Towards Automating the Recruitment Process

By

Usman Shahbaz
Master of research

A thesis submitted to Macquarie University
Master of research
Department of Computing
January 2020

MACQUARIE
University
SYDNEY · AUSTRALIA

Examiner's Copy

Typeset in $\LaTeX\,2_\varepsilon$.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

(Signed) ____________________   Date: 04 - FEB - 2020

Usman Shahbaz

Dedication

This work is dedicated to my family. To my father, who inculcated in me the very nature of being inquisitive. My mother, who taught me that strong personal values in life are more important than the biggest fortunes in the world. To my wife Ayesha, for being my fortress of strength. To my lovely children, Shumail and Inayah, who bring the best of me even during the difficult of days. In essence, my supervisor Dr. Amin Beheshti for his backbone commitment to research. Who believed in my abilities and made a researcher out of me.

Finally, I dedicate this work to the beautiful Australian community who have shown me the real value of diversity.

Acknowledgements

This thesis has required of a lot of strength and dedication for which I have countless people to thank. Working alongside some of the great research members enabled me to not only see things differently but also gave me the power of cutting edge research work in computing. I am thankful to the members of the Data Analytics Research Lab[1], Department of Computing, Macquarie University (MQU) to have helped me shape my research skills.

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Amin Beheshti, for his continued support, encouragement and guidance throughout my study. Amin enabled me on how to question things, to do high quality research and develop an out-of-the-box thinking.

I would also like to thank the teachers and administration members of the Department of Computing at MQU who have been instrumental in shaping my work and shared their advise whenever I required.

Finally, I express my very profound gratitude to my close friends and family members, for providing me with unfailing support and encouragement throughout my study. This thesis would not have been possible without them.

Usman Shahbaz
Sydney, Australia
January 2020

[1]https://data-science-group.github.io/

Publications

- Usman Shahbaz, Amin Beheshti, Sadegh Nobari, Qiang Qu, Hye-Young Paik, Mehregan Mahdavi: iRecruit: Towards Automating the Recruitment Process. ASSRI 2018: 139-152

Abstract

Business world is getting increasingly dynamic. Information processing using knowledge-, service-, and cloud-based systems makes the use of complex, dynamic and often knowledge-intensive activities an inevitable task. Knowledge-intensive processes contain a set of coordinated tasks and activities, controlled by knowledge workers to achieve a business objective or goal. Talent acquisition and recruitment processes - i.e., the process of identifying the jobs vacancy, analyzing the job requirements, reviewing applications, screening, shortlisting and selecting the right candidate - are example of Knowledge-intensive processes. Attracting and recruiting right talent is a key differentiator in modern organizations.

In this thesis, we analyze the state of the art in traditional recruitment model and identify the main gaps when evaluating candidate profile with position description. We put the first step towards automating the recruitment process. We present a framework and algorithms to: imitate the knowledge of recruiters into the domain knowledge, extract data and knowledge from business artifacts, e.g., candidates' CV and position description, and link them to the facts in the domain Knowledge Base. We develop a digital dashboard to help recruiters draw insights in a quick and easy way. We adopt a motivating scenario in recruiting a Data Scientist role in an organization, and conduct a user study to highlight how iRecruit significantly facilitates the knowledge intensive tasks in the recruitment process.

Contents

List of Figures

1
Introduction

The quality of the services any organization provides largely depend on the quality of their processes [3–6]. Organizations are progressively getting a handle on this idea and moving towards process-based ventures. As this occurs, the focal issue (and opportunity) for any organization turns into the recognizable proof, measurement, investigation and improvement of its procedures. In this context, Business Processes (BPs) defined as a set of coordinated tasks and activities, carried out manually or automatically, to achieve a business objective or goal [3, 7, 8]. Over the last decade, many BPs across and beyond the enterprise boundaries have been implemented. Examples of business processes include tasks and activities in Administration, Manufacturing, Operations, Procurement and Customer Service.

Recently, various technologies such as social media and Web 2.0 have made dynamic processes more prevalent. This enables the process workers in the front line to be more proactive and use their knowledge and best practices in the decision making process and to choose the best next steps [9, 10]. Such knowledge-intensive processes, controlled by knowledge workers who have the experience, understanding, information, and skills. In this context, a knowledge-intensive process defined as a set of coordinated tasks and activities, controlled by knowledge workers to achieve a business objective or goal. Recruitment process - i.e., the process of attracting, shortlisting, selecting and appointing suitable candidates for jobs within an organization - is an example of a knowledge-intensive process [11, 12].

Attracting and recruiting right talent is a key differentiator in modern organizations. Recruitment process involves many data-driven, collaborative and knowledge intensive steps to ensure the right fit for an organizational talent requirement. Competing for high-quality resource talent is becoming a prevalent issue for almost all the organizational leaders and at the same time the recruitment process is a very expensive process for organizations. Recruiting for wrong fit can prove exceptionally costly not only due to monetary perspective but also it has serious consequences on employee morale and productivity [1, 13]. Current state of the art in the knowledge intensive recruitment process does not provide data-driven techniques to relate candidate CV related data with the position description and rank candidates based on a score.

We present novel techniques to imitate the knowledge of a recruiter, extract data [14, 15] from candidate's CV and link [16, 17] them to position description. Finally we present a data visualization based dashboard to present the data summaries for quick and easy navigation to data insights. We focus on a motivating scenario in recruitment, where a knowledge worker (e.g., a recruiter) in a knowledge intensive process, will be processing the Data Scientist candidate information from the candidates CV and enrich them with information such as university ranking, specific industry experience, location and more. We compare the candidate's contextualized profile with job descriptions (including job purpose, required skills, required experience and required education level) and rank their profile in terms of suitability for the position. The goal is to provide evidences to the knowledge workers, i.e., recruiters, and help them in decision making.

1.1 Key Research Issues and Contributions Overview

In this section, we outline key research issues tackled in this thesis following by contributions overview. One of the key challenges in the recruitment process, is the analysis of the CV and making inferences on different named entities and comparing these between candidates. Recruiters have limited time evaluating candidates pertinent to a role. To reduce this gap of making decisions based on limited information/data, we propose a novel approach where the relevant information about a candidate is presented in a contextual way and in a consistent manner irrespective of the way the CV is articulated or written. Our approach helps on the challenges in two ways. One that the information relevant to a candidate is extracted and presented as a summary view and secondly it is complemented with linked entity domain knowledge. This will assist recruiters automatically extracting facts, information, and insights from the raw business artifacts, e.g., candidates' Curriculum Vitae (CV[1]) and job advertisements.

In this thesis, we propose an analytical approach to the recruitment process. We propose that within the process there needs to be a feedback from the pool of talent stage to the position description. Secondly we propose techniques within knowledge graphs, linking two graph nodes to automate the manual process of matching each of the candidates' CV to the position description and give a comparison score. We call this comparison score as 'Position Description Match Score (PDMS)'. We also present a framework to imitate the knowledge of recruiters into a recruitment domain Knowledge Base (rKB), i.e., a knowledge base that consists of a set of concepts organized into the recruitment taxonomy (e.g., universities, organizations, jobs and best practices), instances for each concept and relationships among them. Finally we present a data visualization based dashboard to present the summarized candidate data for the recruitment analysts and present findings in a user-friendly way to draw insights.

Our focus builds on a motivating scenario in recruitment, where a knowledge worker (e.g., a recruiter) in a knowledge intensive process, will be processing the Data Scientist candidate information from the candidates CV and augment that information with linked entities (e.g., university ranking, specific industry experience, location etc.). Compare the candidate's enriched profile with the job description data (e.g., job purpose, required skills, required experience, required education level etc.) and then rank all candidates from highest match to the lowest match between candidates enriched profile and job description. This thesis includes offering:

- We put the first step towards automating the recruitment process and presents a framework

[1]A Curriculum Vitae (CV) is a written overview of a person's experience and other qualifications for a job opportunity. It is akin to a résumé in North America.

(namely *iRecruit*) to imitate the knowledge of recruiters into a recruitment domain Knowledge Base (rKB), i.e., a knowledge base that consists of a set of concepts organized into the recruitment taxonomy (e.g., universities, organizations, jobs and best practices), instances for each concept and relationships among them.

- We present a set of domain specific algorithms to extract data and knowledge from business artifacts and link them to the facts in the domain knowledge.
- A 'Position Description Match Score (PDMS)' which is a scalable and extensible scoring function to automate the manual process of matching candidates' CV to the position description and give a comparison score.
- A data-visualization based data summary to enable recruitment analysts with enriched candidate's CV, filter relevant candidates data along with Position Description Match Score and generate quick insights in an easy way.

The remainder of this thesis is organized as follows. In chapter 2 we present the background and analysis of the current state of the art in traditional recruitment model, impact of changes in technology on the recruitment model. Then we discuss the data-driven and knowledge-intensive processes. Afterward, in chapter 3, we present the details of our framework to imitate the knowledge of a recruiter, extract data from candidate artifacts such as a CV or Linkedin[2] profile and link them to position description. We also present a data visualization based dashboard to present the linked data [18] to enable recruitment analysts to draw insights in a quick and easy way. In chapter 4, we provide the implementation and discussion about the evaluation results. Finally, in chapter 5, we conclude our remarks and elaborate future directions to build on our work.

[2]https://www.linkedin.com/

2

Background and State-of-the-Art

Business world is getting increasingly dynamic. Information processing using knowledge-, service-, and cloud-based systems makes the use of complex, dynamic and often knowledge-intensive activities an inevitable task [3, 19]. Knowledge-intensive processes contain a set of coordinated tasks and activities, controlled by knowledge workers to achieve a business objective or goal [9]. Recruitment process [11] - i.e., the process of attracting, shortlisting, selecting and appointing suitable candidates for jobs within an organization - is an example of a knowledge-intensive process, where recruiters (i.e., knowledge workers who have the experience, understanding, information, and skills) control various tasks from advertising positions to analyzing the candidates' Curriculum Vitae. Attracting and recruiting right talent is a key differentiator in modern organizations.

Screening right candidates is not only time consuming, but it is also resource intensive, since it demands knowledge intensity to make the correct selection decisions. Recruiters spend on average 6.25 seconds on a CV before they make Yes/No decision [20]. This is a very short time to make an informed decision on a candidate's profile with the relevant business skill-set gap. Moreover, there are only six elements which are mostly looked at by the recruiters. These includes Name, current Title/organization, previous Title/organization, previous Position's Start/End Dates, current Position's Start/End Dates and Education [20].

This chapter illustrates core concepts and the current state-of-the-art in Talent Acquisition, Recruitment-Processes, Knowledge-Intensive Processes and Data-Driven Processes.

2.1 Talent Acquisition and Technology-Led Recruitment

Talent acquisition - i.e., identifying organizational talent requirement and then recruiting the best candidates to fulfil the organizational talent requirements is considered a major organizational leverage. Competing for high-quality resource talent is becoming a prevalent issue for almost all the organizational leaders [21]. The recruitment process however is a very expensive process for organizations being able to select the best fit and competent employees on the market is becoming increasingly hard amongst the job market competition [22, 23]. Organizations have to go through multiple screening phases to shortlist the right candidates.

There is significant financial and time cost associated to each of these phases [24]. Recruiting for wrong fit can prove exceptionally costly not only due to monetary perspective but also it has serious consequences on employee morale and productivity [1, 13, 24]. Advances in technology within the businesses has also demanded more research being conducted within recruitment in general and applicant and employee behavior in particular in recent times [22].

2.1.1 Human Resource Management(HRM) and Recruitment in HRM

Human Resource Management (HRM) is an integral function of any modern organization and plays an important strategic role in an organizations future. HRM, as defined by Storey [25], looks at an organizational process which acquires and maintains new capabilities and competencies through workforce planning and by applying traditional management techniques. HRM has evolved over time and has shifted from an old model of managing labour costs to a more strategic enabler [26]. Human resources are an ever changing part of an organization and hence requires a more strategic lens and plays a more integral role [27, 28].

Stoilkovska et.al [29] define recruitment as the process of finding the right candidates which make up a candidate pool which fits an open job vacancy that a organization might have. Recruitment becomes integral component to any HRM strategy since it helps build firstly a strong and competent workforce [24] but also it initiates the other HRM policies and processes since these are applied to the same workforce that is recruited [30].

Selecting the right employees is most commonly conducted by interviewing the candidates. However, this is evolving in recent years where organizations are looking at other methods to help improve the selection pool since interviews are costly and time consuming activity [31]. One of the most important things during recruitment is to ensure equal opportunity is provided to all candidates and that the selection pool is based on candidate skills [29].

2.1.2 Traditional Recruitment Model

Many researchers have approached the traditional recruitment process differently and while there is no one determined model many of the theories have very similar constructs [32]. One such recruitment process model, as proposed by Breaugh [1], consists of five different interconnected steps as elaborated in Figure 2.1. Breaugh considers the first step begins with the organization establishing recruitment objectives i.e., specification of jobs or positions required by the organization and what skills, education, work experience traits are expected of the potential workforce. While position or job descriptions can help identify and list the skills, background and experience requirements, it is equally important to document any budget constraints for these pertinent jobs. Once the organization's requirements are built the next step is to come up with a strategy plan which includes whether the sources are classified as external, walk-ins or internal [33]. During this step organization plans on how to build an effective message that is used to attract potential candidates and of how many positions should be filled and what characteristics, such as skills, work experience and education that is required by the candidates [1].

According to Breaugh, third step focuses on which recruitment method is best for the pertinent roles. This includes identifying which recruiters would be responsible, which platforms would be used and time duration that the job roles would be open for candidates to apply. Once recruitment methods are finalised, the next step looks at the interest of the applicants and determines how

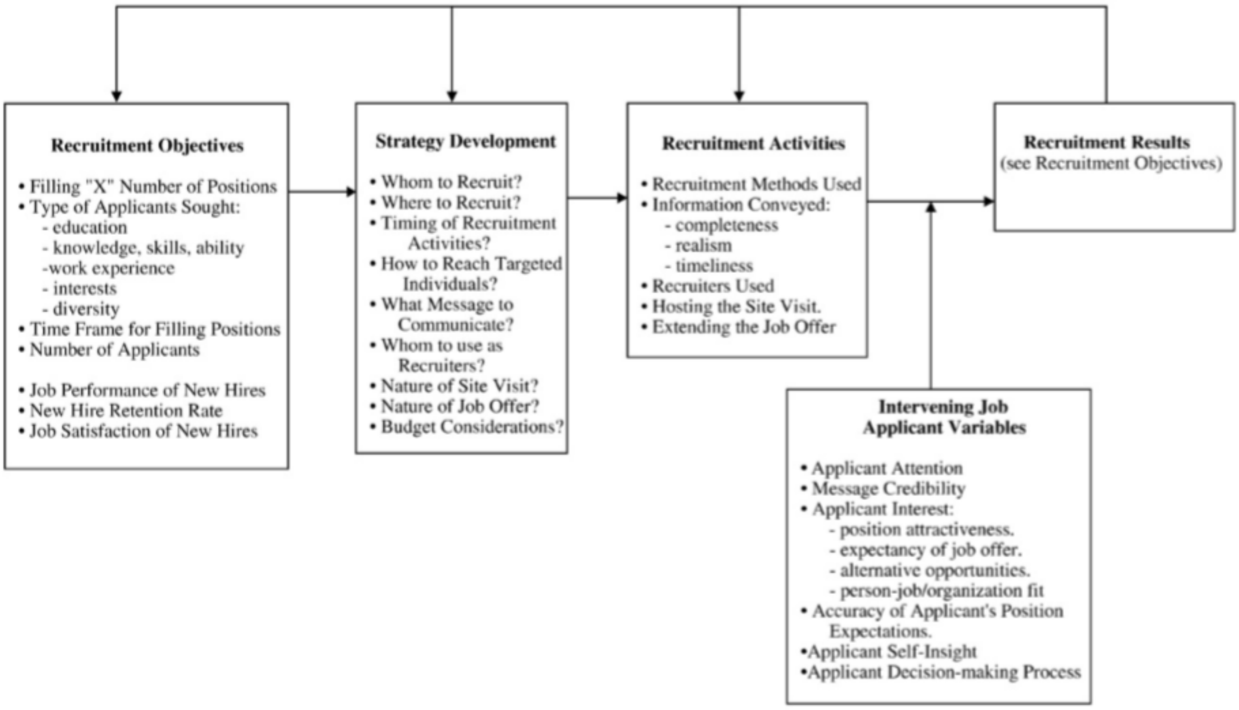

FIGURE 2.1: Model of Recruitment Process - Breaugh [1]

the applicants feels about the role as compared to any other job offers they might have and how interesting is the role and applicants own expectations from the role. During this step the applicants also reflects on their decision making process whether the role aligns to their objectives and career ambitions. The final step is the recruitment results. This is the final results of the whole recruitment process, which should be connected with the organizational recruitment objectives, the recruitment methods that were employed during the process and which activities were used and their effectiveness [1]. After these five steps, selected candidate is recruited for the vacant position and they fill the organization's skill gap.

Similarly to the suggested recruitment process model by Breaugh [1], are the ideas proposed by Compton [2], who debates that for an effective recruitment plan, an organization needs to link its human resource strategy to its strategic business plan in conjunction with its vision and getting inputs from its key stakeholders [2]. Figure 2.2 illustrates the HR processes in line with organization's strategic business plan.

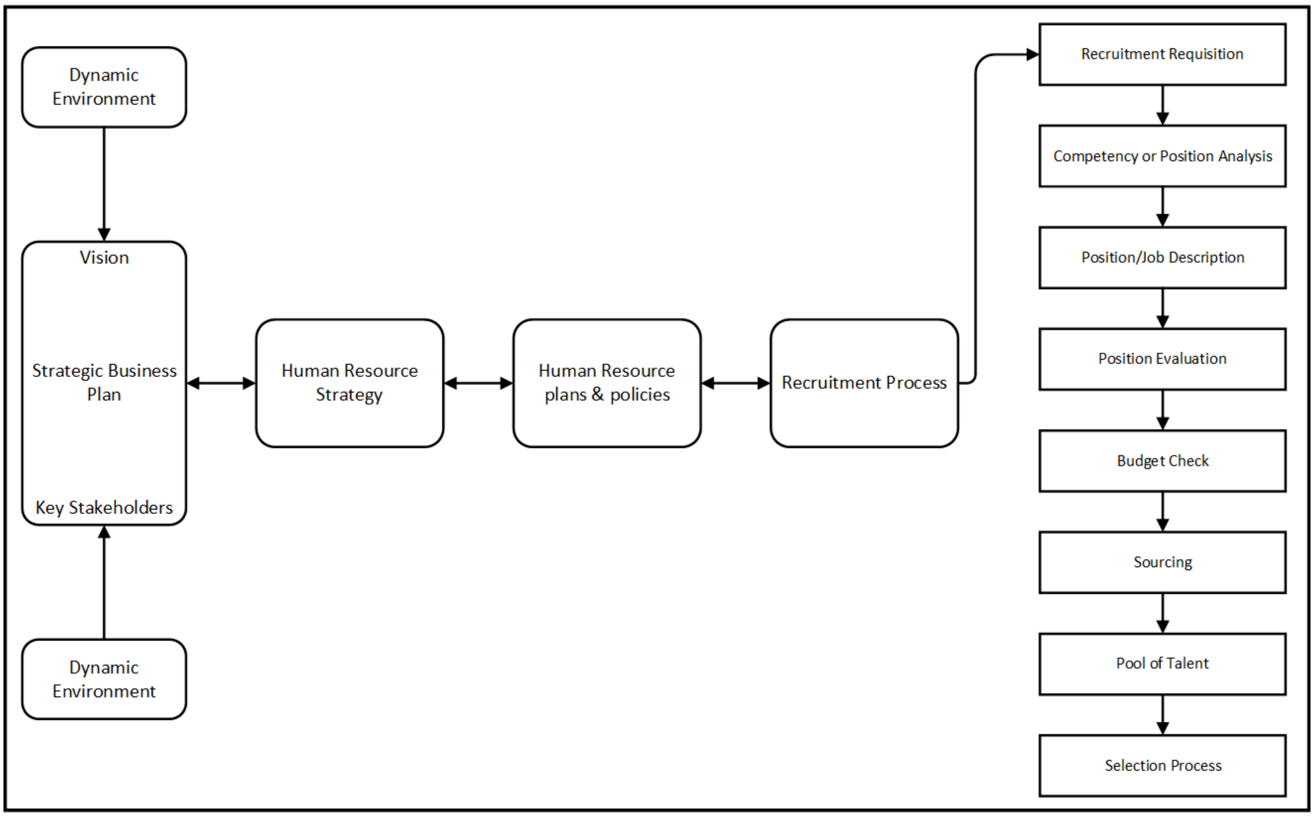

FIGURE 2.2: Effective Recruitment and Selection Process [2]: an illustration on the HR processes in line with organization's strategic business plan.

2.1.3 Assessment Centers and Psychometric Testing

Some organizations run assessment centers to identify candidates who can better handle certain situations such as pressure situations, team building, and conflict management. Assessment centers are run by experts in organizational behaviour and try to focus only on the behaviours organizations are mostly looking for [34]. Assessment centers are costly to run and require a major investment in time from both the potential candidates and business as well. While assessment centers identify people with respect to their behaviour under certain conditions, the real business like situations are very hard to recreate hence assessment centers can't guarantee people would behave exactly as they did during the assessment center [35].

Psychometric and behaviour tests are not as costly to run as compared to an assessment center however substantial time is invested from the candidates and the business to conduct these tests. Within a psychometric and behaviour test a candidate has to answer questions based on a certain scenario that is presented and they choose the most likely action they feel is best under that situation. Organizations use such psychometric tests to rule out any unwanted behaviours that people exhibit under certain situations [36]. They also assess how people might perceive other team members around them and would interact with them during regular business. More and more businesses use predefined psychometric tests to gauge the behaviours [37, 38] they are looking for in their employees to ensure a safe and sustainable business [39].

2.1.4 Job Interviews and Campus Recruitment

Job interviews are the most common assessment method used by businesses to assess the candidates against the vacant positions. Job interviews are sometimes structured but most of the businesses rely on unstructured format. An unstructured interview is where the hiring manager asks questions to the candidate without a specific, pre-defined pattern. Job interviewers can ask question pertaining the candidate's background to a technical question based on a skill that is expected in the role [40]. In contrast, structured interviews expect interviewers to ask questions in a defined pattern and same questions are repeated for each candidate to ascertain responses and rank candidates [40].

Many businesses run campus recruitment to attract talented and bright people who are finishing university degrees. This involves presenting the business profile to the students and highlighting the skills that the organization is looking for against the vacant roles. Students compare different businesses and assess which companies best suite their career ambitions and show their interest [41]. Campus recruitment is very popular in identifying junior roles within organizations and then groom them to more senior roles. Many businesses also run their own management trainee programs which induct people straight out of universities and then progress them within different parts of the business to finally positioning the candidates in one business unit of their choice.

2.1.5 Influence of Job Boards and Online Recruitment

Technology has played an important role in the evolution of recruitment process. Online recruitment or e-recruitment has been adopted by many organizations to help automate some of the tasks within recruitment [11]. While there is no single framework, most of the organizations have used these advances in technology towards automating HR evaluation processes and HR Reward processes [42]. While these advances in technology can be used by any organization, most of the early adaptors are large businesses who have used automation in different areas within the recruitment processes and HR processes in general [43].

While online recruitment techniques have been effectively employed by many businesses to increase the pool of interested potential candidates against any open jobs. This also introduces some major challenges in managing a large amount of applications for the relevant jobs. This becomes a prevalent issue since now finding the right candidates out of large amounts of applications becomes difficult. Managing a large number of applications becomes essentially more time consuming hence there has been focus by many organizations on how to manage this effectively [44].

Dhamija [42] debates that e-recruitment is one of the most popular non-traditional recruitment ways to attract potential job candidates. While many organizations have used online recruitment to improve their recruitment process both in terms of quality of candidates and the time required for applications against open job roles. However in certain circumstances groups of people might get discriminated since they rely on non-internet based methods. Some of the contractual workers who rely on word-of-mouth may be an example of this [42]. Many businesses these days utilize online and electronic recruitment tools to advertise the open position roles to attract potential candidates through different online job boards and social media Websites. Some business use recruitment banks and apply bots, i.e., an autonomous program on a network which can interact with systems or users and runs automated tasks [45], to scan through online candidate applications [46].

2.1.6 Influence of Social Media Recruitment

Social media has revolutionized the way people interact with each other individually or in groups. It has impacted the way people perceive businesses as well and where they see themselves working. Many organizations are now developing their brand pages within the social media space (e.g., LinkedIn[1] and Facebook[2]). organizations see this as a huge opportunity to showcase their vision to the large groups of people and engage with them on the work they are conducting [4]. Similarly potential candidates see this an opportunity to keep a track of the work organizations are doing and get involved where they feel they can contribute [47].

More and more businesses these days use social media to not only post news about their work

[1] https://www.linkedin.com/feed/
[2] https://www.facebook.com/

but also as a tool to post jobs and attract talented people for various positions within the business. Some of the social media platforms like LinkedIn have created paid subscriptions for organizations where organization's recruitment teams can assess and evaluate the market for potential skill and identify most matched potential candidates. They can also reach out to the potential candidates thus identified, to share the job opportunity and see if the relevant person is looking for an opportunity to move. Recruitment through social media has not only increased the reach for the organizations but also has helped in improving the overall time to fill the vacant positions. While some of the traditional methods like internal employee referral might be quicker way to fill a position, social media has certainly improved the overall time to engage right candidates.

2.1.7 Influence of Diversity and Inclusion

As more human communities embrace diversity in people backgrounds, governments and business are finding ways on creating diverse and inclusive culture at workplaces. People expect businesses to create workplaces which appreciate talent and skill and do not discriminate against any person's beliefs or background [48]. This becomes very important for businesses when it comes to recruitment processes that they do not marginalise people based on their opinions, beliefs or backgrounds etc. Organizations need to ensure they are following ethical practices and their employees are creating diverse and inclusive culture. In some cases there are government regulations which ensure businesses to follow ethical standards during recruitment process. Hence businesses are now liable to make diversity and inclusion part of their recruitment process and that no personal biases are introduced unknowingly [49].

2.2 Recruitment Business Processes

2.2.1 Business Processes

Business processes are central to any organization's operations. The tasks and coordinated activities that are performed essentially dictate whether an organization reaches its objectives effectively [3, 19]. Recruitment process, i.e., competing for high-quality resource talent is becoming a prevalent issue for almost all organizations. Technology, social behaviour changes and changes in the way organizations see human resources have all contributed to changes in recruitment process. Recruitment processes adapt and change based on the changing business requirements and organization's demand for growing skills to stay relevant in the market. Standard process workflows fall short in providing the flexibility for such data-driven recruitment processes where screening right candidates is not only time consuming, but it is also resource intensive.

Recent state-of-the-art within Business-Process-Management has been considered as: technologies and tools, applications and practices used to enable process analytics from querying to analyzing process-data [4, 7, 8, 50]; a wide range of process-mining methods have been contributed ranging from unstructured, semi-structured to fully structured processes [3, 51]; the novel concepts of rule based [52–54] Business Process Management, object based [54] BPM, case-based BPM processes [4]; and the up-coming trend focused around Business-Process-Management are: within crowdsourcing [55, 56], cloud based process-management [57], social Business-Process-Management [58], Business process enrichment [59] and process data analytics [3].

2.2.2 Recruitment Process

Competing for high-quality resource talent is becoming a prevalent issue for almost all the organizational leaders [21]. At the same time the recruitment process is a very expensive process for organizations. Recruiting for wrong fit can prove exceptionally costly not only due to monetary perspective but also it has serious consequences on employee morale and productivity [1, 13].

Screening right candidates is a resource intensive and time consuming process. Recruiters are expected to show high knowledge intensity to make correct selection decisions on potential candidates [20]. Recruiters are also challenged with very limited amount of time that they can spend on a single CV in order to go through large amounts of applications against the roles [20].

Specifically, this may generate a significant challenge whether all the relevant information is properly presented in the CV for the recruiter to make the right decision in limited time they can spend on the candidate's CV. If the candidates have positioned their relevant information in different segments within the CV, chances are that the recruiters might miss this. Some of the recruitment techniques (e.g., psychometric tests etc) can potentially help organizations fine tune matching right candidates with the job [60]. However these techniques can be used or abused by organizations in equal measure [61] as well; moreover these are costly and often misused by untrained personnel through the availability of internet-based testing.

2.2.3 Data-Driven Recruitment Processes

In the past few years there has been a significant focus on organizations to understand the data behind their organizational business processes. It is evident with a marked uptake in the development of tools that enable analysis of the process executions, system dependencies and interactions through data mining, data warehousing techniques [62]. There are many facets on the data that is analyzed on the business processes. One such approach is to focus on warehousing business process data and use it more like business reports [63]. Similarly data services and data spaces are created to make use of process data for active business decisions [64]. Another key area of focus for recent researchers has been applying analytical approaches over related process data [8, 53] where business decisions are improved by applying big data techniques on the process data. Some researchers found that by applying data analytics techniques over process mining [62] and process spaces [65] could help understand and complement process data with the rest of the organizational data points.

2.2.4 Data Curation

Recognizing business needs and deciding answers for business problems requires the examination of business process information, scattered over different frameworks [3]. In this context, preparing the raw business data for analytics requires significant investigation over enormous crossover assortments of heterogeneous and mostly unstructured process related execution data. Data curation, i.e., the way toward transforming raw information into contextualized data and knowledge [16, 17, 66], has been introduced fill the gap between persisting the raw data and analytics tasks.

Data Lakes [67] have been designed to encourage the business to leverage big-data with: wide physical circulation, variety of formats, non-standard data-models, self-managed and heterogeneous semantics. The idea of Knowledge Lake [68] acquainted with encourage to the disclosure and correspondence of significant patterns in data. Specifically, a Knowledge Lake can be viewed as a concentrated store containing for all intents and purposes unlimited measures of data and *contextualized data* that is promptly made accessible to anyone approved to perform analytical exercises. The term *Knowledge* here alludes to a set of facts, data, and knowledge drawn from the raw social information utilizing information curation procedures used to move an Information-Item into a Featurized-, Semantic-and Contextualized-Items [17, 69].

Feature engineering is a core concept in building contextualized data elements. There are different types of features which are of interest while dealing with information items. Lexical-based type of features are made up of keywords, phrase, topic, slangs, informal language and abbreviations. Natural-language-based type of features are based on part-of-speech, named entity (e.g., person, organization, product etc). While location-based features are based on the mentions of the location of an item such as country, city and postcode. [17, 69].

In this thesis, we extend knowledge lakes with services to extract the raw data from business artifacts such as CVs, job advertisements, candidates profiles, Universities and Organization's Job descriptions, and job search engine websites such as indeed.com and theladders.net. These services will persist the data in the knowledge lake. We also enhance the knowledge lake by imitating the knowledge of recruiters and inspired by Google Knowledge Graph[3], we focused on constructing a recruitment domain Knowledge Base (rKB).

2.2.5 Knowledge-Intensive Processes

Processes which require an intervention within their life-cycle are defined as scenario based processes, where a knowledge worker essentially intervenes during the process and make decisions based on their professional knowledge skills. In this context, HR-recruitment process is considered a type of knowledge-intensive process and Human Resource recruiters are considered as knowledge workers. HR-recruitment process is heavily reliant on the recruiter's professional knowledge and hence cannot be automated through traditional sense of systems workflow [4].

One needs to collect and present a wide-range of process objects and the knowledge worker's activities around these process objects, elaborating the artifact-centric nature of knowledge intensive processes. Different ways [70–73] have been used previously to incorporate business objects

[3]https://developers.google.com/knowledge-graph/

that used both data as well as process objects. Similarly a portion of these works utilized a variation of limited state machines to determine lifecycles [70] and some academic works investigated explanatory ways to deal with determining the object lifecycles following an occasion oriented style [71] while another work focused on querying object driven processes [74].

Comparable and related work is object-driven processes [71] where the process model is characterized as the document lifecycle. Some of existing methodologies [75] for demonstrating impromptu processes concentrated on supporting specially appointed work processes through user input. Some different methodologies [75, 76] concentrated on smart user help to manage end users during impromptu procedure execution by giving recommendations on conceivable subsequent steps. While other works [77], concentrated on demonstrating and querying methods for knowledge intensive activities and focused about entities (e.g., players, objects and activities) as top notch residents and spotlights on the development of business artifacts over time.

2.2.6 AI and Technology in Recruitment

Artificial Intelligence (AI) has disrupted many different industries and has enabled organizations create business leverage by applying cutting edge automation and optimization techniques. AI has been seen to have a major impact on the HR processes and has helped improve many time consuming tasks. Upadhyay and Khandelwal [78] highlight this massive trend that most of the recruiters are looking at ways AI can help in improve the overall quality and effectiveness of the recruitment process. One of the most tedious task within the recruitment process is to scan and extracting relevant information from candidate's CV automatically [79]. AI can help with aggregation of different candidate evaluations and relevant information.

HR department keeps multiple candidate evaluations like written tests and interview evaluations, these can be aggregated together and presented back in a unified way. Similarly another area where AI can help is ranking [80] candidates based on some score that the recruiters can apply based on their manual scanning of the candidate's CV and help in maintaining a good talent pool. AI can help build effective ranking algorithms that can help maintain a priority list of talent pool [81].

AI driven chat-bots work as recruitment assistants that utilize personal and updated contact information to communicate with candidate through preliminary and standard notifications. These automated notifications can include automated messages once an application has made some progress within the process or answering candidate basic questions pertaining office location and time of interview. Chat-bots can communicate via emails, text messages or dialogue box and share relevant information with candidates and help automate many of the recruitment basic tasks [78]. Job matchmaking techniques packaged as computer software and backed by computer algorithms can also be utilized to help sort candidate CVs and using learning based techniques [82].

Applying AI in recruitment can be very beneficial for organizations. One of the main applications of AI can be found in matching the behaviours of best talented people in their organizations and try to find similar traits in the job applicants. IBM Watson is an AI application developed by IBM and is used in many different avenues. IBM Watson Recruitment helps businesses recruit the right candidates for open jobs by aggregating information from different sources and then ranking them based on their overall score [83]. IBM studies that 66 percent of CEOs consider cognitive computing can help drive significant HR value [84]. IBM Watson Recruitment helps the recruitment process effective by Understanding what makes the pertinent role successful and secondly Reasoning which candidates are would be best fit for the role by providing holistic and unbiased screening [84].

Similarly HireVue [85] is another AI application that help businesses improve their recruitment process. HireVue using computer vision algorithms can assess interviews of the potential candidates and compares them to the best organization's talent and makes recommendations [85]. HireVue was able to leverage recruitment and scenario based games and video interviews to improve recruitment process from 42 days down to only 5 day [86].

Faliagka et al.2012 debate that by leveraging AI based tools organizations can improve their recruitment process by acquiring knowledge from candidate's CV, their blog posts or social media posts. These tools can aggregate multiple data points and generate candidate's personality traits, emotions and moods [81].

Researcher studied that AI has impacted recruitment processes effectively in many organizations and helped create business leverage. Upadhyay and Khandelwal [78] debate that AI can help extract personality traits and applicant's attitude from social media sites which was traditionally only possible through interviews [78]. Traditional formal style interviews are costly for the business and previously they were the only way to understand the candidate's personality traits and their attitude [81]. AI is able to help in this area and by extracting information through social media, it is able to create a good estimation of candidate's personality traits and in way less time as compared to the traditional interviews.

Another area that Upadhayay and Khandelwal [78] highlighted is the bias that is sometimes introduced unconsciously through the traditional manual process can be reduced through AI. AI acts unbiased based on the trained algorithm rules to screen the best candidates accordingly [78]. This way no bias is introduced due to candidate's background, personal or religious beliefs but focuses on their skills and experience bases on the job description. Objective of recruitment systems is to save organizational cost through modernizing the processes. Organization's time to fill job roles improve by effectively automating pre-screening of CVs and then sorting and ranking candidates based on their skill match [87]. Modern organizations are forced to look at ways to improve and automate the traditional recruitment process for them to deal with changing ways of work life. As technology and social media impacts common human interactions, businesses who want to operate successfully must adopt new technology trends [88].

Many studies indicate towards the importance of technology adoption and big data being applied to recruitment process [89]. However for a thorough and impactful transformation, technology adaption must move beyond just basic recruitment analytics and become a continual part of HR decision making based on AI [90]. While automation through AI has eased a lot of manual tasks in recruitment, it has also introduced a few challenges as well. The algorithms that are used to sort candidates CV and rank them are now required to go through testing and evaluation so that they do not introduce any unfavourable bias towards certain group of people. Similarly understanding the ethical use of personal information about candidates is now becoming very sensitive issue. It impacts both the HR professionals and the candidates themselves [91].

3

Towards Automating Recruitment Process

Knowledge acquisition has become cornerstone in many organizations to enable effective automation. Acquiring the knowledge accurately from a knowledge worker becomes cumbersome especially when the knowledge is often 'made-up' by the expert as the occasion demands [92]. Moreover, while explaining a rule an expert might simplify the rule if they are explaining it to somebody they feel already has some subject background [92]. Processing the information from knowledge and cloud-bases systems is fast becoming the foundations of human life. Recently, there has been significant focus on understanding process based data captured through various data systems that support processes [4, 7]. An accurate information capture system thus becomes a backbone structure to any information process.

Business-processes (BPs) - i.e., set of coordinated tasks and activities, carried out manually or automatically, to achieve business goal or objective - are core to enterprise operations [3, 7]. Over the last decade, many BPs across and beyond the enterprise boundaries have been implemented.

Various technologies such as social media and Web 2.0 have made dynamic processes more prevalent. This enables the process workers in the front line to be more proactive and use their knowledge and best practices in the decision making process and to choose the best next steps.

Such knowledge-intensive processes, controlled by knowledge workers who have the experience, understanding, information, and skills are utilised to achieve business objective or goal. Recruitment process - i.e., the process of attracting, shortlisting, selecting and appointing suitable candidates for jobs within an organization - is an example of a knowledge-intensive process. Attracting and recruiting right talent is a key differentiator in modern organizations. Recruitment process involves many data-driven, collaborative and knowledge intensive steps to ensure the right fit for an organizational talent requirement.

The recruitment process as highlighted in Figure 2.2 starts with the requisition step based on the organizational need for a specific skill against the strategic business plan and runs linear till the selection process. Talent pool that is established as an outcome of this process is generally compared relative to each other rather than the position description. Current literature within the recruitment processes do not provide enough data driven rule based techniques to process related data from expert knowledge to improve recruitment decisions in a knowledge intensive process. There are generally two main contributing factors:

- Firstly, it is time consuming and difficult to compare each of the candidates with the position description; and
- Secondly, since there is inherit requirement of building a rank order of candidates, so it sometimes takes precedence in developing the candidates ranking with respect to each other rather than relative to the position description.

To overcome the above two challenges, in this thesis, we propose an analytical approach to the recruitment process. We propose that within the process there needs to be a feedback from the pool of talent stage to the position description where we can compare a candidates' CV to the position description that has been established based on the organizational skills requirements. This new process is shown in Figure 3.1. Secondly we propose techniques within knowledge graphs, linking two graph nodes to automate the manual process of matching each of the candidates'

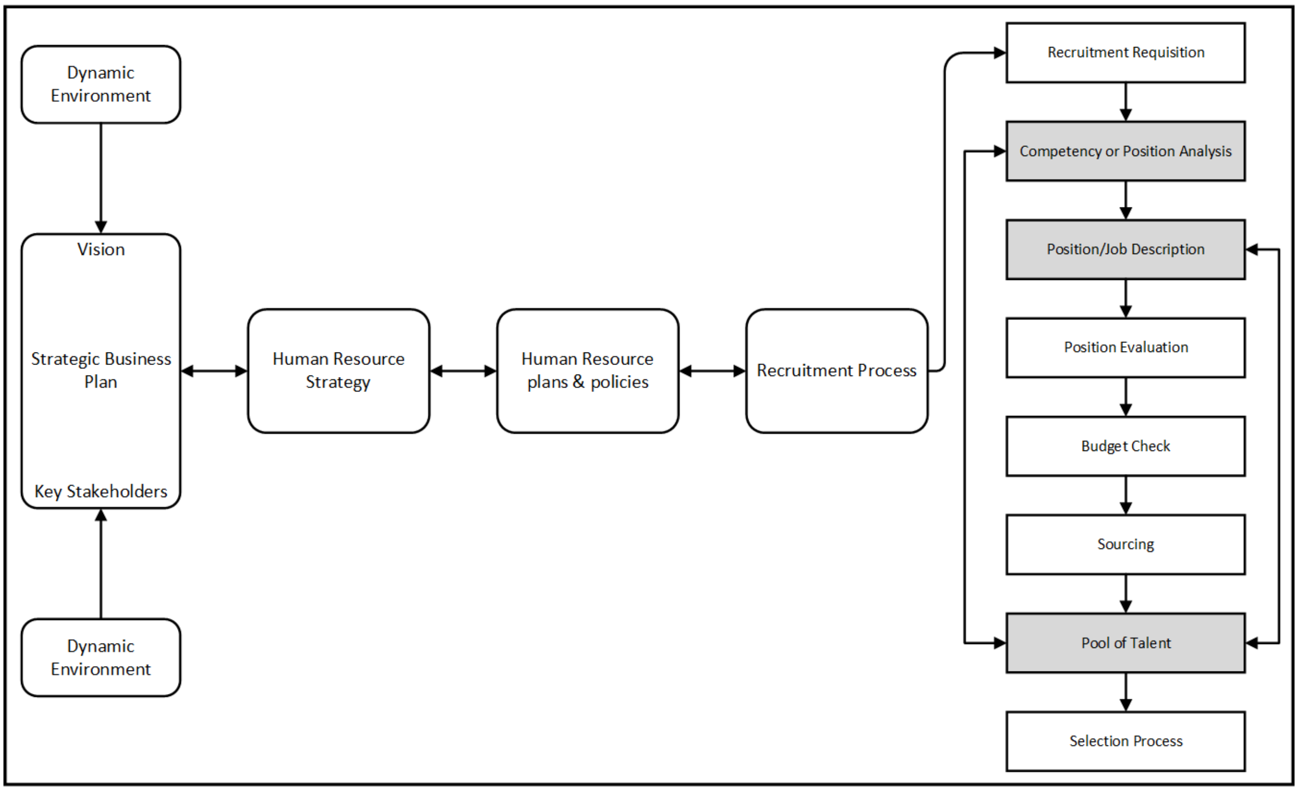

FIGURE 3.1: Updated effective recruitment and selection process [2].

CV to the position description and give a comparison score. We call this comparison score as 'Position Description Match Score (PDMS)'. We also present a framework to imitate the knowledge of recruiters into a recruitment domain Knowledge Base (rKB), i.e., a knowledge base that consists of a set of concepts organized into the recruitment taxonomy (e.g., universities, organizations, jobs and best practices), instances for each concept and relationships among them. Finally we present a data visualization based dashboard to present the summarized candidate data for the recruitment analysts and present results in a user-friendly way to draw insights.

Our motivating scenario revolves around in recruitment, where a knowledge worker (e.g., a recruiter) in a knowledge-intensive process (e.g., recruitment), shall be processing the Data Scientist candidate information from the candidates CV and augment that information with linked entities (e.g., university ranking, specific industry experience, location etc.). Compare the candidate's enriched profile with the job description data (e.g., job purpose, required skills, required experience, required education level etc.) and then rank all candidates from highest match to the lowest match between candidates enriched profile and job description.

The main contributions, presented in this chapter, include:

- We put the first step towards automating the recruitment process and presents a framework (namely *iRecruit*) to imitate the knowledge of recruiters into a recruitment domain Knowledge Base (rKB), i.e., a knowledge base that consists of a set of concepts organized into the recruitment taxonomy (e.g., universities, organizations, jobs and best practices), instances for each concept and relationships among them.
- We present a set of domain specific algorithms to extract data and knowledge from business artifacts and link them to the facts in the domain knowledge.
- A 'Position Description Match Score (PDMS)' which is a scalable and extensible scoring function to automate the manual process of matching candidates' CV to the position description and give a comparison score.
- A data-visualization based dashboard to enable analysts interact with enriched candidate's CV, filter relevant candidates data along with Position Description Match Score and generate quick insights in a less cumbersome way.

The remainder of the chapter is organized as follows. In Section 3.1 we provide a motivating scenario. We present the recruitment analytical pipeline in Section 3.1.1. We present the evaluation and implementation in Chapter 4.

3.1 Motivating Scenario: Recruiting Data Scientists

As the motivating scenario, we focus on the recruitment process for the *data scientist* role. Data science is a multi-disciplinary subject that integrates scientific methods, procedures, algorithms and data-systems to generate knowledge and insights from structured, semi-structured and unstructured data. Data science is related to computer programming, statistics and data-mining. A data scientist is someone who utilizes tools and techniques to extract insights and interpret data. These methods require knowledge of methods from machine-learning, and business knowledge. Responsibilities include: identification, integration, curation, analyzing and visualizing the (big) data. For this role, there is an estimated growth of 39% [20] by 2020. Data science skill shortage is projected to increase, in US alone there is a shortage of 100,000 new data science roles forecasted for the next 2 years [93].

In Australia, there are more than 500 jobs advertised on LinkedIn which is 15% higher than compared with 2018 [94]. Skill matching is difficult for a data science role as well and on average data science roles remain open 5 days longer than other roles. From the business perspective a good match between the candidates skills with the highlighted business deficiency ensures that business can achieve their strategic ambitions. A bad match can not only cost a business millions but also puts their business relevance at risk as well [95].

3.1.1 Imitating the Knowledge of Recruiters

We put the first step towards automating the recruitment process and presents a framework (namely *iRecruit*) to imitate the knowledge of recruiters into a recruitment domain Knowledge Base (rKB), i.e., a knowledge base that consists of a set of concepts organized into the recruitment taxonomy (e.g., universities, organizations, jobs and best practices), instances for each concept and relationships among them. We explain the techniques to construct the rKB domain knowledge.

The rKB knowledge base includes the important entities such as organizations and their relevant information, universities and their profile including location and relevant information, countries and cities and relevant skills entities. To build this knowledge-base, we first identified the list of recruitment related categories and their related types/sub-types provided by recruitment guides [1]. Then we have focused on the recruitment for the computer science category and more specifically the data scientist role; and identified popular concepts and instances related to this category on the Web.

For example, we have identified: concepts such as Programming Languages, Educations, Organizations and Technical Skills. Moreover, we have identified the best practices that recruiters follow and identified concepts such as Seniority Levels, Team Leadership, Statistical modeling (that data scientists are expert in), as well as reporting skills. Figure 3.2 shows a small snippet of the formulated rKB, which illustrates the above notions.

[1]https://builtvisible.com/recruitment-seo-guide/

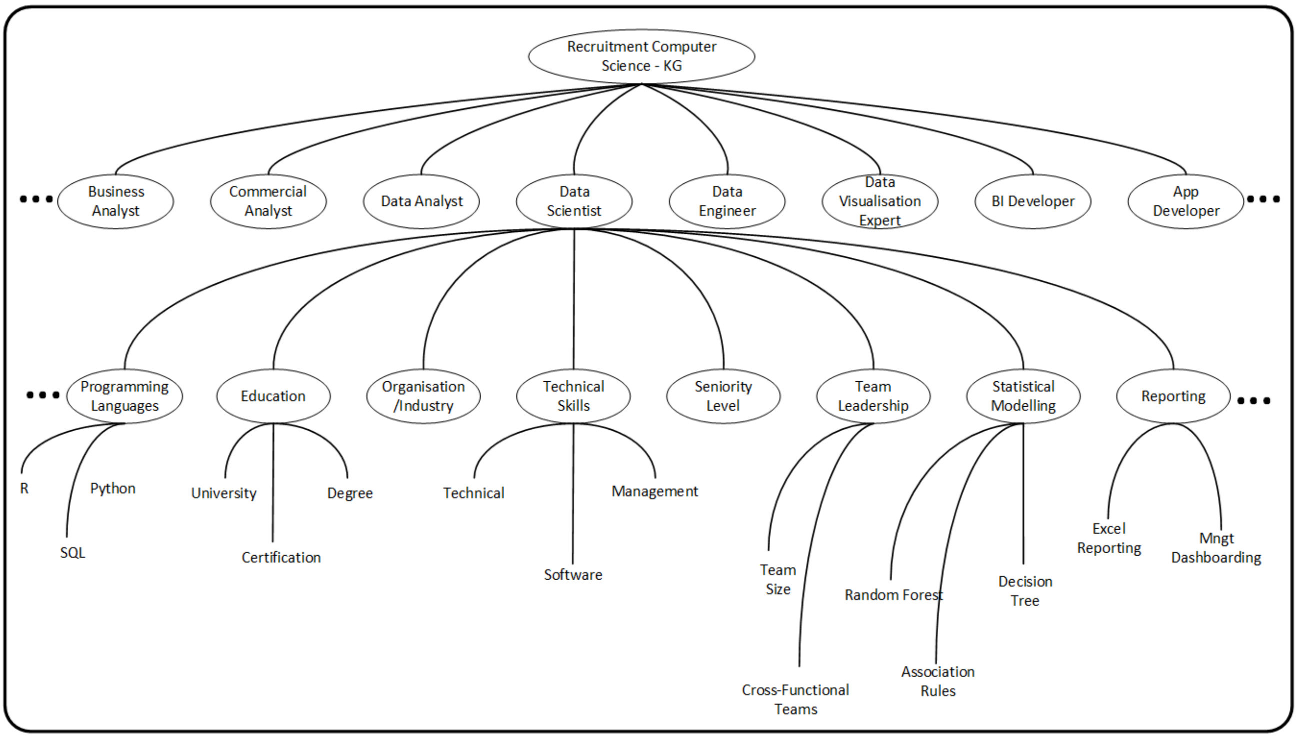

FIGURE 3.2: A sample fragment of the rKB..

3.1.2 Recruitment Knowledge Graph

Let $R = (V, E)$ be a Resource Description Framework (RDF) graph where V is a set of nodes and $E \subseteq (V \times V)$ is a set of ordered pairs called edges [96]. Let $G = (V, E)$ be an Entity-Relationship (ER) attributed graph where V is a set of nodes and $E \subseteq (V \times V)$ is a set of ordered pairs called edges. An ER graph $G = (V_G, E_G)$ with n number of entities is illustrated as $G \subseteq R$, $V_G = V$ and $E_G \subseteq E$ such that G is a non-symmetric directed graph with no directed cycles. A resource in an ER graph is defined recursively where: (i) $\in$ is a resource; (ii) The sets V_G and E_G are resources; and (iii) The set of Entity-Relationship graphs are closed under intersection, union and set difference: let G_1 and G_2 be two ER graphs, then $G_1 \cup G_2$, $G_1 \cap G_2$ and $G_1 - G_2$ are resources.

(Entity) *An entity E is defined having a unique identity and as a data object. Entities exist separately and are described by a set of* attributes. *Entities can be simple such as position description keywords, skills and topics (e.g., coding skills, behavioural skills, data science topics). Entities can be composed of atomic-information item such as named entity (country, university, industry type, organization)extracted from structured data source such as position description, candidate's CV or* *semi-structured data source such as Wikidata, but might not conform to an entity type. Entities can also be such as position rules, rules based on temporal conditions (e.g., last years' university ranking, countries within a geographic distance, industries with highest skill) which are complex Data Sources. One way would be to define "temporal entities" could be university ranking within a certain time period or industry vertical championing a skill-set within a certain time period.*

*(**Relationship**) A* relationship *is defined as a direct or indirect link between two or more entities. These entities are related with a predicate set of traits of entities that uniquely recognizes the relationship. Relationships can be: Time-based, Rule-based and Activity-based [4, 97] We model the contextualized data and knowledge as a graph [96, 97] of typed nodes (e.g. raw information and generated features) and edges (relationships) such as:*

- $keyword \xrightarrow{(extractedFrom)} CV$: express that a keyword (e.g., skills, tools, country) is extracted from candidate CV
- $keyword \xrightarrow{(similar\text{-}keywords)} keyword-set$: express that a keyword is similar to other keywords (e.g., universities offering data science degrees, universities within the same location) within the keyword-set
- $Topic \xrightarrow{(extractedFrom)} CV$: express that a topic (e.g., data science conferences, machine-learning algorithms) is extracted from candidate CV
- $Topic \xrightarrow{(related\text{-}topics)} topic-set$: express that a topic is related to other topics(e.g., similar machine-learning algorithms, similar data-visualization tools) within the topic-set
- $University \xrightarrow{(extractedFrom)} CV$: express that a university name is extracted from the candidate CV
- $University \xrightarrow{(rank[timestamp])} University-profilerank-per-subject$: express a rank of a university as compared to similar universities at a certain time
- $Company \xrightarrow{(extractedFrom)} CV$: express a company(e.g., Google, Amazon) is extracted from candidate CV
- $Company \xrightarrow{(industry)} Company-Industry-Location$: express that a company belongs to a certain industry vertical (e.g., Microsoft within ICT Industry)

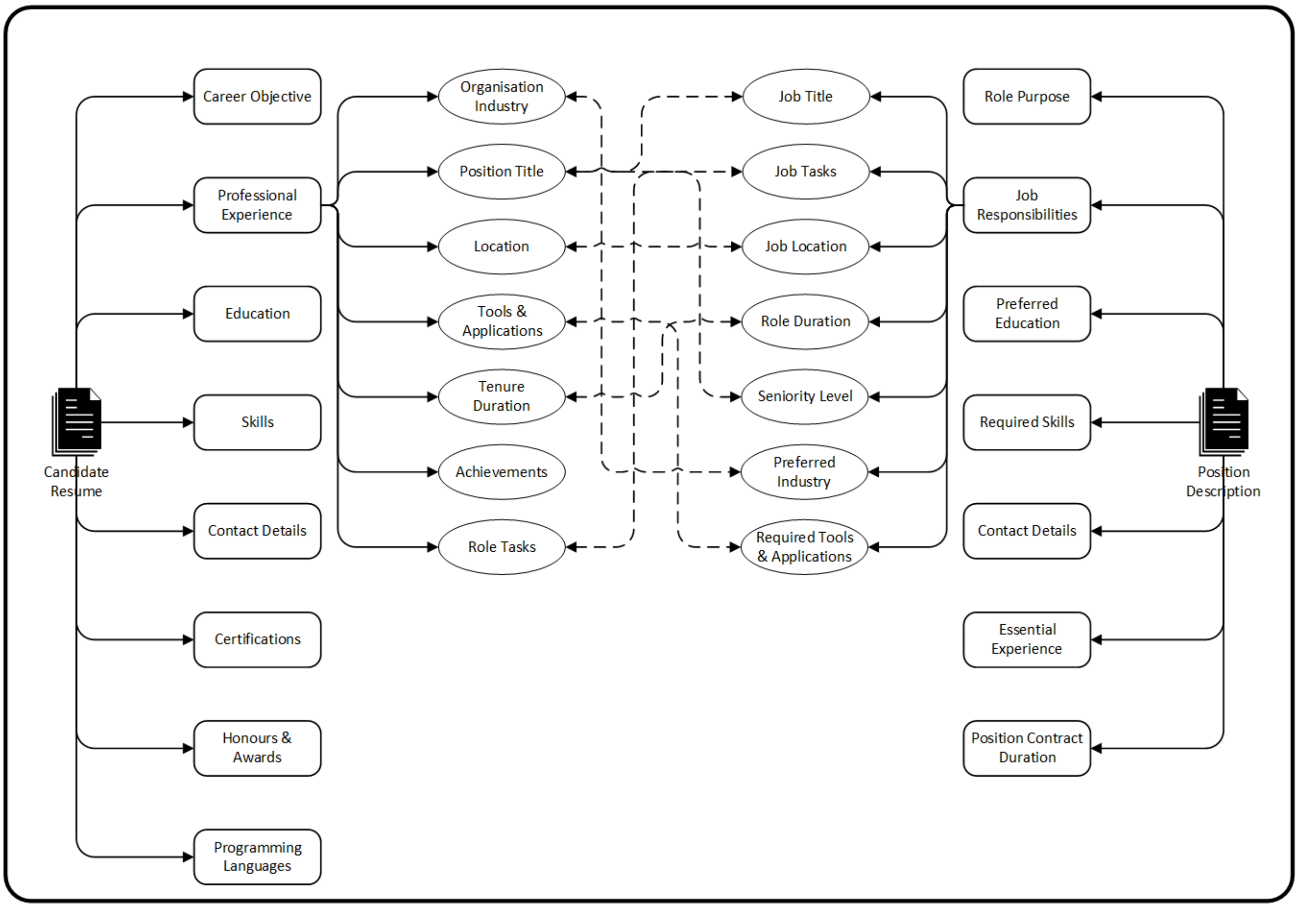

FIGURE 3.3: Linking entity nodes between candidate CV/resume with position description.

- $Organization \xrightarrow{\textit{(worked-in)}} Position-Description$: express an organization rule (e.g., candidates who are currently or previously worked with Google) from the position description
- $University \xrightarrow{\textit{(graduated-from)}} Position-Description$:express a university rule is (e.g., candidates who graduated from Macquarie University) from the position description

We then utilise curation services [16, 98] to automatically: (a) Extract features such as keyword, part-of-speech, and named entities such as Persons, Locations, Organizations, Software Tools and more; (b) Enrich the extracted features by providing synonyms and stems leveraging lexical knowledge bases for the English language such as WordNet; (c) Link the extracted enriched features to external knowledge bases (such as Google Knowledge Graph[2] and Wikidata[3]) as well

[2]https://developers.google.com/knowledge-graph/
[3]https://www.wikidata.org/

as the contextualized data islands; and (d) Annotate the items in a data island by information about the similarity among the extracted information items, classifying and categorizing items into various types, forms or any other distinct class.

To enable querying this large graph, we leverage already available work [99], a SPARQL [100] query engine for analyzing large graphs, to organize the data and extracted-enriched-linked features. Technical details of these services and how we organize and query the data in the Knowledge Lake, can be found in [68][4].

3.1.3 Extracting Data and Knowledge from Business Artifacts

We leverage already available resources [14, 16, 68] to extract data and knowledge from the business artifacts, including CVs in PDF, Candidates Website, and job advertisements. Figure 3.4 illustrates the features that our approach aims to extract from one side on the position description/business requirement and other side on the knowledgebase of the candidate's CV.

We also make use of available work in data curation and enrichment [17], where authors have presented the notions of information-item (raw business artifact, e.g., a candidate CV or jub description), featurized-item (enabling automatic extraction of various features from Schema-based to Natural-Language-based and Metadata-based features), semantic-item (automatically enriched extracted features) and contextualized-item (automatically linking extracted data to the Domain Knowledge). Figure 3.5 illustrates how we generate featurized, semantic and contextualized items from raw business artifacts in the recruitment processes based on this technique.

Algorithm 1 illustrates how we transform business artifacts (such as candidate CV and Job Description) into contextualized and curated items. We leverage existing extraction algorithms [101] for PDF[5] Documents, to extract information from major sections such as career summary/ambition, Professional Experience, Education and Skills in candidates' CV.

The features extracted from these CV sections can then be further grouped based on feature

[4] https://github.com/unsw-cse-soc/CoreKG

[5] The Portable Document Format (PDF) is a file format developed by Adobe in the 1990s to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems. A curriculum vitae (CV) provides a summary of candidates experience and skills and mainly prepared in PDF format.

Data: Business Artifact [CV, Job Advertisement]

Result: Contextualized Business Artifact

Step1: Retrieve Business Artifact Schema and Pre-Process;

Step2: Build Set-of-Features: **for** *Every Business Artifact* **do**

    Generate Features(based on schema, based on lexical and Natural-Language, based on Time and Location);

end

Step3: Build Enrichment-Set: **for** *Every Item in Set-of-Features* **do**

    Append Item with features based on schema, based on lexical and Natural-Language, based on Time and Location

end

Step4: Build Linking-Set: **for** *Every Item in Enrichment-Set* **do**

    for *Every Entity in rKB.* **do**

        Generate Similarity;

        Append the Item and Entity;

        Update Training-Data;

    end

end

Step5: Build Short-Listed-Candidates;

Algorithm 1: Recruitment data-curation algorithm.

type such as Lexical/Natural Language based, Temporal and Geo. This enables us to enrich the extracted information from the CV with external data sources (such as WikiData) and similar services such as: University Rankings[6] [7], company reviews[8] [9], educational certifications [10] [11] [12]. The set of features that are extracted from the CV sections include:

- **Lexical-based features:** This type is related to the vocabulary of a language such as topic,

[6] https://www.timeshighereducation.com
[7] https://www.topuniversities.com/qs-world-university-rankings
[8] https://www.glassdoor.com.au/Reviews/index.htm
[9] https://greatplacetowork.com.au/
[10] https://www.acs.org.au
[11] https://www.webopedia.com/quick_ref/computer-certifications.html
[12] https://en.wikipedia.org/wiki/Academic/_certificate

keyword, phrase, abbreviation, special characters, slangs, informal language.

- **Natural-Language-based features:** This category of features relates to entities that can be extracted by the analysis and synthesis of natural language (NL) and parts of speech, named entities like organization type, industry vertical, job role, job tasks etc.

- **Temporal-based features:** This feature category is related to time space. Estimating total professional experience from the CV, number of years at a specific organization, duration of the educational degree etc.

- **Location-based features:** This category is related to the mentions of locations in the CV of the candidate. (e.g., in Universities the text may contain 'Sydney'; a city in Australia, or worked at an organization based in Melbourne, a city in Australia).

Next, we define a set of enrichment functions to enrich the extracted items. For instance, if education section contains Bachelor Of Computer Science, the enrichment function 'Synonym' can be used to enrich this keyword with its synonyms such as Bachelor of Science from knowledge sources such as Wikidata[13]. The result (e.g., set of synonyms) will be stored in the Enrichment Set. The proposed enrichment functions are built against the Lexical-based features, using knowledge sources such as WordNet[14] and dictionaries to enrich with their Synonyms, Stems, Hypernyms Hyponyms and more. Figure 3.3 illustrates the result of applying Algorithm 1 on a CV to Link entity nodes between candidate CV and position description.

Next step is to compute the similarity among the items extracted from the business artifatcs and the entities in the domain knowledge base. In this phase we define the *position description match score (PDMS)*, which is the scoring function applied on the extracted entities from the candidates CV and their likelihood match with the relevant position description sections. *PDMS* is a crucial component firstly, since this is used to determine the match between candidate CV and position description sections and secondly it ranks the candidates based on their scores in a descending order.

[13]https://www.wikidata.org/

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

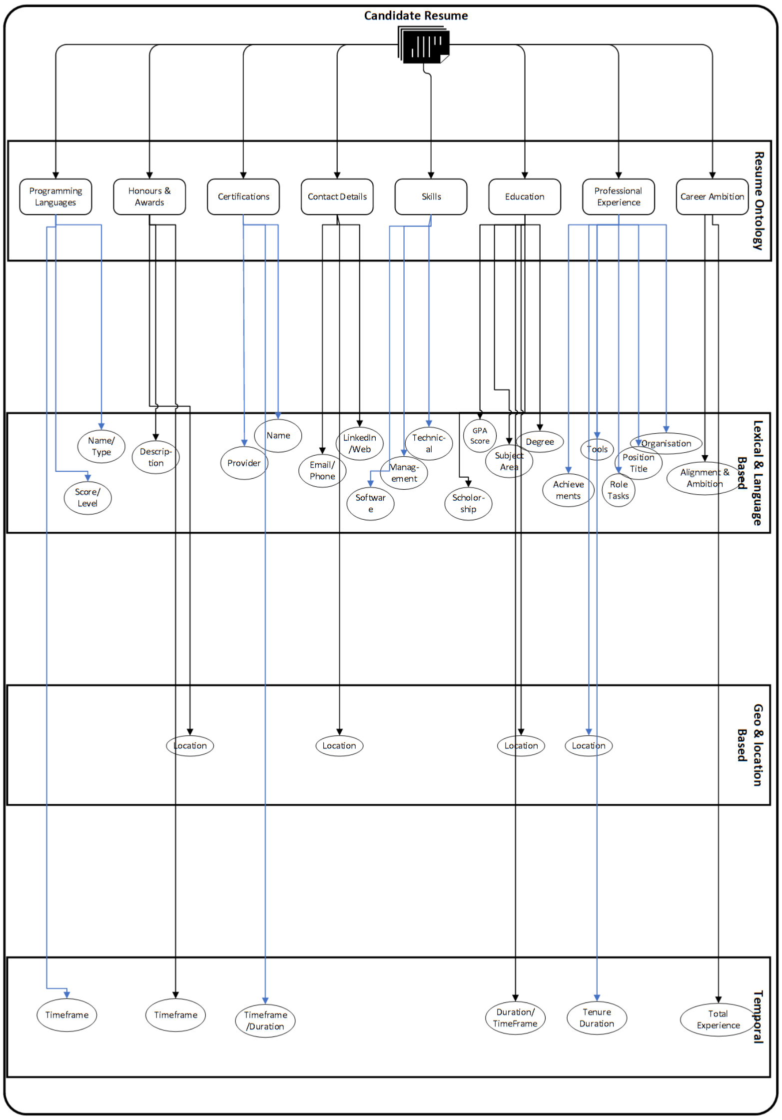

FIGURE 3.4: Contextualizing the candidate CV.

(Scoring Function) A Scoring Function S assigns a score to value V for an entity E extracted from the sub-section D of candidate's CV. This scoring function S can be then represented as: $(S, E, V, D) = \sum_{d \in D} \gamma_{d,e}.conf_x(e, a, d)$, where $\gamma_{d,e}$ represents the importance of d for E, and $conf_x(e, a, d)$ represents the confidence of (e, a) as extracted from the document d by extraction system X.

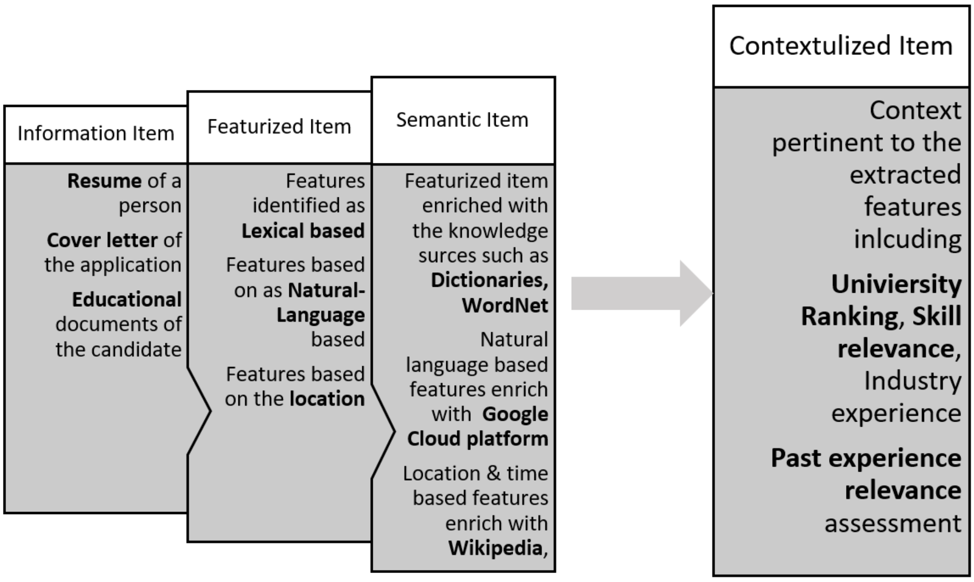

FIGURE 3.5: Constructing a contextualized-item from a raw business process.

3.1.4 Data Visualization Based Summary

In this phase we propose a data visualization based data summary approach. We leverage Microsoft Power BI [102] to present our data summaries and built a dashboard. Once the data has been extracted from the candidates CV and enriched with pertinent context. We then score this contextualised data with position description based on PDMS. Candidates are ranked based on the PDMS and then presented in the dashboard. Figure 3.6 illustrates the *Candidate Ranking Dashboard* developed in Power BI. The dashboard shows total number of candidates being evaluated and their extracted data from different sections of their CV. The users also have the ability to filter candidates based on certain criteria applied on the enriched entities. For example, a user might want to filter candidates specialising in a certain machine learning platform. Figure 3.6(A) shows how such a filter is applied and relevant candidates are filtered and presented in the dashboard. Similarly users also have the ability to filter candidates based on a certain type of experience. Figure 3.6(B) highlights how a user can filter candidates based on machine learning skills. Figure 3.6(C) highlights how a user can filter candidates based on experience type.

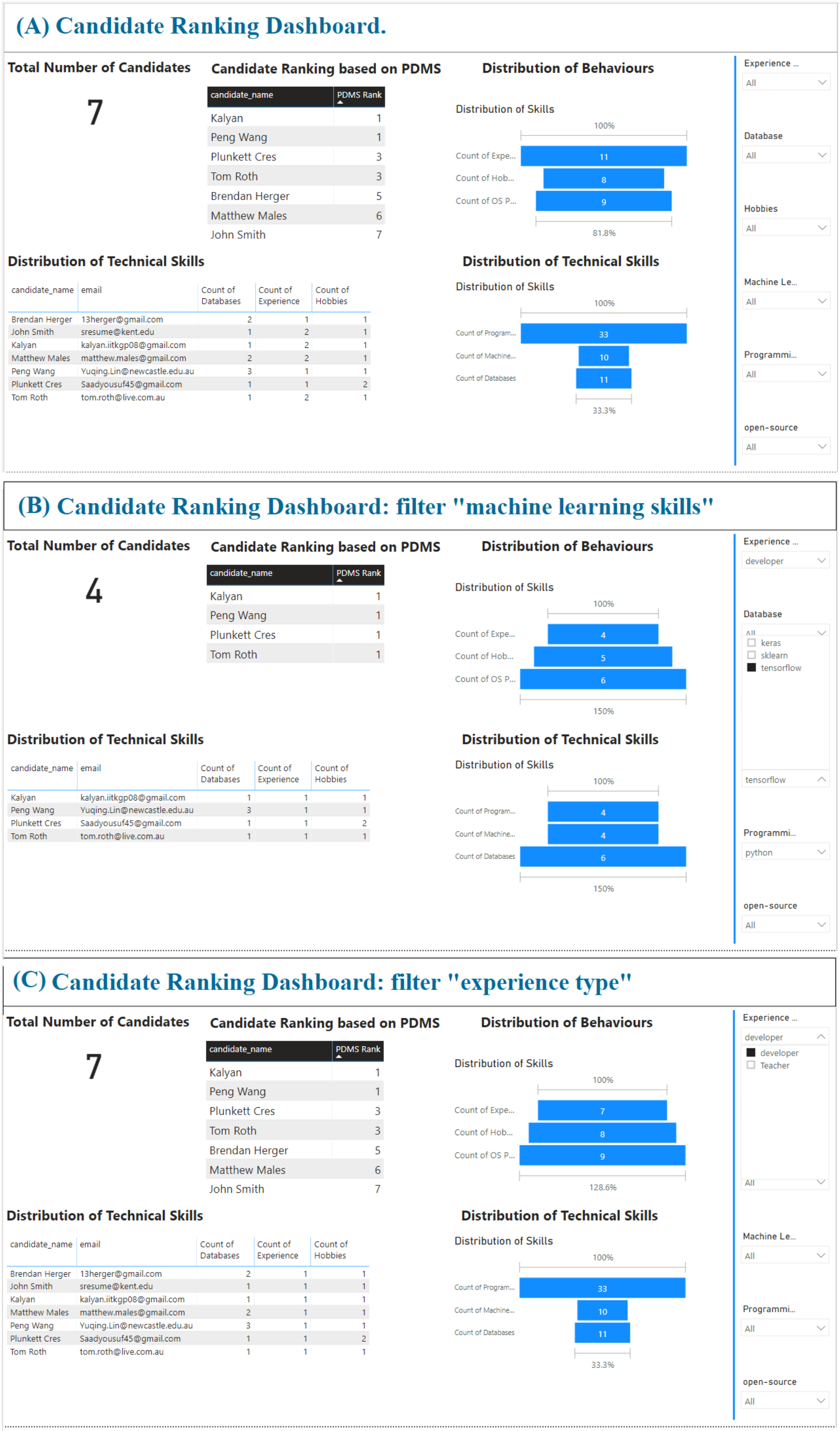

FIGURE 3.6: Screenshots of the Candidate Ranking Dashboard (A), filter on right data science candidates based on their machine learning skills (B) and experience type (C).

4

Experiment and Evaluation

While Recruiting talented people is paramount to organizational sustainability and success, recruiting wrong people is a prevailing issue not only in terms of financial cost but also from a organizational culture perspective as well. In Chapter3 we proposed a framework to imitate the knowledge of the recruiters into a recruitment domain Knowledge Base (rKB) which consists of a set of concepts organized into recruitment taxonomy. We also explained the techniques to construct the domain Knowledge Base (rKB). We focused on a motivating scenario of Recruitment of Data Scientist positions and linking the position description data to the candidate CV data and ranking the potential candidates. In this chapter we present the experiment and evaluation dimension for our approach.

4.1 Implementation and Evaluation

We develop services to extract the raw data from business artifacts such as CVs, job advertisements, candidates profiles, Universities and Companies Job descriptions, and job search engine websites such as indeed.com and theladders.net. These services will persist the data in the knowledge lake [68]. To imitate the knowledge of recruiters and inspired by Google Knowledge Graph[1], we focused on constructing a recruitment domain Knowledge Base (rKB). There are many systems that can be used at this level including curation APIs [16], Google Cloud Platform[2], and Microsoft Computer Vision API[3] to extract information items from artifacts (such as CVs, personal home pages and company Websites).

We have identified many useful machine learning algorithms and leverage these algorithms as service to enable us to summarize the constructed recruitment knowledge base [10, 12]. Machine Learning (ML) combines techniques from statistics and artificial intelligence to create algorithms that can learn from empirical data and generalize to solve problems in various domains. One of the main challenges in Machine Learning is to enable users to subscribe and use ML application software in the cloud. This task is challenging as building ML services or AI-based application is different from building traditional SaaS services.

For example, for training models, each training problem is different and analysts need a toolbox to explore different algorithms and pick the best ones that apply to building a particular model. In this context, it is important to make these models available as a service so that others can easily replicate the training as well as the test environments. This will enable knowledge workers of all skill levels (e.g., an end-user with limited computer science background to data scientists) to access and reuse ML services in their processes. The evaluation of accuracy and performance of the Knowledge Lake and knowledge extraction services demonstrated in [16, 68].

Extracting data and knowledge and linking them to the domain-knowledge. We have leveraged the in-house developed extraction APIs [66] to extract various information from the business artifacts, including CVs and Job advertisements. To extract information from CV's in PDF format, it

[1] https://developers.google.com/knowledge-graph/
[2] http://cloud.google.com/
[3] http://azure.microsoft.com/

is possible to use PDF extraction and convertor API's available in ProgrammableWeb[4] Website [14]. There are various open and external knowledge sources such as indeed [5], seek [6], glassdoor [7], collegegrad [8], and government open data [9]. In the next section, we evaluate the usability of the approach regarding the intended application audience, i.e., the recruiters and expert users.

4.1.1 Evaluation Hypothesis

In order to evaluate our approach and framework we conducted a user study. We focus on validating the following hypothesis with the help of this study:

1. (H1): The iRecruit framework is reliable in linking CV and position description with less training effort.
2. (H2): The summaries offered by iRecruit are comprehensible with little recruitment domain knowledge.

4.1.2 Experiment

We conducted a controlled environment experiment to study our framework. Participants were recruited with mainly two backgrounds: People with and without background in recruitment processes or people management.

The participants were first instructed on our motivating scenario and then given an overview on the usage of our framework through a presentation which followed the below sequence:

- *(Part 1) Imitating Recruiter's Knowledge and Extracting Information:* first, we highlighted why building a recruitment Knowledge Base (rKB) based on a recruiter's knowledge, was crucial to link the relevant information between candidate's CV and position description. For example, a recruiter might look at candidate's education profile and evaluate the University and its

[4] https://www.programmableweb.com/search/pdf
[5] www.indeed.com/
[6] www.seek.com.au/
[7] www.glassdoor.com/
[8] www.collegegrad.com/
[9] https://data.gov.au/search?q=job

rank. Afterwards we demonstrated how data would be extracted through our framework by using the rKB.

- *(Part 2) Data Curation:* by using enrichment functions on the extracted entities we highlighted the relevant context around the individual information items so that they can used more meaningfully.

- *(Part 3) Linking Data and Ranking Candidates:* we presented the extracted-enriched-linked data from candidate's CV with position description and presented the data in a data visualization based dashboard for easy insight generation. We also ranked candidates based on their Position Description Match Score (PDMS).

4.1.3 Questionnaire

The questionnaire was designed to be able to validate the evaluation hypothesis that were defined for the user study. The questionnaire comprised of 7 questions with all having multiple choice rating scale answers to choose from. Participants were instructed to choose only one option which they thought was most suitable in their opinion. Most questions expected a user to select a rating scale on a 5 point scale from Strongly Disagree to Strongly Agree. There were three questions each in the questionnaire that pertain to H1 and H2 validation while one question asked the participants about their background experience in recruitment or people management.

4.1.4 Evaluation of Hypothesis

To properly evaluating the Hypothesis, it would be important to find, attract, screen, and shortlist suitable candidates. Traditional ways used by recruiters include creating a shortlist scorecard to list out each criteria with the goal to assign a rating for the candidate features. An interesting future work in this category, would be to understand how a software can use AI to shortlist candidates. For example, techniques such as Reinforcement Learning can be used to monitor existing resume database to learn which candidates moved on to be successful and unsuccessful in the same role. This will enable the learning algorithm to learn from candidates' experience, education, and other

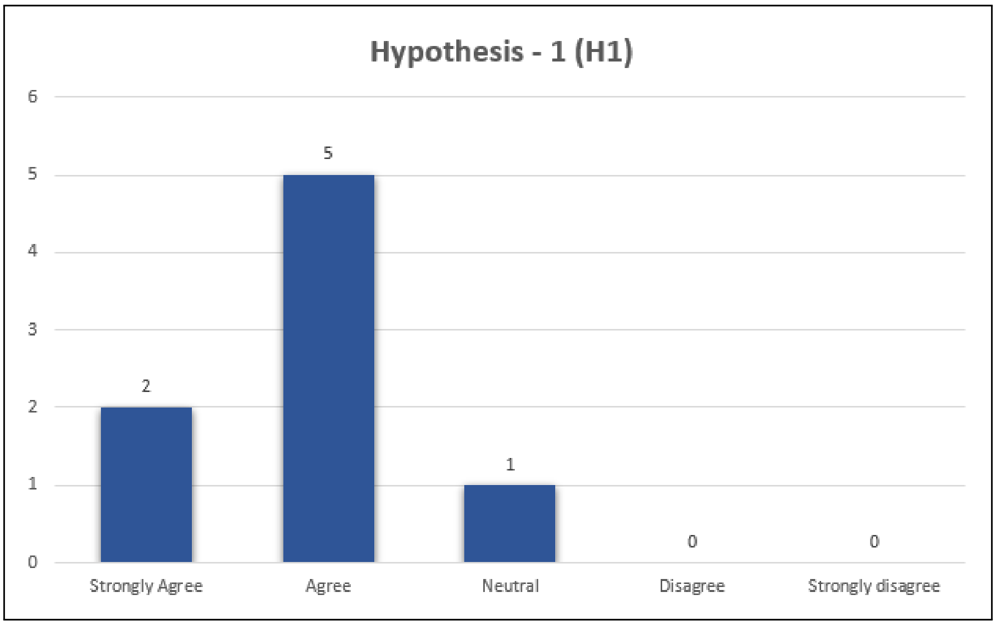

FIGURE 4.1: Evaluation of Hypothesis with 8 participants: (H1): The iRecruit framework is reliable in linking CV and position description with less training effort;

qualifications over time. In our evaluation, we relaid on the knowledge and expertise of the recruiters to help us analyzed the extracted information and use them on the suitability of CVs to the related positions. The study was conducted on 8 participants residing in Sydney, Australia. Figures 4.1 and 4.2 shows the results of evaluation of both the hypothesis.

Findings on H1

The experiment results indicate that overall all the participants except one found that the framework was able to link the CV and position description with less training effort. We also observed from Fig 4.1 that all the participants that had some people management or recruitment experience found that the approach was able to link the candidate CV and position description reliably. Except one participant all the participants either agreed or strongly agreed that the framework was able to rank candidates based on the positions description match score (PDMS) effectively and better as compared to the traditional methods of comparing candidate's CV. Moreover, majority of the participants found that the framework was effectively able to link the skills and experience section from candidate's CV with the position description requirements.

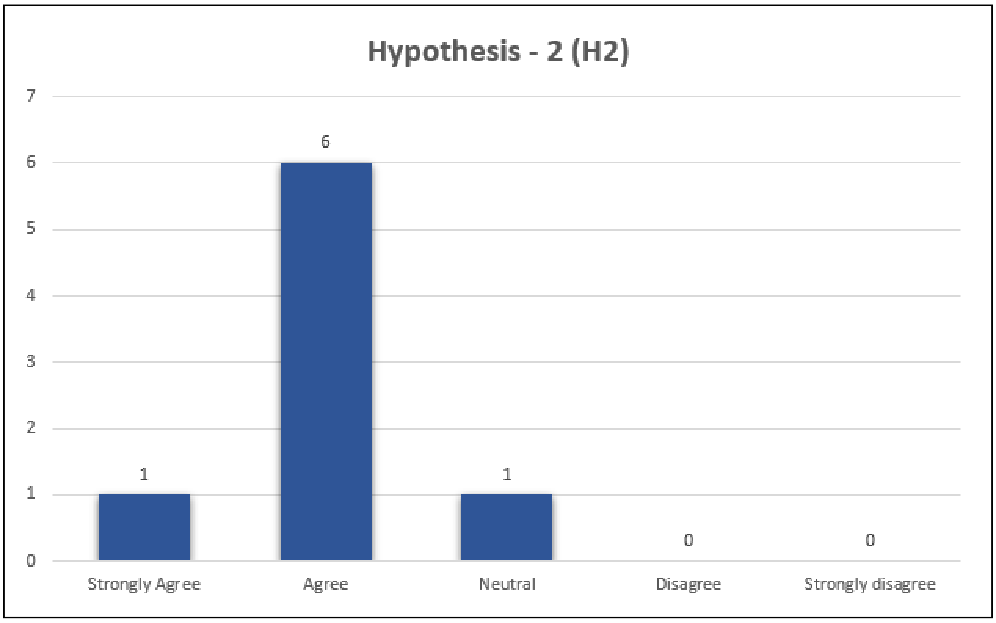

FIGURE 4.2: Evaluation of Hypothesis with 8 participants: (H2): The summaries offered by iRecruit are comprehensible with little domain knowledge;

Findings on H2

All the participants considered the data visualization based candidate summaries were very helpful and easy to comprehend. They also agreed that there is little recruitment domain knowledge required to comprehend the results. As observed from Fig 4.2, majority of the participants also found that the visual representation of the candidates CV sections were easier to draw insights from as compared to traditional methods of comparing candidates CVs. Moreover, most of the participants found the visual data summaries presented in the framework to be very easy and intuitive to compare multiple candidates and rank against each other.

5

Conclusion and Future Directions

In this chapter, we summarize the contributions of this thesis and discuss future research directions to develop on this work.

5.1 Concluding Remarks

Attracting and retaining human talent is cornerstone for every organization's success. organizations spend enormous amount of time and effort in championing the right organization's culture and horn out human skills that are required for it to sustain in the future. Most of the organizations by adapting traditional recruitment processes do not completely compare and rank candidates as compared to the position description or spend more time ranking candidates amongst each other. To overcome these challenges, we presented a framework namely iRecruit which helps the recruiters extract relevant information from a candidate's CV, enrich this information with context data and then link with position description. We present the ranking function Position Description Match Score (PDMS) which ranks candidates based on their match with position description and present these data summaries in a Data visualization based dashboard for easy insight generation. We presented the evaluation of our framework through a user study.

One of the possible future areas of focus for this thesis would be to evaluate the validity of the PDMS and subsequently improve this validity through optimisation techniques. Another major possible area of future work is to work on data extraction through rule based systems. These possible future works are elaborated in following sections.

5.2 Artificial Intelligence and Future of Recruitment

5.2.1 Rule-Based Knowledge Extraction

In order to build a domain specific knowledge-base, one may need to extract a subject matter expert's knowledge under certain conditions and save it in a knowledge base. Recruitment analysts (recruiters) make various decisions which are only understandable given certain conditions (rules) are *true*. Building a rule based recruitment knowledge-base is a potential future work for this thesis, since this defines what information is considered important from a recruiters perspective within a candidate's CV or from a open dataset. We would also like to explore how we can leverage *Adaptive Ripple Down Rules* [103], a Ripple Down Rules (RDR) approach to extract knowledge from recruitment experts (recruiters). RDR is an incremental approach to knowledge acquisition, i.e., the process used to define the rules and ontologies required for a knowledge-based system. A future line of work will be to link RDR to Intelligent Knowledge Lakes [104] to enable intelligent summarization of the large recruitment graph.

The main limitation of the current knowledge graph represented in this dissertation, includes the knowledge graph usage framework. This is quite important and enables the introduction of knowledge graph features to support Trust and Transparency concerns [105–109] as well as Privacy issues [110] regarding the use of knowledge graph contents by applications, such as in recommender systems [111, 112] and crowdSourcing of large knowledge graphs [113–115]. A future line of work will introduce a time-aware Blockchain-enhanced RDR to address the above mentioned shortcomings.

5.2.2 Ethical AI and Biases in Recruitment

Recruitment processes in the modern world focuses on diversity and inclusion and purely based on the skills required for the job openings. Many organizations highlight their effort to build more inclusive corporate culture on their job boards and company pages. Organizations realize that they could be missing out on top talent if they do not offer a more inclusive culture within their organizations. Discrimination of any kind can be very disruptive both for the employees and organizations alike and can trigger major regulatory problems for the organizations as well [116].

While employing AI brings a lot of benefit to the recruitment process, it also brings a major issue as well. Ethical AI as defined by House of Lords select committee on AI, UK is that AI techniques and application of algorithms in a manner that they are not allowed to hurt, destroy or deceive human beings [117]. The report also highlights that only a few companies dominate the AI technology and highlights to the UK government to use competition laws to prevent monopolization of data [117]. One of the other major concerns by introducing AI is the algorithm bias. Which occurs when a computer system makes decisions based on the data that might have prejudices of the people who designed it [118]. This can become a major concern since decoding the design mechanism of the algorithm is not that easy and requires back end design understanding. Within the next 5 years it is estimated that 55 percent to the households would own a voice assistant worldwide. Similarly medical diagnostics and cyber-security applications rely heavily on AI [119].

Ethical regulations are a must to ensure that risk are highlighted properly to avoid any prejudice towards any set of individuals or societies [119] Many large organizations are getting more aware of these programmatic biases that might be introduced within the recruitment process. Most of the studies have found that these biases are introduced towards the minority groups and they are disadvantaged since the training data for these AI based algorithms are not trained for smaller minority sets [120].

A
Appendix

Appendix A (page 43) removed from Open Access version as they may contain sensitive/confidential content.

References

[1] J. A. Breaugh. *Employee recruitment: Current knowledge and important areas for future research*. Human Resource Management Review **18**(3), 103 (2008). xv, 2, 6, 7, 8, 13

[2] R. L. Compton. *Effective recruitment and selection practices* (CCH Australia Limited, 2009). xv, 8, 9, 21

[3] S. Beheshti, B. Benatallah, S. Sakr, D. Grigori, H. R. Motahari-Nezhad, M. C. Barukh, A. Gater, and S. H. Ryu. *Process Analytics - Concepts and Techniques for Querying and Analyzing Process Data* (Springer, 2016). URL `https://doi.org/10.1007/978-3-319-25037-3`. 1, 5, 12, 13, 14, 19

[4] A. Beheshti, B. Benatallah, and H. R. Motahari-Nezhad. *Processatlas: A scalable and extensible platform for business process analytics*. Softw., Pract. Exper. **48**(4), 842 (2018). URL `https://doi.org/10.1002/spe.2558`. 11, 13, 15, 19, 25

[5] B. Benatallah, S. Sakr, D. Grigori, H. R. Motahari-Nezhad, M. C. Barukh, A. Gater, S. H. Ryu, *et al.* *Business process data analysis*. In *Process Analytics*, pp. 107–134 (Springer, 2016).

[6] B. Benatallah, S. Sakr, D. Grigori, H. R. Motahari-Nezhad, M. C. Barukh, A. Gater, S. H. Ryu, *et al.* *Business process paradigms*. In *Process analytics*, pp. 19–60 (Springer, 2016). 1

[7] S. Beheshti, B. Benatallah, H. R. M. Nezhad, and S. Sakr. *A query language for analyzing business processes execution*. In *Business Process Management - 9th International Conference, BPM 2011, Clermont-Ferrand, France, August 30 - September 2, 2011. Proceedings*, pp. 281–297 (2011). 1, 13, 19

[8] S. Beheshti, B. Benatallah, and H. R. Motahari-Nezhad. *Scalable graph-based OLAP analytics over process execution data*. Distributed and Parallel Databases **34**(3), 379 (2016). 1, 13, 14

[9] A. Beheshti, F. Schiliro, S. Ghodratnama, F. Amouzgar, B. Benatallah, J. Yang, Q. Z. Sheng, F. Casati, and H. R. Motahari-Nezhad. *iprocess: Enabling iot platforms in data-driven knowledge-intensive processes*. In *Business Process Management Forum - BPM Forum 2018, Sydney, NSW, Australia, September 9-14, 2018, Proceedings*, pp. 108–126 (2018). 2, 5

[10] F. Schiliro, A. Beheshti, S. Ghodratnama, F. Amouzgar, B. Benatallah, J. Yang, Q. Z. Sheng, F. Casati, and H. R. Motahari-Nezhad. *icop: Iot-enabled policing processes*. In *Service-Oriented Computing - ICSOC 2018 Workshops - ADMS, ASOCA, ISYyCC, CloTS, DDBS, and NLS4IoT, Hangzhou, China, November 12-15, 2018, Revised Selected Papers*, pp. 447–452 (2018). 2, 34

[11] U. Shahbaz, A. Beheshti, S. Nobari, Q. Qu, H. Paik, and M. Mahdavi. *irecruit: Towards automating the recruitment process*. In *Service Research and Innovation - 7th Australian Symposium, ASSRI 2018, Sydney, NSW, Australia, September 6, 2018, and Wollongong, NSW, Australia, December 14, 2018, Revised Selected Papers*, pp. 139–152 (2018). 2, 5, 10

[12] F. Amouzgar, A. Beheshti, S. Ghodratnama, B. Benatallah, J. Yang, and Q. Z. Sheng. *isheets: A spreadsheet-based machine learning development platform for data-driven process analytics*. In *Service-Oriented Computing - ICSOC 2018 Workshops - ADMS, ASOCA, ISYyCC, CloTS, DDBS, and NLS4IoT, Hangzhou, China, November 12-15, 2018, Revised Selected Papers*, pp. 453–457 (2018). 2, 34

[13] B. Shulman and G. Chiang. *When to use an executive search firm and how to get the most out of the relationship*. Employment Relations Today **34**(1), 13 (2007). 2, 6, 13

[14] R. Rastan, H. Paik, J. Shepherd, S. H. Ryu, and A. Beheshti. *TEXUS: table extraction system for PDF documents*. In *Databases Theory and Applications - 29th Australasian Database Conference, ADC 2018, Gold Coast, QLD, Australia, May 24-27, 2018, Proceedings*, pp. 345–349 (2018). 2, 27, 35

[15] S. Beheshti, B. Benatallah, S. Venugopal, S. H. Ryu, H. R. Motahari-Nezhad, and W. Wang. *A systematic review and comparative analysis of cross-document coreference resolution methods and tools*. Computing **99**(4), 313 (2017). URL https://doi.org/10.1007/s00607-016-0490-0. 2

[16] S. Beheshti, A. Tabebordbar, B. Benatallah, and R. Nouri. *On automating basic data curation tasks*. In *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*, pp. 165–169 (2017). 2, 14, 26, 27, 34

[17] A. Beheshti, B. Benatallah, A. Tabebordbar, H. R. Motahari-Nezhad, M. C. Barukh, and R. Nouri. *Datasynapse: A social data curation foundry*. Distributed and Parallel Databases **37**(3), 351 (2019). URL https://doi.org/10.1007/s10619-018-7245-1. 2, 14, 15, 27

[18] E. Rajabi *et al.* *Interlinking big data to web of data*. In *Big Data Optimization: Recent Developments and Challenges*, pp. 133–145 (Springer, 2016). 4

[19] Y. J. Sun, M. C. Barukh, B. Benatallah, and S. Beheshti. *Scalable saas-based process customization with casewalls*. In *Service-Oriented Computing - 13th International Conference, ICSOC 2015, Goa, India, November 16-19, 2015, Proceedings*, pp. 218–233 (2015). URL https://doi.org/10.1007/978-3-662-48616-0_14. 5, 12

[20] *Ladders, inc.; job search engine*. https://cdn.theladders.net/static/images/basicSite/pdfs/TheLadders-EyeTracking-StudyC2.pdf. Accessed: 2019-06-07. 5, 6, 13, 22

[21] R. E. Ployhart. *Staffing in the 21st century: New challenges and strategic opportunities*. Journal of management **32**(6), 868 (2006). 6, 13

[22] M. S. Taylor and C. J. Collins. *Organizational recruitment: Enhancing the intersection of research and practice* (2000). URL http://digitalcommons.ilr.cornell.edu/articles/568. 6

[23] D. ODonovan. *HRM in the Organization: An Overview*, pp. 75–110 (Springer International

Publishing, Cham, 2019). URL https://doi.org/10.1007/978-3-030-13229-3_4. 6

[24] *Managing human resources personnel management in transition* (Blackwell, Malden, Mass. u.a, 2005). 6, 7

[25] D. J. Storey. *Exploring the link, among small firms, between management training and firm performance: a comparison between the uk and other oecd countries*. The International Journal of Human Resource Management **15**(1), 112 (2004). URL http://www.tandfonline.com/doi/abs/10.1080/0958519032000157375. 6

[26] B. Becker and B. Gerhart. *The impact of human resource management on organizational performance: progress and prospects.(special research forum: Human resource management and organizational performance)*. Academy of Management Journal **39**(4) (1996). 6

[27] P. Bibi. *Hrm practices and empoyees' rentention: The job embeddedness theory prospective*. ASIAN JOURNAL OF MULTIDISCIPLINARY STUDIES **4**(5) (2016). URL http://www.ajms.co.in/sites/ajms2015/index.php/ajms/article/view/1745. 6

[28] A. Bas. *Strategic hr management: Strategy facilitation process by hr*. Procedia - Social and Behavioral Sciences **58**(C), 313 (2012). 6

[29] A. Stoilkovska, J. Ilieva, and S. Gjakovski. *Equal employment opportunities in the recruitment and selection process of human resources*. UTMS Journal of Economics **6**(2), 281 (2015). 7

[30] B. Griepentrog, C. Harold, B. Holtz, R. Klimoski, and S. Marsh. *Integrating social identity and the theory of planned behavior: Predicting withdrawal from an organizational recruitment process*. Personnel Psychology **65**(4), 723 (2012). 7

[31] A. . B. . Elsevier/Pergamon. *Recruitment and selection.* (Amsterdam ; Boston : Elsevier/Pergamon : ELEARN, ©2009, 2009). URL https://www.worldcat.org/title/recruitment-and-selection/oclc/474930987. 7

[32] Y. Acikgoz. *Employee recruitment and job search: Towards a multi-level integration*. Human Resource Management Review **29**(1), 1 (2019). 7

[33] K. Moser. *Recruitment sources and post-hire outcomes: The mediating role of unmet expectations*. International Journal of Selection and Assessment **13**(3), 188 (2005). 7

[34] P. Jansen and B. Stoop. *Assessment center graduate selection: Decision processes, validity, and evaluation by candidates*. International Journal of Selection and Assessment **2**(4), 193 (1994). 9

[35] J. Kottke and K. Shultz. *Using an assessment center as a developmental tool for graduate students: A demonstration*. Journal of Social Behavior and Personality **12**(5), 289 (1997). URL http://search.proquest.com/docview/1292310988/. 9

[36] K. Rogers. *Putting online; psychometrics to the test.(pros and cons of online psychometric testing as part of recruitment process)*. Personnel Today p. 1 (2002). 10

[37] A. Beheshti, V. M. Hashemi, S. Yakhchi, H. R. Motahari-Nezhad, S. M. Ghafari, and J. Yang. *personality2vec: Enabling the analysis of behavioral disorders in social networks*. In *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, pp. 825–828 (2020). URL https://doi.org/10.1145/3336191.3371865. 10

[38] A. Beheshti, V. M. Hashemi, and S. Yakhchi. *Towards context-aware social behavioral analytics*. In P. D. Haghighi, I. L. Salvadori, M. Steinbauer, I. Khalil, and G. Anderst-Kotsis, eds., *MoMM 2019: The 17th International Conference on Advances in Mobile Computing & Multimedia, Munich, Germany, December 2-4, 2019*, pp. 28–35 (ACM, 2019). URL https://doi.org/10.1145/3365921.3365942. 10

[39] J. Cartwright. *To ask or not to ask ... that's the big question: from personality profiling to psychometric testing, pre-employment screening is becoming an increasingly important part of the recruitment process. is it relevant, insightful or just plain intrusive?(randstad column)*. Australian Banking and Finance **2**(7), 25 (2010). 10

[40] B. Farago. *Applicant reactions to structured and unstructured recruitment interviews*. Ph.D. thesis (2010). Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2019-10-19, URL http:

//simsrad.net.ocs.mq.edu.au/login?url=https://search-proquest-com.simsrad.net.ocs.mq.edu.au/docview/756453947?accountid=12219. 10

[41] B. L. KAROL. *RELATIONSHIP OF RECRUITER BEHAVIOR, PERCEIVED SIMILARITY, AND PRIOR INFORMATION TO APPLICANTS' ASSESSMENTS OF THE CAMPUS RECRUITMENT INTERVIEW*. Ph.D. thesis (1977). Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2019-10-16, URL http://simsrad.net.ocs.mq.edu.au/login?url=https://search-proquest-com.simsrad.net.ocs.mq.edu.au/docview/302816371?accountid=12219. 10

[42] P. Dhamija. *E-recruitment: A roadmap towards e-human resource management* (2012). 10, 11

[43] N. Anderson. *Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research*. International Journal of Selection and Assessment **11**(2-3), 121 (2003). 11

[44] J. Reingold, L. Armstrong, W. Zellner, and E. Baig. *Executive job searches going high tech*. Health Progress **80**(4) (1999). URL http://search.proquest.com/docview/274594246/. 11

[45] A. Uszok and P. Kunstman. *Intelligent software agent system architecture* (2010). US Patent 7,861,252. 11

[46] L. Panayotopoulou, M. Vakola, and E. Galanaki. *E-hr adoption and the role of hrm: evidence from greece*. Personnel Review **36**(2), 277 (2007). 11

[47] H. K. Aggerholm and S. E. Andersen. *Social media recruitment 3.0*. Journal of Communication Management **22**(2), 122 (2018). 11

[48] A. Garcia. *Diversity inclusion in silicon valley technology companies*. International Journal of Innovation and Applied Studies **12**(3), 761 (2015). URL http://search.proquest.com/docview/1703411559/. 12

[49] L. Strickland, B. Blonder, E. Klein, M. Kohl, E. Mcgee, M. Quintana, R. Ridley, B. Tellman, and L. Gerber. *Without inclusion, diversity initiatives may not be enough*. Science **357**(6356), 1101 (2017). URL http://search.proquest.com/docview/1974996234/. 12

[50] S. Beheshti, B. Benatallah, H. R. Motahari Nezhad, and M. Allahbakhsh. *A framework and a language for on-line analytical processing on graphs*. In *Web Information Systems Engineering - WISE 2012 - 13th International Conference, Paphos, Cyprus, November 28-30, 2012. Proceedings*, pp. 213–227 (2012). 13

[51] W. M. Van der Aalst. *Process mining in the large: a tutorial*. In *European Business Intelligence Summer School*, pp. 33–76 (Springer, 2013). 13

[52] A. Tabebordbar, A. Beheshti, B. Benatallah, and M. C. Barukh. *Adaptive rule adaptation in unstructured and dynamic environments*. In *Web Information Systems Engineering - WISE 2019 - 20th International Conference, Hong Kong, China, November 26-30, 2019, Proceedings*, pp. 326–340 (2019). 13

[53] A. Tabebordbar and A. Beheshti. *Adaptive rule monitoring system*. In *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services, SE4COG@ICSE 2018, Gothenburg, Sweden, May 28-2, 2018*, pp. 45–51 (2018). 14

[54] S. Beheshti, B. Benatallah, and H. R. Motahari Nezhad. *Enabling the analysis of cross-cutting aspects in ad-hoc processes*. In *Advanced Information Systems Engineering - 25th International Conference, CAiSE 2013, Valencia, Spain, June 17-21, 2013. Proceedings*, pp. 51–67 (2013). 13

[55] A. Beheshti, K. Vaghani, B. Benatallah, and A. Tabebordbar. *Crowdcorrect: A curation pipeline for social data cleansing and curation*. In *Information Systems in the Big Data Era - CAiSE Forum 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings*, pp. 24–38 (2018). 13

[56] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, E. Bertino, and N. Foo. *Reputation management in crowdsourcing systems*. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2012, Pittsburgh, PA, USA, October 14-17, 2012*, pp. 664–671 (2012). 13

[57] E. Duipmans and D. L. F. Pires. *Business process management in the cloud: business process as a service (bpaas)*. University of Twente (2012). 13

[58] K. D. Swenson. *Social BPM: work, planning and collaboration under the impact of social technology* (Future Strategies Inc., 2011). 13

[59] Z. Maamar, S. Sakr, A. Barnawi, and S. Beheshti. *A framework of enriching business processes life-cycle with tagging information*. In M. A. Sharaf, M. A. Cheema, and J. Qi, eds., *Databases Theory and Applications - 26th Australasian Database Conference, ADC 2015, Melbourne, VIC, Australia, June 4-7, 2015. Proceedings*, vol. 9093 of *Lecture Notes in Computer Science*, pp. 309–313 (Springer, 2015). URL https://doi.org/10.1007/978-3-319-19548-3_25. 13

[60] T. Melamed and D. Jackson. *Psychometric instruments: potential benefits and practical use*. Industrial and Commercial Training **27**(4), 11 (1995). 13

[61] R. Kwiatkowski. *Trends in organisations and selection: an introduction*. Journal of Managerial Psychology **18**(5), 382 (2003). 14

[62] W. Van Der Aalst and et al. *Process mining manifesto*. In *International Conference on Business Process Management*, pp. 169–194 (Springer, 2011). 14

[63] F. Casati, M. Castellanos, U. Dayal, and N. Salazar. *A generic solution for warehousing business process data*. In *Proceedings of the 33rd international conference on Very large data bases*, pp. 1128–1137 (VLDB Endowment, 2007). 14

[64] M. J. Carey, N. Onose, and M. Petropoulos. *Data services*. Communications of the ACM **55**(6), 86 (2012). 14

[65] H. R. M. Nezhad, R. Saint-Paul, F. Casati, and B. Benatallah. *Event correlation for process discovery from web service interaction logs*. VLDB J. **20**(3), 417 (2011). 14

[66] S. Beheshti, A. Tabebordbar, B. Benatallah, and R. Nouri. *Data curation apis*. CoRR **abs/1612.03277** (2016). 1612.03277, URL http://arxiv.org/abs/1612.03277. 14, 34

[67] A. Beheshti, B. Benatallah, R. Nouri, V. M. Chhieng, H. Xiong, and X. Zhao. *Coredb: a data lake service*. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pp. 2451–2454 (2017). 14

[68] A. Beheshti, B. Benatallah, R. Nouri, and A. Tabebordbar. *Corekg: a knowledge lake service*. PVLDB (2018). 14, 27, 34

[69] A. Tabebordbar, A. Beheshti, and B. Benatallah. *Conceptmap: A conceptual approach for formulating user preferences in large information spaces*. In *Web Information Systems Engineering - WISE 2019 - 20th International Conference, Hong Kong, China, November 26-30, 2019, Proceedings*, pp. 779–794 (2019). 15

[70] C. Gerede and J. Su. *Specification and verification of artifact behaviors in business process models*. In *ICSOC*, pp. 181–192 (2007). 15, 16

[71] K. Bhattacharya, C. E. Gerede, R. Hull, R. Liu, and J. Su. *Towards formal analysis of artifact-centric business process models*. In *BPM*, pp. 288–304 (2007). 16

[72] Y. Sun, J. Su, and J. Yang. *Universal artifacts: A new approach to business process management (BPM) systems*. ACM Trans. Management Inf. Syst. **7**(1), 3:1 (2016).

[73] S. Beheshti, H. R. M. Nezhad, and B. Benatallah. *Temporal provenance model (TPM): model and query language*. CoRR **abs/1211.5009** (2012). 15

[74] J. Kuo. *A document-driven agent-based approach for business processes management*. Information and Software Technology **46**(6), 373 (2004). 16

[75] H. Reijers, J. Rigter, and W. Aalst. *The case handling case*. Int. J. Cooperative Inf. Syst. **12**(3), 365 (2003). 16

[76] H. Schonenberg, B. Weber, B. F. van Dongen, and W. M. P. van der Aalst. *Supporting flexible processes through recommendations based on history*. In *BPM*, pp. 51–66 (2008). 16

[77] S. Beheshti, B. Benatallah, and H. R. M. Nezhad. *Enabling the analysis of cross-cutting aspects in ad-hoc processes*. In *CAiSE*, pp. 51–67 (2013). 16

[78] A. K. Upadhyay and K. Khandelwal. *Applying artificial intelligence: implications for recruitment*. Strategic HR Review **17**(5), 255 (2018). 16, 17, 18

[79] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach* (Prentice Hall Press, Upper Saddle River, NJ, USA, 2009), 3rd ed. 16

[80] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, E. Bertino, and N. Foo. *Collusion detection in online rating systems*. In *Web Technologies and Applications - 15th Asia-Pacific Web Conference, APWeb 2013, Sydney, Australia, April 4-6, 2013. Proceedings*, pp. 196–207 (2013). 16

[81] E. Faliagka, K. Ramantas, A. Tsakalidis, and G. Tzimas. *Application of machine learning algorithms to an online recruitment system* (2012). 16, 17, 18

[82] P. Montuschi, V. Gatteschi, F. Lamberti, A. Sanna, and C. Demartini. *Job recruitment and job seeking processes: How technology can help*. IT Professional **16**(5), 41 (2014). 17

[83] *Ibmwatson; watson recruitment*. https://www.ibm.com/talent-management/hr-solutions/recruiting-software. Accessed: 2019-09-15. 17

[84] *Ibmwatson; datasheet2019*. https://www.ibm.com/downloads/cas/1AKXPKVV. Accessed: 2019-09-15. 17

[85] *Hirevue; job search engine*. https://www.hirevue.com/products/video-interviewing. Accessed: 2019-09-14. 17

[86] *Hirevue; case study 2017*. https://cdn2.hubspot.net/hubfs/464889/Hilton%20Aug%202017/2017_12_SuccessStory_Hilton_CustomerMarketing3.pdf?__hstc=&__hssc=&hsCtaTracking=b76cefe9-cece-4631-bef5-53084aa900e5%7Ca26c5bca-5fbe-46e3-b0e4-c1f79894f72c. Accessed: 2019-09-14. 17

[87] D. R. Dickson and K. Nusair. *An hr perspective: the global hunt for talent in the digital age*. Worldwide Hospitality and Tourism Themes **2**(1), 86 (2010). 18

[88] I. Martincevic and G. Kozina. *The impact of new technology adaptation in business*. In *Economic and Social Development: Book of Proceedings*, pp. 842–848 (Varazdin Development and

Entrepreneurship Agency (VADEA), Varazdin, 2018). URL http://search.proquest.com/docview/2176205516/. 18

[89] M. Y. Siyu Zang. *Human resource management in the era of big data*. Journal of Human Resource and Sustainability Studies pp. 41–45 (2015). 18

[90] T. Rasmussen and D. Ulrich. *Learning from practice: how hr analytics avoids being a management fad*. Organizational Dynamics **44**(3), 236 (2015). 18

[91] T. Bondarouk and C. Brewster. *Conceptualising the future of hrm and technology research*. International Journal of Human Resource Management **27**(21), 2652 (2016). 18

[92] P. Compton and R. Jansen. *A philosophical basis for knowledge acquisition* (1989). 19

[93] *Knowledge discovery, inc.; ai, data science, and machine learning site*. https://www.kdnuggets.com/2018/09/how-many-data-scientists-are-there.html. Accessed: 2019-16-07. 23

[94] *Linkedin, inc.; linkedin u.s. workforce report jan-2019.; professional network site*. https://economicgraph.linkedin.com/resources/linkedin-workforce-report-january-2019. Accessed: 2019-16-07. 23

[95] *Linkedin, inc.; identifying apec's ai talent.; professional network site*. https://economicgraph.linkedin.com/research/identifying-apec-ai-talent. Accessed: 2019-16-07. 23

[96] M. Hammoud, D. A. Rabbou, R. Nouri, S. Beheshti, and S. Sakr. *DREAM: distributed RDF engine with adaptive query planner and minimal communication*. PVLDB **8**(6), 654 (2015). URL http://www.vldb.org/pvldb/vol8/p654-Hammoud.pdf. 24, 25

[97] O. Batarfi, R. E. Shawi, A. G. Fayoumi, R. Nouri, S. Beheshti, A. Barnawi, and S. Sakr. *Large scale graph processing systems: survey and an experimental evaluation*. Cluster Computing **18**(3), 1189 (2015). URL https://doi.org/10.1007/s10586-015-0472-6. 25

[98] S.-M.-R. Beheshti, S. Venugopal, S. H. Ryu, B. Benatallah, and W. Wang. *Big data and cross-document coreference resolution: Current state and future opportunities*. arXiv preprint arXiv:1311.3987 (2013). 26

[99] S. Beheshti, B. Benatallah, and H. R. Motahari-Nezhad. *Galaxy: A platform for explorative analysis of open data sources*. In *Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016.*, pp. 640–643 (2016). 27

[100] S.-M.-R. Beheshti, S. Sakr, B. Benatallah, and H. R. Motahari-Nezhad. *Extending sparql to support entity grouping and path queries*. arXiv preprint arXiv:1211.5817 (2012). 27

[101] R. Rastan, H. Paik, J. Shepherd, S. H. Ryu, and A. Beheshti. *TEXUS: table extraction system for PDF documents*. In *Databases Theory and Applications - 29th Australasian Database Conference, ADC 2018, Gold Coast, QLD, Australia, May 24-27, 2018, Proceedings*, pp. 345–349 (2018). 27

[102] *Power bi desktop; data visualisation tool*. https://powerbi.microsoft.com/en-us/. Accessed: 2019-06-07. 31

[103] T. Yoshida, H. Motoda, T. Washio, and T. Wada. *Adaptive ripple down rules method based on minimum description length principle*. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, pp. 530– (IEEE Computer Society, Washington, DC, USA, 2002). URL http://dl.acm.org/citation.cfm?id=844380.844772. 40

[104] A. Beheshti, B. Benatallah, Q. Z. Sheng, and F. Schiliro. *Intelligent knowledge lakes: The age of artificial intelligence and big data*. In *Web Information Systems Engineering - WISE 2019 - 20th International Conference, Hong Kong, China, November 26-30, 2019, Proceedings* (2019). 40

[105] S. M. Ghafari, S. Yakhchi, A. Beheshti, and M. A. Orgun. *SETTRUST: social exchange theory based context-aware trust prediction in online social networks*. In H. Hacid, Q. Z. Sheng, T. Yoshida, A. Sarkheyli, and R. Zhou, eds., *Data Quality and Trust in Big Data - 5th International Workshop, QUAT 2018, Held in Conjunction with WISE 2018, Dubai, UAE, November*

12-15, 2018, Revised Selected Papers, vol. 11235 of *Lecture Notes in Computer Science*, pp. 46–61 (Springer, 2018). URL https://doi.org/10.1007/978-3-030-19143-6_4. 40

[106] S. M. Ghafari, A. Joshi, A. Beheshti, C. Paris, S. Yakhchi, and M. A. Orgun. *DCAT: A deep context-aware trust prediction approach for online social networks*. In P. D. Haghighi, I. L. Salvadori, M. Steinbauer, I. Khalil, and G. Anderst-Kotsis, eds., *MoMM 2019: The 17th International Conference on Advances in Mobile Computing & Multimedia, Munich, Germany, December 2-4, 2019*, pp. 20–27 (ACM, 2019). URL https://doi.org/10.1145/3365921.3365940.

[107] S. M. Ghafari, S. Yakhchi, A. Beheshti, and M. A. Orgun. *Social context-aware trust prediction: Methods for identifying fake news*. In H. Hacid, W. Cellary, H. Wang, H. Paik, and R. Zhou, eds., *Web Information Systems Engineering - WISE 2018 - 19th International Conference, Dubai, United Arab Emirates, November 12-15, 2018, Proceedings, Part I*, vol. 11233 of *Lecture Notes in Computer Science*, pp. 161–177 (Springer, 2018). URL https://doi.org/10.1007/978-3-030-02922-7_11.

[108] B. Abu-Salih, P. Wongthongtham, S. Beheshti, and D. Zhu. *A preliminary approach to domain-based evaluation of users' trustworthiness in online social networks*. In B. Carminati and L. Khan, eds., *2015 IEEE International Congress on Big Data, New York City, NY, USA, June 27 - July 2, 2015*, pp. 460–466 (IEEE Computer Society, 2015). URL https://doi.org/10.1109/BigDataCongress.2015.74.

[109] B. Abu-Salih, P. Wongthongtham, S. Beheshti, and B. Zadjabbari. *Towards a methodology for social business intelligence in the era of big social data incorporating trust and semantic analysis*. In J. H. Abawajy, M. Othman, R. Ghazali, M. M. Deris, H. Mahdin, and T. Herawan, eds., *Proceedings of the Second International Conference on Advanced Data and Information Engineering, DaEng 2015, Bali, Indonesia, April 25-26, 2015*, vol. 520 of *Lecture Notes in Electrical Engineering*, pp. 519–527 (Springer, 2015). URL https://doi.org/10.1007/978-981-13-1799-6_54. 40

[110] U. Salama, L. Yao, X. Wang, H. Paik, and A. Beheshti. *Multi-level privacy-preserving access control as a service for personal healthcare monitoring*. In I. Altintas and S. Chen, eds., *2017*

IEEE International Conference on Web Services, ICWS 2017, Honolulu, HI, USA, June 25-30, 2017, pp. 878–881 (IEEE, 2017). URL https://doi.org/10.1109/ICWS.2017.111. 40

[111] S. Yakhchi, S. M. Ghafari, and A. Beheshti. *CNR: cross-network recommendation embedding user's personality*. In H. Hacid, Q. Z. Sheng, T. Yoshida, A. Sarkheyli, and R. Zhou, eds., *Data Quality and Trust in Big Data - 5th International Workshop, QUAT 2018, Held in Conjunction with WISE 2018, Dubai, UAE, November 12-15, 2018, Revised Selected Papers*, vol. 11235 of *Lecture Notes in Computer Science*, pp. 62–77 (Springer, 2018). URL https://doi.org/10.1007/978-3-030-19143-6_5. 40

[112] S. Yakhchi, A. Beheshti, S. M. Ghafari, and M. A. Orgun. *Enabling the analysis of personality aspects in recommender systems*. CoRR **abs/2001.04825** (2020). 2001.04825, URL https://arxiv.org/abs/2001.04825. 40

[113] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, N. Foo, and E. Bertino. *Representation and querying of unfair evaluations in social rating systems*. Comput. Secur. **41**, 68 (2014). URL https://doi.org/10.1016/j.cose.2013.09.008. 40

[114] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, N. Foo, and E. Bertino. *Detecting, representing and querying collusion in online rating systems*. CoRR **abs/1211.0963** (2012). 1211.0963, URL http://arxiv.org/abs/1211.0963.

[115] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, N. Foo, and E. Bertino. *An analytic approach to people evaluation in crowdsourcing systems*. CoRR **abs/1211.3200** (2012). 1211.3200, URL http://arxiv.org/abs/1211.3200. 40

[116] M. Bell, M. Özbilgin, T. Beauregard, and O. Sürgevil. *Voice, silence, and diversity in 21st century organizations: Strategies for inclusion of gay, lesbian, bisexual, and transgender employees*. Human Resource Management **50**(1), 131 (2011). URL http://search.proquest.com/docview/851363675/. 41

[117] *Houseoflordsselectcommittee; ai in the uk ready, willing and able*. https://www.hirevue.com/products/video-interviewing. Accessed: 2019-09-16. 41

[118] Anonymous. *What's all the fuss about... ethical ai*. Computer Act!ve (527), 49 (2018). URL http://search.proquest.com/docview/2120136700/. 41

[119] J. Fahrenkamp-Uppenbrink. *An ethical way forward for ai*. Science **361**(6404), 763.17 (2018). 41

[120] G. Beattie and P. Johnson. *Possible unconscious bias in recruitment and promotion and the need to promote equality*. Perspectives: Policy and Practice in Higher Education **16**(1), 7 (2012). URL http://www.tandfonline.com/doi/abs/10.1080/13603108.2011.611833. 41