment systems, none of the researches found in the literature has analyzed whether a résumé ranker can work without job offers or past selection processes. In our consideration, to analyze only résumés is a possible method to use in a e-Recruitment ranking system. We saw that the résumés have a great quantity of information that it is used by humans to determine the future of the applicants in a selection process. Having say that, it is only necessary to know how to detect and use it with computers. Not forgetting, to verify the consequences and the performance of this kind of analysis.

To conclude, the résumés are an important element in the e-Recruitment systems. The information located in these documents have an impact over the recruiters' decisions but also

on the perception of the applicants. Despite the fact that the e-Recruitment systems have existed for 15 years and that the résumés are an important source of information for the selection process, few researches have study the possibility to use them all alone. Therefore, in the following chapters we will present a set of experiments done to determine whether it is possible to develop an e-Recruitment system using only the résumés, without job offers or past selection process.

Chapter 3

Processing Automatically a Human Resources Corpus

To develop and test an e-Recruitment system, it is, frequently, used sets of documents, or corpus. The corpora to train and test e-Recruitment systems can be classified into two types. The first type consists in the employ of real data, especially résumés, which are not linked to any recruitment or selection process, like it was done by Faliagka et al. (2013) and, Senthil Kumaran and Sankar (2013). In other words, the authors recollect documents over the Internet, like from on-line job boards or LinkedIn, to create training and/or test corpora. The second type corresponds in using real data but actually linked to a recruitment and/or selection process, as in Kessler et al. (2012). In spite of the importance of the corpora in an e-Recruitment system development and test, few authors, like Kessler et al. (2008), describe the complete procedure employed to process a human resources' corpus, even if it is not an evident task as we will show along this chapter.

In this thesis, we make use of the *Recruitment Corpus*, a large and heterogeneous collection of documents coming from real recruitment and selection processes. The processes were done by the staff of *Adoc Talent Management* between November 2008 and March 2014. More specifically, the Recruitment Corpus is a set of résumés, motivation and recommendation letters, diplomas, interview minutes, social networks invitations (e.g. LinkedIn, Twitter and Facebook) and other documents that were sent by job applicants or were created by the HRM during the recruitment and selection processes of a job.

The Recruitment Corpus is organized by jobs. These belong to different domains, e.g. engineering, business, health, economy, marketing and biology, and position levels, like technical or executive. All the jobs are divided into applications, i.e. the candidates' *dossiers* containing the documents used along the recruitment and selection processes. Each application is linked to a meta-data file which indicates the applicant's unique ID, the name of the job position, the creation date of each application in the server and the last selection phase reached by the applicant.

According to the information given by the recruiters, there are 5 sequential phases that an applicant can reach:

1. *Not Analyzed*. It is the first phase and consists in the arrival of the application to the server. Thus, all the persons who applied for a job pass through it. In some occasions, a certain quantity of applicants do not get through the following phases, the reason is their late arrival to the selection process which, in fact, is ending.

2. *Analyzed*. This phase corresponds to the decision of whether the applicant will be contacted. The pronouncement is done by the job recruiter(s) after reading the applicant's résumé.

3. *Contacted*. The third phase consists in approaching, by telephone or e-mail, the applicant that was chosen by the recruiters due to their features. In this phase, the recruiter can decide whether the supplementary information given by the applicant fulfilled the searched profile or not. As well, the applicants can decline their participation in the process.

4. *Interviewed*. It consists in interview the applicant that was contacted previously. The objective is to know better the applicant, through questions or basic activities, and to know how they cope with job situations. The applicants can pass through multiple interviews, for example with the recruiter, the job manager, future colleagues, etc. To this phase, only a reduced group of applicants arrive.

5. *Hired*. The last phase consists in hiring the candidate that successfully passed the previous phases. Nevertheless, not all the instances arrive at this point. In some cases, the recruiters or companies are not convinced about the final applicant(s), or these last ones decide to decline the job proposal.

Since this thesis does not have for objective to determine who has to be hired, other than identify who should be contacted in first place, we classified the selection phases into two classes: *Selected* and *Rejected*. The first class, corresponding to the phases *Contacted*, *Interviewed* or *Hired*, represents the candidates that were approached by a recruiter. The second class, contains the candidates that were *Not Analyzed* and *Analyzed*, in other words, the applicants that were not approached by the recruiters. These two classes will be used along all the chapters. To understand better the corpus, we present in Figure 3.1 the histogram of Selected candidates, measured by percentage, in the corpus. As it can be seen, not all the jobs contain the same proportion of Selected applicants, in fact the median is 40.94% and the mean $42.93\% \pm 1.44$.

The heterogeneity of the Recruitment Corpus not only resides on the different kinds of documents, located in the applicants' *dossiers*, but also on their language. As Adoc Talent Management is located in France, the main language of the Recruitment Corpus' applicantions is French, nonetheless, due to the globalization and the capacity of European Union citizens to work freely in France, it is possible to find some documents in languages like English, Spanish and German.

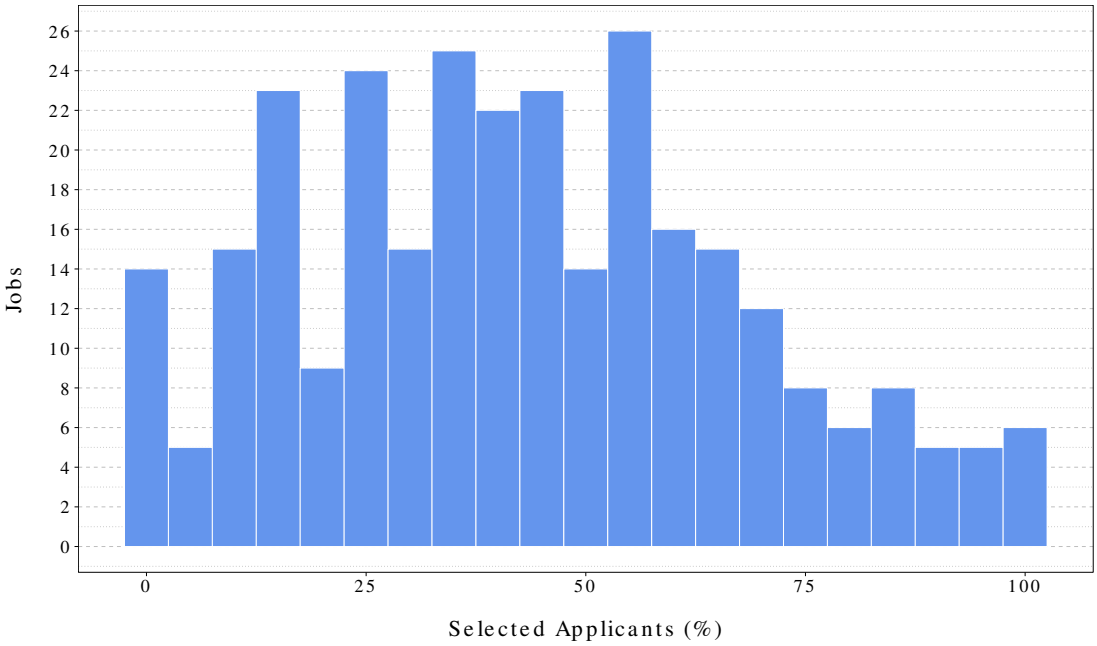

Figure 3.1 *– Percentage of Selected candidates by the number of job offers.*

The size of the Recruitment Corpus is not negligible as we can see in Table 3.1, where the number of job positions, job applicants and documents, without considering the language, are presented. Most of the e-Recruitment systems, presented in the state of the art, were created or evaluated with a small amount of data. For example, Kessler et al. (2012) used 1,917 applications divided into 12 job positions; Faliagka et al. (2013) collected 200 LikendIn IT-related profiles that were considered as résumés, from these 100 were used to train their algorithms, the rest was used on the evaluation considering 3 real job offers; Senthil Kumaran and Sankar (2013) assembled a corpus of 478 résumés found in the Internet. Actually, it is very difficult to find complete sets of data related to a recruitment and selection process or similar résumés at the Web. In some cases, it is due to legislative normative or the specificity of the data needed.

Table 3.1 *– Number of job offers, job applications and files in the Recruitment Corpus.*

Job positions	Job applicants	Files
296	29,368	47,388

Because of the Recruitment Corpus' characteristics, heterogeneity and size, a manual analysis to find the résumés was impossible. Thus, we present the methodology developed to process this collection of documents.

3.1 Objectives

The main objective of this chapter is to present the development of a methodology to process automatically the résumés from the applications located in the Recruitment Corpus. This automatic processing should consist in:

- Detecting French documents.
- Determining, from French documents, those that are résumés.
- Representing the résumés in an appropriate format for NLP tasks.

3.2 Methodology

In this section, we will present the different steps to process the résumés in the Recruitment Corpus.

3.2.1 Document Conversion

In order to apply for a job proposed by Adoc Talent Management, the applicants have to e-mail their résumés, motivation letters and other documents demanded by recruiters. Therefore, it should be noted, that all the files located in the Recruitment Corpus do not have always the same file format.

Of all the formats located in the Recruitment Corpus, in this thesis only four types of documents were of our interest: PDF (*.pdf*), Microsoft Word (*.doc* and *.docx*), OpenDocument Text (*.odt*) and Rich Format Text (*.rtf*). The reason was that these file formats are the most frequently used by the applicants to send their résumé and motivation letter. Moreover, they represent 80.52% (38, 161 files) of the Recruitment Corpus. The rest of the Recruitment Corpus, 19.48%, corresponds mainly to web and image files, like PNG (*.png*), JPEG (*.jpg* and *.jpeg*) and HTML files (*.html* and *.htm*). These last formats were used to send for example social networks invitations, diplomas and identity documents.

The documents with the chosen formats were converted into plain UTF-8 text using the following tools:

- *Calibre Ebook Management*[1] for files having a *.pdf*, *.docx*, *.odt* or *.rtf* extension. The accentuated letters of texts coming from PDF files were verified that they were correctly coded; see Cabrera-Diego et al. (2013) for a discussion.
- *Catdoc*[2] for files with a *.doc* extension.

[1] http://calibre-ebook.com/
[2] http://site.n.ml.org/info/catdoc/

3.2.2 Language Recognition

Although the majority of the recruitment and selection processes available in the Recruitment Corpus are for French job positions and in consequence, most documents are in French, it is possible to find documents in other languages, especially English, German and Spanish. This mixture of languages and the lack of information in the meta-data about the application language, drove us to apply a language recognition step.

We used the Google's *Compact-Language-Detector (CLD2)*[3] through its Perl module[4] to determine the language of the documents converted previously into plain text. The CLD2 is a free tool that makes use of probabilities and 4-grams of letters to predict the language of documents.[5]

3.2.3 Résumé Detection

In the Recruitment Corpus, each application is composed of different kinds of documents, for instance, a résumé, the last diploma and an interview minute. The constitution of the applications depends on the documents sent by the candidate and the last phase in which they arrived.

Despite the existence for every application of a meta-data file, this does not indicate the type of documents located in each application. Thus, it was impossible to know *a priori* which documents were contained in the applicant's *dossier*.

In order to determine automatically which document of the applications was a résumé and which was not, we developed a Résumé Detector. This tool is based on a Support Vector Machine algorithm (SVM) (Cortes and Vapnik, 1995) and it was built using the API LIBSVM (Chang and Lin, 2011).

A SVM, is a supervised machine learning method where data is represented through vectors mapped in high dimensional spaces (Cortes and Vapnik, 1995). This method consists in searching the function that generates the best hyperplane that allows the vectors to be classified into different classes (Cristianini and Shawe-Taylor, 2000). We will talk about the training and evaluation of the SVM created to detect the résumés in the following sections.

[3]https://code.google.com/p/cld2/
[4]"Lingua::Identify::CLD" https://github.com/ambs/Lingua-Identify-CLD
[5]Documentation available at: https://code.google.com/p/cld2/wiki

SVM Training

We created a training corpus through a manual classification of French résumés and other French documents (motivation letters, publication lists, spam and interview minutes). All the used documents, excepting the interview minutes, came from a collection of spontaneous applications, i.e. job applications that do not belong to the Recruitment Corpus. The interview minutes were sent by the recruiters and corresponded to processes done by Adoc Talent Management between june and december 2014. In total, the training corpus for the SVM was composed of 699 résumés and 635 other documents.

To process the training corpus, all the files were converted into plain UTF-8 text, using the tools described in Section 3.2.1. In each file we deleted the stop-words[6] and applied the French Porter's Algorithm. The Porter's Algortithm (Porter, 1980) is a stemming method that consists in truncating the words following rules in order to create the closest dictionary form. For example: the words in the French sentence *j'ai deux chats* (I have two cats) will be truncated into *j'ai deux chat* (I have two cat); the "s" from *chats* is deleted as this letter the at the end of a word means plural.

To build the SVM of the Résumé Detector, we followed the procedure proposed by Hsu et al. (2003) which consists in testing, in first place, two different SVM kernels, the linear and the radial one. The other kernels are only test if the results obtained by the linear and radial one are not performing enough. The two kernels used were tuned up through a grid-search and a five-fold cross-validation.

The tunning of the SVM consisted in searching the parameters that give the best performance according to a cross-validation. Being more specific, we divided into 5 equal parts the training corpus. Then it was used 4 of the 5 parts to train a SVM with a parameters set. The precision of the resulting SVM was obtained using the 5th part. The same process was done 4 times more, changing one of the 4 parts used to train the SVM. Once the evaluation was done for the 5 SVM having the same parameters, a new set of parameters was tested. The objective was to create a grid, where it could be determined the best combinations of parameters.

In the case of the linear kernel, it was necessary to determine only the parameter C, which controls the trade-off between the classification errors during the training process (Rychetsky, 2001, page 82). For the radial kernel, in addition to C, the parameter γ had to be set; the value of γ defines the influence radius of the samples during the training. During the grid-search, we started C with the value of 1 and we decreased and increased it by exponents of 10.[7] The value of γ started with the recommended value that is 1 divided by the numbers of features of the training corpus, in our case, words. We changed the value of γ by exponents of 10. As we started to get better results, the exponents were fixed and values

[6] List taken from the Perl's module "Lingua::StopWords".
[7] The value of C was always greater of zero.

of the parameters were increased or decreased.

After the use of the grid-search, we found that for both kernels the best $C = 1$. In the case of $\gamma = 1 \times 10^{-4}$.

Table 3.2 presents the results of the cross validation of the SVM using the best parameters found for the linear and radial kernel. These results are presented in terms of precision, recall and F-score.

Table 3.2 *– Evaluation, in terms of precision, recall and F-score, of the linear and radial SVM through a five-fold cross-validation*

Linear ($C = 1$)						
Subcorpora	L_1	L_2	L_3	L_4	L_5	Mean[8]
Precision	0.972	1.00	0.950	0.971	0.971	0.979
Recall	0.986	0.971	0.971	0.992	0.992	0.982
F-score	0.979	**0.985**	0.960	**0.982**	**0.978**	**0.977**
Radial ($C = 1; \gamma = 1 \times 10^{-4}$)						
Subcorpora	L_1	L_2	L_3	L_4	L_5	Mean[9]
Precision	0.979	0.963	0.964	0.951	0.932	0.952
Recall	0.986	0.949	0.971	0.985	0.992	0.977
F-score	**0.982**	0.956	**0.967**	0.968	0.961	0.964

As seen in Table 3.2, the best results were those obtained with the linear kernel; its average F-score was 0.977 which is slightly better than the F-score of radial kernel (0.964). This performance was the expected one because the number of features, in this case words, is much greater than the 2 possible classes (Résumés and Other documents) and, in these cases, the linear kernel has always a better performance than the radial one (Hsu et al., 2003).

It is of interest to indicate that the performance of our SVM is similar to the one developed by Kessler et al. (2008).[10] They obtained an average F-score of 0.980 during the cross-validation using a linear kernel.

The final version of the Résumé Detector was implemented using a SVM with a linear kernel, $C = 1$ and the complete training corpus.

[8] Mean F-score is obtained from the average precision and recall.
[9] Idem.
[10] We could not make use of the tool developed by them as it is not available to the public.

SVM Evaluation

To estimate the Résumé Detector performance we created an evaluation corpus which consisted in a multilingual and heterogeneous set of 240 documents (résumés, motivation letters, publications lists, diplomas, etc.), divided into 4 groups of 60 documents:

- French résumés
- Résumés in other languages
- Other French documents
- Other documents in other languages

The files for the evaluation corpus were collected randomly from the Recruitment Corpus and were classified into the 4 possible groups by a non-expert recruiter.

As the creator of the evaluation corpus was a non-expert recruiter, we asked, separately, to two expert recruiters to classify the files from the evaluation corpus into the same 4 groups.

To determine whether the behavior of both recruiters was similar, we measured the agreement of both expert recruiters. Their accord was calculated with *Cohen's Kappa* and *Kendall's W*. For the former, the experts recruiters obtained a $\varkappa = 93\% \pm 0.04$, which according to the literature can be interpreted as an excellent agreement (Altman, 1990, page 404; Landis and Koch, 1977). For the latter agreement measure, the Kendall's W, the recruiters obtained a $W = 0.905$ with a *p-value* $= 2.58 \times 10^{-13}$; according to Schmidt (1997) a value greater than 0.90 is "unusually strong agreement"[11]. We can see, in consequence, that both recruiters have a similar definition of what it is a résumé.

In order to use the evaluation corpus, all its documents passed through the same preprocessing steps used for the training corpus (Section 3.2.3). As the SVM was designed for French and not all the files from the evaluation corpus were in French, a language recognition step (Section 3.2.2) was done.

To measure the Résumé Detector's performance, we used the standard formulas of precision, recall and F-score. Table 3.3 shows the confusion matrix according to the evaluation corpus annotated by the Recruiter 1. Table 3.4 shows the confusion matrix according to the evaluation corpus annotated by the Recruiter 2. Table 3.5 shows the results of this evaluation in terms of precision, recall and F-score; a mean for each measures is presented as well.

As it can be seen in Table 3.5, the Résumé Detector's performance is better considering the evaluation corpus annotated by the Recruiter 2. This means that the *definition* of résumé calculated by the SVM, during the training step, is more similar to the one of the Recruiter 2 than with the one of the Recruiter 1. Despite these differences in the results the Résumé

[11] The word unusually is used in its sense of remarkable or interesting.

Table 3.3 – *Confusion matrix of the Résumé Detector evaluation according the evaluation corpus annotated by Recruiter 1.*

		Résumé Detector		
		Résumés	Other Documents	Total
Recruiter 1	Résumé	55	2	57
	Other Documents	5	62	67
	Total	60	64	124

Table 3.4 – *Confusion matrix of the Résumé Detector evaluation according the evaluation corpus annotated by Recruiter 2.*

		Résumé Detector		
		Résumés	Other Documents	Total
Recruiter 2	Résumé	56	1	57
	Other Documents	2	65	67
	Total	58	66	124

Detector reaches a good performance over the evaluation corpus with an average F-score of 0.956.

A deep analysis of the documents where the Résumé Detector and the recruiters did not agree revealed that the problematic cases corresponded to bilingual résumés and motivation letters attached to short résumé (contact information, last job and education, languages and some skills). In the case of the bilingual résumés, the problem has its origin in Language Recognition step, that considered French as the main document language while Recruiters did not. With respect to the motivation letters attached to short documents, we saw that either the two recruiters had a clear definition whether those document were résumés.

3.2.4 Résumé Uniqueness

We found that, in the Recruitment Corpus, there were candidates which have more than one résumé for the same job. This happened either because the applicant attached several résumés in one same application or because they sent multiple applications for the same

Table 3.5 – *Evaluation of the Résumé Detector in terms of precision, recall and F-score.*

Corpus	Precision	Recall	F-score
Recruiter 1	0.964	0.916	0.939
Recruiter 2	**0.982**	**0.965**	**0.973**
Mean	0.973	0.940	0.956

job. In both cases these résumés may or not be exactly the same; occasionally, the applicants change somewhat their résumé to fit better the searched job profile.

To avoid false or biased results from this kind of cases, we validated the résumé uniqueness in each job offer. The validation was done using 3 tests over all the possible couples of résumés in a job offer.[12] These tests were created in order to detect the résumés that could be the same or with small modifications:

- One résumé by candidate: both résumés had to come from two different applicants. In other words, inside the applicant's *dossier*, could not exist more than one résumé.
- Résumés with different content: the Linux's tool *diff*[13] was used to determine whether both résumés were identical.[14] Finding equal documents would mean that they came from the same user.
- Not equal e-mails: All the e-mails addresses located inside the couple of résumés had to be different.[15] The objective of this test was to find the résumés that were from the same applicant but that were changed somewhat and would be hard to detect them with the previous tests.

Each couple not surpassing at least one of the tests was considered as problematic. In this case, we set two possible solutions:

1. If the couple of résumés belonged to the same class, Selected or Rejected, the résumé that arrived firstly to the server, according to the meta-data, was the deleted one from the corpus. The reason is that we considered that the most recent résumés was the one expected, for the applicants, to be read by the recruiters.
2. If both résumés belonged to different classes, i.e. one Selected and another Rejected, we deleted from the corpus the résumé marked as Rejected. We decided to give a priority to the résumés of Selected candidates because this happens when two *dossiers* were sent by a same applicant. When one of the duplicated applications was read and marked as Selected by a recruiter the second one, in consequence, was classed as Rejected after reading.

3.2.5 Data Representation

To represent the résumés of the Recruitment Corpus we made use of the Vector Space Model or VSM (Salton et al., 1975). The VSM is a paradigm where documents are represented as vectors in a vectorial space (Manning et al., 2008). Each vector consists in the weights, like

[12] The number of possible couples for a certain job offer is given by all the possible combinations of two résumés (C_2^n) taken from the total number of résumés (n).

[13] http://www.gnu.org/software/diffutils/

[14] The tool *diff* was configured to ignore the multiple white spaces and empty lines but to be also case-insensitive.

[15] The e-mails were detected using a regular expression.

absolute frequency, of the expressions (words, *n*-grams or phrases) in which a document can be indexed (Moens, 2006). Put differently, a vector in the VSM is the numerical representation of the documents according to a the weights of their expressions.

For the vectors of the VSM, we chose to use 4 types of words *n*-grams: unigrams, bigrams, SU3 and SU4. These last 2 *n*-grams are called skip bigrams (Huang et al., 1992; Lin, 2004) and we decided to use them as a way to increase the coverage of the document structure without increasing the size of the data (Huang et al., 1992). The SU3 is a bigram where there is a gap of one word between two words; the SU4 is a bigram where the gap has the size of two words. For example, from the phrase "je travaille ici depuis cinq ans" we can obtain four SU3 (*je ici*; *travaille depuis*; *ici cinq* and *depuis ans*) and three SU4 (*je depuis*; *travaille cinq*; *ici ans*).

To reduce the possible noise coming from the text, the size of the Vector Space Model and the curse of dimensionality (Bellman, 1961), all the résumés passed through a text normalization. First, we lowercased the text of all documents. Then, we removed all the punctuation marks, numbers and stop-words[16] from each document. Finally, we reduced the documents' lexicon through the lemmatization of the words using the tool Freeling 3.1 (Padró and Stanilovsky, 2012). The lemmatization is the action of replacing the words of a text to their dictionary form, or lemma, using lexicons and a morphological analysis (Moens, 2006; Manning et al., 2008); e.g. the phrase in French *Ce sont de nouvelles tasses* (They are new mugs) is lemmatized to *Ce être de nouveau tasse* (It be new mug).

Although for the detection of résumés we made use of a stemming method to reduce the documents' lexicon richness, we decided that for the experimentation the lemmatization would be a better method. In languages like French, Italian or Spanish, the use of stemming is a fast method to reduce the vocabulary, but not the best. The reason is that these languages are highly inflected, i.e. the words change greatly in order to express different grammatical categories, like tense, gender and number. For example, *entrée* in French can mean entrance, but also the past participle of the 2nd person in feminine of the verb to enter. In both cases, the stemming would be *entré*, something which does not correspond to their respective lemmas, *entrée* and *entrer*. Therefore, rules based on the replacement of suffixes are not enough to obtain correctly the dictionary form of an inflected language word.

3.3 Results and Discussion

From 38, 161 files passed through the document conversion tools, 64 were not converted correctly according to the tools' logs due to errors. We analyzed these 64 cases and we found that the reasons of the errors were mainly corrupted files or a wrong extensions. Therefore,

[16] List taken from the Perl's module "Lingua::StopWords".

from the Recruitment Corpus, we obtained 38,097 files in plain text.

At the language recognition step, we found that from 38,097 files, 32,845 were French documents. This number represents 69.31% of the total of the Recruitment Corpus and 86.21% of the converted file formats.

The Résumé Detector was applied over the documents of the corpus that were detected previously in French (32,845). From these files, our tool detected 22,439 documents as French résumés. This number represents 47.35% of the total corpus and 68.31% of the documents converted into text.

Finally, from the step of résumé uniqueness, we found that 2,854 résumés of 22,439, were not unique. These were deleted.

In total, the Recruitment Corpus is composed of 19,585 French résumés; these were converted into vectors to use a VSM in the following chapters. In Table 3.6 we present the results of these four preprocessing steps.

Table 3.6 – *Number of résumés used during the processing of the Recruitment Corpus*

Preprocessing step	Files			Difference Reason
	Input	Output	Difference	
Document Conversion	38,161	38,097	64	Could not being converted
Language Recognition	38,097	32,845	5,252	Not in French
Résumé Detector	32,845	22,439	10,406	Were not detected as résumés
Résumé Uniqueness	22,439	19,585	2,854	Were not unique

3.4 Conclusions

In order to create an e-Recruitment system, especially those dedicated to rank job applicants, it is necessary to have a corpus to develop and/or test the system. Although, the processing of the corpus is not a obvious task, few authors describe the necessary steps to do it.

At this chapter, we presented the Recruitment Corpus, a heterogeneous collection of HRM documents created during the recruitment and selection processes done by Adoc Talent Management. As well, we presented the tasks done by us in order to retrieve and treat French résumés. These tasks consisted in the conversion of the documents into plain text, the language detection, the résumé detection and the résumé uniqueness.

In the case of the résumé detection, we developed our own tool using a Support Vector Machine (SVM), a machine learning algorithm. The evaluation done to this tool showed a very good performance and equivalent to the one developed by Kessler et al. (2008).

We also presented at this chapter the methodology used to represent French résumés. In the following chapters we present the experiments done with the set of French résumés.

Chapter 4

Textual Proximity Measures Applied to Résumés

A *résumé* is a personal document sent by a job seeker in order to apply for a job (American Heritage Dictionary of the English Language, 2014). It contains a range of information that goes from the work experience and academic qualifications to the extracurricular activities and skills (Nemanick Jr and Clark, 2002; Cole et al., 2003, 2007, 2009; Tsai et al., 2011). The résumés are written to attract the recruiters and to show that the applicants' capacities and abilities are linked to the job specifications (Huggins, 1977).

It is well known that résumés are an important component during the selection process, as we indicated in Chapter 1. The reason is that résumés have the capacity to change the recruiters' perception about an applicant (Thoms et al., 1999). This is done through the prediction of the applicant's behavior (Mael, 1991; Owens and Schoenfeldt, 1979) and future job performance (Hunter and Hunter, 1984; Owens, 1976).

For the e-Recrutiment systems the résumés are an important component but the job offers too. The reason is that résumés represent the applicants and the job offers the recruiters or employers. They contain different kinds of information and points of view that can be used to create systems to rank applicants according to a job offer, e.g. Kessler et al. (2012), Singh et al. (2010) and Senthil Kumaran and Sankar (2013).

While it can be though, *a priori*, that job offers are a basic element for e-Recruitment system, we consider that it might not be the case. In the literature, we can find at least one e-Recruitment system that make use of only résumés to do a ranking of them (Faliagka et al., 2013). To achieve its goal, the authors asked to a group of recruiters to analyze manually a set of résumés. The analysis consisted in finding the most relevant résumés of previous selection processes. Then, the results from this manual analysis were utilized to create a model through a supervised machine learning algorithm. The evaluation of the system was done through 100 LinkedIn profiles searched randomly. A recruiter analyzed the profiles, as

if they were résumés and submitted to 3 different IT jobs.

Although the results obtained by Faliagka et al. (2013) show a correlation between the rankings done by the machine learning algorithms and by a human, in our opinion, it is difficult to apply this kind of methods in the real life. The reason is that the authors consider that recruiters always judge the same information and in the same way. To our knowledge, this is not the case. The recruiters analyze the applications according to the needs of the employer and, in occasions, they take into account differently the information. Nonetheless, the results of this study show us that the résumés could be used alone in a e-Recruitment system.

In consequence, we chose to study the creation of a new method based only on the analysis of résumés, but without having to use previous selection processes. The reason is the fact that studies in psychology have already probed that the résumés have enough data in order to know whether an applicant has the possibility of being hired (Tsai et al., 2011). We have only to check whether and how computers should be utilized to detect this information.

For our methodology, we decided to use proximity measures as most e-Recruitment systems do. However, before the selection of one in particular, we decided to analyze whether and how a set of these measures should be used in a HRM context. The reason is that to our knowledge, in the literature, none of the developed systems or studies about the résumés have verified these questions.

The chapter is divided in 5 sections. We start with the objectives of this chapter (Section 4.1). In Section 4.2, we talk about the methodology used for the analysis. The results and the discussion of them are presented in Section 4.3 and Section 4.4, respectively. We end with the conclusions in Section 4.5.

4.1 Objectives

For this chapter, we have defined two objectives:

- Determine whether proximity measures can be used only over résumés to infer part of the selection process.
- Analyze how should be employed the proximity measures over résumés to get their best performance in a HRM context.

4.2 Methodology

We can consider, that in each selection process there are two types of applicants: *Selected*, i.e. those that were approached by a recruiter, and *Rejected*, candidates that were not contacted by a recruiter. The type of each applicant is determined by the recruiters after reading the respective résumé and according to several things, for instance, the job fit[1] (Cole et al., 2004) and job-related knowledge (Chen et al., 2011). It can be inferred that, in consequence, each type of applicant may share a particular set of characteristics.

Furthermore, from this inference, it can be considered that, in average, Selected résumés share more characteristics with themselves than and with respect to the Rejected ones. This can be founded on the idea that a résumé is selected due to its similarity with the job requirements, while the cause to reject one is its dissimilarity with the researched features. At the end, a set of common characteristics will always be smaller than a group of dissimilar ones.

This inference excludes the idea that in some occasions during the selection process the recruiters find *raræ aves*, i.e. candidates that may not match the searched profile though approaching them could be a window of opportunity. However, our inference is focused more on the general behavior than in these atypical cases.

In the following section we present the model used to represent and test our inference.

4.2.1 Model

Let us consider J a job composed of n résumés, $J = \{r_1, r_2, \ldots, r_n\}$. Each résumé of J belongs to a class, either Selected ($S : S \subseteq J$) or Rejected ($R : R \subseteq J$). As well, let us take into account σ, the *proximity* of two documents, i.e. a measure that indicates how similar or dissimilar are two files.[2] Thus, the set of proximities between résumés can be set according to the class of its elements:

- Between only Selected candidates, $\sigma(r_x, r_y : r_x \wedge r_y \in S, r_x \neq r_y)$.
- Between only Rejected applicants, $\sigma(r_x r_y : r_x \wedge r_y \in R, r_x \neq r_y)$.
- Between Selected and Rejected candidates, $\sigma(r_x, r_y : r_x \in S, r_y \in R, r_x \neq r_y)$.

In Figure 4.1 we present a visual example of the calculation of σ between 3 selected résumés and 3 rejected ones.

[1] This means how well an applicant fits in a certain position.

[2] In order to simplify the inference, we will bear in mind, that σ is a symmetrical measure with an interval between 0 and 1. Where 0 means completely similar and 1 totally different. In the model application, the interval of σ and its interpretation can change according to the measure used.

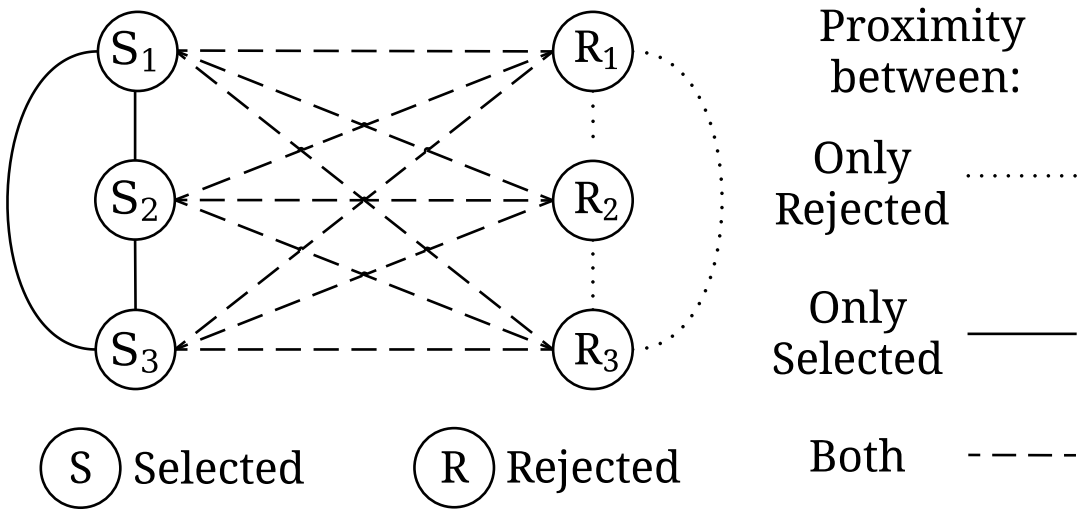

Figure 4.1 *– Example of the possibles proximity calculation between 3 Selected résumés and 3 Rejected ones.*

Then, let us consider our inference "in average, selected résumés share more characteristics with themselves than and with respect to the rejected ones" and split it into three parts to create their respective models. The first part, that concerns the fact that Selected résumé share more characteristics between themselves, can be modeled as Equation 4.1:

$$\overline{\sigma}(J_S, J_S) = \frac{1}{|S|} \sum \sigma(r_x, r_y : r_x \wedge r_y \in S, r_x \neq r_y) \tag{4.1}$$

where $\overline{\sigma}$ is the average proximity, J_S is the subset of J that contains all the Selected résumés, and $|S|$ is the number of Selected résumés.

The characteristics between only Rejected résumés can be considered as in Equation 4.2:

$$\overline{\sigma}(J_R, J_R) = \frac{1}{|R|} \sum \sigma(r_x, r_y : r_x \wedge r_y \in R, r_x \neq r_y) \tag{4.2}$$

where $\overline{\sigma}$ is the average proximity, J_R is the subset of J that contains all the Rejected résumés, and $|R|$ is the number of Rejected résumés.

The characteristics between Selected and Rejected résumés can be modeled as Equation 4.3:

$$\overline{\sigma}(J_S, J_R) = \frac{1}{|S| * |R|} \sum \sigma(r_x, r_y : r_x \in S, r_y \in R, r_x \neq r_y) \tag{4.3}$$

where $\overline{\sigma}$ is the average proximity, J_R is the subset of J that contains all the Rejected résumés, J_S is the subset of J that contains all the Selected résumés, $|S|$ is the number of Selected résumés and $|R|$ is the number of Rejected résumés.

Thus, the complete inference can be modeled as Equation 4.4:

$$\overline{\sigma}(J_S, J_S) > \overline{\sigma}(J_R, J_R) + \overline{\sigma}(J_S, J_R) \tag{4.4}$$

where $\overline{\sigma}(J_S, J_S)$ is the average proximity between only the Selected résumés, $\overline{\sigma}(J_R, J_R)$ the average proximity between only the Relected résumés and $\overline{\sigma}(J_S, J_R)$ the average proximity between both types of résumés. Equation 4.4 express that the average proximity of Selected résumés is greater than the sum of the average proximity of the Rejected résumés and the average proximity between the Selected and Rejected résumés.

This model was the one used along this chapter to test our inference. In the next section we discuss the method used to calculate the proximity of the résumés.

4.2.2 Calculation of the Résumé Proximity

According to the literature, a proximity measure is the one that determines how similar or dissimilar are two elements (Green and Rao, 1969). Therefore, there are different ways to calculate the proximity. For this chapter, we decided to use similarity and distance measures to determine the proximity of résumés.[3] The objective is to know how the use of them affects our inference.

We selected 3 similarity measures which will be defined by taking into account the existence of two vectors $X = \{x_1, x_2, \ldots, x_i\}$ and $Y = \{y_1, y_2, \ldots, y_j\}$. The similarity measures are:

- *Cosine Similarity*

$$\frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} \tag{4.5}$$

- *Jaccard's Index*

$$\frac{|X \cap Y|}{|X \cup Y|} \tag{4.6}$$

- *Dice's Coefficient*

$$2\frac{|X \cap Y|}{|X| + |Y|} \tag{4.7}$$

As it can be seen in Equation 4.6 and Equation 4.7, the measures of Jaccard's Index and Dice's Coefficient respectively, are expressed in terms of sets instead of vectors. In order to express them in terms of vectors, we needed to redefine each part of their equations. This redefinition allowed us, as well, to settle a non-binary version of the Jaccard's Index and Dice's Coefficient. In other words, a variant where was not taken into account the presence

[3] In the case of distance measures, these determine the dissimilarity between two elements.

or absence elements but the weigh of each vector component.

Each part of the Equation 4.6 and Equation 4.7 has been redefined as:

$$|X \cap Y| = \sum \begin{cases} min(x^\beta, y^\beta) & \text{if } x \neq 0 \wedge y \neq 0 \\ 0 & \text{if } x = 0 \vee y = 0 \end{cases} \tag{4.8}$$

$$|X \cup Y| = \sum \begin{cases} max(x^\beta, y^\beta) & \text{if } x \neq 0 \vee y \neq 0 \\ 0 & \text{if } x = 0 \wedge y = 0 \end{cases} \tag{4.9}$$

$$|X| + |Y| = \sum \begin{Bmatrix} x^\beta & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{Bmatrix} + \begin{Bmatrix} y^\beta & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{Bmatrix} \tag{4.10}$$

where min represents the function that returns the smaller element, max is the function returning the largest element and β, the exponent that allowed us to generalize the binary and non-binary version of the measures. If $\beta = 0$ then the calculation will be done using a binary model but if $\beta = 1$ the measurement will be the one which takes into account the vector component's weight.

For the distance measures, we have selected two of them which are defined, as well, by taking into account the existence of two vectors $X = \{x_1, x_2, \ldots, x_i\}$ and $Y = \{y_1, y_2, \ldots, y_j\}$:

- *Euclidean Distance*

$$\sqrt{\sum |x - y|^2} \tag{4.11}$$

- *Manhattan Distance*

$$\sum |x - y| \tag{4.12}$$

We decided to choose these 5 proximity measures as they are the most frequently used ones in NLP tasks.

All these measures were applied by couple of résumés. The number of résumés couples in a job was given by the combination C_2^n, where n is the number of résumés in a job. In addition, to accelerate the calculation of the résumé proximity, this task was parallelized using GNU Parallel (Tange, 2011). The GNU Parallel is a shell tool created to run multiple times the same task but with different inputs.

Vector Weight

Besides the 5 different measures of proximity, we decided to use as well 3 different vector weights. These were: *absolute frequency*, *relative frequency* and *Term Frequency Inverse Document Frequency (TF-IDF)*.

The Term Frequency Inverse Document Frequency is the joint of two factors. The first one is the *Term Frequency (TF)* which is a coefficient obtained from the relative frequency of each expression (n-grams, phrases, words, etc.) in a document. It is based on the idea that the most frequent expressions have a certain relation to the document content (Salton and McGill, 1986). The second coefficient is the *Inverse Document Frequency (IDF)* (Spärck-Jones, 1972), which takes into account the inverse proportion of the number of documents containing a construction regarding the size of the corpus. The idea of the IDF is funded that frequent expressions that do not appear in all the documents should be considered more important than those that exist in most of the corpus (Salton and Yang, 1973).

The TF-IDF formula used in this chapter is presented in Equation 4.13:

$$\text{TF-IDF}(n, d_x) = \frac{C(n, d_x)}{|d_x|} \cdot \log_{10} \frac{|D|}{|d \in D : n \in d|} \tag{4.13}$$

where n is an n-gram, d_x is the x document of the corpus D, $C(n, d_x)$ is the number of occurrences of n in d_x, $|d_x|$ is the size in terms of n-grams of d_x and $|D|$ is the total number of documents in the corpus D. The TF-IDF was calculated with respect to the résumés of each job.

We decided to do the calculation of the proximity using four n-gram vectors (unigrams, bigrams, SU3 and SU4) with 3 different weights (absolute frequency, relative frequency and TF-IDF). Nevertheless, in the case of Cosine Similarity, we only used the absolute frequency and TF-IDF vectors. The reason is that the results would be the same using relative frequency or absolute frequency (Manning et al., 2008, page 111).

Merging the Proximity

The proximity between two résumés was calculated using vectors of 4 different n-grams representations: unigrams, bigrams, SU3 and SU4. Therefore, four values of proximity were obtained by each couple of résumés. In order to have only one value of proximity by measure and vector weight, we decided to merge these 4 values into one.

We chose as merging method the average of the 4 proximity values. Although this merge method is quite *naïve* and simple, our purpose was to follow an *a fortiori* principle. If this combination method leaded us to good results, the use of more sophisticated merging methods, like a weighted arithmetic mean, would lead to better results.

4.2.3 Statistical Test

We tested our inference through the statistical test of the *Analysis of Variance*, also known as *ANOVA*. An ANOVA is a test based on the analysis of the data variability and the null hy-

pothesis that the means of two or more groups are equal. Put differently, this statistical test is used to determine whether the means of several groups are statistically different (Howell, 2013, page 397).

In this chapter, we used a parametric two-tailed ANOVA for independent groups to test whether $\overline{\sigma}(J_S, J_S)$ and $\overline{\sigma}(J_R, J_R) + \overline{\sigma}(J_S, J_R)$ were statistically different. The ANOVA is parametric as each group, had to fulfill two conditions: have a normal distribution and homogeneous variances. It is two-tailed[4] as we want to know if the means of both groups are different or equal. In other words, the alternative hypothesis of the ANOVA is $\overline{\sigma}(J_S, J_S) \neq \overline{\sigma}(J_R, J_R) + \overline{\sigma}(J_S, J_R)$. The ANOVA is for independent groups as $\overline{\sigma}(J_S, J_S)$ and $\overline{\sigma}(J_R, J_R) + \overline{\sigma}(J_S, J_R)$ do not share the same elements, i.e. one proximity value only belongs to one group. It is important to note that in order to use an ANOVA an *alpha* must be set, which is a value that indicates at what point we will consider refuted the ANOVA's null hypothesis; in our case we defined the $\alpha = 0.05$.

It is common in the statistical data to have values with an atypical magnitude respecting the total dataset, these values are what so-called *outliers* (Mason et al., 2003, page 70). Therefore, before the application of the statistical test, we suppressed the outliers of both groups. We defined an outlier as the value that is 1.5 times the interquartile range ($IQR = Q_3 - Q_1$) below the first quartile (Q_1) or above the third quartile (Q_3) (Montgomery and Runger, 2010, page 208).

For testing the normal distribution of the groups, we used a Shapiro-Wilk Test (Shapiro and Wilk, 1965) with an $\alpha = 0.05$. As this test can only be used in groups that contain between 3 and 5,000 elements (Royston, 1995), the groups having less than 3 proximity values were not considered Gaussian. The groups with more than 5,000 elements were considered as normal even if they violate the normality assumption. However, as stated by Howell (2013, page 424), the ANOVA is a robust statistical test where the normality assumption can be discarded with minor effects.

With respect to test of the variance homogeneity, we made use of the Bartlett's Test (Bartlett, 1937) using an $\alpha = 0.05$. In case of a resulting heterogeneous variance, the ANOVA was done only if the biggest variance of a group was not greater than 4 times the smallest one (Howell, 2013, page 354).

The jobs with at least one group not surpassing the ANOVA's parameters, were considered as not analyzable.

[4] An ANOVA is one-tailed if the alternative hypothesis only considers that the mean of one group is lesser than the other one.

4.2.4 Data

The experimentation, in this chapter, was done over the vectorized résumés of 224 jobs belonging to the Recruitment Corpus (Chapter 3). The résumés of each job were divided into two groups, Selected and Rejected, according to the meta-data file linked to each applicant. We did the analysis of the groups job by job, i.e. we tested our experimental hypothesis 224 times, using only the résumés that belonged to each job.

Although the Recruitment Corpus is composed of 296 jobs, not all of them could be used in our experiments. We found that there were 63 jobs where at least one of the classes, selected or rejected, did not exist. In other words, there was not at least two French résumé, in one or both groups, to apply a proximity measure. These 63 cases are represented in Figure 4.2, where it can be seen that in some jobs there were only Selected or Rejected applicants. In addition to these 63 jobs, we found 9 more jobs where the Selected group did not have at least three proximity measures.[5] Therefore, we took out from the experimentation these 72 jobs.

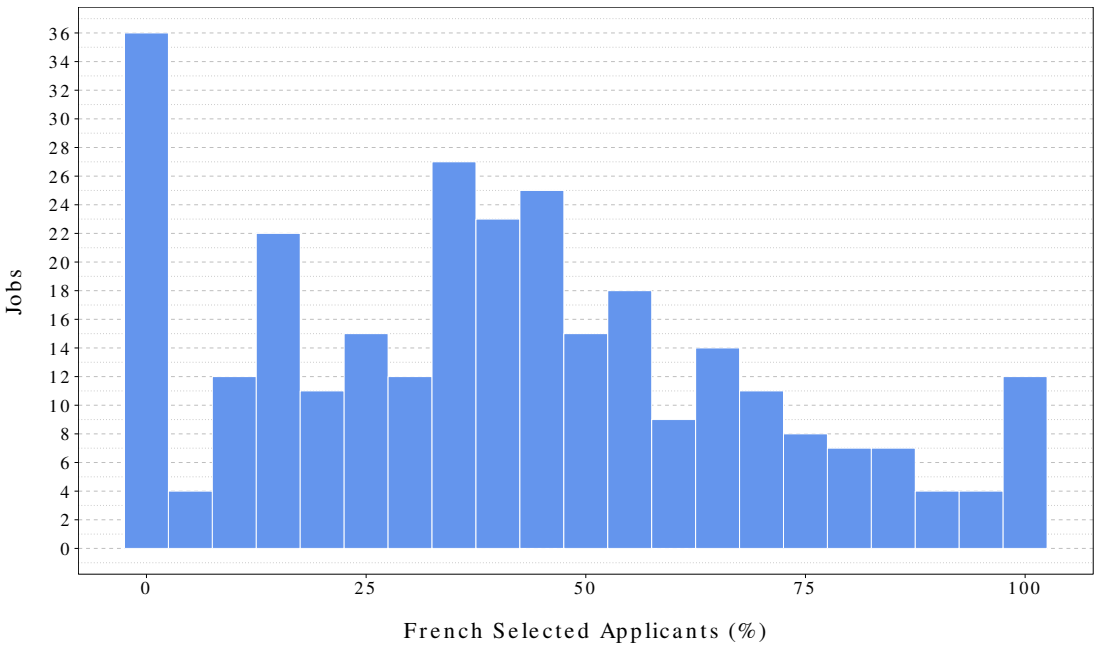

Figure 4.2 *– Percentage of French Selected candidates by the number of job offers.*

[5]We decided to remove from the test these jobs due to two reasons. The first one is that we could not determine the normality of the group. The second one is that the impossibility to use the ANOVA is not related to the measures or vector weight application but to the data.

4.3 Results

We present in table 4.1 the results of the ANOVAs done for each proximity measure according to the different vector weights used. For each result it is shown, in the first column, the number of jobs that refuted the null hypothesis of the ANOVA, i.e. the means of both groups were statistically different. The second column indicates the number of jobs where the average proximity of both groups was statistically equal. In the third column, we present the number of jobs that did not have the conditions necessary to do an ANOVA. The total number of analyzable jobs in the corpus was 224.

Table 4.1 – *Results of the Analysis of Variance (ANOVA) for each proximity measure.*

Measure	Weight	Statistically		Not Analyzable
		Different	Equal	
Cosine Similarity	Frequencies	**181**	**38**	**5**
	TF-IDF	171	48	5
Binary Jaccard's Index	-	168	54	2
Binary Dice's Coefficient	-	168	53	3
Non-binary Jaccard's Index	Abs. Frequency	178	41	5
	Rel. Frequency	**181**	**38**	**5**
	TF-IDF	165	57	2
Non-binary Dice's Coefficient	Abs. Frequency	177	41	6
	Rel. Frequency	**181**	**38**	**5**
	TF-IDF	164	58	2
Manhattan Distance	Abs. Frequency	140	66	18
	Rel. Frequency	**181**	**38**	**5**
	TF-IDF	137	68	19
Euclidean Distance	Abs. Frequency	144	61	19
	Rel. Frequency	151	68	5
	TF-IDF	144	62	18

It can be seen in Table 4.1 that the number of jobs which were not analyzable and statistically different varied according to the measure and vector weight. The use of distance measures (Euclidean and Manhattan) with absolute frequency and TF-IDF gave the fewest number of jobs where the average proximity is statistically different. When we made use of these measure, the number of not analyzable jobs increased too. It should be taken into account that for each result, there is a 5%, in total, of false concluding either for statistically different or equal result. This is according to the $\alpha = 0.05$ used for the ANOVA tests.

In order to know how large was the difference between the groups which were statistically different, we measured the effect size. More specifically, we used the *omega-squared (ω^2)*, a low biased measure for the effect size (Howell, 2013, page 425).

In Figure 4.3 we show the boxplots[6], in natural logarithm scale, of the effect size for each measure according to the different vector weights.

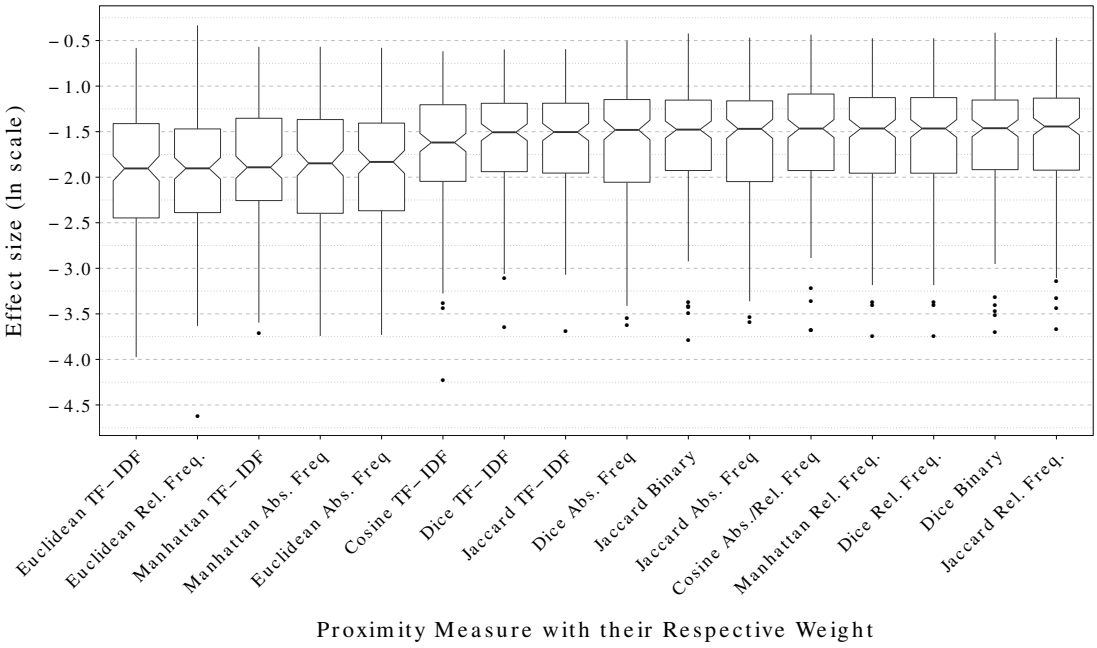

Figure 4.3 *– ANOVAs' Effect Size for each measure in natural logarithm scale. Only the jobs with statistically different groups were taken into account.*

As stated by Kirk (1996), the effect size of the omega-squared can be classified in three levels: low ($\omega^2 = -4.605$), medium ($\omega^2 = -2.830$) and high ($\omega^2 = -1.980$).[7] The larger the effect, the greater the difference between the groups of each job. In our case, in terms of the median, the largest effect size is the one given by Jaccard's Index with relative frequency (-1.443), while the shortest is the one of Euclidean Distance with TF-IDF (-1.903). This means that for all the measures, the difference between groups, according to the median, have a high effect size ($\omega^2 > -1.980$).

Having presented the results of the ANOVA, in Table 4.2, we present by measure and vector weight, the number of jobs, with statistically different groups, that had a positive or negative difference. In other words, those jobs with groups statistically different, where $\overline{\sigma}(J_S, J_S) - [\overline{\sigma}(J_R, J_R) + \overline{\sigma}(J_S, J_R)]$ was positive or negative.[8] The results are organized by measure and vector weight.

[6]See the Annex A which talks about the boxplots.

[7]These values are in natural logarithm scale. The original figures are: low ($\omega^2 = 0.010$), medium ($\omega^2 = 0.059$) and high ($\omega^2 = 0.138$).

[8]In strict sens, the interpretation of this difference changed according to the proximity measure used. Similarity measures consider the 0 to be totally different, while for distance measures it means completely equal. Thus, to simplify the understanding of the results, we take the sens of proximity. More proximal values minus less proximal ones gives a positive difference and vice-versa.

Table 4.2 – *Number of jobs supporting and refuting the experimental hypothesis about the proximity of selected and reject résumés.*

Measure	Weight	Difference	
		Positive	Negative
Cosine Similarity	Frequencies	174	7
	TF-IDF	164	10
Binary Jaccard's Index	-	165	**3**
Binary Dice's Coefficient	-	165	**3**
Non-binary Jaccard's Index	Abs. Frequency	172	6
	Rel. Frequency	**176**	5
	TF-IDF	160	5
Non-binary Dice's Coefficient	Abs. Frequency	172	5
	Rel. Frequency	**176**	5
	TF-IDF	159	5
Manhattan Distance	Abs. Frequency	51	89
	Rel. Frequency	**176**	5
	TF-IDF	54	83
Euclidean Distance	Abs. Frequency	66	78
	Rel. Frequency	114	37
	TF-IDF	64	80

As it can be seen in Table 4.2, not all the proximity measures supported with the same degree our hypothesis. While the similarity measures (Cosine Similarity, Jaccard's Index and Dice's Coefficient) supported greatly our inference, the distance measures (Euclidean and Manhattan Distances) gave mixed results. In this last case, only the use of relative frequency generated the results that we expected, the other vector weights refuted in most cases our experimental hypothesis.

With the purpose of comparing easily the results between the measures and vector weights, we designed 3 different rates and one global score. The ratios take into account the values presented in Table 4.1 and Table 4.2 in order to generate a value between 0 and 1. The 0 indicates a bad performance and 1 an excellent functioning.

The *Significance rate* (Equation 4.14) is the ratio between the number of jobs with groups statistically different and the total number of analyzable jobs in the corpus.

$$SR = \frac{\text{Total}_{\text{Statistically different}}}{\text{Total}_{\text{Analyzable jobs}}} \tag{4.14}$$

It indicates how well is the performance of the proximity measure with a certain weight in order to create statistically different groups.

The *Testing rate* (Equation 4.15) is the proportion of jobs tested with an ANOVA regarding the total number of analyzable job.

$$TR = \frac{\text{Total}_{\text{Statistically different}} + \text{Total}_{\text{Statistically equal}}}{\text{Total}_{\text{Analyzable jobs}}} \tag{4.15}$$

This ratio expresses how wellW is the performance of a proximity measure along a vector weight in order to have the lesser number of not analyzable jobs.

The cases supporting our experimental hypothesis are measured with the *Inference rate* (Equation 4.16), which uses the number of job offers supporting the inference per the number of jobs with groups statistically different.

$$IR = \frac{\text{Total}_{\text{Positive difference}}}{\text{Total}_{\text{Statiscally different}}} \tag{4.16}$$

The Inference rate helps us to know which proximity measure and vector weight is better in order to detect résumés groups which are statistically different.

The *Global score* (Equation 4.17) is a value which allow us to rank the measures according to their Significance, Testing and Inference rates. The score is also restricted to the same interval like the ratios.

$$GS = \sqrt[3]{SR * TR * IR} \tag{4.17}$$

This score tries to determine which of all measures, with their respective weight, was the best to address our needs [0, 1].

We present in Table 4.3 the values of the 3 rates and the Global score for each measure using the different kinds of vector weights.

We can observe in Table 4.3 that not all the measures indicate, in the same degree, that the analyzed groups are statistically different. Moreover, the performance of a measure changes according to the vector weight used to represent the résumés. The best value of the Significance rate was the one given by Cosine Similarity with frequencies, Manhattan Distance, Jaccard's Index and Dice's Coefficient, these last three with relative frequency (0.808). The worst value was the one obtained by Manhattan Distance with TF-IDF (0.612).

In the case of Testing rate, the best value is the one obtained by the Binary Jaccard's Index, and Jaccard's Index and Dice's Coefficient with TF-IDF (0.991). The value was close to one as for the 3 cases, only 2 jobs did not surpass the parameters necessary to do an analysis of variance. However, as all the values are greater the 0.900, we can determine that the measures and the vector weights do not affect considerably the existence of not analyzable jobs.

Table 4.3 – *Significance rate (SR), Testing rate (TR), Inference rate (IR) and Global score (GS) for each proximity measure.*

Measure	Weight	SR	TR	IR	GS
Cosine Similarity	Frequencies	**0.808**	0.978	0.961	0.912
	TF-ID	0.777	0.973	0.943	0.893
Binary Jaccard's Index	-	0.750	**0.991**	**0.982**	0.900
Binary Dice's Coefficient	-	0.750	0.987	**0.982**	0.899
Non-binary Jaccard's Index	Abs. Frequency	0.795	0.978	0.966	0.909
	Rel. Frequency	**0.808**	0.978	0.972	**0.916**
	TF-IDF	0.737	**0.991**	0.970	0.891
Non-binary Dice's Coefficient	Abs. Frequency	0.790	0.973	0.972	0.907
	Rel. Frequency	**0.808**	0.978	0.972	**0.916**
	TF-IDF	0.732	**0.991**	0.970	0.889
Manhattan Distance	Abs. Frequency	0.625	0.920	0.364	0.594
	Rel. Frequency	**0.808**	0.978	0.972	**0.916**
	TF-IDF	0.612	0.915	0.394	0.604
Euclidean Distance	Abs. Frequency	0.643	0.915	0.458	0.646
	Rel. Frequency	0.674	0.978	0.755	0.792
	TF-IDF	0.643	0.920	0.444	0.640

For the Inference rate, the measures with the best performance were the binary ones (Jaccard's Index and Dice's Coefficient) with a value of 0.982. In this case, even if these cases do not give the largest number of jobs that support the hypothesis (165 jobs for both cases), their ratio, according to the jobs refuting the hypothesis, was the best one. In this case, we can say that the measure and the vector weights have an impact on the number of cases that support and reject our experimental hypothesis.

Finally, the measures with the best Global score were Manhattan Distance, Jaccard's Index and Dice's Coefficient, the three with Relative Frequency (0.916). This would mean that these 3 proximity measures were the best to evaluate our hypothesis over the French résumés.

4.4 Discussion

Along the Section 4.3, we saw that the similarity measures, with all the vector weighs tested, supported in a greater or lesser degree our inference. This could be seen in Table 4.3, where this kind of proximity measures obtained a Global Score always greater than 0.880.

However, concerning the distance measures, the support given to our inference greatly varied according to the vector weight used, as we saw in Table 4.1, Table 4.2 and Table 4.3.

In our opinion, the reason of the results obtained by the distance measures is due to their mathematical characteristics and properties. Unlike the chosen similarity measures, which have a closed interval $[0,1]$, the chosen distance measures can have a half-open interval $[0,\infty)$. This kind of interval in these distance measures is caused due to following reasons:

- Lack of a normalization factor. One thing in common that have the similarity measures chosen in this experiment is a denominator in their formulæ. This denominator works as a data normalization, as it takes into account the size of the documents.
- Documents of different sizes. As the chosen distance measures do not have a normalization factor, the size of the documents, in terms of vocabulary or length, have an effect over their resulting value. In other words, the maximum value of a distance measure interval changes according to the documents' size. The large files push farther the maximum value. In our case, the résumés of the Recruitment Corpus are not limited neither in length nor in vocabulary, therefore, the measures were affected.

The effect of the half-open interval in our experiments consisted in making the proximity values, calculated between two couples of résumés, from the same job, not comparable. In other words, two values calculated in the same job and with the same measure and vector weight, between two couples of résumés, may have different interpretations.

For example, we have 3 documents A, B and C. The documents A and B do not have any word in common; A and C shares half the words, the other half is different. The distance calculation between A and B may have the same value that A and C, however their scale and in consequence their interpretation are different. In the first case, the value of $\sigma(A,B)$ would mean that A and B are completely different while the other value, $\sigma(A,C)$, means half similar. Therefore, when we do an average of values, we do a mixture of different scales and interpretations, causing a biased result. In our case, the results got distorted and made most of the jobs to reject our experimental hypothesis.

For the Manhattan Distance, this effect can be discarded making use of relative frequency, as we saw. The reason is that this vector weight, for the Manhattan Distance, works like a document normalization and it closes the interval between 0 and 2. This is because for two completely different vectors using relative frequency as weight, $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_{n+1}, b_{n+2}, \ldots, b_m\}$, their Manhattan Distance (σ) is equal to 2 as expressed in Equation 4.18.

$$\sigma(A,B) = \sum_{1}^{n} |a_i - b_i| + \sum_{n+1}^{m} |a_i - b_i| = \sum_{1}^{n} a_i + \sum_{n+1}^{m} b_i = 1 + 1 = 2 \tag{4.18}$$

Therefore, the scale and the interpretation of the proximity values obtained by this measure and vector weight, were the same for all the couples of résumés in a job. In other words, the values closer to zero meant equal documents and those near to 2 meant totally different.

For the Euclidean Distance, we presume that the use of relative frequency do not have the same effect as it did for the Manhattan Distance. We think that it makes the interval to be still semi-open if both vectors have more than one component, however, this time as $[0, \sqrt{2})$. The verification of this idea must be done.[9] However, if it is true, as all the résumés are larger than one n-gram, the results obtained using Euclidean Distance and relative frequency are biased.

The performance of the binary variants of Dice's Coefficient and Jaccard's Index exceeded our expectation. Even though these measures only take into account the presence or absence of n-grams they are sensible enough to support our hypothesis. This must be due to the fact that the frequency of terms or concepts is not relevant for recruiters, instead of it, the most important thing for recruiters is the appearance or not of terms related to the job requirements.

Finally, the use of TF-IDF with the similarity measures did not overpass the results obtained using absolute or relative frequency as we expected. The reason to expect better results is that the TF-IDF was designed to weight the components of a vector by their importance (Salton and Buckley, 1988). Nevertheless, we think that the performance gotten by the similarity measures is not boosted enough due to the way we calculated the IDF. We chose to get the IDF by job and not by the corpus; if we had used the corpus, the n-grams would have been weighted better. For example, some common words in all résumés, which are not stop-words, would have a lesser weight in the vector, while specific job n-grams would have a greater one.

4.5 Conclusions

This chapter presented the first set of experiments of this thesis. It consisted in determine whether and how we could use the proximity measures and résumés in a HRM context. More specifically, this first experiment was done by testing whether the selected résumés, i.e those approached by a recruiter, had more in common with themselves than and with respect to rejected résumés. The experimentation consisted in using different proximity measures and vector weights, in order to know how their use could affect our results. We evaluated the results using the statistical test of analysis of variance (ANOVA). The results showed us that proximity measures applied uniquely to résumés can be used in a HRM context. Nonetheless, in order to have performing results, we determined that it is important to select correctly the proximity measure and the vector weight. Otherwise, we can face to false or biased results.

The experiments were done over a set of 224 jobs, containing only French résumés; this group of jobs came from the Recruitment Corpus (Chapter 3). With respect to the proximity measures, we made use of three similarity measures: Cosine Similarity, Dice's Coefficient

[9] In our opinion $\sum(x : 0 < x < 1)^2 < 1$ when $\sum(x : 0 < x < 1) = 1$

and Jaccard's Index. For these last two measures, we tested a binary version, that only took into account the presence or absence of n-grams. As well, we utilized two dissimilarity measures: Manhattan and Euclidean Distances. For the vector weights, we used Term Frequency Inverse Document Frequecy (TF-IDF), relative and absolute frequency.

We evidenced that similarity measures have a great performance regardless of the vector weight used to do the proximity measurement. However, the selection of a vector weight had an impact on the resulting performance in each case. We observed, as well, that the distance measures of Manhattan and Euclidean are greatly affected by the vector weight. Only the use of relative frequency with Manhattan Distance gave us not only the expected results, but ones that were not also biased.

We found that the binary variants of Jaccard's Index and Dice's Coefficient have an excellent performance in this HRM context. Furthermore, the good results obtained open the door to the analysis of the Selected résumés in order to find the vocabulary that attracted the recruiters. In other words, the lexical set that consists more in the relevance to the recruiter, rather than in their frequency of appearance, which is not taken into account by the humans. We think that the resulting vocabulary could be exploited in order to create new resources that would ameliorate e-Recruitment systems.

Although the results obtained in this chapter confirmed our inference, it should be interesting to determine whether other types of inferences about the proximity of résumés work as well. It may be the case, that studying other inferences, will let us to understand better the information located in the résumés and its automatic analysis.

To end the conclusions, the results obtained in this chapter allowed us to think in a new method for an e-Recruitment system. It consists in the use of the average proximity between résumés to calculate a score which would express the relevance of an applicant in a job. This new methodology is presented in Chapter 5.

Chapter 5

Automatic Assistance for Processing Large Quantities of Résumés

The massification of computers and the Internet changed the way people look and apply for a job (Rafter et al., 2000a; De Meo et al., 2007; Kessler et al., 2012; Parry and Wilson, 2009). Nowadays, job seekers look for a job using the Internet and they apply sending an e-mail or filling an on-line form containing a digital résumé (Elkington, 2005).

For the Human Resources Managers (HRM), the computers and the Internet have brought, as well, a transformation in the way they deal with job applicants (Bizer et al., 2005). At the present time, they make use of the e-mail to receive the applications, they explore the Internet to find interesting people that will recruit, or use the computers to manage the applications and to be in contact with the persons who made them.

Nonetheless, the use of digital resources and technologies have also brought out undesirable consequences, specially to HRM (Barber, 2006, page 11). For instance, to post a job offer in the Internet not only makes it spread faster and wider (Montuschi et al., 2014; Arthur, 2001, page 126), it also attracts a great number of applicants that cannot be managed manually by the recruiters (Noelle, 2005; Faliagka et al., 2011). Moreover, several applications are of job seekers which are unqualified and in consequence irrelevant for the recruiters (Rafter et al., 2000b; Trichet et al., 2004).

Therefore, to reduce the effects of the Internet and the computers over the recruitment and selection processes, there have been developed the e-Recruitment systems along the last 15 years. There are different kinds of e-Recruitment systems, however we can highlight those that assist the recruiters telling them who should be approached.

According to the literature, most of these e-Recruitment systems are based on the proximity analysis between résumés and job offers, e.g. Montuschi et al. (2014) and Kessler et al.

(2012). Only one of the systems found make use uniquely of résumés (Faliagka et al., 2013). However, it consisted in training a system, based on selection processes previously done, to replicate the recruiters' behavior. Although it is an interesting approach, it is difficult to replicate and implement in large scale, from our point of view.

Thus, in Chapter 4, we presented whether and how different proximity measures could be used effectively only over résumés to find details about the recruitment process of applicants. This without having to use previous selection processes or job offers. More specifically, we tested 5 proximity measures, with 3 different vector weights, to determine which were the most adequate for a HRM context when the information came only from résumés. From the experimentation we found that the résumés of selected applicants, i.e. those that were approached by a recruiter, had more in common with themselves than or with respect to the résumés of rejected candidates. This meant that some proximity measures, used with certain vector weights, were sensible enough to detect that the selected candidates shared common characteristics and that were different from those located in the rejected résumés.

In this chapter we apply the idea that selected résumés have more in common with themselves than or with respect to the rejected ones. The application is done through the proximity calculation between résumés and by the ranking of résumés according to their average proximity.

As well, in this chapter we introduce two relevance feedback methods in order to improve the results of the proximity between résumés. The use of the relevance feedback was motivated by the work of Kessler et al. (2009), where the recruiters were asked, through a simulation, to give their opinion about a small set of résumés. Then, the input given by the recruiters was used to improve the ranking of résumés, in consequence, made the system's output to be closer to the recruiter's needs.

This chapter is divided in 6 parts, we start with the objective (Section 5.1). Then we present in Section 5.2 the methodology, where it is explained in detail how we used and calculated the proximity between résumés but also the relevance feedback. The results and their discussion are located in Section 5.3 and Section 5.4 respectively. We finish with the conclusions at Section 5.5.

5.1 Objective

To develop a method that allow us to create a résumé priority list for each job. At each list, the résumés at the top are the ones that should be read, in first place, by a recruiter, and those at the bottom the ones that could be avoided.

In this way the recruiters could find faster the interesting résumés and avoid, at some point and if wanted, those résumés that may come from unqualified applicants.

5.1.1 Hypothesis

We infer that the calculation of the average proximity between résumés can be used to rank the résumés according to their relevance with a specific job. Moreover, we hypothesize that at the top of the ranking, the average proximity between résumés will position the documents that correspond to those candidates that were approached by a recruiter (Selected). At the end, the method will place those résumés that were not contacted by a recruiter (Rejected).

5.2 Methodology

Let us consider J a job composed of n résumés $J = \{r_1, r_2, \ldots, r_n\}$, σ a proximity measure and λ the score of each résumé. For each résumé in J we calculated a λ according to an average proximity done between itself and the rest of résumés in J. Equation 5.1, presents the formula used to calculate the lambda of each résumé:

$$\lambda(r) = \frac{1}{n-1} \sum_{x=1}^{n} \sigma(r, r_x : r \neq r_x) \tag{5.1}$$

where r is a résumé from J, r_x are all the résumés of J excepting r and n is the size of J in terms of résumés. The résumé priority list for job J is created according to the sorted values of λ, from the greater average proximity to the lesser one.

We have called the calculation of λ as *AIRP* or *Average Inter-Résumé Proximity*, a method that only makes use of résumés in order to score them. In Figure 5.1, we can see an example of the application of the AIRP methodology and the generation of a priority list for a job of 4 résumés.

In this section we will present the proximity measure used in this chapter (Section 5.2.1) and the relevance feedback (Section 5.2.2), a method used along AIRP to improve our results. As well, we will introduce the evaluation methodology (Section 5.2.3), the experiments done in this chapter (Section 5.2.4) and the data utilized in our experiments (Section 5.2.5).

5.2.1 Proximity Measure

To calculate AIRP we decided to use Dice's Coefficient with relative frequency. The reason was its performance during the Chapter 4.

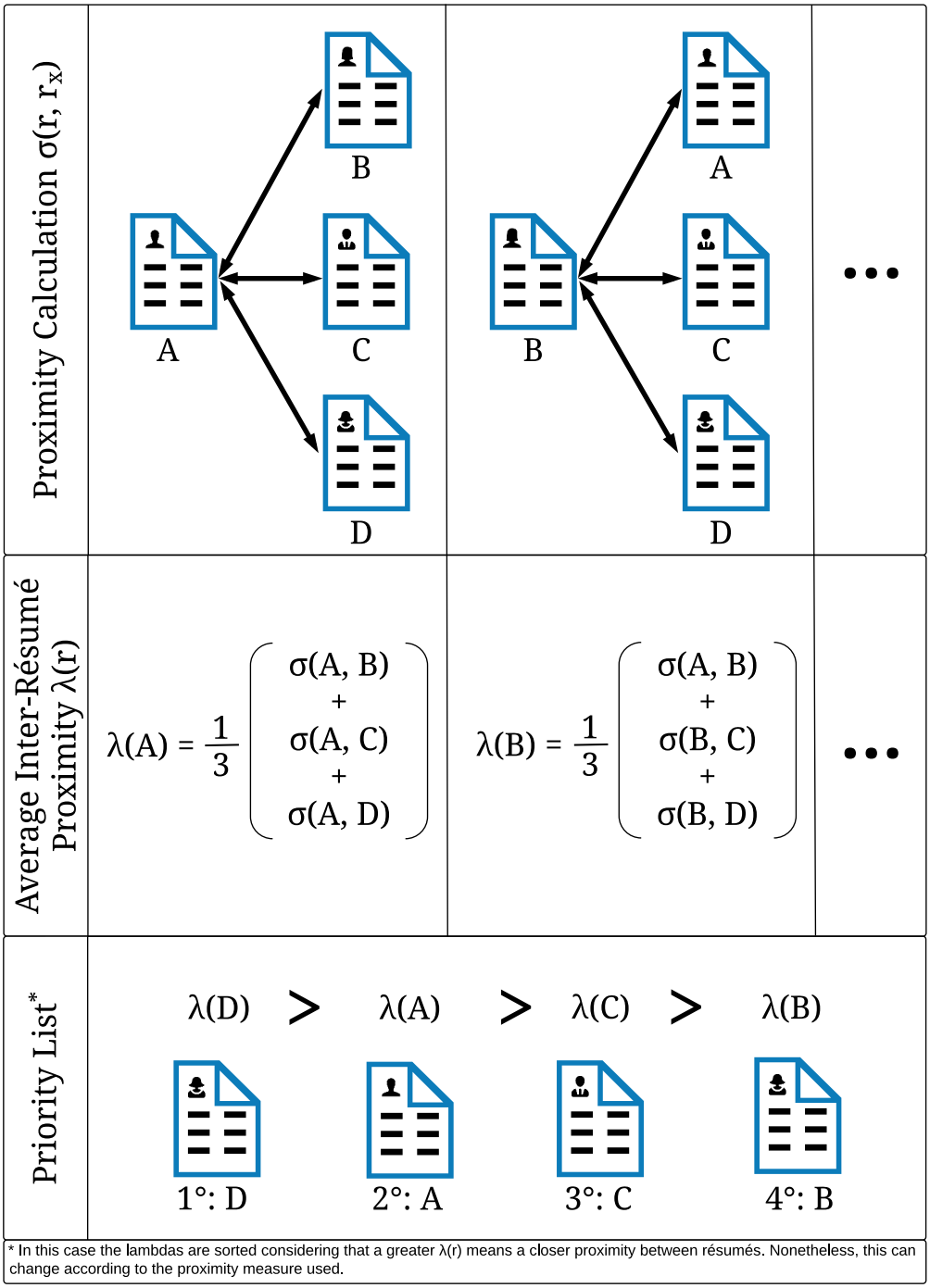

Figure 5.1 *– Example of the application of AIRP methodology and the generation of a résumé priority list for a job of 4 résumés.*

5.2.2 Relevance Feedback

The relevance feedback is the interaction of a human user over an information retrieval system, where the user evaluates the retrieval operation and modifies the request in order to improve the search of data (Rocchio, 1971). In our case, the information retrieval system is the creation of the résumé priority list. The relevance feedback is the interaction done by a recruiter (or another entity) over the system to enhance the ranking of applicants.

For the experiments done in this chapter, we have defined two kinds of relevance feedback. One explicit, where the recruiter must participate in order to ameliorate the results. One other implicit, where the system uses the data obtained by ARIP, to improve automatically the results of the ranking. Both types of relevance feedback will be explained more precisely in the following sections.

Explicit Relevance Feedback

The *Explicit Relevance Feedback* is the one where the recruiters assist directly the system in order to improve the results. It is based on the ideas of Rocchio (1971) where a user has to introduce an error signal into the system in order to get a result closer to their needs. This error signal is produced by the user indicating which results, from a first system output, were relevant and which were not according to their needs.

In our case, the Explicit Relevance Feedback corresponds in asking the recruiter whether a set of résumés were interesting enough to contact their respective applicants. If it was not the case, the system would modify the λ values, using the feedback, in order to give to the recruiter a new résumé priority list which would be, hopefully, closer to the recruiter's needs.

In spite of the fact that the relevance feedback of Rocchio (1971) was done over only the top of the retrieved documents, we decided to test whether other positions of the priority lists worked as well to do a relevance feedback. In total, we tested 4 positions:

- Top. Following the original method of the relevance feedback, we tested whether to choose documents from the beginning of the priority lists worked to do a relevance feedback in a HRM context.
- Bottom. We decided to test whether the résumés located at the end of the priority lists could be used to do a relevance feedback. We inferred that finding a relevant résumé at the end of the list would be a more powerful error input than detecting an irrelevant one at the top. In addition, we did this owing to the fact that leaving a pertinent résumé at the bottom would leave apart the objective of the priority list. At

the end, the recruiter would have to analyze all the résumés in order to find all the interesting résumés.

- Random. This position consisted in asking the recruiter whether a résumé, chosen randomly from the ranking, was relevant or not. We wanted to determine whether generating a priority list to apply after the Explicit Relevance Feedback was better than choosing aleatory résumés from the beginning.
- Top and Bottom (Both). This position consists in asking to the recruiter whether a résumé from the top and another from the bottom were relevant. The goal was to use in some extent the original idea of the relevance feedback, but at the same time try to find if any interesting résumés was ranked at the bottom. In the case that an odd number of résumés was used for the relevance feedback, the documents at the top had priority.[1]

The original relevance feedback was designed to be used in a Vector Space Model, modifying the weights of the documents vectors (Jackson and Moulinier, 2002). Nonetheless, in our case we decided to modify the proximity values. This modification was done through the multiplication of a factor to each Average Inter-Résumé Proximity (λ); the formula is presented in Equation 5.2.

$$\Lambda(r) = \lambda(r) \cdot EF(r) \tag{5.2}$$

where r is a résumé from the job, $\Lambda(r)$ is the modified Average Inter-Résumé Proximity for the document r and EF is the modifying factor or Explicit Factor.

The *Explicit Factor (EF)* is a value which takes into account the number of relevant and irrelevant résumés found by the recruiter in each position tested. Its objective is to change the rank of résumés in the priority list.

Let us consider, $J = \{r_1, r_2, \ldots, r_n\}$ the set of résumés of a job used to create a priority list, S the set of résumés marked as relevant by the recruiter ($S \subseteq J$, $|S| = k$) and R the set of résumés considered as irrelevant ($R \subseteq J$, $|R| = l$, $R \cap J = \emptyset$). The Explicit Factor is calculated following the formula presented in Equation 5.3:

$$EF(r) = \frac{\left[\omega + \sum \sigma(r, r_x : r \neq r_x, r_x \in S) + CF(r : r = r_x)\right] \cdot (l + \omega)}{(k + \omega) \cdot \left[\omega + \sum \sigma(r, r_x; r \neq r_x, r_x \in R) + CF(r : r = r_x)\right]} \tag{5.3}$$

where ω is an offset to avoid indeterminate values in the formula whenever $S = \emptyset$ or $R = \emptyset$. For the experiments we set $\omega = 1 \times 10^{-10}$. The $CF(r)$ is a correction factor for r that will be described in the next paragraph. The value of the EF will be higher as the résumé r is more similar to the relevant résumés (S), and will be lower if r is more proximal to the irrelevant ones (R).

The correction factor (CF) is a value that allows us to keep at the top or the bottom of the rankings the résumés that were already evaluated by a recruiter. In consequence, it allows

[1] For example, if 5 résumés were used as relevance feedback, three of them came from the top of the priority list and 2 from the bottom of it.

us not to lose the information previously obtained from the recruiter, in the new priority list. The Equation 5.4 shows the possible values of the correction factor.

$$CF(r) = \begin{cases} 50 & \text{if } r \in S \wedge \text{ERF Position: Top or Random} \\ 1,000 & \text{if } r \in S \wedge \text{ERF Position: Bottom} \\ 1,000 & \text{if } r \in R \wedge \text{ERF Position: Bottom} \end{cases} \tag{5.4}$$

where r is a résumé from a job, S is the set of relevant résumés and R the set of the irrelevant ones. These figures where obtained by experimentation.

As it can be seen in Equation 5.4, the value of the correction factor changes according to the position used for the Explicit Relevance Feedback. As well, it must be noted that the correction factor can only appear once in each $EF(r)$, either at the numerator or the denominator. The reason is that a résumé can only be either relevant or irrelevant ($r : r \in S \vee r \in R$).

Implicit Relevance Feedback

We infer for this chapter that, at the top of a résumé priority list were the relevant résumés for a recruiter. Thus, based on this, we implemented a relevance feedback where it was not necessary for a recruiter to evaluate the résumés, in other words an Implicit Relevance Feedback.

The *Implicit Relevance Feedback* consists in stating that résumés located at the top of a priority list were always relevant for a recruiter. Then, to calculate a modified proximity score (Λ) for each résumé, but with respect to the chosen top résumés. The value of Λ is used to build a new priority list for a recruiter, following the same perspective explained at the beginning of the methodology (Section 5.2).

Considering $J = \{r_1, r_2, \ldots, r_n\}$ the set of résumés of a job used to create a priority list and I the set of résumés considered automatically as relevant ($I \subseteq J$, $|I| = k$). The value of k, i.e. the size of I, is set by the user. We can define Λ as Equation 5.5:

$$\Lambda(r) = \frac{1}{|I|} \sum_{x=1}^{k} \begin{Bmatrix} \sigma(r, r_x : r \neq r_x, r_x \in I) & \text{if } r \notin I \\ 1 & \text{if } r \in I \end{Bmatrix} \tag{5.5}$$

where r is a résumé of the job, $\Lambda(r)$ is the modified Average Inter-Résumé Proximity for the document r, I is the set of résumés chosen to do the Implicit Relevance Feedback and σ is a proximity measure.

It can be seen in Equation 5.5, that the right part of the formula changes according whether r belongs to I. This was done because we wanted to boost the value of Λ for the résumés that were the ones used for Implicit Relevance Feedback. The value of 1 was chosen because is the result given by Dice's Coefficient, when two documents are the same ($r = r_x$).

5.2.3 Evaluation Methodology

The evaluation methodology is divided in two parts. In first place, we present the measure used to evaluate the performance of our methods. In second place, the baselines employed to compare the results obtained by our methodology.

Mean Average Precision

To evaluate the performance of the résumé priority list, we decided to use the *Mean Average Precision* or *MAP* (Buckley and Voorhees, 2000), a measure used since 2001 in the *Text REtrieval Conference (TREC)*. The *MAP*, as its name express, is the mean of the *Average Precision (AP)*, which in turn is composed of the average calculation of the *Precision (P)*.

In the MAP, the Precision (P) is defined as the quotient between the number of relevant documents previously found in a ranking and the document position from which is calculated the Precision. If the evaluated file is not relevant, the value of $P = 0$. The Average Precision (AP) is the sum of all the values of the Precision and divided by the number of relevant documents in the ranking.

In our case, let us consider a priority list calculated for a job $L = \{r_1, r_2, \ldots, r_k\}$ which is composed of résumés (r) ordered by rank (from 1 to k). As well, let us take into account S, the set of all the relevant résumés of L ($S \subseteq L$). The precision, P, for a résumé in a certain position ($r_x : r \in L, 1 \leq x \leq k$) is defined at Equation 5.6. The AP for the priority list L is defined in Equation 5.7:

$$P(r_x) = \begin{cases} 0 & \text{if } r_x \notin S \\ \frac{1}{x} \sum_{z=1}^{x} \begin{cases} 0 & \text{if } r_z \notin S \\ 1 & \text{if } r_z \in S \end{cases} \end{cases} \tag{5.6}$$

$$AP = \frac{1}{|S|} \sum_{x=1}^{k} P(r_x) \tag{5.7}$$

The Mean Average Precision (MAP) is the average of all the values of the AP obtained for all the jobs analyzed.

In Table 5.1 we show an example about the calculation of the P, the AP and the MAP considering two jobs, each one with a priority list and their respective relevant résumés. The first job is composed of $L_1 = \{r_1, r_2, \ldots, r_5\}$ and $S_1 = \{r_1, r_2, r_5\}$; the second one is constituted by $L_2 = \{r_1, r_2, \ldots, r_6\}$ and $S_2 = \{r_1, r_2, r_4\}$.

We did not use the common measures of precision and recall, because these are not useful to evaluate rankings. The reason is that in rankings, like the priority lists, the most relevant results should appear at the top but neither the precision nor the recall take into account

Table 5.1 *– Example of the calculation of the Precision (P), the Average Precision (AP) and the Mean Average Precision (MAP) for two priority lists.*

Job	Résumé	$r \in S$	P	AP	MAP
1	r_1	YES	$\frac{1}{1} = 1$	$\frac{2.600}{3} = 0.866$	$\frac{0.866 + 0.916}{2} = 0.891$
	r_2	YES	$\frac{2}{2} = 1$		
	r_3	NO	0		
	r_4	NO	0		
	r_5	YES	$\frac{3}{5} = 0.600$		
2	r_1	YES	$\frac{1}{1} = 1$	$\frac{2.750}{3} = 0.916$	
	r_2	YES	$\frac{2}{2} = 1$		
	r_3	NO	0		
	r_4	YES	$\frac{3}{4} = 0.750$		
	r_5	NO	0		
	r_6	NO	0		

the results' order (Manning et al., 2008, page 145). However, according to Voorhees and Harman (2001), the *AP* is a measure which takes into account the precision and the recall at the same time:

- Recall. The number of pertinent documents that were retrieved according to the total number of relevant documents.
- Precision. Whether the pertinent documents where positioned at the top of the ranking or not.

Baselines

We decided to compare our results with two baselines, in other words, with basic methods used to create a résumé priority list.

The first baseline consisted in creating a résumé ranking following a random order. The objective is to verify that the results using the different methods were better than the values obtained by coincidence. This baseline was done generating 100 random priority lists and calculating the average MAP of them.

The second baseline corresponded in ordering the résumés by the creation date of each application in the server, as it is indicated by the applicants' meta-data (Chapter 3). The oldest résumés were at the top while the most recent at the bottom.

5.2.4 Experiments

For this chapter, we decided to do 4 experiments which are explained as follows:

1. AIRP versus Baselines. We tested whether the generation of a résumé priority list was better, in terms of the MAP, using AIRP than more basic methods (see Section 5.2.3).
2. AIRP with Implicit Relevance Feedback. We experimented whether the use of Implicit Relevance Feedback over the results given by AIRP could be improved. The size of the Implicit Relevance Feedback was determined by a percentage calculated from the number of résumés in the job. We variated this percentage from 10% to 100% of résumés; the step was of 10%. In addition, we decided to apply it 5 times successively. The reason was that we did not have to ask to the recruiters their opinion about the résumés several times. Moreover, an iterative method would sum up the inferred feedback.
3. AIRP with Explicit Relevance Feedback. We used the Explicit Relevance Feedback over the values of AIRP to determine how this method could ameliorate the results. We tested the four possible positions described in Section 5.2.2: Top, Bottom, Both and

Random[2]. As well, for each position we tested from 1 to 10 résumés for the Explicit Relevance Feeback.

4. AIRP using the best parameters of relevance feedback. We decided for the last experiment, to use the best parameters of the 2 previous experiments, in order to see whether both Relevance Feedback could be merged.

5.2.5 Data

For this chapter, we made use of 185 jobs from the Recruitment Corpus (Chapter 3) with their respective French résumés. We chose the 185 jobs according to 2 characteristics, that were set in order to reduce the bias and/or false results during the experimentation:

- Minimum number of résumés. Each job analyzed had at least 20 French résumés. As we decided to use up to 10 résumés for the Explicit Relevance Feedback, we considered that jobs having less than 20 résumés would have biased results.
- Minimum number of Selected and Rejected résumés. The jobs had at least 5 French Selected résumés and 5 French Rejected Résumés. The objective was to avoid the effect of "looking for a needle in a haystack".

It is important to note, that instead of asking to the recruiters whether a résumé for a job was relevant or not, we made use of the meta-data of each applicant. In this case, the résumés classed as Selected were taken as relevant for the recruiters, while those classed as Rejected were considered to be irrelevant. This information was used for the test of the Explicit Relevance Feedback, but as well for the evaluation of the experiments.

5.3 Results

In this section, we present the results obtained for the 4 experiments done:

1. AIRP versus Baselines.
2. AIRP with Implicit Relevance Feedback.
3. AIRP with Explicit Relevance Feedback.
4. AIRP using the best parameters of relevance feedback.

5.3.1 AIRP versus Baselines

The Experiment 1 consisted in comparing the results of the résumé priority list done by the Average Inter-Résumé Proximity (AIRP) with the results given by the two baselines described

[2]The position *random* used in this case is not the same as the one used in the baseline.

in Section 5.2.3. The Figure 5.2 shows the mean, in terms of the MAP, for AIRP and the baselines; the three have their respective 95% confidence interval bars.

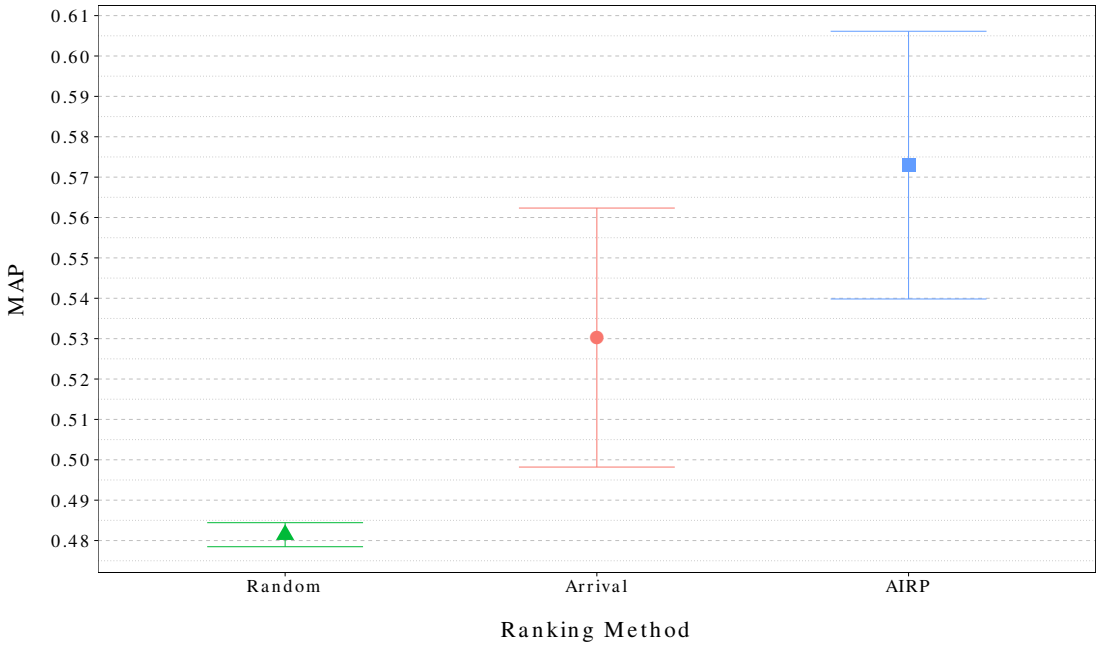

Figure 5.2 *– Results, in terms of the Mean Average Precision, of AIRP and the two baselines: résumés sorted randomly by applicants' arrival date.*

The MAP obtained by AIRP is 0.5729, while the MAP for the baseline based on the arrival order and the one based on the random order is 0.5302 and 0.4814, respectively.

In order to validate the three values of the MAP, we decided to verify whether they were statistically different. We made use of an analysis of variance, or ANOVA (See Chapter 4), from which we obtained a p-value $= 1.250 \times 10^{-10}$. We did, as well, a Tukey's HSD test, i.e. a *post hoc* analysis, in order to know where the differences occurred between the methods. The *post hoc* results are presented in Table 5.2.

Table 5.2 *– Results of the Tukey's HSD test for AIRP versus the baselines.*

Methods	Adjusted p-value	Statistically different ($\alpha = 0.05$)
Random - Arrived	0.0040	YES
AIRP - Arrived	0.1159	NO
AIRP - Random	0	YES

The results from the Tukey's HSD test (Table 5.2) indicated that the methods of AIRP and arrival order were not statistically different between them. This confirms the results

presented in Figure 5.2, where we can see that the 95% confidence interval of the Arrival and AIRP methods overlapped, while the ones of Random did not.

5.3.2 AIRP with Implicit Relevance Feedback

In the Experiment 2, we tested whether the priority lists generated by AIRP with the Implicit Relevance Feedback were better, in terms of the MAP, to the ones created only using AIRP.

As stated in Section 5.2.4, the experiments were done job by job and through the evaluation of different sizes of the set used for the Implicit Relevance Feedback. The size of the set was defined by a percentage of résumés, that varied between 10% and 100% with steps of 10%. The Implicit Relevance Feedback was tested 5 times iteratively, i.e. 6 times in total, one after other.

Figure 5.3 and Figure 5.4 show the MAP of the original AIRP (point 0) and of AIRP with the Implicit Relevance Feedback through their 6 consecutive applications (point 1 to 6). The results are presented according to each tested percentage of résumés and with their respective bars of the 95% confidence interval.[3]

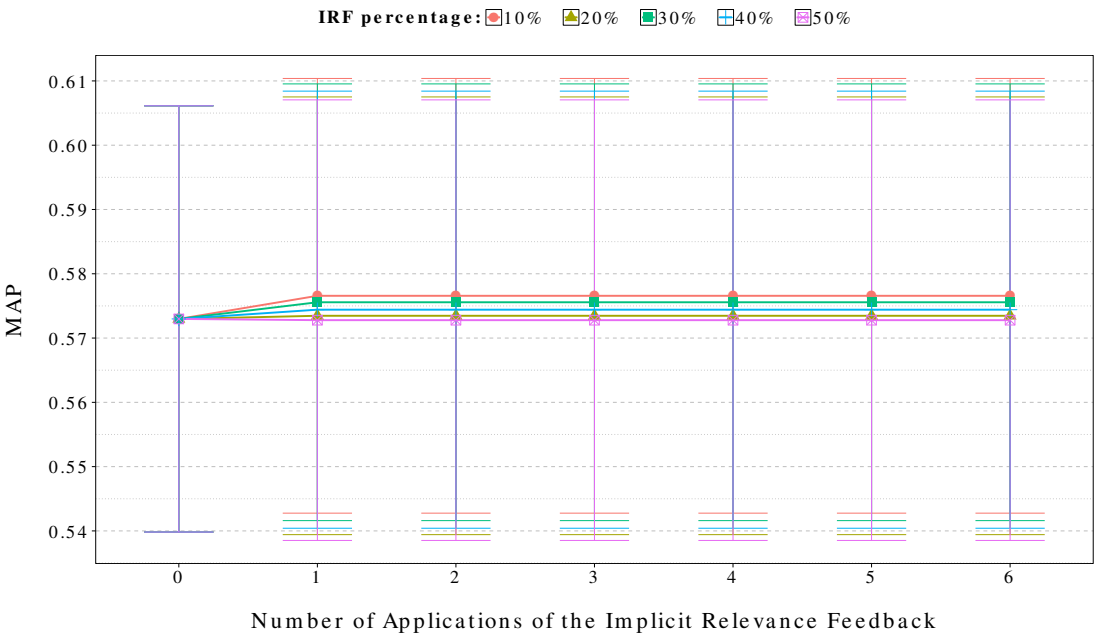

Figure 5.3 *– Results of the values for AIRP with Implicit Relevance Feedback for percentages from* 10% *to* 50%.

[3]In Appendix B we present the same results but without the 95% confidence interval bars.

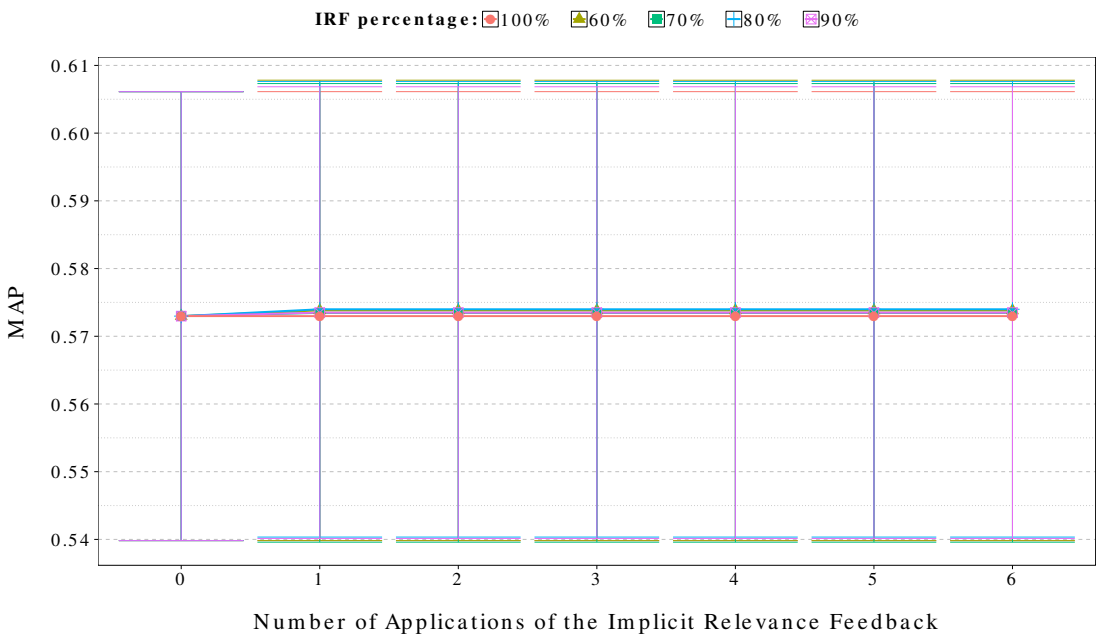

Figure 5.4 *– Results of the values for AIRP with Implicit Relevance Feedback for percentages from* 60% *to* 100%.

In the first application of the Implicit Relevance Feedback, the results obtained varied for most of the percentage of résumés. The exception was the 100% of résumés, which always gave the same value of the original AIRP (0.5729). We can observe, as well, that the differences obtained in the MAP are minimal. They passed from 0.5729 to 0.5765 in the best of the cases, which was using 10% of the résumés, while in the worst, using 50% of the résumés, the MAP decreased to 0.5727.

Although, the difference of the resulting MAP were minimal, we tested if these values were statistically different. In other words, we verified if the use of the Implicit Relevance Feedback gave us a real difference, positive or negative, with respect to the original method of AIRP. We made use of an analysis of variance (ANOVA) from which we obtained a p-value = 0.893. As the resulting p-value is greater than 0.05, this mean that the results are statistically equal, i.e. neither the gain nor the loss in the MAP are significant.

With respect to the iterative application of the method, the resulting MAP did not change after the first application, i.e. they converged.

5.3.3 AIRP with Explicit Relevance Feedback

For the 3rd experiment, we evaluate whether AIRP could be improved with the use of the Explicit Relevance Feedback.

This experiment, as indicated in 5.2.4, consisted in creating a résumé priority list for each job using the method of AIRP. Then to choose a block of résumés according to a certain position (Top, Bottom, Random or Both; see Section 5.2.2). Finally, to use this block of résumés to calculate the Explicit Relevance Feedback in order to generate a new résumé priority list. The size of the blocks went from 1 to 10 résumés.

Figure 5.5 presents the average MAP of the original AIRP (point 0) and the ones obtained using AIRP with Explicit Relevance Feedback for the four positions chosen (points 1 to 10). Each value has their own bar of the 95% confidence interval.

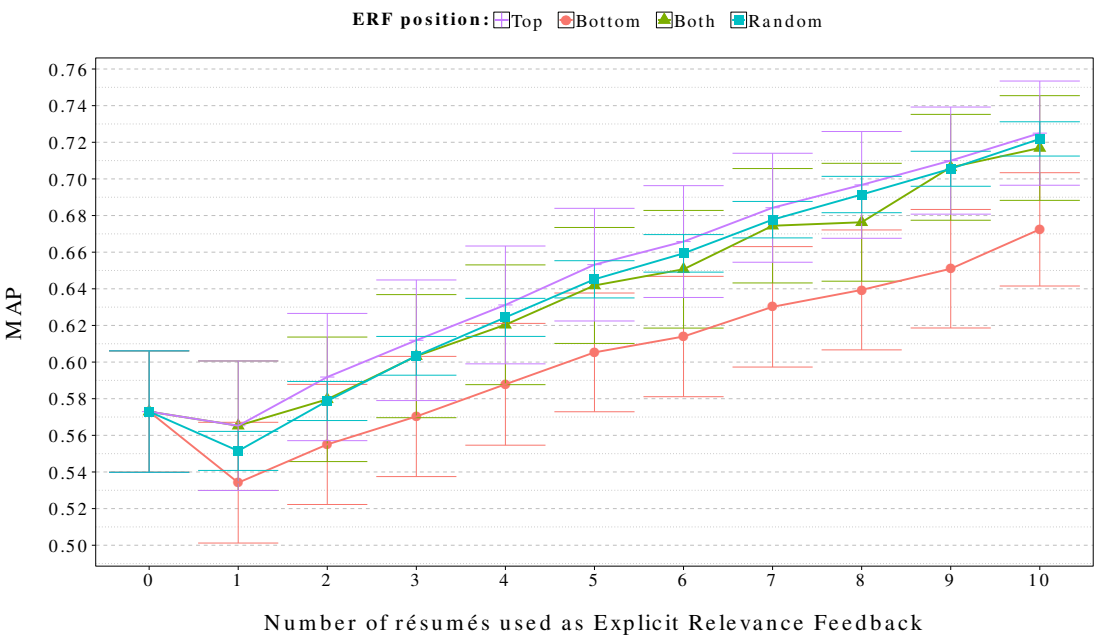

Figure 5.5 *– Results of the values for AIRP with Explicit Relevance Feedback.*

We can observe, in Figure 5.5, that the performance of the Explicit Relevance Feedback changed according to the position used. However, all the values of the MAP increased as we rose the size of the résumé block. Nevertheless, in all the cases, it can be seen that the value of the MAP decreased, in comparison to the original AIRP, when we used only one résumé for the Explicit Relevance Feedback.

The Top position gave always the best value of the MAP, however, the Random position become closer to the Top one as we increased the size of the block. The worst performance

was the one given by the Bottom position. In summary, the worst value for the MAP, 0.5341, is given by the Bottom position when 1 résumé is used; the best MAP, 0.7249, is obtained by the Top position when the size of the block is equal to 10 résumés.

It should be noted that for a block of one résumé, the Top and Both positions, share the same value of the MAP. The reason, as we indicated in Section 5.2.2, is that for odd numbers, the Both position takes one extra résumé from the beginning of the priority list. Therefore, when the block is equal to one résumé, the Both position worked like the Top one.

We tested whether the MAP values of the 4 methods were statistically different when 10 résumés were used as Explicit Relevance Feedback. To do that, we utilized an ANOVA that gave us a p-value = 0.0184, i.e. at least one pair of the results were statistically different. To determine which pairs were the ones with this characteristic we applied a Tukey's HSD test as *post hoc* analysis; the results obtained from this test are presented in Table 5.3.

Table 5.3 *– Results of the Tukey's HSD test for AIRP with Explicit Relevance Feedback when 10 résumés are used.*

ERF Position	Adjusted p-value	Statistically different ($\alpha = 0.05$)
Bottom - Top	0.0654	NO
Both - Top	0.9813	NO
Random - Top	0.9971	NO
Both - Bottom	0.1599	NO
Random - Bottom	0.0095	YES
Random - Both	0.9894	NO

As it can be seen in Table 5.3, only the results obtained by the Random and Bottom positions are statistically different between them. For the other cases the results are statistically equal. This confirms the behavior of the 95% confidence interval bars shown in Figure 5.5, only the ones of the Random and Bottom position did not overlap between them.

5.3.4 AIRP Using the Best Parameters

For the 4th experiment we decided to continue to use the Explicit Relevance Feedback however this time applied iteratively, as we did with the Implicit Relevance Feedback. The objective was to ameliorate the system's performance faster.

Being more concise, we decided to use for the Explicit Relevance Feedback a block of 2 résumés from the Top position. The reasons to do this were two. On one hand, during the Experiment 3 (Section 5.3.3), the block of 2 résumés from the Top position showed the

best performance and its difference, in the terms of the MAP, with other positions was the greatest. On the other, two résumés is a quantity small enough in order to apply the method iteratively, without the risk of having to ask the recruiters to analyze several résumés.

As we say previously, we chose to apply the Explicit Relevance Feedback iteratively. The idea came from the basis used for the Experiment 2 (Section 5.3.2), in each continuous application of the method, we can sum up the feedback given by the recruiters and in consequence to improve faster the resulting priority list. Due to the characteristics of the Explicit Relevance Feedback, we decided to verify that in each iteration the chosen block of 2 résumés were not composed of documents analyzed previously by the recruiter.[4] For example, the first two résumés used in the Explicit Relevance Feedback were marked as relevant, thus, in the second application of the method, the résumés to use are the third and the fourth of the priority list. We decided to apply the Explicit Relevance Feedback 5 times in total, giving us at the end 10 résumés analyzed by job.

As it was explained in Section 5.2, the Explicit Relevance Feedback consisted in the generation of a modified lambda (Λ). This is done by the multiplication of the Explicit Factor with the value of AIRP (λ) of each résumé. Due to the iterative application of the Explicit Relevance Feedback, there were in consequence, two ways in which the Explicit Factor could be applied:

- The *Original lambdas* method consisted in multiplying, for each application, the obtained Explicit Factor by the λ gotten for each résumé at the Experiment 1 (Section 5.3.1).
- The *Iterative lambdas* method corresponded in the multiplication of the Explicit Factor, obtained in each application, by the lambda calculated in the previous step, i.e. λ in the first application and Λ in the following ones. At the end, after 5 applications of the Explicit Relevance Feedback, the original scores, λ, would have been modified 5 times consecutively.

In both cases, the total number of résumés analyzed by the recruiter in each application is used to calculate the Explicit Factor. The results of this experiment are presented in Figure 5.6.

As it can be observed in Figure 5.6, the most performing method, in terms of the MAP, is the one using the Original lambdas. When 10 résumés have been used as feedback the Original lambdas obtained a MAP of 0.7554, while the Iterative lambdas had a MAP of 0.7020.

We decided to verify whether the two results were statistically different when we used 10 résumés as feedback. We made use of a non-parametric test of Kruskal-Wallis[5] (Kruskal

[4]In the Explicit Relevance Feedback, the résumés already analyzed by the recruiters were positioned always at the top (relevant) or at the bottom (irrelevant) of the priority list. Therefore, the iterative application of the method could select the same résumés at the Top, if they were not marked as analyzed.

[5]We could not use an ANOVA because both methods did not have a normal distribution and homogeneous

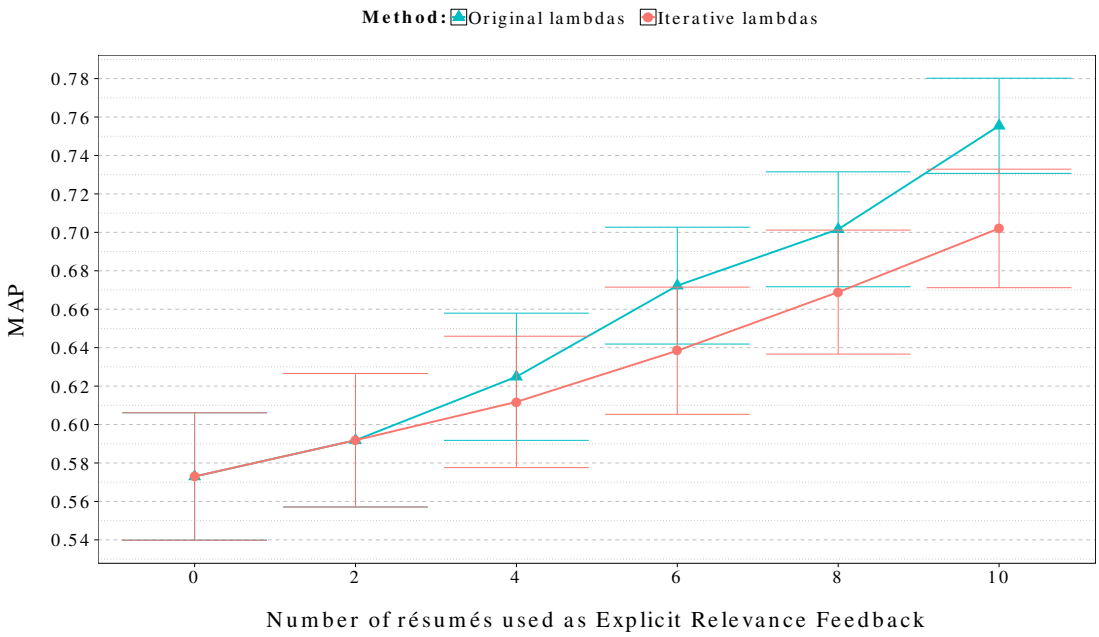

Figure 5.6 *– Results of the values for AIRP with iterative Explicit Relevance Feedback using blocks of résumés.*

and Wallis, 1952), from which we obtained a p-value = 0.0465. In normal conditions, this would mean that the 2 results were statistically different (p-value < 0.05). Nonetheless, both distributions were not equal[6]. Thus, the Kruskal-Wallis test can only indicate that, in average, the performance obtained by the Original lambdas method is greater than the one got by the Iterative lambdas method.

5.4 Discussion

It may seem interesting that for the Experiment 1, the arrival order has a better MAP than the one obtained by the random ranking, considering that relevant candidates can arrive at any time. We could believe that this results is caused due to a bias in the recruiters' behavior and that they select on purpose the first candidates. However, this may not be the case. What happens, according to the recruiters, is that at the beginning of each recruitment process, even before the post of the job offer, they have already started looking for potential applicants in HR databases or the Internet. The recruiters, therefore, began populating the applicants' database with selected résumés in some cases even before the arrival of external applications. Then, when the MAP was calculated over the rankings based on the arrival

variances.

[6]If the set of values have different distributions, the interpretation of the result given by the test of Kruskal-Wallis changes.

date, it found, in several cases, than at the top of the list there were only selected résumés.

Although the outcome of the statistical test between the results of AIRP and the arrival ranking may discourage the usefulness of the latter, it must be considered that not all the jobs fulfill the characteristic that the first arrived are always selected candidates. Moreover, this behavior may not be true in other HR databases. Therefore, we consider that the results given by the arrival ranking are more due to the characteristics of the Recruitment Corpus than a possible method to exploit in a e-Recruitment system.

The results obtained in the Experiment 2, where we tested the Implicit Relevance Feedback, did not give us neither a significant loss or gain in the system performance. We believe that this lack of performance was caused by two factors. The first one is that we did not always get at the top of the rankings the résumés that were Selected, as we could see in the MAP value of the first experiment (0.5729). The second factor is the use of percentages to choose which résumés was going to be used as the Implicit Relevance Feedback. In the Recruitment Corpus the size of the jobs (quantity of résumés) is not homogeneous; there are large and small jobs. Therefore, the use of a percentage of résumés as a measure to choose the Implicit Relevance Feedback may not always give enough information to improve the results. For instance, taking into account that the smallest job had to contain 20 résumés, selecting only 2 résumés (10%) may not give enough data about what was searched. While for large jobs, choosing, for example 50% of the total résumés, may give several résumés that were Rejected, making the Implicit Relevance Feedback to have a poor quality. This problem may be solved with a methodology based on a percentage of résumés, but also on thresholds, for example a minimum or maximum number of résumés to choose, which should be set according to the job size.

Despite the results in the Experiment 2, it may look interesting that for 100% we got the same MAP as the original AIRP without any relevance feedback. However, this is caused by the formula used to calculate the Implicit Relevance Feedback. When we set the 100% of résumés, the summation of Equation 5.1 and Equation 5.5 are the same, they only differ in the factor which multiplies them. Therefore, it was obvious that the results would give the same priority list and in consequence obtain the same MAP.

We did not expect that the iterative application of the Implicit Relevance Feedback would converge in the first iteration. Nevertheless, after an analysis, we determined that the cause was simple. We designed the formula of the Implicit Relevance Feedback to always keep the set of chosen résumés at the top of the ranking. In consequence, when we iterated and used the same percentage of Implicit Relevance Feedback, we came up again with the same résumés chosen previously. This problem may be dealt with the implementation of incremental or decremental percentages of résumés, making to vary the set of the chosen ones in each iteration.

Another possibility to improve the results obtained in the Experiment 2 consists in the

modification of the formula used by the Implicit Relevance Feedback. Nowadays, the method used is more similar to the one of AIRP's calculation than to the one used by the Explicit Relevance Feedback. Therefore, we could modify the formula to generate a modified lambda (Λ) as the one presented in Equation 5.8:

$$\Lambda(r) = \lambda(r) \cdot \frac{1}{|I|} \sum_{x=1}^{k} \sigma(r, r_x : r \neq r_x, r_x \in I) \tag{5.8}$$

where r is a résumé, λ is the original value of AIRP which is multiplied by a factor. This factor is the sum of the average proximity calculated between the résumé r and all the résumés chosen to be used in the Implicit Relevance Feedback.

Respecting the Experiment 3, where we tested the Explicit Relevance Feedback, we saw that the system performance increased as the number of résumés used as feedback rose as well. However, we saw that the performance of the systems, no matter the position (Top, Bottom, Both, Random), decreased, in comparison to the original value of AIRP, when we used only one résumé as feedback. For example, for the Bottom the difference was of -0.03882, while for the Both or the Top the difference was only of -0.00772.

After a deep analysis of the results, we found the problem. The Explicit Factor affected in great amount the value of AIRP (λ) when it was used only with one résumé, especially if it is a Rejected one. It must be rememberer that the Explicit Factor is calculated as the proximity between the résumés chosen for the Explicit Relevance Feedback and a résumé to rank. In the case of the Top and Both positions, we found that some priority lists began with a Rejected résumé but were followed by Selected ones. When it was applied the Explicit Factor, using as element the first reject résumé, it created an *upside-down* effect. Thus, the topped résumés, which most were Selected ones, finished at the bottom and vice-versa; this created, at the end, a reduction of the MAP value.

With respect to the bottom position, we found that most of the priority list ended, as expected, with Rejected résumés. However, when these Rejected résumés were used as Explicit Relevance Feedback, without any Selected one, they penalize in great amount all the résumés which had a proximity to the Rejected ones, although they were not close to the bottom. In other words, the Rejected résumés at the bottom, decreased the Λ of résumés nearer to the top but that had some relation to those at the bottom. As a consequence, the performance was affected and the resulting MAP was reduced. This behavior explains, as well, the reason of the low performance of the Bottom position: there were not to many selected résumés at the bottom at least for the last 4 positions.

To reduce the effect of using one résumé as Explicit Relevance Feedback, we propose to add two exponents to each summation of Equation 5.2. One with the objective to decrease the influence of rejected résumés, for example a square root, and one other to increase the strength of those that were considered as selected, e.g. 2nd power. In the tests done to determine the viability of this idea, we observed an increment of the MAP over the problematic

cases. However, we still have to study the impact of these exponents on the none-problematic cases and on the use of more than 1 résumé as Relevance Feedback.

In the Experiment 3, the results got by the Explicit Relevance Feedback using the Bottom position were the less performing. We can see, that in order to reach the performance obtained by the Top position we should have to use at least 2 more résumés as feedback. This less performing behavior was due to the low number of Selected résumés at the end of the rankings. Nonetheless, despite its performance, we think that this position for the Explicit Relevance Feedback is important. First, it may be of interest to recruiters because, in our opinion, it is easier and faster to classify irrelevant résumés than pertinent ones. Second, to leave behind a relevant résumé would contradict the objective of the résumés priority list, to reduce the time necessary for the recruiters to find the applicants that must be approached. Finally, an interesting applicant marked as irrelevant, may not have a second chance in the selection process, as their opposite have during an interview.

We thought that the use of résumés of the Bottom and the Top (Both) at the same time would be better than using résumés from the Top position. The reason is that it would allow us to merge the benefits of the two positions, to find the interesting and irrelevant résumés at the top and at the bottom of the priority lists. However, that was not the case, instead of it, this position had a performance which was below the Top position. This particular behavior of the results can be explained with 2 reasons. There were enough relevant résumés at the first places of the priority lists to improve the results. But, there were not enough pertinent résumés at the bottom in order to boost the performance and overtake the results obtained with only the Top position.

The Random position, in Experiment 3, was used to determine at what point the Explicit Relevance Feedback could be done without caring of the position from where the résumés were taken to do the feedback. We saw that as the number of résumés chosen randomly in the Explicit Relevance Feedback got closer to 10, to chose the first résumés of a priority list became irrelevant in the process. Therefore, we know that for an e-Recruitment system, based on our ideas, the best performance of the Explicit Relevance Feedback would be gotten by choosing the n top résumés if $n \leq 10$. If $n > 10$, we infer that the system performance will be equal or better choosing the résumés randomly from the priority list.

In the case of the Experiment 4, we observed that the application of the Explicit Relevance Feedback iteratively and using blocks of 2 résumés can boost the results more and faster than using large blocks of résumés. Nonetheless, the way to change the Explicit Factor, Orignal lambdas or Iterative lambdas, had an impact on the results. We saw that the Original lambdas had a much better performance, reaching a MAP value of 0.7554, than the Iterative Lambdas, with a MAP of 0.7020. Furthermore, the value obtained by the Original lambdas at 10 résumés surpassed the one obtained by the Top position of the Experiment 3 (0.7249).

The cause of the low performance of the Iterative lambdas, in the Experiment 4, may be related due to the fact that the lambda of the nth application (Λ_n) is at the end the original one (λ) multiplied several times by n Explicit Factors; this is presented in Equation 5.9:

$$\Lambda_n(r) = \lambda(r) \cdot \prod_{x=1}^{n} EF_x(r) \tag{5.9}$$

where n is the number of applications done, r is a résumé from the analyzed job in the Experiment 4 and EF_x the xth Explicit Factor. In consequence, the fact of multiplying several Explicit Factors at the same time may reduce their efficiency to create a better priority list.

To obtain better results for the Experiment 4, we think we should find another way to calculate the lambdas in each iteration, i.e. a way that would be independent from the precedent results, but also from the original ones.

5.5 Conclusions

In this chapter we presented a set of experiments based on the idea that as the selected résumés have more in common with themselves than and with respect to the rest of résumés, then it should be possible to create priority lists of résumés according to a proximity measurement. We did these experiments using a method that we developed called the Average Inter-Résumé Proximity or AIRP. The method consists in obtaining for each résumé of a job, its average proximity with respect to the other ones of the same job.

Four experiments were done over a set of 185 jobs from the Recruitment Corpus (Chapter 3). We made use of the Dice's Coefficient with relative frequency as measure of AIRP. To improve the results of AIRP, we put into test the Relevance Feedback in two modalities. An Explicit Relevance Feedback which is the traditional one, where a user gives an input to improve the request; and an Implicit Relevance Feedback, where the feedback is inferred from th results of AIRP. This last method works as unsupervised input. All the experiments were evaluated using the Mean Average Precision (MAP) and were compared with 2 baselines, one based on the arrival time and one in a random sorting.

In general we obtained good results in our experiments, for example, we determined that the use of a priority list, generated with AIRP would be a better tactic than analyzing résumés randomly.

We found that the use of the Explicit Relevance Feedback can improve the resulting priority lists but, it should be taken into account the quantity of résumés and the position from where they are chosen. We found that, the Explicit Relevance Feedback works better using blocks between 2 and 10 résumés chosen from the top of the priority lists. However, for

more than 10 résumés, the best option would be to choose randomly the résumés from the priority list. This means that the performance of the Explicit Relevance Feedback using a block of more than 10 résumés, can have a similar performance than choosing randomly the résumés rather than from a specific position, like the top.

In addition, we tested an iterative Explicit Relevance Feedback but using small blocks of two résumés. We found that this method improves more and faster the performance of AIRP. In other words, we can get better priority list for the recruiters, using an iterative Explicit Relevance Feedback and small blocks of résumés. Different sizes of blocks should be tested to determine how the size of the them affect the performance of the iterative application of the Explicit Relevance Feedback.

Not all the experiments showed the expected performance, like the Implicit Relevance Feedback or the Explicit Relevance Feedback using résumés from other positions, e.g. bottom of the list. However, we found the reasons and the possible solutions to overcome the unexpected problems.

We believe that to improve the results obtained by the relevance feedback, in its two modalities, we need to modify the way they work. Nowadays, the implemented relevance feedback changes the average proximity by itself, instead of the weights of the vectors used to calculate the proximity. Adapting the average proximity only affects the totality of elements and not the elements by themselves, those that made more similar or different two résumés. If we redefine the vector weights, we can recalculate the proximity of résumés based on the elements that were important to the recruiter regardless the existence of less relevant elements. In other words, in the Selected résumés there must be features that even if they coexisted with a great number of irrelevant ones, these last did not influence the recruiter to make the résumé a Rejected one.

To conclude, the use of AIRP confirms the idea that it is not necessary to use the job offer in order to classify résumés by their priority. Nonetheless, at this point, we ask our selves how the Average Inter-Résumé Proximity could be used along with the proximity between a job offer. Furthermore, in which degree merging both proximities calculations could improve the performance of an e-Recruitment system. This problematic will be presented and discussed in the following chapter.

Chapter 6

Using the Proximity Between a Résumé, a Job Offer and others Résumés

Throughout thesis, we have presented that most of the e-Recruitment systems, which works on the ranking of résumés, are based on the comparison between résumés and job offers. More specifically, they make use of proximity measures, in most cases, to determine how similar or dissimilar is the information located in both kinds of documents. The résumés more proximal to a certain job offer are those that are marked as the most relevant for the recruiters.

Although this methodology is valid to rank résumés, we decided to test whether the résumés could be used alone in this kind of systems as it was done by Faliagka et al. (2013) However, the difference between us and the work of Faliagka et al. (2013) is that we did not want to make use of previous selection process to train a system.

In consequence, in Chapter 4, we started analyzing whether a set of proximity measures, applied only over résumés, were sensible enough to detect details about the selection process of candidates. As well, we studied which were the implications and effects of using the proximity measures in certain ways over résumés. From this analysis we found that the résumés from selected candidates, i.e. those that were approached by a recruiter, were not only different to those from the rejected ones, they were more proximal between themselves. In other words, the selected candidates shared more in common with themselves than or with respect to their counterpart.

From the results and observations obtained from Chapter 4, we decided to test, in Chapter 5, whether the average proximity between résumés could be used to create résumés priority lists. In other words, we studied whether it was possible to rank the résumés by relevance for the recruiters according to the average proximity of the information located

only in the résumés. Among the different outcomes obtained from the experimentation, we concluded that the average proximity between résumés is a feasible method in order to rank the résumés by their relevance.

Arriving to this point, we asked us how to create a method which could take the concept of the average proximity between résumés and at the same time the basis of several e-Recruitment systems: the proximity between a job offer and the résumés. The reason is that we consider that the résumés and a job offer contain different kinds of information which are presented through several points of view, but that converge in a common point, a same job. The merging of these two types of information may be a way to improve the performance of the creation of priority lists in a HRM context and to include intrinsically at the relevance of each résumé, different points of view of the job.

Although there are several methods available in the literature to merge the results of different approaches, we decided to use in this chapter one method called the trivergence. The trivergence is not only a merging approach but an innovative method which allows us to compare the information given by the proximity of one résumé with respect to other ones, as we did in Chapter 5, but also with a job offers.

This chapter has been divided in 4 parts. First, we start with the objectives of this chapter (Section 6.1), which are followed by the Methodology in Section 6.2. There we present in depth the Trivergence. Then, in Section 6.3 we will introduce the results which will be discussed in Section 6.4. Finally, the conclusions are shown in Section 6.5.

6.1 Objectives

For this chapter, we have set 2 objectives:

- Determine how to use the trivergence in order to have an adequate proximity method which can be applied in a HRM context.
- Determine whether the trivergence is a performing method to create a résumé priority list for the recruiters.

6.2 Methodology

In this section, we present the methodology used in this chapter to test and apply the idea of trivegergence in a HRM context.

We start with the definition of the Kullback-Leibler Divergence, an important component of the trivergence. Then we continue with the definition of the trivergence and the trivergence applied to the HRM. We end with the description of the experiments and the data utilized to do them.

6.2.1 Kullback-Leibler Divergence

The idea of *divergence* was introduced by Mahalanobis (1930) and it is based on the difficulty to differentiate two probabilistic distributions. If they are very similar it is harder to distinguish them and vice-versa. Therefore, a divergence measure is defined as the measure that has the property of increasing its value as two distributions become more dissimilar and in consequence easier to recognize (Pardo, 2005).

Since the studies of Mahalanobis, several persons have designed and studied the divergences. For instance, Balakrishnan and Sanghvi (1968), Lin (1991), Rathie and Kannappan (1972) and, Kullback and Leibler (1951). This last one will be presented along this section.

The Kullback-Leibler divergence (KL) is an information-theoretic measure that determines how dissimilar are two probabilistic distribution (X, Y) in the same space (Baker and McCallum, 1998; Bigi, 2003). In other words, it measures the quantity of information of X that is contained in Y (Cover and Thomas, 2005, page 19). The divergence of Kullback-Leibler is not a metric as it is not symmetric, in other words $\mathrm{KL}(X, Y) \neq \mathrm{KL}(Y, X)$.[1]

For two probabilistic distributions X and Y, the Kullback-Leibler (KL) distribution is given by Equation 6.1.

$$\delta_{KL}(X, Y) = \sum_{\omega \in X} X_\omega \log_2 \frac{X_\omega}{Y_\omega} \tag{6.1}$$

where X_ω (Equation 6.2) is defined as the number of occurrences (C) of element ω in X with respect to the cardinality of X (the total number of different elements of X).

$$X_\omega = \frac{C(\omega, X)}{|X|} \tag{6.2}$$

In the case of Y_ω, it is defined as Equation 6.3.

$$Y_\omega = \begin{cases} \dfrac{C(\omega, Y)}{|Y|} & \text{if } \omega \in Y \\ \dfrac{1}{T} & \text{if } \omega \notin Y \end{cases} \tag{6.3}$$

[1] Although there is a symmetric version of this divergence, in this chapter we will focus on the non-symmetric one.

where $C(\omega, Y)$ is the number of occurrences of element ω in Y, $|Y|$ is the cardinality of Y and $1/T$ is a *smoothing*[2] value only given when ω does not exist in Y; $T = |X| + |Y|$. There are several ways to calculate a smoothing like Kat's back-off model (Katz, 1987) and Good-Turing estimation (Good, 1953). In our case, the formula presented is the smoothing proposed by Torres-Moreno (2015) for the Kullback-Leibler Divergence.

The Kullback-Leibler divergence and its variants have been used in several Natural Language Processing tools, for example, to do text classification (Baker and McCallum, 1998; Bigi, 2003), spoken-document retrieval (Liu et al., 2007) and automatic summarization evaluation (Saggion et al., 2010).

6.2.2 Trivergence

Based on the divergence, Torres-Moreno (2015) introduced the idea of *trivergence* a statistical measure which compares three elements at the same time through the evaluation of their probabilistic distributions. The trivergence is based on the application multiple divergences, like the one of Kullback-Leibler (Kullback and Leibler, 1951) or Jensen-Shannon (Lin, 1991), in order to get the dissimilarity of 3 elements.

According to Torres-Moreno (2015), the trivergence can be done through the multiplication[3] of 3 divergences as presented in Equation 6.4:

$$\tau(X, Y, Z) = \delta(X, Y) \cdot \delta(X, Z) \cdot \delta(Y, Z) \tag{6.4}$$

where X, Y and Z are three probabilistic distributions, τ is the trivergence and δ is a divergence measure. To use this formula, Torres-Moreno (2015) set a cardinality condition: $|X| > |Y| > |Z|$; this was done to always measure the dissimilarity of a small probabilist distribution with respect to a larger one.

For this chapter, we decided to use the divergence of Kullback-Leibler to calculate the trivergence. The formula used in the experimentation will be presented as follows.

Multiplicative Trivergence of Kullback-Leibler

The trivergence based on the multiplication of Kullback-Leibler divergences has been defined by Torres-Moreno (2015) as Equation 6.5:

$$\tau(X, Y, Z) = \sum_{\omega \in X} X_\omega \log_2 \frac{X_\omega}{Y_\omega} \cdot \sum_{\omega \in X} X_\omega \log_2 \frac{X_\omega}{Z_\omega} \cdot \sum_{\omega \in Y} Y_\omega \log_2 \frac{Y_\omega}{Z_\omega} \tag{6.5}$$

[2] A smoothing is the reevaluation of zero values into non-zero values (Jurafsky and Martin, 2000, page 205) for expressions (n-gram, word, etc.) that do not exist in a document (Jackson and Moulinier, 2002, page 39)

[3] In the article of Torres-Moreno (2015) it is presented another method to do the trivergence. This consist in the composition of two divergences.

where X, Y and Z are three probabilistic distributions fulfilling the cardinality condition of $|X| > |Y| > |Z|$. For this formula, X_ω was defined as Equation 6.2, Y_ω as Equation 6.3 and Z_ω as Equation 6.6:

$$Z_\omega = \begin{cases} \dfrac{C(\omega, Z)}{|Z|} & \text{if } \omega \in Z \\ \dfrac{1}{T} & \text{if } \omega \notin Z \end{cases} \tag{6.6}$$

where C is the number of occurrences of ω in the one of the three probabilistic distributions and T is the smoothing value to avoid indeterminate values. For all the cases $T = |X| + |Y| + |Z|$.

Following the characteristics of the divergences, in the case of the multiplicative trivergence of Kullback-Leibler, low values of τ indicate less dissimilarity between the 3 distributions.

6.2.3 Trivergence Applied to HRM

In this chapter, we decided to model the creation of the résumé priority lists as a problem to solve using the trivergence of distributional probabilities. Along this section, we explain the modifications and considerations took into account to apply the trivergence in a HRM context.

Let us consider a job J composed of n résumés, $J = \{r_1, r_2, \ldots, r_n\}$, and a job offer linked to the job J. We have defined the three elements of a trivergence as:

- X: the Job Offer
- Y: a résumé $r_y \in J$
- Z: the concatenation of all the résumés from J without r_y, in other words $J \setminus r_y$

We decided to represent each element of the trivergence in a n-grams of words quadruplet (unigrams, bigrams, SU3 and SU4; see Section 3.2.5). Therefore, a trivergence of a résumé was done 4 times.

To obtain only one trivergence value by résumé, we decided to do the average of the 4 resulting calculations. This process is similar to the one done by Fresa (Saggion et al., 2010), where an average of the divergence of n-grams is done to determine how similar is a summary and its source.

In addition, we chose not to take into account the existence of the cardinality condition, as indicated Torres-Moreno (2015). The reason is that we cannot ensure that the cardinality size will be satisfied in all cases. We know that Z is larger than X and Y, nonethe-

less, we do not know either $X \leq Y$ or $X > Y$. Therefore, the multiplicative trivergence of Kullback-Lleibler can generate 8 combinations.[4] In Equation 6.7, we present the 8 possibles combinations according to our elements:

$$\tau(X,Y,Z) = \begin{cases} \delta(X,Y) \cdot \delta(X,Z) \cdot \delta(Y,Z) \\ \delta(Y,Z) \cdot \delta(Y,X) \cdot \delta(Z,X) \\ \delta(Z,Y) \cdot \delta(Z,X) \cdot \delta(Y,X) \\ \delta(X,Y) \cdot \delta(Y,Z) \cdot \delta(Z,X) \\ \delta(X,Z) \cdot \delta(Z,Y) \cdot \delta(Y,X) \\ \delta(Z,X) \cdot \delta(Z,Y) \cdot \delta(X,Y) \\ \delta(Y,X) \cdot \delta(Y,Z) \cdot \delta(X,Z) \\ \delta(X,Z) \cdot \delta(X,Y) \cdot \delta(Z,Y) \end{cases} \tag{6.7}$$

Although we considered that the trivergence can be represented by 8 possible combinations, we inferred that not all of them would have the same performance in a HRM context. Therefore, it should be tested each combination.

6.2.4 Experiments

For this chapter, we decided to do two experiments. The first one consisted in finding which is the most performing combination, in average, of the multiplicative trivergence of Kullback-Leibler in a HRM context. More specifically, for a set of jobs, we calculated the trivergence of Kullback-Leibler of their hired applicant's résumé using the 8 possible combinations. Then we compared the 8 resulting values to chose the one that gave us the lowest average trivergence, i.e. the one that showed the less dissimilarity between the hired applicants' résumé', the rest of résumés of the job and the job offer.

The second experiment consisted in the application of the most and less performing combination of the multiplicative trivergence of Kullback-Leibler in a set of jobs. We compared the results obtained with 3 baselines which are described in the following section. For this experiment, the performance was measured with the Mean Average Precision or MAP (Section 5.2.3).

Baselines

The first baseline consisted in sorting randomly the résumés in the priority list as we did in (Chapter 5). The second baseline corresponded in the use of Average Inter-Résumé Proximity (AIRP), following the same methodology presented in Chapter 5 as well. We chose for the last baseline, the calculation of the proximity between the résumés and the job offer

[4] Some of these combinations can be deleted if we would use the symmetric version of Kullback-Leibler or other symmetric divergence.

used to attract the applicants.

In the case of the last baseline, we used as proximity measure, the Dice's Coefficient with Relative Frequency as we did in Chapter 4 and Chapter 5. The only difference, in this chapter, was its application between résumés and a job offer, instead of only between résumés.

6.2.5 Data

The experiments of this chapter were done over a set of 64 jobs of the Recruitment Corpus (Chapter 3) from which we had access to their correspondent French job offer[5]. From this set, 15 were used in the first experiment, as their meta-data indicated who was the hired applicant.[6] The rest, 49 jobs, were used for the second experiment.

The job offers used in this chapter, were the documents posted on-line to invite job seekers to applicate for their correspondent position. These documents describe the job, the searched features (education, experience, skills) and, in some cases, the enterprise or company that will hire the person, the city where will be held the position, the salary and the tasks that will be done. The job offers were created by the recruiters with the advice of the employer in order to describe the job and present the needs of the company.

To use the job offers in the same way that we use the résumés, we converted the jobs offers into plain UTF-8 text using the method described in Section 3.2.1. As well, they were transformed into a vectorial model following the methodology indicated in Section 3.2.5.

6.3 Results

In this section we will present, in first place, which combination of the multiplicative trivergence of Kullback-Leibler was the most suited for a HRM context (Section 6.3.1). In second place, we will present the application of the trivegernce into a HRM context (Section 6.3.2).

6.3.1 Most Adequate Trivergence in a HRM Context

As we explained in Section 6.2.3, the multiplicative trivergence of Kullback-Leibler generates 8 possible combinations. However, we did not infer that all these combinations would have the same performance in a HRM context. Therefore, we tested the 8 possible combinations

[5]Note all the job offers are posted in French.
[6]The meta-data of the jobs do not always indicate who was/were the hired applicant(s).

using a set of 15 jobs. These jobs had the characteristic that were linked to a job offer and we knew who was the hired applicant.

We considered for the calculation of each combination of the trivergence $\tau(X, Y, Z)$, that X was the job offer, Y was the hired applicant's résumé and Z was all the résumés of a job excepting the one of the hired applicant. Figure 6.1 presents the notched box plots[7] of the average value, in natural logarithm scale, of the trivergence of each possible combination.

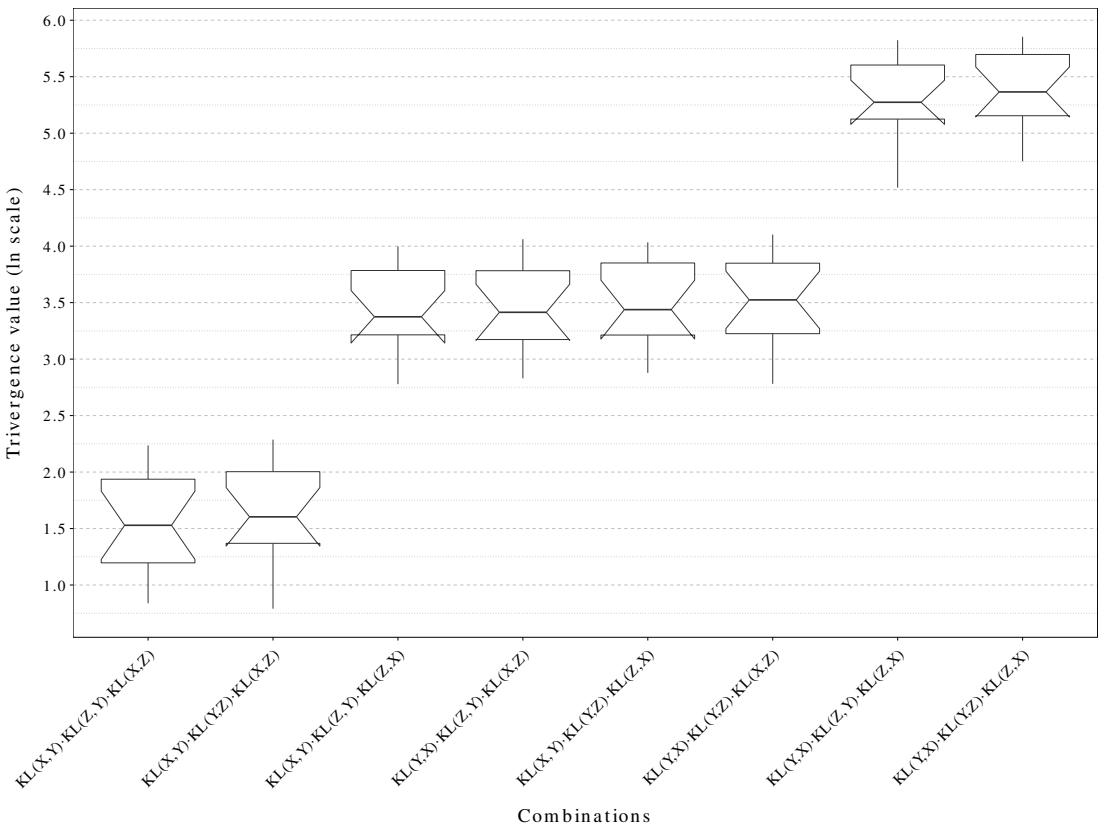

Figure 6.1 *– Boxplots of each possible combination of the trivergence's elements.*

To determine whether the average values obtained for each combination were statistically different, we used an analysis of variance (ANOVA) using a *Welch's t test*.[8] From this test, we obtained a p-value $= 2.2 \times 10^{-16}$ which means that the average values are statistically different.

According to the boxplots presented in Figure 6.1, there are two combinations which are the most appropriated for a HRM context:

$$\tau(X, Y, Z) = KL(X, Y) \cdot KL(Z, Y) \cdot KL(X, Z)$$

[7]See Appendix A for understanding the box plots

[8]We had to use a Welch's t test for the ANOVA as the values of each combination did not have homogeneous variance.

$$\tau(X,Y,Z) = KL(X,Y) \cdot KL(Y,Z) \cdot KL(X,Z)$$

These, according to an ANOVA, are statistically equal with a p-value = 0.541. As well, based on Figure 6.1, there are two combinations which are the less performing in a HRM context:

$$\tau(X,Y,Z) = KL(Y,X) \cdot KL(Y,Z) \cdot KL(Z,X)$$

$$\tau(X,Y,Z) = KL(Y,X) \cdot KL(Z,Y) \cdot KL(Z,X)$$

An analysis of variance indicated that these two combinations are statistically equal with a p-value = 0.460.

6.3.2 Application of the Trivergence into a HRM Context

For the second experiment, we made use of the results obtained in the previous experiment of this chapter and we tested the multiplicative trivergence of Kullback-Leibler over 49 jobs. More specifically, we used the trivergence defined as:

$$\tau(X,Y,Z) = KL(X,Y) \cdot KL(Z,Y) \cdot KL(X,Z)$$

to create a résumé priority list for each 49 jobs. We chose this trivergence combination because it was the one that, according to the results of the previous experiment, were the most appropriate for a HRM context.

As stated in Section 6.2.4, we decided to compare the results of the most appropriated trivergence combination with the less performing one but also against AIRP (Section 5), a random ranking and, the proximity between résumés and a job offer. In the case of the less appropriate trivergence combination, we decided to use:

$$\tau(X,Y,Z) = KL(Y,X) \cdot KL(Y,Z) \cdot KL(Z,X)$$

With respect to AIRP, we applied this method over the corpus of 49 jobs. For the random priority lists, we generated for the 49 jobs 100 rankings sorted randomly. Finally, for each of the 49 jobs we calculated the proximity between their belonging job offer and set of résumés; the proximity was done using Dice's Coefficient and relative frequency. Figure 6.2 presents the results of the most and less appropriated combination of trivergence compared with the three baselines.

As it can be seen in Figure 6.2, the most performing trivergence combination gave a MAP of 0.7074, which is over the result given by the proximity between the résumés and the job offer (0.7014), AIRP (0.6878), the less appropriate trivergence combination (0.6389) and the random sorting of applicants (0.5952).

We can observe, as well, that the 95% confidence interval bars of all the methods overlap at least once. Thus, we tested whether all the results were statistically different using

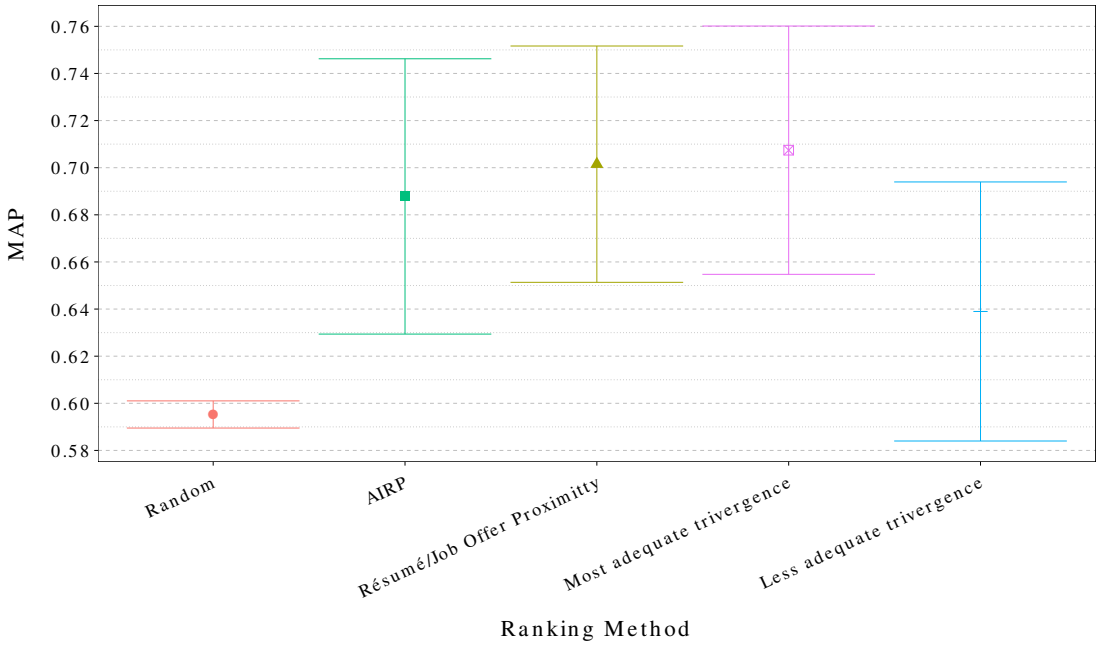

Figure 6.2 – *Results of the analysis of 49 jobs with AIRP, with random sorting, with the proximity between résumés and a job offer, and with the most and the less adequate combination of trivergence.*

an analysis of variance (ANOVA) and we obtained a p-value $= 9.59 \times 10^{-8}$, which is below $\alpha = 0.05$, meaning that at least one mean is statistically different.

To determine which resulting average were different, we used a Tukey's HSD test as *post hoc* analysis. The results of the test are presented in Table 6.1.

As presented in Table 6.1, the *post hoc* analysis indicates that only the results of the random method are statistically different from the most adequate trivergence, AIRP and, the proximity between the résumés and the job offer. The other methods are statistically equal between them.

6.4 Discussion

Although at first sight the results from the first experiment (Section 6.3.1) do not express nothing more than the performance of the trivergence combinations, in fact they have an HRM interpretation as we inferred previously. In the case of the most adequate trivergence combination, according to the results:

$$\tau(X,Y,Z) = KL(X,Y) \cdot KL(Z,Y) \cdot KL(X,Z)$$

presented as diagram in Figure 6.3, we can interpret it as:

Table 6.1 *– Results of the Tukey's HSD test from the analysis of the application of the trivergence and its comparison with a set of baselines.*

Methods	Adjusted p-value	Statistically different ($\alpha = 0.05$)
Less adequate trivergence Most adequate trivergence	0.4658	NO
AIRP Most adequate trivergence	0.9897	NO
Résumé/Job offer Most adequate trivergence	0.9999	NO
Random Most adequate trivergence	0.0013	YES
AIRP Less adequate trivergence	0.7649	NO
Résumé/Job offer Less adequate trivergence	0.5583	NO
Random Less adequate trivergence	0.5742	NO
Résumé/Job offer AIRP	0.9974	NO
Random AIRP	0.0147	YES
Random Résumé/Job offer	0.0029	YES

- $KL(X,Y)$: the hired applicant (Y) is defined by the job offer (X).
- $KL(Z,Y)$: the hired applicant (Y) is determined by all the résumés different from the one of the hired applicant (Z).
- $KL(X,Z)$: the résumés different from the one of the hired applicant (Z) are delineated by the job offer (X).

This interpretation is due to the fact that, as we indicated in Section 6.2.1, the divergence of Kullback-Leibler determines how dissimilar are two probabilistic distributions A and B by measuring the quantity of information of A that is contained in B. In other words, there is a great quantity of information of the job offer (X) that is located in the résumé of the hired applicant (Y) and in the rest of résumés (Z). As well, that most of the information of all the résumés of a job, excepting the one of the hired applicant, (Z) are represented inside the résumé of the hired candidate (Y).

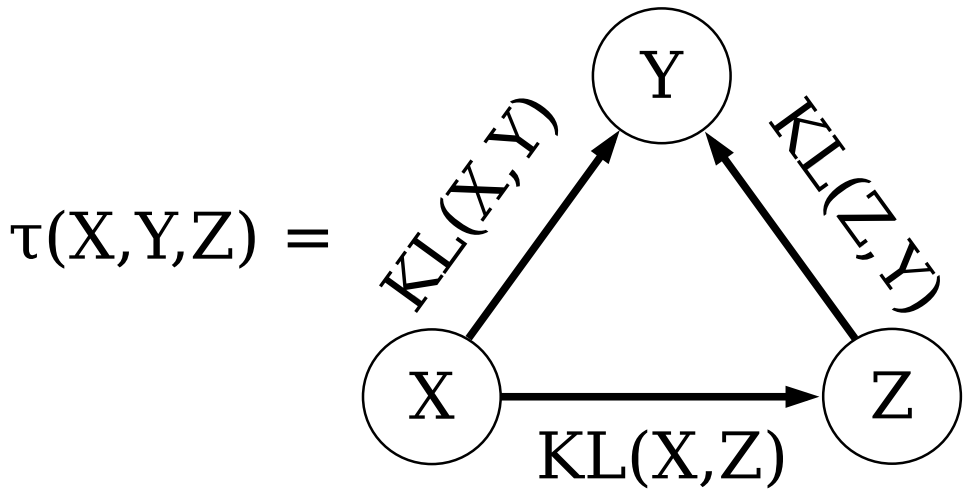

Figure 6.3 *– The trivergence $\tau(X,Y,Z) = KL(X,Y) \cdot KL(Z,Y) \cdot KL(X,Z)$ presented as a diagram, where X is the job offer, Y the résumé of the hired applicant and Z the rest of résumés sent to a job.*

In the case of the second most adequate combination:

$$\tau(X,Y,Z) = KL(X,Y) \cdot KL(Y,Z) \cdot KL(X,Z)$$

presented as diagram in Figure 6.3, we can see a similar HRM interpretation. The difference with the previous combination resides in the use of the divergence $KL(Y,Z)$ instead of $KL(Z,Y)$. This means, that most of the information of the hired applicant (Y) is located in all the résumés of a job different from the hired applicant (Z).

The less performing combinations, which in fact are the complement of the most adequate combinations, do not have a logical HRM interpretation:

$$\tau(X,Y,Z) = KL(Y,X) \cdot KL(Y,Z) \cdot KL(Z,X)$$

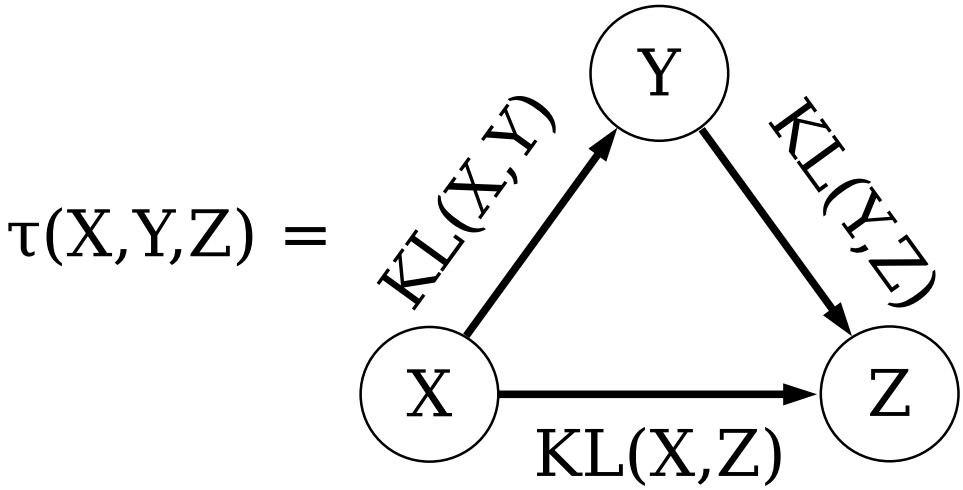

Figure 6.4 *– The trivergence $\tau(X,Y,Z) = KL(X,Y) \cdot KL(Y,Z) \cdot KL(X,Z)$ presented as a diagram, where X is the job offer, Y the résumé of the hired applicant and Z the rest of résumés sent to a job.*

$$\tau(X,Y,Z) = KL(Y,X) \cdot KL(Z,Y) \cdot KL(Z,X)$$

In both cases, they could be interpreted that a job offer (X) is not only defined by the hired applicant (X) but also by the rest of applied résumés (Z). This goes against the purpose of a job offer, to attract applicants.

At the second experiment, without considering the Tukey's HSD *post hoc* analysis, we observed that the most and less adequate combination of the trivergence have a better a performance with respect to the random sorting of applicants. This means that the trivergence, even the less adequate combination, can sort the résumés by pertinence better than choosing randomly the résumés of a job. As well, that the most adequate trivergence has a better performance that the less adequate one.

With respect to the results given by the Tukey's HSD test, we saw that all the methods are statistically different from the random sorting of résumés, excepting the less adequate trivergence. The rest of the results were considered as statistically equal. This validates the behavior of the 95% confidence interval bars shown in Figure 6.2

We certainly expected for the most adequate trivergence a performance that would be higher than the one obtained by AIRP and the proximity between résumés and job offers. In other words, a performance that would express the merging of two ideas. The reason is that the trivergence takes into account the dissimilarity between one résumé and the rest of them, and at the same time the dissimilarity between the job offer and a résumé. Nonetheless, we got similar performances for these three methods.

We do not consider that the obtained results make less relevant the use of the trivergence in a e-Recruitment system. In fact, we must consider that the trivergence represents a dissimilarity measure, while the other two methods used correspond to a similarity one, i.e. they treat different the information to calculate the proximity of documents. It may be possible to get better results changing of measure, e.g. the Jensen-Shanon divergence, of method to do the trivergence, like composition of divergences, normalization of the data or even the smoothing for the unseen n-grams.

As well, it is important to say, that we do not know in which proportion the results of the trivergence are due to the proximity among résumés or due to the dissimilarity between the job offer and the résumés. This makes difficult to do a deep analysis of the real performance of the trivergence in a HRM context. However, we can say that this method is feasible for an e-Recruitment system.

In this chapter, we compared indirectly the methods of AIRP and, the proximity between résumés and job offers. Both methods obtained similar performances that according to a *post hoc* test were statistically equal. The results made us think, in first place, that we can get a similar ranking using whether a set of résumés or the job offer. In second place, that a set of résumés, sent for the same position, may represent indirectly the job offer. Moreover, as these two methods do not utilize completely the same information it may be interesting to merge them to improve or reinforce the generation of priority lists.

6.5 Conclusion

In this chapter, we presented an innovative method to compare the proximity of the résumés with a job offer and the proximity between only résumés. This comparison was done through a method called the trivergence. The trivergence is a probabilistic measure where 3 probability distributions are compared at the same time by the multiplication of divergences. The goal of this method is to determine how dissimilar are the three probabilistic distributions.

At this chapter, we made use of the trivergence to generate a résumé priority list, i.e. a list that ranks the résumés according to their relevance for a job. The calculation of the trivergence was done over the information located in the résumés and the job offer. Two experiments where done, one to adequate the trivergence to the HRM context and one other to evaluate its performance.

The results obtained in this chapter have shown us that the trivergence could be used in HRM context. Nevertheless, we found that it is important to set in a specific order the elements of the trivergence. Not all the possible combinations to set the elements are adequate for a HRM context because they do not give the same performance. Moreover, only a pair of combinations have a logical HRM interpretation. In other words, we found that

the most performing trivergences are those where the job offer describes the résumé of all the applicants. We observed, as well, that the résumé of the hired applicant describes the résumés of the rest of candidates and vice-versa. Nonetheless, this does not look to affect the performance of the combinations greatly.

Although the performance of most adequate trivergence was similar to the one of AIRP, and, the proximity between résumés and job offers, we consider that the trivergence can contribute to develop a different e-Recruitment system. The reason is that this method takes into account two different ways to rank the résumés. It may be only necessary to redefine some aspects of the trivergence, like the smoothing or normalization, in order to get a better performance. As well, we consider that using the relevance feedback, as we did for AIRP in Chapter 5, could be a way to boost the results.

To conclude, we can say that the results given by the trivergence showed us that it is a method that could be used for an e-Recruitment system and that can have a logical HRM interpretation. Furthermore, that the output obtained is similar to other methods. Nonetheless, we consider that we have to explore deeper how the trivergence works in order to find the best way to improve its results and being able to merge them with other outputs.

Chapter 7

Conclusion and Perspectives

In this thesis we presented the study of different and innovative methods for e-Recruitment systems. The objective was to set new ideas and algorithms that would allow us to create, in a future, a prototypical e-Recruitment system. A set of methods that could assist recruiters by telling them which résumés should be read in first place.

To be more conscience, we studied the creation of methods capable to rank, by relevance, job applicants but using, mainly, the proximity between résumés. Moreover, the studied methods had the particularity that were not necessary to know previous selection processes to do the ranking of newer ones.

Three big experiments were done in this thesis, one to determine how to use the proximity of résumés, another to apply the proximity of résumés and one more where the proximity between résumés was applied along to the proximity between résumés and job offers.

The first experiment, Chapter 4, was the base of this thesis. It had for objective to determine whether and how the calculation of the proximity between résumés could be done in a HRM context. We tested different proximity measures and vector weights, over an inference model, to determine which of them were the most performing to analyze résumés. The proximity measures were Cosine Similarity, Manhattan and Euclidean Distance, Dice's Coefficient and Jaccard's Index. While the used vector weights were TF-IDF and, relative and absolute frequency. The model consisted in the inference that, in average, the résumés of Selected applicants, i.e. those approached by a recruiter, had more in common with themselves than and with respect to the rest of résumés in a job. The experiment was done over 224 jobs.

The results of the first experiment were evaluated through an Analysis of Variance or ANOVA. This was done to know whether the results were statistically different or not. As well, we made use of the ANOVA to measure the effect of the proximity measures and vector weight over the inference. The results showed us that our inference was correct, the Selected résumés have more in common with themselves than and with respect to the rest of

résumés. Nonetheless, the results indicated us that it is important to choose correctly the proximity measure and vector weight. The reason is that not all the measures with a certain vector weight work equally over the résumés; furthermore, some of them can lead to false or biased results. For example, while we saw that Dice's Coefficient with relative frequency is an excellent measure to be used with résumés, on the opposite side we found that Manhattan Distance with absolute frequency was the less performing one.

In the second experiment we used the results obtained previously and we developed the method AIRP or Average Inter-Résumé Proximity. It corresponded to the calculation of the average proximity between résumés to rank each applicant. The idea was founded on the fact that, sharing the Selected résumés more common characteristics with themselves, it could be possible to rank them at the top of a list using the average proximity. The analysis was done over 185 real jobs and consisted in measuring how AIRP was capable to rank at the top, the candidates that were considered as Selected by recruiters.

In addition, during the second experiment, we tested the use of the relevance feedback. The relevance feedback consisted in generating an error input, either by asking to a recruiter or by inferring it, to adequate the AIRP results to users' needs.

The evaluation of the second experiment was done through the Mean Average Precision (MAP), a measure that allowed us to determine the precision and recall of a ranking. The results indicated that using AIRP is better than analyzing randomly the résumés of a job. Moreover, the use of Relevance Feedback can improve the results, however, it is essential to ask the recruiters' opinion and not inferring it.

For the last experiment, we decided to use the proximity between résumés along with the proximity between résumés and job offers. To do this, we made use of a model called the trivergence. The trivergence is a probabilistic dissimilarity measure, based on multiple divergences. It is apt to compare three different elements at the same time. The use of the trivergence was supported due to the fact that we considered that résumés and job offers contain different types of information and point of views. Thus, these had to be compared simultaneously.

We decided to use, for the last experiment, the multiplicative trivergence of Kullback-Leibler. In order other words, a trivergence based on the mutliplication of 3 divergences of Kullback-Leibler.

The experimentation was composed of two parts. In the first one we determined the most and less adequate trivergence for an HRM context. This was done beacuse the trivergence of Kullback-Leibler is not symmetric, thus, multiple combinations existed. In the second part, we used the most and less adequate trivergence over 49 jobs to rank their applicants.

The most and less adequate trivergence were evaluated with the MAP and compared to the performance of AIRP, a random sorting and the proximity between the résumés and the job offer. The results, showed that the most adequate trivergence is a better method than reading randomly the résumés of a job. However, its performance was similar to the one obtained by AIRP and the proximity between résumés and the job offer. Despite that, we consider that it is necessary only to change some parts of the methodology in order to increase the performance of the most adequate trivergence.

From the three experiments we can conclude, in first place, that it is feasible to use only the résumés in e-Recruitment systems, without having to know which was the recruiters' behavior in past selection processes. This opens the door to new researches where the field is the automatic analysis of résumés, e.g. which elements should be taken into account by the computers to consider that an applicant should be contacted. As well, it creates the possibility to develop new e-Recruitment systems where the job offers can be left aside during the ranking of applicants.

We can also conclude that it is necessary to study whether the proposed methodology, especially those related to the proximity calculation, are adequate enough for the HRM context. We saw along this thesis, that not all proximity methods work in the same way with the résumés. In some occasions, they can lead to false or biased results.

To end the conclusions, although we consider that the developed methods here are interesting and open several doors to new researches, there is a question that we could not answer for the moment: is this method feasible during real-time selection processes? Our experiments were done over already finished selection processes. In other words, the methods were tested over processes that contained all the applicants that participated, by invitation or initiative. Nonetheless, during a real scenario the applicants do not arrive all at once, in some occasions the arrival of the applicants can take weeks, even months, since the post of a job offer. We consider that the methods like AIRP, the relevance feedback and trivergence are timeless and could be applied several times during the selection processes, however, it is necessary to validate this with a new methodology.

7.1 Perspectives

In addition to the perspectives presented in each chapter, we have some of them which are general for this thesis.

The first place is the publication of the results done in this thesis. At this moment, we have only presented the results of Chapter 4 at an international conference. Through the publication our results, we will not only spread faster our methods, we will get an input from

the scientific community.

We would like to apply the same tests done in this thesis but to other languages, like English or Spanish. Although we consider that our methodology is language independent, to do the same tests over different languages would consolidate our results.

Finally, we would like to develop an e-Recruitment prototype, using the methods developed in this thesis, in order to test whether the methods work as expected in real cases.

Appendix A

Box plots

Box plots (Tukey, 1977) are a way to display batch of data. In this thesis, we make use of a variant form called *notched box plots* (McGill et al., 1978). They are composed of 7 elements which are shown in Figure A.1 and described as follows:

1. Line: it represents the median of the values.
2. Lower quartile: it corresponds to the first quartile ($Q1$).
3. Upper quartile: it corresponds to the third quartile ($Q3$).
4. Lower whisker: it is the first quartile ($Q1$) minus 1.5 times the interquartile range ($IQR = Q_3 - Q_1$).
5. Upper whisker: it is the third quartile ($Q1$) plus 1.5 times the interquartile range ($IQR = Q_3 - Q_1$).
6. Outliers: they are the values greater than the upper whisker or lesser than the lower whisker.
7. Notch: it represents the confidence interval of the median. Their range is the median $\pm 1.57 \times IQR$.

According to (Chambers, 1983, page 62), when the notches of two box plots do not overlap, as in Figure A.1, it is a strong evidence, although not a formal test, that the medians are statistically different with a 95% of confidence.

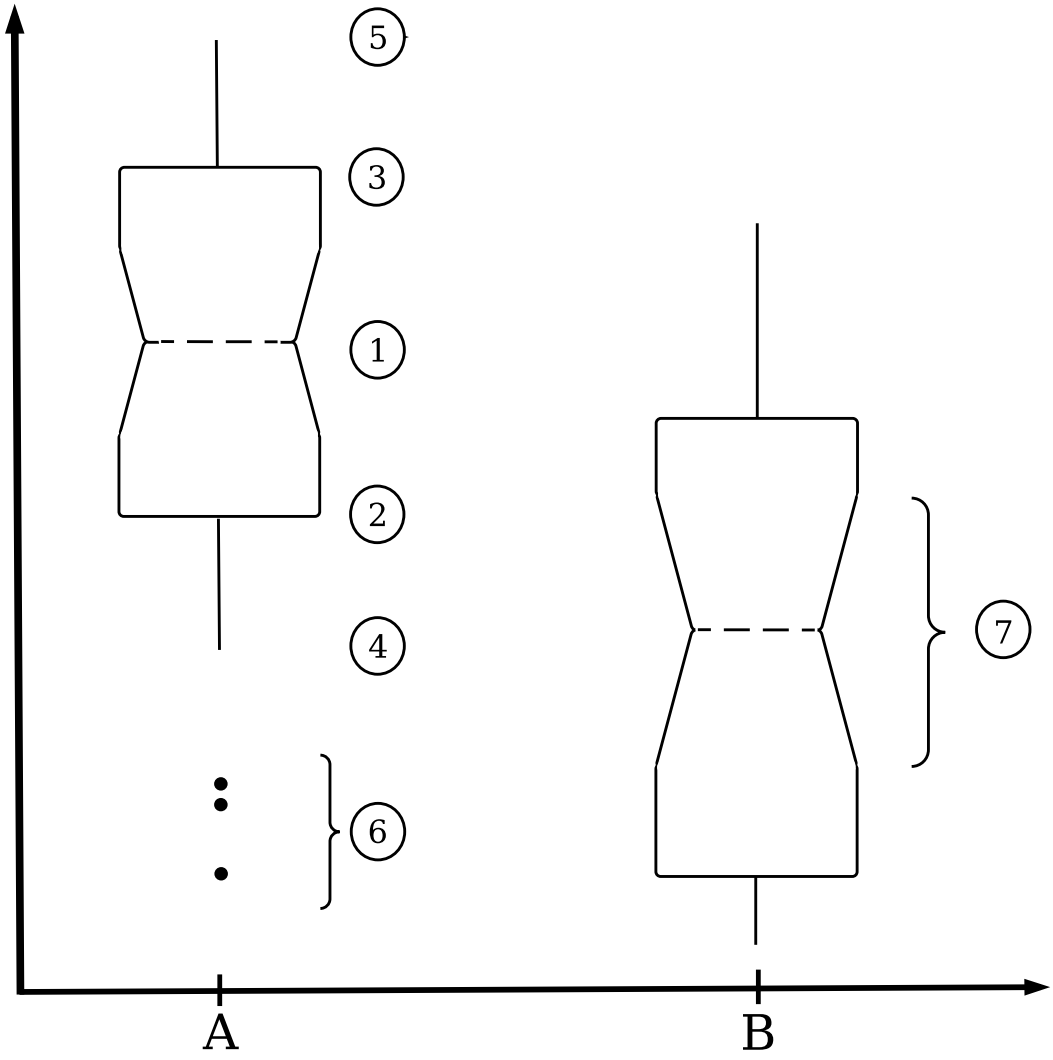

Figure A.1 *– Two examples of boxplots.*

Appendix B

Supplement Results of the Implicit Relevance Feedback

We present in Figure B.1 and Figure B.2 the results obtained by the Implicit Relevance Feedback but without the 95% confidence interval bars. The objective is to observe clearer the difference between the results.

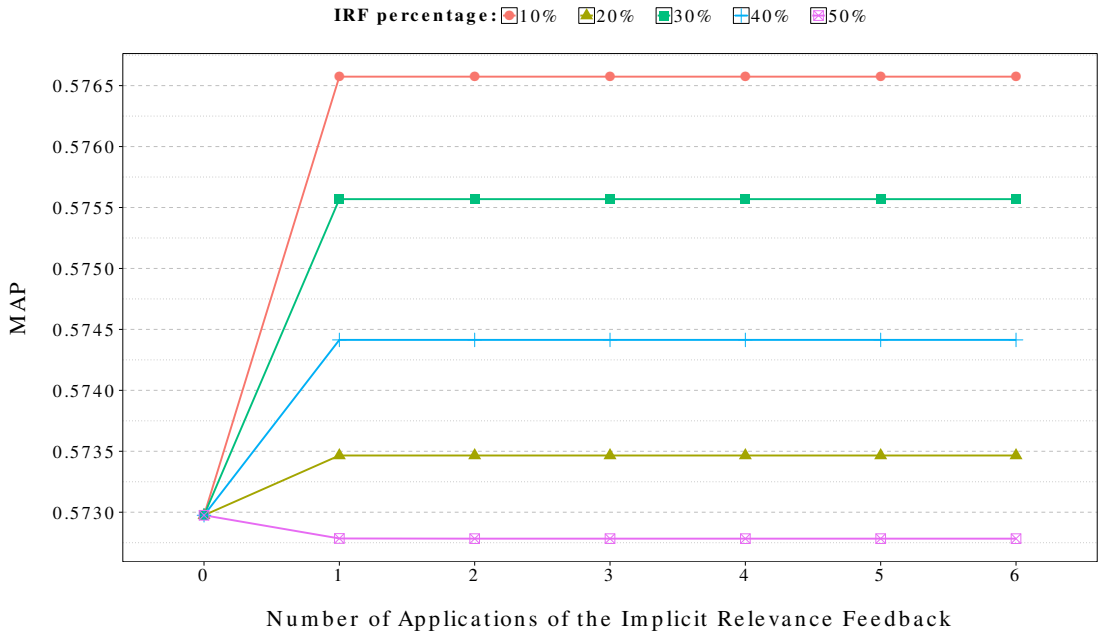

Figure B.1 *– Results of the values for the AIRP with Implicit Relevance Feedback for percentages from* 10% *to* 50%.

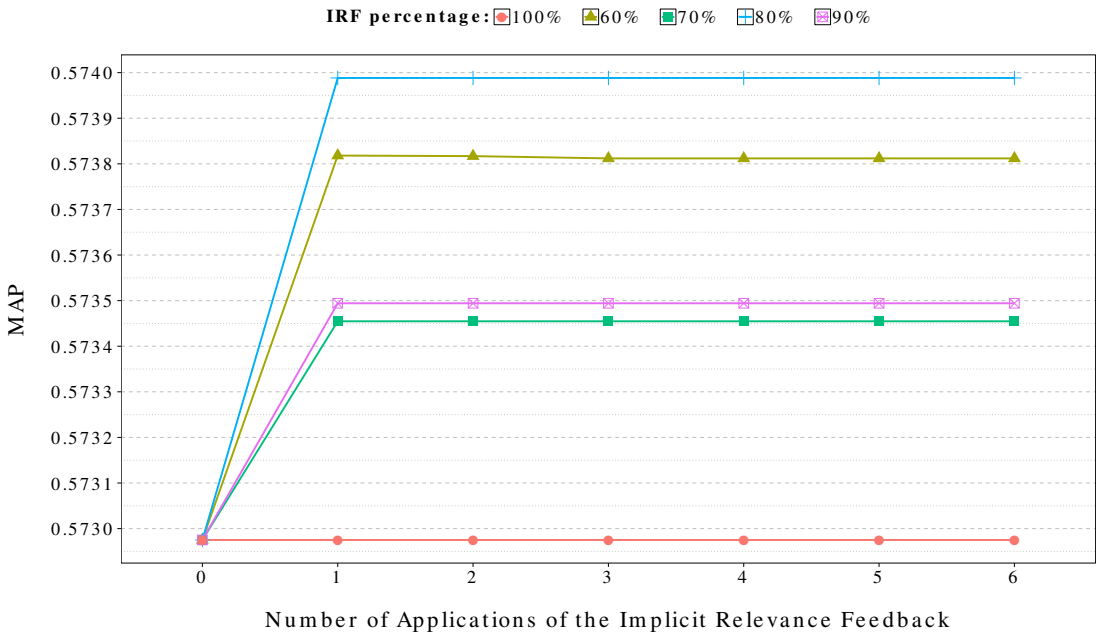

Figure B.2 *– Results of the values for the AIRP with Implicit Relevance Feedback for percentages from* 60% *to* 100%.

Bibliography

(Altman, 1990) D. G. Altman, 1990. *Practical Statistics for Medical Research*. Chapman & Hall/CRC Texts in Statistical Science. CRC press.

(American Heritage Dictionary of the English Language, 2014) American Heritage Dictionary of the English Language, 2014. Résumé. [Online; accessed 8-January-2015].

(Armstrong and Taylor, 2014) M. Armstrong and S. Taylor, 2014. *Armstrong's handbook of human resource management practice*. Kogan Page Publishers.

(Arthur, 2001) D. Arthur, 2001. *The Employee Recruitment and Retention Handbook*. AMACOM.

(Baker and McCallum, 1998) L. D. Baker and A. K. McCallum, 1998. Distributional Clustering of Words for Text Classification. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98)*, New York, NY, USA, 96–103. ACM.

(Balakrishnan and Sanghvi, 1968) V. Balakrishnan and L. Sanghvi, 1968. Distance between populations on the basis of attribute data. *Biometrics*, 859–865.

(Barber, 1998) A. E. Barber, 1998. *Recruiting employees: Individual and organizational perspectives*, Volume 8. Sage Publications.

(Barber, 2006) L. Barber, 2006. *E-recruitment Developments*. Institute for Employment Studies.

(Bartlett, 1937) M. S. Bartlett, 1937. Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 160*(901), 268–282.

(Bellman, 1961) R. E. Bellman, 1961. *Adaptive control processes: a guided tour*. Princeton, New Jersey: Princeton University Press.

(Bigi, 2003) B. Bigi, 2003. Using Kullback-Leibler Distance for Text Categorization. In F. Sebastiani (Ed.), *Advances in Information Retrieval*, Volume 2633 of *Lecture Notes in Computer Science*, 305–319. Springer Berlin Heidelberg.

(Bizer et al., 2005) C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, and R. Eckstein, 2005. The impact of semantic web technologies on job recruitment processes. In *Wirtschaftsinformatik 2005*, 1367–1381. Springer.

(Boxall and Purcell, 2003) P. F. Boxall and J. Purcell, 2003. *Strategy and Human Resource Management*. Management, work and organisations. Palgrave Macmillan.

(Buckley and Voorhees, 2000) C. Buckley and E. M. Voorhees, 2000. Evaluating evaluation measure stability. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, 33–40. ACM.

(Cabrera-Diego et al., 2013) L. A. Cabrera-Diego, J.-M. Torres-Moreno, and M. El-Bèze, 2013. SegCV : traitement efficace de CV avec analyse et correction d'erreurs. In *Actes de TALN 2013*, 707–714.

(Cappelli, 2001) P. Cappelli, 2001. Making the Most of On-Line Recruiting. *Harvard Business Review*.

(Chambers, 1983) J. Chambers, 1983. *Graphical methods for data analysis*. Chapman & Hall Statistics Series. Wadsworth International Group.

(Chang and Lin, 2011) C.-C. Chang and C.-J. Lin, 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2*(3), 27:1–27:27.

(Chapman and Webster, 2003) D. S. Chapman and J. Webster, 2003. The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment 11*(2-3), 113–120.

(Chen et al., 2011) C.-C. Chen, Y.-M. Huang, and M.-I. Lee, 2011. Test of a model linking applicant résumé information and hiring recommendations. *International Journal of Selection and Assessment 19*(4), 374–387.

(Clech and Zighed, 2003) J. Clech and D. A. Zighed, 2003. Data mining et analyse des cv: une expérience et des perspectives. *Revue d'intelligence artificielle 17*(1-3), 189–200.

(Cole et al., 2003) M. S. Cole, H. S. Feild, and W. F. Giles, 2003. Using recruiter assessments of applicants' resume content to predict applicant mental ability and big five personality dimensions. *International Journal of Selection and Assessment 11*(1), 78–88.

(Cole et al., 2004) M. S. Cole, H. S. Feild, W. F. Giles, and S. G. Harris, 2004. Job type and recruiters' inferences of applicant personality drawn from resume biodata: Their relationships with hiring recommendations. *International Journal of Selection and Assessment 12*(4), 363–367.

(Cole et al., 2009) M. S. Cole, H. S. Feild, W. F. Giles, and S. G. Harris, 2009. Recruiters' inferences of applicant personality based on résumé screening: Do paper people have a personality? *Journal of Business and Psychology 24*(1), 5–18.

(Cole et al., 2005) M. S. Cole, H. S. Feild, and J. O. Stafford, 2005. Validity of resume reviewers' inferences concerning applicant personality based on resume evaluation. *International Journal of Selection and Assessment 13*(4), 321–324.

(Cole et al., 2007) M. S. Cole, R. S. Rubin, H. S. Feild, and W. F. Giles, 2007. Recruiters' perceptions and use of applicant résumé information: Screening the recent graduate. *Applied Psychology 56*(2), 319–343.

(Cortes and Vapnik, 1995) C. Cortes and V. Vapnik, 1995. Support-vector networks. *Machine learning 20*(3), 273–297.

(Cover and Thomas, 2005) T. M. Cover and J. A. Thomas, 2005. *Elements of Information Theory* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

(Cristianini and Shawe-Taylor, 2000) N. Cristianini and J. Shawe-Taylor, 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

(De Meo et al., 2007) P. De Meo, G. Quattrone, G. Terracina, and D. Ursino, 2007. An XML-Based Multiagent System for Supporting Online Recruitment Services. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans 37*(4), 464–480.

(Di Meglio et al., 2007) E. Di Meglio, M. G. Grassia, and M. Misuraca, 2007. The Ideal Candidate. Analysis of Professional Competences through Text Mining of Job Offers. In *Effectiveness of University Education in Italy*, 261–275. Springer.

(Dipboye and Jackson, 1999) R. L. Dipboye and S. L. Jackson, 1999. Interviewer Experience and Expertise Effects. In R. W. Eder and M. M. Harris (Eds.), *The Employment Interview Handbook*, 259–279. SAGE Publications, Inc.

(Edgley, 1995) K. Edgley, 1995. The best job in the world may be one click away. *The Times*.

(Elkington, 2005) T. Elkington, 2005. Bright future for online recruitment. *Personnel Today*, 9.

(Faliagka et al., 2013) E. Faliagka, L. Iliadis, I. Karydis, M. Rigou, S. Sioutas, A. Tsakalidis, and G. Tzimas, 2013. On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed CV. *Artificial Intelligence Review*, 1–14.

(Faliagka et al., 2011) E. Faliagka, L. Kozanidis, S. Stamou, A. Tsakalidis, and G. Tzimas, 2011. A personality mining system for automated applicant ranking in online recruitment systems. In *Web Engineering*, 379–382. Springer.

(Forbes, 2003) C. Forbes, 2003. Video résumé: Seeing is believing. *Canadian HR Reporter 16*, 12–13.

(García-Sánchez et al., 2006) F. García-Sánchez, R. Martínez-Béjar, L. Contreras, J. T. Fernández-Breis, and D. Castellanos-Nieves, 2006. An ontology-based intelligent system for recruitment. *Expert Systems with Applications 31*(2), 248–263.

(Giaretta and Guarino, 1995) P. Giaretta and N. Guarino, 1995. Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases: knowledge building & knowledge sharing 25*.

(Good, 1953) I. J. Good, 1953. The population frequencies of species and the estimation of population parameters. *Biometrika 40*(3-4), 237–264.

(Green and Rao, 1969) P. E. Green and V. R. Rao, 1969. A Note on Proximity Measures and Cluster Analysis. *Journal of Marketing Research 6*(3), pp. 359–364.

(Guarino et al., 2009) N. Guarino, D. Oberle, and S. Staab, 2009. What is an ontology? In *Handbook on ontologies*, 1–17. Springer.

(Howell, 2013) D. Howell, 2013. *Fundamental statistics for the behavioral sciences* (8 ed.). Cengage Learning.

(Hsu et al., 2003) C.-W. Hsu, C.-C. Chang, and C.-J. Lin, 2003. *A practical guide to support vector classification*.

(Huang et al., 1992) X. Huang, F. Alleva, H.-W. Hon, M.-Y. Hwang, and R. Rosenfeld, 1992. The SPHINX-II Speech Recognition System: An Overview. *Computer, Speech and Language 7*, 137–148.

(Huggins, 1977) D. Huggins, 1977. Job hunting: The way to get the foot in the door. *Canadian Business 35*.

(Hunter and Hunter, 1984) J. E. Hunter and R. F. Hunter, 1984. Validity and utility of alternative predictors of job performance. *Psychological bulletin 96*(1), 72.

(Hutchinson, 1984) K. L. Hutchinson, 1984. Personnel administrators' preferences for resume content: A survey and review of empirically based conclusions. *Journal of Business Communication 21*(4), 5–14.

(Jackson and Moulinier, 2002) P. Jackson and I. Moulinier, 2002. *Natural Language Processing for Online Applications: Text Retrieval, Extraction & Categorization*. Amsterdam: John Benjamins.

(Johnson, 2009) P. Johnson, 2009. HRM in changing organizational contexts. In David G. Collings and G. Wood (Eds.), *Human Resource Management A critical approach*, 19–37. Routledge.

(Jurafsky and Martin, 2000) D. Jurafsky and J. H. Martin, 2000. *Speech & language processing*. Pearson Education India.

(Karaa and Mhimdi, 2011) W. B. A. Karaa and N. Mhimdi, 2011. Using ontology for resume annotation. *International Journal of Metadata, Semantics and Ontologies 6*(3), 166–174.

(Katz, 1987) S. M. Katz, 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing 35*(3), 400–401.

(Kerrin and Kettley, 2003) M. Kerrin and P. Kettley, 2003. *E-recruitment: Is it Delivering?* Institute for Employment Studies.

(Kessler et al., 2008b) R. Kessler, N. Béchet, M. Roche, M. El-Bèze, and J. M. Torres-Moreno, 2008b. Automatic profiling system for ranking candidates answers in human resources. In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, 625–634. Springer Berlin Heidelberg.

(Kessler et al., 2012) R. Kessler, N. Béchet, M. Roche, J.-M. Torres-Moreno, and M. El-Bèze, 2012. A hybrid approach to managing job offers and candidates. *Information Processing & Management 48*(6), 1124–1135.

(Kessler et al., 2009) R. Kessler, N. Béchet, J.-M. Torres-Moreno, M. Roche, and M. El-Bèze, 2009. Job offer management: how improve the ranking of candidates. In *Foundations of Intelligent Systems*, 431–441. Springer Berlin Heidelberg.

(Kessler et al., 2007) R. Kessler, J. M. Torres-Moreno, and M. El-Bèze, 2007. E-gen: automatic job offer processing system for human resources. In *MICAI 2007: Advances in Artificial Intelligence*, 985–995. Springer Berlin Heidelberg.

(Kessler et al., 2008) R. Kessler, J. M. Torres-Moreno, and M. El-Bèze, 2008. E-Gen : Profilage automatique de candidatures. *TALN 2008, Avignon, France*, 370–379.

(Kim and O'Connor, 2009) S. Kim and J. G. O'Connor, 2009. Assessing Electronic Recruitment Implementation in State Governments: Issues and Challenges. *Public Personnel Management 38*(1), 47–66.

(Kirk, 1996) R. E. Kirk, 1996. Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement 56*(5), 746–759.

(Knouse, 1994) S. Knouse, 1994. Impressions of the resume: The effects of applicant education, experience, and impression management. *Journal of Business and Psychology 9*(1), 33–45.

(Knouse, 1989) S. B. Knouse, 1989. The role of attribution theory in personnel employment selection: A review of the recent literature. *The Journal of General Psychology 116*(2), 183–196.

(Kruskal and Wallis, 1952) W. H. Kruskal and W. A. Wallis, 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association 47*(260), 583–621.

(Kullback and Leibler, 1951) S. Kullback and R. A. Leibler, 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics 22*(1), pp. 79–86.

(Lacasta et al., 2010) J. Lacasta, J. Nogueras-Iso, and F. J. Z. Soria, 2010. *Terminological Ontologies: Design, Management and Practical Applications*, Volume 9. Springer.

(Landis and Koch, 1977) J. R. Landis and G. G. Koch, 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics 33*(1), 159–174.

(Lin, 2004) C.-Y. Lin, 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, 74–81. Association for Computational Linguistics.

(Lin, 1991) J. Lin, 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory 37*(1), 145–151.

(Liu et al., 2007) P. Liu, F. Soong, and J.-L. Thou, 2007. Divergence-Based Similarity Measure for Spoken Document Retrieval. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Volume 4, IV–89–IV–92.

(Looser et al., 2013) D. Looser, H. Ma, and K.-D. Schewe, 2013. Using formal concept analysis for ontology maintenance in human resource recruitment. In *Proceedings of the Ninth*

Asia-Pacific Conference on Conceptual Modelling-Volume 143, 61–68. Australian Computer Society, Inc.

(Mael, 1991) F. A. Mael, 1991. A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology 44*(4), 763–792.

(Mahalanobis, 1930) P. C. Mahalanobis, 1930. On test and measures of group divergence, Part I: Theoretical formulae.

(Manning et al., 2008) C. D. Manning, P. Raghavan, and H. Schütze, 2008. *Introduction to information retrieval*. Cambridge University Press.

(Mason et al., 2003) R. L. Mason, R. F. Gunst, and J. L. Hess, 2003. *Statistical Design and Analysis of Experiments : With Applications to Engineering and Science.* Wiley Series in Probability and Statistics. Wiley-Interscience.

(McGill et al., 1978) R. McGill, J. W. Tukey, and W. A. Larsen, 1978. Variations of Box Plots. *The American Statistician 32*(1), pp. 12–16.

(Moens, 2006) M.-F. Moens, 2006. *Information Extraction: Algorithms and Prospects in a Retrieval Context*, Volume 21 of *The Information Retrieval Series*. Springer.

(Montgomery and Runger, 2010) D. C. Montgomery and G. C. Runger, 2010. *Applied Statistics and Probability for Engineers*. John Wiley & Sons.

(Montuschi et al., 2014) P. Montuschi, V. Gatteschi, F. Lamberti, A. Sanna, and C. Demartini, 2014. Job recruitment and job seeking processes: how technology can help. *IT Professional 16*(5), 41–49.

(NACE, 1995) NACE, 1995. Planning... A Valuable Guide to Your Career. In *Planning Job Choices* (39th ed.). Bethlehem, PA: National Association of Colleges and Employers.

(Nemanick Jr and Clark, 2002) R. C. Nemanick Jr and E. M. Clark, 2002. The differential effects of extracurricular activities on attributions in resume evaluation. *International Journal of Selection and Assessment 10*(3), 206–217.

(Noelle, 2005) M. Noelle, 2005. Log on, log in: corporate online recruitment in the FTSE 100. *IRS Employment Review 3*(829), 40–48.

(Owens, 1976) W. A. Owens, 1976. Background data. *Handbook of industrial and organizational psychology 1*, 609–644.

(Owens and Schoenfeldt, 1979) W. A. Owens and L. F. Schoenfeldt, 1979. Toward a classification of persons. *Journal of Applied Psychology 64*(5), 569.

(Padró and Stanilovsky, 2012) L. Padró and E. Stanilovsky, 2012. FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey. ELRA.

(Pande and Basak, 2012) S. Pande and S. Basak, 2012. *Human Resource Management: Text and Cases*. Pearson Education India.

(Pardo, 2005) L. Pardo, 2005. *Statistical inference based on divergence measures*. CRC Press.

(Parry and Wilson, 2009) E. Parry and H. Wilson, 2009. Factors influencing the adoption of online recruitment. *Personnel Review 38*(6), 655–673.

(Pennebaker et al., 2001) J. W. Pennebaker, M. E. Francis, and R. J. Booth, 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates 71*, 2001.

(Porter, 1980) M. F. Porter, 1980. An algorithm for suffix stripping. *Program 14*(3), 130–137.

(Quinlan, 1993) J. R. Quinlan, 1993. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

(Radevski and Trichet, 2006) V. Radevski and F. Trichet, 2006. Ontology-Based Systems Dedicated to Human Resources Management: An Application in e-Recruitment. In R. Meersman, Z. Tari, and P. Herrero (Eds.), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, Volume 4278 of *Lecture Notes in Computer Science*, 1068–1077. Springer Berlin Heidelberg.

(Rafter et al., 2000a) R. Rafter, K. Bradley, and B. Smyth, 2000a. Automated Collaborative Filtering Applications for Online Recruitment Services. In P. Brusilovsky, O. Stock, and C. Strapparava (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems*, Volume 1892 of *Lecture Notes in Computer Science*, 363–368. Springer Berlin Heidelberg.

(Rafter et al., 2000b) R. Rafter, B. Smyth, and K. Bradley, 2000b. Inferring relevance feedback from server logs: A case study in online recruitment.

(Rathie and Kannappan, 1972) P. N. Rathie and P. Kannappan, 1972. A directed-divergence function of type β. *Information and Control 20*(1), 38–45.

(Rocchio, 1971) J. J. Rocchio, 1971. Relevance feedback in information retrieval. *The SMART Retrieval System: Experiments in Automatic Document Processing*, 313–323.

(Roche and Kodratoff, 2006) M. Roche and Y. Kodratoff, 2006. Pruning terminology extracted from a specialized corpus for cv ontology acquisition. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, 1107–1116. Springer.

(Royston, 1995) P. Royston, 1995. Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Applied Statistics*, 547–551.

(Rychetsky, 2001) M. Rychetsky, 2001. *Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches*. Shaker.

(Saggion et al., 2010) H. Saggion, J.-M. Torres-Moreno, I. d. Cunha, and E. SanJuan, 2010. Multilingual Summarization Evaluation Without Human Models. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, Stroudsburg, PA, USA, 1059–1067. Association for Computational Linguistics.

(Salton and Buckley, 1988) G. Salton and C. Buckley, 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management 24*(5), 513–523.

(Salton and McGill, 1986) G. Salton and M. J. McGill, 1986. *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc.

(Salton et al., 1975) G. Salton, A. Wong, and C.-S. Yang, 1975. A vector space model for automatic indexing. *Communications of the ACM 18*(11), 613–620.

(Salton and Yang, 1973) G. Salton and C.-S. Yang, 1973. On the specification of term values in automatic indexing. *Journal of Documentation 29*(4), 351–372.

(Sarada, 2005) M. Sarada, 2005. *The Complete Guide to Resume Writing*. Sterling Publishers Pvt. Ltd.

(Schmidt, 1997) R. C. Schmidt, 1997. Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences 28*(3), 763–774.

(Schneider, 1987) B. Schneider, 1987. The people make the place. *Personnel Psychology 40*(3), 437–453.

(Searle, 2009) R. Searle, 2009. Recruitment and selection. In David G. Collings and G. Wood (Eds.), *Human Resource Management A critical approach*, 151–168. Routledge.

(Sen et al., 2012) A. Sen, A. Das, K. Ghosh, and S. Ghosh, 2012. Screener: a System for Extracting Education Related Information From Resumes Using Text Based Information Extraction System. *International Proceedings of Computer Science & Information Technology 54*.

(Senthil Kumaran and Sankar, 2013) V. Senthil Kumaran and A. Sankar, 2013. Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT). *International Journal of Metadata, Semantics and Ontologies 8*(1), 56–64.

(Shapiro and Wilk, 1965) S. S. Shapiro and M. B. Wilk, 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika 52*(3/4), 591–611.

(Singh et al., 2010) A. Singh, C. Rose, K. Visweswariah, V. Chenthamarakshan, and N. Kambhatla, 2010. PROSPECT: a system for screening candidates for recruitment. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 659–668. ACM.

(Sowa, 2000) J. F. Sowa, 2000. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA, USA: Brooks/Cole Publishing Co.

(Spärck-Jones, 1972) K. Spärck-Jones, 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation 28*(1), 11–21.

(Tange, 2011) O. Tange, 2011. GNU Parallel - The Command-Line Power Tool. *;login: The USENIX Magazine 36*(1), 42–47.

(Thompson, 2000) M. A. Thompson, 2000. *The global resume and CV guide*. Wiley.

(Thoms et al., 1999) P. Thoms, R. McMasters, M. R. Roberts, and D. A. Dombkowski, 1999. Resume characteristics as predictors of an invitation to interview. *Journal of Business and Psychology 13*(3), 339–356.

(Torres-Moreno, 2015) J.-M. Torres-Moreno, 2015. Trivergence of Probability Distributions, at glance. *CoRR abs/1506.06205*.

(Torres-Moreno et al., 2001) J.-M. Torres-Moreno, P. Velázquez-Morales, and J.-G. Meunier, 2001. Cortex: un algorithme pour la condensation automatique de textes. *ARCo 2*, 365.

(Trichet et al., 2004) F. Trichet, M. Bourse, M. Leclère, and E. Morin, 2004. Human resource management and semantic web technologies. In *Information and Communication Technologies: From Theory to Applications, 2004. Proceedings. 2004 International Conference on*, 641–642. IEEE.

(Tsai et al., 2011) W.-C. Tsai, N.-W. Chi, T.-C. Huang, and A.-J. Hsu, 2011. The effects of applicant resume contents on recruiters' hiring recommendations: The mediating roles of recruiter fit perceptions. *Applied Psychology 60*(2), 231–254.

(Tukey, 1977) J. W. Tukey, 1977. *Exploratory data analysis*. Reading Massachusetts: Addison-Wesley.

(Voorhees and Harman, 2001) E. M. Voorhees and D. Harman, 2001. Overview of TREC 2001. In *Proceedings of the Text REtrieval Conference (TREC 2001)*.

(Watson, 2010) T. J. Watson, 2010. Critical social science, pragmatism and the realities of HRM. *The International Journal of Human Resource Management 21*(6), 915–931.

(Wyant and Vise, 1979) J. F. Wyant and R. Vise, 1979. Resume writing: Form and function. *Special Libraries 70*(8), 328–32.

(Yahiaoui et al., 2006) L. Yahiaoui, Z. Boufaïda, and Y. Prié, 2006. Semantic annotation of documents applied to e-recruitment. In *SWAP*.

(Çelik et al., 2013) D. Çelik, A. Karakas, G. Bal, C. Gultunca, A. Elçi, B. Buluz, and M. C. Alevli, 2013. Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs. In *Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual*, 333–338. IEEE.