International Journal of Manpower

Semantic matching of job seeker to vacancy: a bidirectional approach

Sisay Adugna Chala, Fazel Ansari, Madjid Fathi, Kea Tijdens,

Semantic matching of job seeker to vacancy: a bidirectional approach

Received 15 October 2018
Accepted 15 October 2018

Sisay Adugna Chala
Institute of Knowledge Based Systems and Knowledge Management, University of Siegen, Siegen, Germany
Fazel Ansari
Department of Industrial and Systems Engineering, Institute of Management Sciences, Vienna University of Technology, Vienna, Austria and Institute of Knowledge Based Systems and Knowledge Management, University of Siegen, Siegen, Germany
Madjid Fathi
Institute of Knowledge Based Systems and Knowledge Management, University of Siegen, Siegen, Germany, and
Kea Tijdens
Amsterdam Institute for Advanced Labour Studies, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Purpose – The purpose of this paper is to propose a framework of an automatic bidirectional matching system that measures the degree of semantic similarity of job-seeker qualifications and skills, against the vacancy provided by employers or job-agents.

Design/methodology/approach – The paper presents a framework of bidirectional jobseeker-to-vacancy matching system. Using occupational data from various sources such as the WageIndicator web survey, International Standard Classification of Occupations, European Skills, Competences, Qualifications, and Occupations as well as vacancy data from various open access internet sources and job seekers information from social networking sites, the authors apply machine learning techniques for bidirectional matching of job vacancies and occupational standards to enhance the contents of job vacancies and job seekers profiles. The authors also apply bidirectional matching of job seeker profiles and vacancies, i.e., semantic matching vacancies to job seekers and vice versa in the individual level. Moreover, data from occupational standards and social networks were utilized to enhance the relevance (i.e. degree of similarity) of job vacancies and job seekers, respectively.

Findings – The paper provides empirical insights of increase in job vacancy advertisements on the selected jobs – Internet of Things – with respect to other job vacancies, and identifies the evolution of job profiles and its effect on job vacancies announcements in the era of Industry 4.0. In addition, the paper shows the gap between job seeker interests and available jobs in the selected job area.

Research limitations/implications – Due to limited data about jobseekers, the research results may not guarantee high quality of recommendation and maturity of matching results. Therefore, further research is required to test if the proposed system works for other domains as well as more diverse data sets.

Originality/value – The paper demonstrates how online jobseeker-to-vacancy matching can be improved by use of semantic technology and the integration of occupational standards, web survey data, and social networking data into user profile collection and matching.

Keywords Recommender systems, Job description, Bidirectional matching, Job seeker modelling, Semantic matching, Vacancy recommendation

Paper type Research paper

The authors would like to acknowledge the financial support of the Eduworks Marie Curie Initial Training Network Project (PITN-GA-2013-608311) which is part of the European Commission's 7th Framework Program.

0143-7720
DOI 10.1108/IJM-10-2018-0331

1. Introduction

The present day industries require a different set of skills from the labor force as compared to the previous ones due to prevalence of automation. As automation is changing the landscape of skill requirements in industries, studies to analyze the requirements of industries, to provide labor market skills set demand, and ultimately to find the jobseeker-vacancy "best (optimal) match" is of paramount importance. So far, job seekers look for job vacancy advertisements and study the details of job requirements to decide whether they are suitable for their level of expertise which they have stipulated on their resume. Though vacancies are publicly available, due to an overwhelming volume and veracity of data, job seekers are not able to easily find relevant vacancy for their skill set or are unable to analyze the requirements to estimate its relevance. On the other hand, vacancies are not often prepared with desired skill set required by employers (i.e. job provider). It is also becoming customary that recruiters look for online profiles of potential employees from professional networking sites, e.g. Linkedin® and/or via recommendations from networks, such as partners and alliances (Sacchetti, 2013; Rafi and Shaikh, 2013; Hernandez, 2015; Godliman, 2009).

Besides, employees' job descriptions are prepared independently of job vacancy requirements and stored in professional networking sites or collected by job designers/analyzers who do job classifications (ILO, 2012), or who do job analysis research like WageIndicator (WageIndicator, 2015). However, resumes ́ of job seekers fail short of portraying the information available on those sites to cover all required skill sets in accordance with the job descriptions in vacancies.

Belloni *et al.* (2016) emphasized the need to improve the quality of job descriptions. If there is a way that job vacancies are prepared by getting heuristic information from analysis of job descriptions (developed by occupational analysts[1]), taking into account the heuristic information from job vacancies (like what job seekers do), the likelihood of getting the best match of job description for a vacancy will be improved.

The fundamental rationale for studying bidirectional job vacancy matching and recommendation, from the semantic relations perspective, is that job vacancies are not filled due to employees leaving or getting fired because they landed in a job that they are not fitting to, in the first place. The measurement errors associated with the way people provide their job descriptions in relation to its effect on occupational coding are discussed by Belloni *et al.* (2016). Moreover, the internet provides a conducive environment for interaction between organizations and individuals, i.e., employers and job seekers, respectively, in this case.

In this research, we propose the conception and realization of an automatic bidirectional matching system (Kucherov *et al.*, 2014; Muderedzwa and Nyakwende, 2010) that measures the degree of semantic similarity of job descriptions provided by job-seeker, job-holder or job-designer against the vacancy provided by employer or job-agent. This similarity can then provide a feedback to improve job descriptions based on the requirements of vacancies. It also does feed-forward suggestions for the improvement to the preparation of accurate vacancies based on up-to-date job descriptions. We seek to extend the current literature to address conceptual semantic relationships by incorporating detailed descriptions on occupational standards as well as social network data into the process of matching jobseekers' skills relevance to vacancies' requirements. This involves the use of data from open access online sources such as user's self-assessment web survey data from WageIndicator (Wageindicator, 2015), social networking data from open access online sources, standard job description data from International Standard Classification of Occupations and European Skills, Competences, Qualifications, and Occupations (EC, 2015), and vacancy data scraped from the web. The result of the study will be a robust model of jobseeker to vacancy matching that more accurately suggests the top most relevant list of jobseekers possessing skills required by a particular vacancy and vice versa. This study is part of the research on the development of person-job matching system that involves collection and analysis of unstructured text data for: job vacancy modeling enhanced by data from occupational standards (Chala *et al.*, 2016a); improved data collection interface for job seeker skills self-assessment (Chala *et al.*, 2016b); improved job seeker modeling through social network analysis; and matching the job vacancy to the job seeker.

The rest of this paper is organized as follows: in Section 2, we examine the state of the practice in text analysis techniques for bidirectional matching system. In Section 3, we discuss the methodology which is dedicated to describing the data and system setup where we elaborate the source and type of data, data pre-processing, conceptual setup of the system and related algorithms. In Section 4, we describe the use case employed for evaluation of the proposed system in the context of Industry 4.0. Section 5, covers evaluation of the proposed system and presents the preliminary results. Finally, we summarize the results and outline the future research works of the study.

2. Related works

Taking into account the challenges of semantic matching such as computational time, lack of available labeled data and difficulty in feature selection (Christen, 2012), a combination of algorithms needs to be employed for clustering of the textual data, measuring the similarity, matching and searching.

Clustering refers to set of methods and algorithms for analysis of objects (e.g. graph, data, document, text or term) to identify related items, and organizing them into groups whose members are similar (Fortunato, 2010; Schaeffer, 2007; Biemann, 2012; Fasulo, 1999). Proper selection of a clustering method (or algorithm) is highly related to the application context, i.e. clustering of graphs, data, document, text or term. In data clustering – (Gan *et al.*, 2007) – communities are sets of points which are close to each other, with respect to a measure of distance or similarity (Fortunato, 2010). The latter is potentially adaptable to our approach (Charu and Zhai, 2012). In the concept of text (or document and term) clustering, there are approaches using hierarchical clustering or text mining methods (Charu and Zhai, 2012). These methods are focused on organizing text data based on similarity or association measures. The approaches are applied in a similar way for document-, text- and term- clustering (Klahold *et al.*, 2014). So we used the term (word-) clustering to refer to such methods. In this context, hierarchical term clustering algorithms (techniques) such as single-link, complete-link, average-link, cliques, and stars (Li, 1990; Rajasekaran, 2005) are applied. The single-link or single-linkage clustering method detects and merges unlinked pair of points in two clusters with the largest similarity (Manning *et al.*, 2008), while complete-link clustering or complete-linkage clustering determines the similarity of the most dissimilar members of the clusters (Manning *et al.*, 2008). In average-link or average-linkage, the average value of all the pairwise links between points (for which each is in one of the two clusters) is a measure for computing the similarity (William and Baeza-Yates, 1992). The clique clustering groups the data into cliques, i.e. identifying subspaces of a high dimensional data space that allow better clustering than original space (Kochenberger *et al.*, 2005). In addition to the clustering methods which are discussed earlier, there are a number of related works using ontology-based framework for text clustering (Hotho *et al.*, 2002; Yang *et al.*, 2008; Tar and S, 2011; Ma *et al.*, 2012).

In string matching, a number of studies have been conducted in the area of both fuzzy and exact matching of patterns (Hussain *et al.*, 2013) in which they applied exact matching algorithm using two pointers (simultaneously) based on the window sliding method where they tried to compare bidirectional algorithm's results with Quick Search, BM Horspool, Boyer-Moore and Turbo BM algorithms that are deemed to be efficient for character comparisons and attempts to complete processing of selected text. They used a bidirectional matching algorithm that compares a given pattern from both sides, starting from right then from left, one character at a time within the text window and produced an algorithm that scans text string from both sides simultaneously against the given pattern. Its analysis shows that it takes O(mn/2) time where m is the length of the given pattern and n is the length of the target text.

Another study (TextKernel, 2016) uses the resumé text of the candidate's profile and automatically creates a search which is performed on multiple and multi-lingual sources of jobs. The search system collects and structures online jobs, matches them to a profile and helps find relevant job for a job seeker. In our study, the approach implements bidirectional document-document matching (i.e. job seeker information and job descriptions) through term similarity measures for text clustering.

Unlike its usage in (Hussain *et al.*, 2013) bidirectional matching, in this study, refers to matching features (i.e. terms or other values of attributes) in job description with job vacancies and vice versa (cf. Figure 2) to produce a unified search space for documents in the same cluster. This study is also different from TextKernel (2016) in that it tries to generate a common terms database to represent job descriptions and vacancies in the same cluster to maximize the likelihood of their appearance in the suggestion list. It does not use active vacancies and user profiles as in TextKernel (2016), rather it uses user profiles, active and historical vacancies, standard job descriptions and social network data. Its result, at this stage, is a representation of job seekers profiles and job vacancies which will be used as an input to job vacancy recommender in later stages.

In addition to the matching methods, a number of researches on whether vacancy data contributes to the study of labor and (un)employment have been conducted, all building upon a conclusion made decades ago on the measurement of vacancy data (Dunlop, 1966). The study related the concerns of vacancy data to theoretical, policy and operational interests. It holds to date *vis-à-vis* understanding the usefulness and thereby utilizing this data (Carnevale *et al.*, 2014). Dunlop (1966) concluded that the job vacancy concepts and data had small role in operations of business enterprises, governments and other labor employers by stating as follows. "Until these data are perfected and enter into internal organizational processes, the regular completion of questionnaires for outsiders will have limited meaning" (Dunlop, 1966). Despite its limited scope, online job vacancies have unique and rich content that can be exploited to provide useful information to policies in various fields (Kureková *et al.*, 2013).

Wagner (2011) studied how jobs are evolving over a period of time because of changes in the way organizations perform, by societal dynamics and other factors in the work environments. Moreover, Wagner (2011) also summarized how new skills help job seekers fitting into the changing nature of skills demanded in job market as: "retrofitting" – incorporating new skills to existing jobs, "blending" – combining skills sets from different existing jobs or industries to create new specialties, and "problem solving" – the changes in the nature of jobs creates the supply of new problems for people to solve. With vacancies data, that contains the current need of employers, we can get new skills and/or new jobs which have not yet been incorporated in standards. Vacancy data analysis not only helps fill the gap in the data of job descriptions available in occupational standards and job seekers' profiles but also helps identify emerging jobs. Studying the trend of job vacancies and the nature of their changes help device methods of helping job seekers adapt in the changes.

Kureková *et al.* (2014) ascertained the importance of data collected via web-based individual voluntary survey in representing the sample population's characteristics. Our bidirectional matching combines these with vacancy data in order to determine the best (i.e. optimally achieved) matching vacancy for a job seeker with a particular skill set.

Giunchiglia *et al.* (2007) studied the implementation of semantic matching using knowledge of structure of the content. In their study, they performed matching rules to match meanings through concept mapping.

While mostly focusing on the trend analysis and theoretical analysis of job vacancies, previous studies lacked the analysis of actual content in vacancy data to figure out the content-level analysis to predict the next trends. In this study, we focus on the trends in the content of vacancy data from the job detail and skill requirements point of view for Internet of Things (IoT) job vacancies advertised online (cf. Section 4). Unlike (Giunchiglia *et al.*, 2007), in our work no knowledge of the structure of data is assumed. Our semantic analysis is entirely based on symbolic computation of lexical parameters, i.e., terms as features and lexical statistics as feature weights.

3. Methodological framework of the recommender system

The bidirectional jobseeker-to-vacancy matching system being studied in this research is part of the framework shown in Figure 1 that addresses the semantic matching of vacancies with job seeker data. The user interface layer is the component that interacts with the user (i.e. jobseeker or recruitment professional or employer). The logic layer is composed of a number of algorithms to fetch vacancy and jobseeker information from the internet and other sources, analyze and model them. The logic layer not only collects, integrates and models the data, but it also mediates the interaction between the data layer and the user interface layer. The data layer on the other hand stores the data and the model for use by the logic layer during interaction with the user.

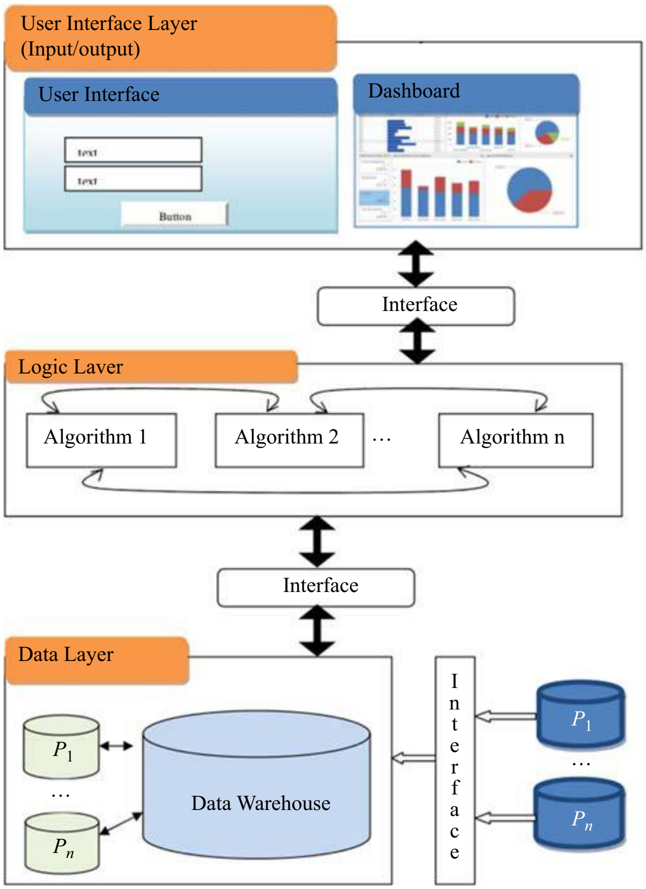

Figure 1.
Conceptual framework of the proposed system

The data set used in this research is online crawled vacancy from various sources. The data obtained from online sources is divided into a training set and test set. Training set is part of the data set that will be used to train the algorithm that indicates the desired result (positive examples) as well as counter examples that show undesired results (negative examples) from which the algorithm learns to discover new rules. Test set, on the other hand, is used to measure the accuracy and effectiveness of the algorithm whether the rules have been learned and to what extent. If the evaluation results are satisfactory, the learned rules can then be used to predict results for new and previously unseen cases. Both data sets are cleaned of formatting text such as html tags, special characters and punctuation marks before reducing its size into a set of statistically significant terms for representation.

In order to improve the quality of matching between job descriptions and job vacancies, a mediating system is required that connects and supports job designers and employers, respectively. After carefully reviewing problems associated with mismatch between job descriptions and vacancies, we studied the application of bidirectional matching with the objective of improving job vacancies and job seeker profiles using social networking data analysis (cf. Section 3.1) and matching of occupational standards with job vacancies (cf. Section 3.2). The system provides suggestions to improve both job descriptions and vacancies using a combination of text mining methods (i.e. representation, clustering and matching).

3.1 Social network analysis in job seeker modeling

Social networking platforms (also called social networking sites) are being utilized by employers to search for their prospective employees. These websites are good sources not only of employee profiles, but also of their connections together with the witness of their connections that contains useful information about job seekers (i.e. prospective employees). This information (i.e. job seekers profiles and connection witnesses) augment the process of assessing whether or not a certain job seeker is suitable match for a vacancy. According to Ben-Karter (2016), which conducted a survey on the extent to which social networking sites are used for recruitment, 45 percent of companies use Twitter®, 80 percent use LinkedIn® and 50 percent use Facebook® to find talent.

Another important source of information to measure the level of expertise of a job seeker for a given skill is to perform self-evaluation where the job seeker's own discretion is used to assign value of skill knowledge in a given scale. Although there are a couple of systems that provide a system for skill self-evaluation such as (ItsYourSkills, 2016), they have not incorporated other means for cross-checking the validity of the self-assessment. The importance of social networking-generated measure of skill of an individual in this regard is to check individual's skill assessment against endorsements of other peers in his/her connection.

For example (ItsYourSkills, 2016), has extensive tree-based profile building system. While this system has an appealing user interface, it fails in a number of ways to address the challenges of user skill profile modeling. First, it suffers from the inherent limitation of close-ended user interfaces – limited options and lack of flexibility (Chala *et al.*, 2016b). Second, its extensive tree-based user interface is only suitable for those who are adept to using websites. A user will be sent to a depth of branching trees that may or may not lead to the skill he/she is going to enter. Third, the system fails to tap the skill profiles of users from social networking sites though social networks are increasingly being used in recruitment and selection processes (Ben-Karter, 2016).

This study is focused on extracting useful knowledge that emanates from relationships between users of social networking sites – the knowledge of each other's skills, expertise, experiences and attitude – and integrating it into recruitment systems *per se* in order to improve and evaluating the match between job seekers and job vacancies.

Social networking data used in this study is obtained from professional interactions in StackExchange.org (2017) where users provide description of their skills and qualifications.

As shown in Figure 2, one row represents one user. In addition the data provided by the uses themselves, users garner other information out of their interaction such as asking or answering questions, and commenting. For example, extracting information about the user in row with id 3187 (cf. Figure 2), we get the user described with skills "software engineer" and "python" which when combining with resumé data better describe the job seeker.

3.2 Bidirectional matching

The way human beings match resumes to job descriptions has been summarized into a series of steps including review of job descriptions, summarization and entry of qualifications for the descriptions' requirements (Jones, 2015).

The automated system, on the other hand, extracts text from multiple documents from various sources and splits the text into words to prepare for the matching process. This method performs a variety of different operations and text analyses related to extraction and matching of several files based on document similarities using indexes. The system is not only helpful to extract and match the specified text from multiple job descriptions but also filters out documents which do not contain a specified text. After the process of matching, the system produces the resulting matched data and stores it into the database for searching.

Once the documents are represented by indexes (i.e. words or terms), we experiment to evaluate *n*-gram string matching using the indexes from the two sets of documents (i.e. job descriptions and job vacancies) to find out the optimal *n*-gram (Jurafsky and Martin, 2009). For example, we use keys from indexes extracted from job descriptions provided by job holders to search for patterns in job vacancies provided by employers, and vice versa. Then we compare the results if they have reasonable matching. The choice of bidirectional matching is because it has been reported to perform well in pattern matching (Chatterjee and Perrizo, 2009).

Moreover, it has a space complexity of O(mn/2), where m and n are the number of characters in the search space (Kucherov *et al.*, 2014). The literature review by Hussain *et al.* (2013) shows that the complexity of bidirectional matching is better than that of all other pattern matching algorithms for text processing.

For instance, considering a job description for a system administrator, one can find a number of vacancies with different job requirements. Let us take part of the job description $D \in \{D_1, D_2, \ldots, D_n\}$ with content "[…] ability to configure systems and networks, manage users, give technical support to users […]" and two example vacancies V_1 and $V_2 \in$


```
and web stuff.</p> " Views="0" UpVotes="5" DownVotes="0" AccountId="2666696"/>
<row Id="3131" Reputation="168" CreationDate="2013-11-04T14:09:25.000" DisplayName="Reflection" LastAccessDate="2013-11-04T16:24:47.950"
Views="0" UpVotes="0" DownVotes="0" ProfileImageUrl="https://i.stack.imgur.com/DBSKd.jpg" AccountId="3301228"/>
<row Id="3150" Reputation="228" CreationDate="2013-11-06T00:53:11.000" DisplayName="LucentMoon"
LastAccessDate="2014-08-26T22:46:12.920" WebsiteUrl="" AboutMe="" Views="0" UpVotes="0" DownVotes="0"
ProfileImageUrl="https://www.gravatar.com/avatar/?s=128&d=identicon&r=PG&f=1" AccountId="3541627"/>
<row Id="3159" Reputation="716" CreationDate="2013-11-06T17:51:54.000" DisplayName="user3159" LastAccessDate="2013-12-03T23:22:41.333"
WebsiteUrl="http://www.wessells.me" Location="Salt Lake City, UT" AboutMe="" Views="0" UpVotes="0" DownVotes="0"
ProfileImageUrl="https://i.stack.imgur.com/FEl2H.jpg" Age="41" AccountId="3545630"/>
<row Id="3160" Reputation="101" CreationDate="2013-11-06T20:51:47.000" DisplayName="new bie" LastAccessDate="2014-07-29T13:38:24.083"
AboutMe="<p>:)</p> " Views="0" UpVotes="0" DownVotes="0" AccountId="362454"/>
<row Id="3163" Reputation="332" CreationDate="2013-11-07T13:34:29.000" DisplayName="user2964971"
LastAccessDate="2014-10-18T11:05:28.253" Views="9" UpVotes="2" DownVotes="0" ProfileImageUrl="https://i.stack.imgur.com/OL6dj.png"
AccountId="3549835"/>
<row Id="3187" Reputation="208" CreationDate="2013-11-11T03:41:21.000" DisplayName="WilHall" LastAccessDate="2013-11-16T20:23:35.290"
WebsiteUrl="http://wilhall.com/" Location="Cambridge, MA" AboutMe="<p>I'm a software engineer. My favorite language is Python.</p> " Views="0"
UpVotes="0" DownVotes="0" Age="24" AccountId="287034"/>
<row Id="3191" Reputation="3858" CreationDate="2013-11-11T20:09:21.000" DisplayName="Ryan Foley"
LastAccessDate="2016-06-09T23:43:41.727" WebsiteUrl="" Location="Colorado Springs, CO" AboutMe="" Views="30" UpVotes="131"
DownVotes="14" ProfileImageUrl="https://www.gravatar.com/avatar/76dcdb2091e5e5d38bc5e331fb8945d3?s=128&d=identicon&r=PG"
AccountId="1825721"/>
<row Id="3198" Reputation="106" CreationDate="2013-11-12T15:41:22.000" DisplayName="Paul" LastAccessDate="2015-12-01T19:33:05.123"
Views="0" UpVotes="0" DownVotes="0" ProfileImageUrl="https://www.gravatar.com/avatar/6e6699562f74433dfef54932f6368bb7?s=128&
d=identicon&r=PG" AccountId="21462"/>
<row Id="3200" Reputation="167" CreationDate="2013-11-12T18:47:10.000" DisplayName="Birdman" LastAccessDate="2014-07-03T20:54:42.920"
Location="Copenhagen, Denmark" Views="0" UpVotes="0" DownVotes="0" ProfileImageUrl="https://i.stack.imgur.com/vul9z.jpg?s=128&g=1"
Age="26" AccountId="966751"/>
<row Id="3210" Reputation="856" CreationDate="2013-11-13T15:18:27.000" DisplayName="John Kennedy"
LastAccessDate="2016-02-11T01:00:17.547" WebsiteUrl="http://www.linkedin.com/pub/john-kennedy/59/724/a97" Location="Los Angeles, CA"
AboutMe="<p>Sr. Network Engineer with experience working on everything from highly secure DoD sattelite networks to large multicast and IPTV
deployments. Currently pursuing CCIE R&amp;S, VCP, NCDA</p> <p>CCNP R&amp;S CCNA Security</p> <p>I'm also a stranger on the internet, so
```


Figure 2.
Raw data from professional social network

$\{V_1, V_2, \ldots, V_m\}$ with content "[…] experience in troubleshooting systems over wide area network and proven knowledge of working on virtual private networks […]" and "[…] skill in corporate email administration and hardware maintenance […]," respectively. It is apparent that some of the key requirements of the vacancies such as troubleshooting in V_1 and maintenance in V_2 are missing in D. Thus, the system, on the one side, provides suggestion for the job designers to enrich D by incorporating troubleshooting from V_1 and maintenance from V_2. On the other side, the system provides recommendations to enrich V_1 and V_2 by incorporating the terms "configure" and "network" from D, respectively (cf. Figures 3 and 4 and cf. Algorithm 1).

Algorithm 1 Algorithm for our Bidirectional Framework.

Require: D and V are two vectors of strings of length m and n

procedure BIDIRECTIONALMATCHING(< D1,…, Dm > , < V1, …, Vn >)

Cluster(< $D_1,\ldots,D_m$ > , < $V_1,\ldots,V_n$) ← Similar(Di;Vj) = > Cluster with similarity

for c ∈ Clusters do = > for each cluster

for $D_i \in D_1,\ldots, D_m$ do = > for each job description

for $t_d \in D_i$ do = > for each keyword in the description

for $V_j \in V_1,\ldots, V_n$ do for each vacancy

if ($t_d \notin V_j$) then append(V_j, t_d) = > append if the term does not exist in vacancy

for $V_j \in V_1,\ldots,V_n$ do = > for each vacancy

for $t_v \in V_j$ do = > for each term in vacancy

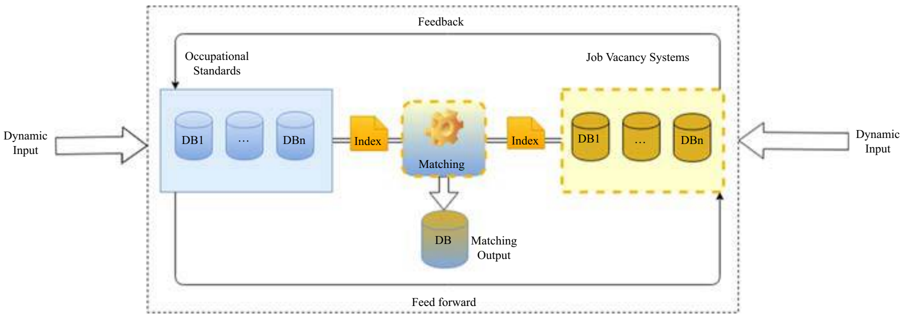

Figure 3.
Bidirectional matching to enrich content of vacancies and job descriptions

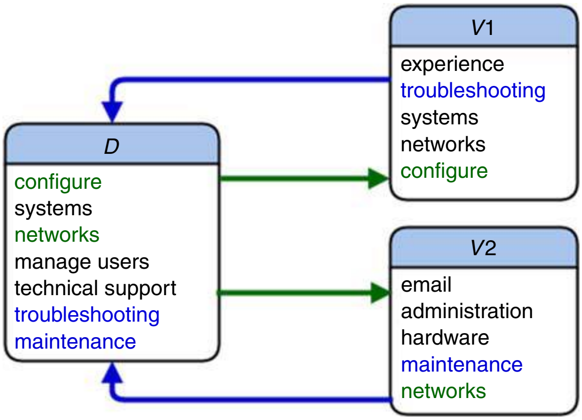

Figure 4.
Bidirectional matching: an example

10: for $D_i \in D_1, \ldots, D_n$ do for each job description
11: if $(t_v \notin D_i)$ then append$(D_i, t_v) = >$ append if the term does not exist in job description
12: return $< D >, < V >$

The algorithm shown in Algorithm 1, clusters similar documents into subgroups (i.e. hierarchies of occupational titles) in order to reduce the search space, before it performs computation of matching and then it returns list of matching jobs. For clustering we used k-means clustering algorithm on the ground that its effectiveness is favored by a number of literatures (e.g. Fahad *et al.*, 2014; Lin *et al.*, 2014).

4. Use case in Industry 4.0 jobs

In this section the applicability of the proposed system is examined in the use-case of job transitions in Industry 4.0. The core idea is to identify the potentials for utilizing bidirectional matching for detection of new trends in the job market through temporal analyzing of job vacancies, and ultimately to provide input for job designers toward improving the quality of job description and keeping them in accord with latest changes. One of the progressive initiatives which imposes changes in the job market, especially in industrialized countries, is Industry 4.0[2]. Industry 4.0 refers to the current trend in production systems which aim to enhance the automatization and the data exchange degree to the depth which enables and accelerates self-organization functions in production and associated business processes (Industrie 4.0 Working Group, 2013; Bartodziej, 2016).

Aiming at the highest level of autonomy in cyber-physical production systems (CPPS) through employing artificial intelligence (AI) solutions, Industry 4.0 triggers a comprehensive transformation in human job description, job roles as well as conditions and requirements of human learning and didactical conceptions (Ansari and Seidenberg, 2016). The transformation toward Industry 4.0 affects the job market with regard to the transition and evolution of existing jobs as well as creating new ones. The German Institute for Employment Research (IAB) estimated that while Industry 4.0 can "yield an improvement in the economic development, […] in ten years there will be 60,000 fewer jobs than in the baseline scenario. At the same time, 490,000 jobs will be lost, particularly in the manufacturing sector, and approximately 430,000 new ones will be created" (Wolter *et al.*, 2015, p. 8). They concluded that "to a great extent, jobs "switch" between sectors, occupations and qualifications" (Wolter *et al.*, 2015, p. 8). In particular, the emergence of new digital technologies, intelligent systems and organizations, such as CPPS, IoT, internet of services (IoS), augmented reality (AR) as well as smart materials, smart products and smart factory, may result in raising job market demands to hire job seekers who may fulfill new tasks and comply with job requirements, i.e. applicants who hold Industry 4.0s skills, knowledge and abilities in association, e.g. with AI, software and hardware cyber-physical systems.

The target question is how quickly job market and job seekers could cope with the changes imposed by Industry 4.0. In this context, we placed the focus on information technology (IT) sector, examine and compare the IoT jobs trends in job posting relative to other job titles as well as trends in job seeker's interests in IoT jobs. To put in a nutshell, IoT refers to "a networked world of connected devices, objects, and people" (Greengard, 2015), colloquially known as industrial internet of things which represents the industrial concept of IoT. The analysis gains benefits from the proposed bidirectional matching approach to determine the trends of IoT, especially in the past two years. The bidirectional analysis has been applied to online crawled job postings and job seeker's interests of using sample data of 65 job seekers and 300 vacancies collected in the period between the beginning of August 2016 and end of November 2016. Figure 5 and Table I show example raw and extracted data, respectively. Figure 6 reveals the growing trends of IoT job posting in the recent years (i.e. in the interval of 2012–2016). This is an indication that IoT jobs are growing and need attention in matching them with job seekers.

```
jk:'28c26d6c9ed398eb' efccid: '8b1a46d7dc7251dd',
srcid:'315799370c9ac95f', cmpid:'c4f5e238ec297fa5', num:'7',
srcname:'CDK Global', cmp:'CDK Global', cmpesc:'CDK Global',
cmplnk:'/q-CDK-Global-jobs.html', loc:'Orange County, CA',
country:'US', zip:'', city:'', title:'Full Stack SR Software Engineer – 6623',
locid:'544e90d8616c87c1',
rd:'eCVGrSY0V901x7uzs1I1iws75F1iBVRKu_x_5sFFEUI'
...
<span class="location" itemprop="address" itemscope
itemtype="http://schema.org/Postaladdress"><span
itemprop="addressLocality">Orange County,
CA</span></span></span><table cellpadding=0 cellspacing=0
border=0><tr> <td class=snip><div><span class=summary
itemprop="description">Expert level design capabilities in multi-tiered,
database driven, <b>internet</b> based applications. You have family,
friends <b>things</b> ...</span>
```


Figure 5.
Sample raw data of vacancy

Table I.
Sample vacancy data for selected features

Job ID	Job title	Job description	Company	Job location	Country	Job post date
28c26d6c9ed398eb	Full Stack SR Software Engineer – 6623	Expert level design capabilities in multi-tiered, database driven, < b > internet < /b > based applications. You have family, friends, sporting events, and lots of < b > things < /b >	CDK Global	Orange County, CA	USA	September 27, 2016

At the same time, Figure 7 depicts the job seekers' interests for IoT jobs which have been drastically increasing, i.e. almost twice, over the past two years (2015–2016). Comparing the findings on job postings and job seekers' interests, one may identify the unmet demand of IoT jobs for job seekers.

As shown in Figure 8, 4.07 postings per job seeker have been detected by September 25, 2016, which reveals the inappropriate adaptation and reaction of the job market to the job seekers tendency, which may be caused on the job market side due to overlooking the job seekers' interests, and on the job seekers side due to lack of attention to the incremental rate of changes in the industrial sectors. To assure proactive reaction and adjustment to the trends, in both cases, the job market and job seekers should be appropriately and timely notified with regard to the trends of changes in job postings and/or job seekers' interests, respectively. Such kind of trend analysis using bidirectional matching, as a semantic-aided occupational analysis method, may support time-based and content-based analysis of web-data. It coherently facilitates adopting job market and job seekers strategies for posting up-to-date vacancies, and for coping with speed of market changes with respect to the technological evolution, respectively.

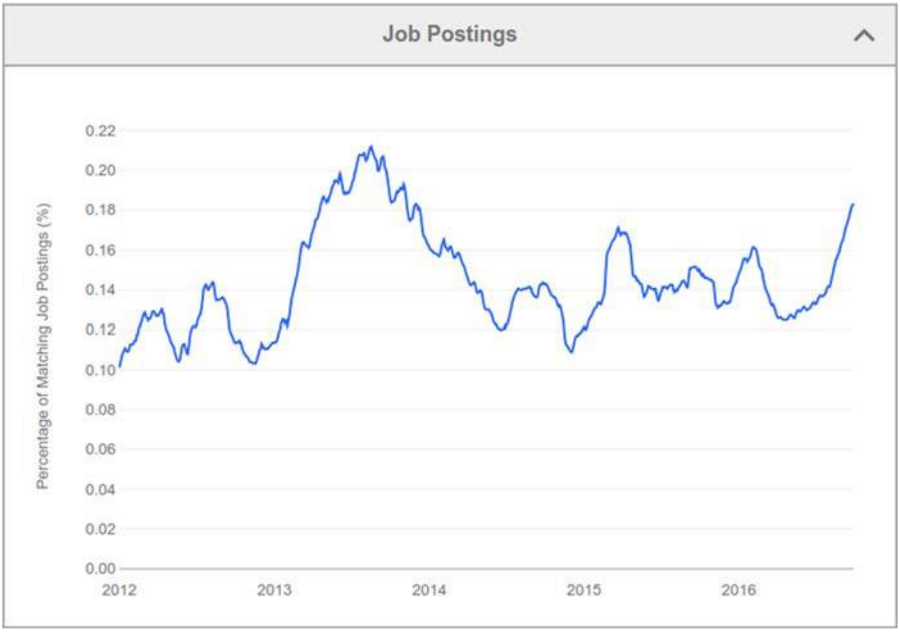

Figure 6.
Growing Trends of IoT Job posting in recent years

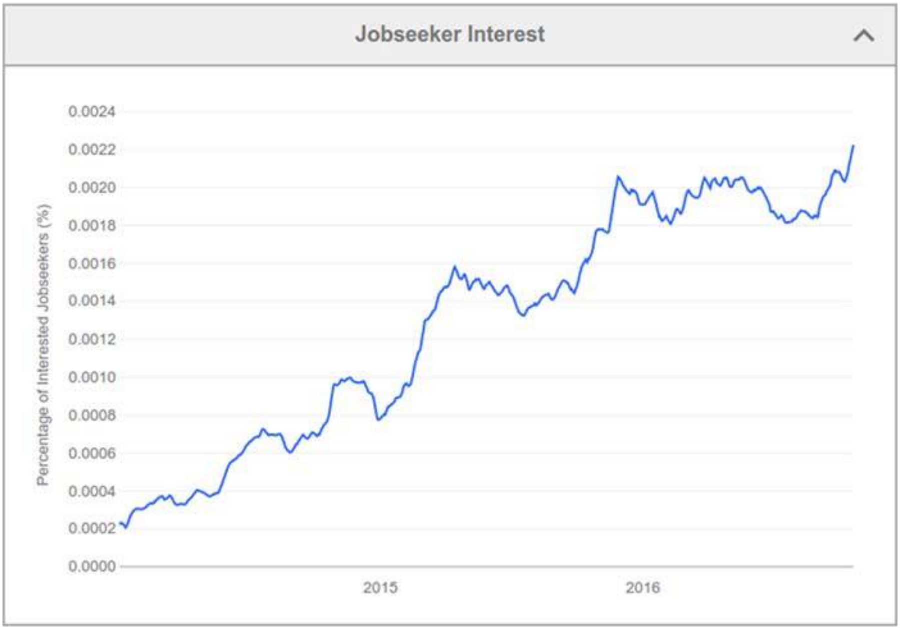

Figure 7.
Increasing job seeker interest for IoT jobs in recent years

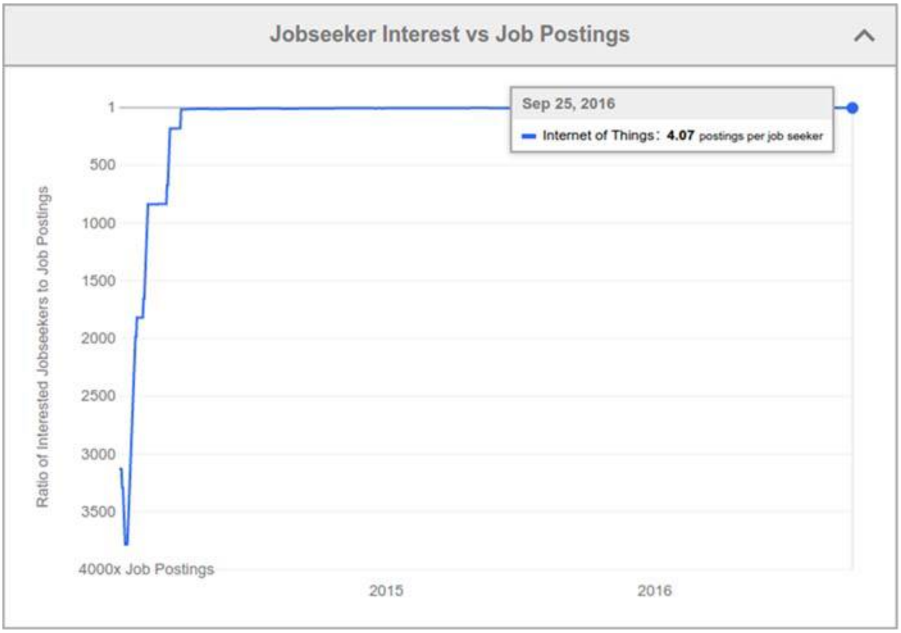

Figure 8. Unmet demand of IoT jobs for job seeker

5. Evaluation

In order to evaluate the proposed bidirectional matching method, we utilized job seeker profiles bolstered by data from social networking sites and vacancy data in the area of IoT. Data of individuals from professional network that contains professional details of persons (i.e. attributes) and information produced due to interaction of these individuals on topics (i.e. relations) is collected from online open access sources. After cleaning noisy data and eliminating individuals that have no data in either attributes, relations or both, we obtained 65 records of persons with data about themselves and professional connections as well as 110 vacancies. From 65 job seekers with relevant skills for the jobs in 110 vacancy positions, to test the system, evaluation is done using accuracy measure as shown in Equation (1) because it is an evaluation metric that gives a single measure of quality (Manning *et al.*, 2008). The accuracy measure which is the ratio of correct matches to total number of records evaluates the effectiveness, such that CM is the number of records that are actually correct matches to the job under consideration, CNM is the number of individuals that have relevant skills but not matched, NCM is the number of individuals that do not have relevant skills but matched and NCNM is the number of individuals that do not have relevant skills and are not matched. Then:

$$\text{Accuracy} = \frac{\text{NM}+\text{NCNM}}{\text{CM}+\text{CNM}+\text{NCM}+\text{NCNM}}. \tag{1}$$

For example, for one vacancy, the result shown in the confusion matrix in Table II, is when matching job seeker with vacancy in the absence of social networking data. Whereas, the result shown in Table III is when matching job seeker with the same job vacancy using social networking data included.

The experiment clearly shows promising result in achieving accuracy improvement that the number of incorrect matches are reduced (from 4 to 3) while the number of correct matches are increased (from 7 to 10) from 47.27 to 50.91 percent as shown in Tables II and III for the example job vacancy.

Thus, inclusion of social network data in skill measurement for bidirectional matching shows consistent improvement positive difference for 65 job seekers, as shown in Figure 9, improving overall matching accuracy by 59.1 percent.

6. Conclusion and outlook

The paper discusses the general framework of online occupational recommender system that utilizes bidirectional matching, occupational analysis method, to improve the accuracy of job matching to job seekers and enrich the quality of recommendations to job seekers and recruiters as well as job designers. It also explored the importance of occupational standards and social networking data in the jobseeker-to-vacancy matching in online recruitment systems and presents an Industry 4.0 use case scenario to showcase the trends in job demands and jobseeker interests with promising insights.

Predicted → Actual ↓	Matched	Not matched
Correct	7	9
Not correct	4	45

Table II. Preliminary Result (without social network)

Predicted → Actual ↓	Matched	Not matched
Correct	10	6
Not correct	3	46

Table III. Preliminary result (with social network)

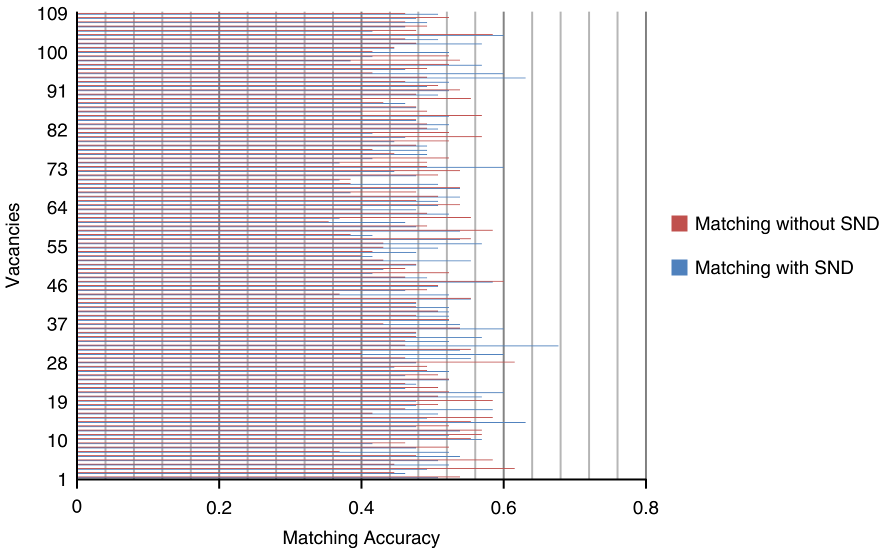

Figure 9. Matching accuracy with and without social network data (SND)

There are a number of future works in this study. First, implementation of a full-fledged bidirectional matching software prototype should be implemented and deployed so as to measure its effectiveness and also to experiment on the varied use cases to measure its actual usability in the real world. Considering "language variation" across Europe as a challenge of occupational (Big) data analysis, another aspect of the future research is to analyze multi-lingual job descriptions and job vacancies and applying it at micro level (i.e. individual level) to support job seekers to improve the quality of their resume for a particular job posting so as to include required and preferred skills.

Notes

1. In this article, occupational analyst and job designer are used interchangeably.
2. It is also known as fourth industrial revolution.

References

Ansari, F. and Seidenberg, U. (2016), "A portfolio for optimal collaboration of human and cyber physical production systems in problem-solving", *Proceedings of the 13th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2016), International Association for Development of the Information Society*, *Mannheim*, *October 28-30*, pp. 311-315.

Bartodziej, C.J. (2016), *The Concept Industry 4.0: An Empirical Analysis of Technologies and Applications in Production Logistics*, Springer.

Belloni, M., Brugiavini, A., Meschi, E. and Tijdens, K. (2016), "Measurement error in occupational coding: an analysis on SHARE data", *Journal of Official Statistics*, Vol. 32 No. 4, pp. 917-945.

Ben-Karter (2016), "What are the most successful job search strategies for 2015-2016?", Bring IT to Life, available at: http://ben-karter.tumblr.com/post/123450223230/what-are-the-most-successful-job-search-strategies (accessed August 8, 2016).

Biemann, C. (2012), *Structure Discovery in Natural Language*, Springer.

Carnevale, A.P., Jayasundera, T. and Repnikov, D. (2014), "Understanding online job ads data: a technical report", Center on Education and the Workforce, Georgetown University, available at: http://cew.georgetown.edu/onlinejobmarket (accessed February 13, 2017).

Chala, S.A., Ansari, F. and Fathi, M. (2016a), "A framework for enriching job vacancies and job descriptions through bidirectional matching", *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, *Rome*, *April 23-25*, pp. 219-226.

Chala, S.A., Ansari, F. and Fathi, M. (2016b), "Towards implementing context-aware dynamic text field for web-based data collection", *International Journal of Human Factors and Ergonomics*, Vol. 4 No. 2, pp. 93-111.

Charu, C.A. and Zhai, C.X. (2012), *Mining Text Data*, Springer.

Chatterjee, A. and Perrizo, W. (2009), "Bidirectional string matching algorithm in text mining", *IADIS Information Systems Conference*, available at: www:cs:ndsu:nodak:edu/~perrizo/saturday/papers/sede09/sede09_arijit1_bidir_string_match.pdf (accessed February 13, 2017).

Christen, P. (2012), *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*, Springer Science & Business Media.

Dunlop, J.T. (1966), "Job vacancy measures and economic analysis", National Bureau of Economic Research, pp. 27-47, available at: www.nber.org/books/unkn66-2 (accessed February 13, 2017).

EC (2015), "ESCO home European commission", available at: https://ec:europa:eu/esco/home (accessed February 13, 2017).

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., Foufou, S. and Bouras, A. (2014), "A survey of clustering algorithms for big data: taxonomy and empirical analysis", *IEEE Transactions on Emerging Topics in Computing*, Vol. 2 No. 3, pp. 267-279.

Fasulo, D. (1999), "An analysis of recent works on clustering algorithms".

Fortunato, S. (2010), "Community detection in graphs", *Journal of Physics Reports*, Vol. 486, pp. 75-174, doi:10:1016/j:physrep:2009:11:002.

Gan, G., Ma, C. and Wu, J. (2007), "Data clustering: theory, algorithms, and applications", ASA-SIAM Series on Statistics and Applied Probability.

Giunchiglia, F., Yatskevich, M. and Shvaiko, P. (2007), "Semantic matching: algorithms and implementation", *Journal on Data Semantics*, Vol. IX, pp. 1-38.

Godliman (2009), "How to manage headhunters for candidates", available at: http://godlimanpartners: com/interface/resources/How_To_Manage_Headhunters_for_Candidates

Greengard, S. (2015), *The Internet of Things*, MIT press.

Hernandez, J.H. (2015), "Ways to make your resume perfect for a job opening", available at: www: careerealism:com/resume-perfectmatch-job-opening/ (accessed February 13, 2017).

Hotho, A., Maedche, A. and Staab, S. (2002), "Text clustering based on good aggregations", *Künstliche Intelligenz*, Vol. 16 No. 4, pp. 48-54.

Hussain, I., Hassan Kazmi, S.Z., Ali Khan, I. and Mehmood, R. (2013), "Improved bidirectional exact pattern matching", *International Journal of Scientific and Engineering Research*, Vol. 4 No. 5, pp. 659-663.

ILO (2012), *International Standard Classification of Occupations ISCO-08. Volume 1 Structure, Group Definitions and Correspondence Tables*, International Labour Organization, Geneva, available at: www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf (accessed August 8, 2016).

Industrie 4.0 Working Group (2013), "Final report of the industrie 4.0 working group: recommendations for implementing the strategic initiative industrie 4.0", acatech – National Academy of Science and Engineering, Sponsored by German Federal Ministry of Education and Research.

ItsYourSkills (2016), "Sample individual skills profile", Individual Skills Profile Sample (Online), available at: www.itsyourskills.com/sample/individual-skills-profile (accessed August 8, 2016).

Jones, L. (2015), "How to match qualifications to a job description in a resume", available at: http:// work:chron:com/match-qualificationsjob-description-resume-8135:html (accessed February 13, 2017).

Jurafsky, D. and Martin, M. (2009), *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, 2nd ed., Prentice-Hall.

Klahold, A., Uhr, P., Ansari, F. and Fathi, M. (2014), "Using word association to detect multitopic structures in text documents", *IEEE Intelligent Systems*, Vol. 29 No. 5, pp. 40-46.

Kochenberger, G., Glover, F., Alidaee, B. and Wang, H. (2005), "Clustering of microarray data via clique partitioning", *Journal of Combinatorial Optimization*, Vol. 10 No. 1, pp. 77-92.

Kucherov, G., Salikhov, K. and Tsur, D. (2014), "Approximate string matching using a bidirectional index", *Lecture Notes in Computer Science*, Vol. 8486, pp. 222-231.

Kureková, L., Beblavý, M. and Thum, A.-E. (2013), "Online job vacancy data as a source for micro-level analysis of employers' preferences: a methodological enquiry", paper prepared for the First International Conference on Public Policy (ICPP), Grenoble, June 26-28.

Kureková, L.M., Beblavý, M. and Thum, A.-E. (2014), "Using internet data to analyse the labour market: a methodological inquiry", Discussion Paper No. 8555.

Li, X. (1990), "Parallel algorithms for hierarchical clustering and clustering validity", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 12, pp. 1088-1092.

Lin, Y.S., Jiang, J.Y. and Lee, S.J. (2014), "A similarity measure for text classification and clustering", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26 No. 7, pp. 1575-1590.

Ma, J., Xu, W., Sun, Y., Turban, E., Wang, S. and Liu, O. (2012), "An ontology-based text-mining method to cluster proposals for research project selection", *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, Vol. 42 No. 3, pp. 784-790.

Manning, C.D., Raghavan, P. and Schütze, H. (2008), *Introduction to Information Retrieval*, Cambridge University Press.

Muderedzwa, M. and Nyakwende, E. (2010), "A framework for improving the effectiveness of it in employment screening", *Research and Development (SCOReD), 2010 IEEE Student Conference IEEE*, available at: http://ieeexplore:ieee:org/xpls/absnall:jsp?arnumber=5703988 (accessed February 13, 2017).

Rafi, M. and Shaikh, M.S. (2013), "An improved semantic similarity measure for document clustering based on topic maps", Computing Research Repository, available at: http://arxiv:org/abs/1303:4087 (accessed February 13, 2017).

Rajasekaran, S. (2005), "Efficient parallel hierarchical clustering algorithms", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 16 No. 6, pp. 497-502.

Sacchetti, L. (2013), "The magic of headhunting: a how-to guide to hunting and closing top candidates", available at: http://ren-network:com/wp-content/uploads/2013/05/The-Magic-of-Headhunting-A-How-to-Guide-to-Hunting-and-Closing-Top-Candidates:pdf (accessed February 13, 2017).

Schaeffer, S.E. (2007), "Survey: graph clustering", *Journal of Computer Science Review*, Vol. 1 No. 1, pp. 27-64, doi:10:1016/j:cosrev:2007:05:001.

StackExchange.org (2017), "Stackexchange data dump", available at: https://archive.org/details/stackexchange (accessed December 19, 2017).

Tar, H.H. and S., N.T.T. (2011), "Ontology-based concept weighting for text documents", *International Conference on Information Communication and Management*.

Textkernel (2016), Online, available at: www.textkernel.com/hr-software-modules/bidirectional-matching/ (accessed March 8, 2016).

WageIndicator (2015), "Salary checks – worldwide wage comparison", available at: www:wageindicator:org (accessed February 13, 2017).

Wagner, C.G. (2011), *Emerging Careers and How to Create Them: 70 Jobs for 2030*, World Future Society.

William, B.F. and Baeza-Yates, R. (1992), *Information Retrieval: Data Structures and Algorithms*, Prentice Hall.

Wolter, M.I., Mönnig, A., Hummel, M., Schneemann, C., Weber, E., Zika, G., Helmrich, R., Maier, T. and Neuber-Pohl, C. (2015), "Industry 4.0 and the consequences for labour market and economy", Scenario calculations in line with the BIBB-IAB qualifications and occupational field projections, Technical report, Institute for Employment Research (IAB).

Yang, X., Guo, D., Cao, X. and Zhou, J. (2008), "Research on ontology-based text clustering", Third International Workshop on Semantic Media Adaptation and Personalization, pp 14-146.

Further reading

Branham, L. (2012), "The 7 hidden reasons employees leave: how to recognize the subtle signs and act before it's too late", AMACOM, American Management Association.

San Jose State University (2013), "Emerging career trends for information professionals: a snapshot of job titles in summer", San Jose State University, San Jose, CA, available at: http://isschool.sjsu.edu/downloads/emerging_trends_2012.pdf (accessed February 13, 2017).

About the authors

Sisay Adugna Chala is Marie Curie Research Fellow in the Institute of Knowledge Based Systems and Knowledge Management (KBS & KM), University of Siegen. He holds MSc and BSc from Addis Ababa University, Ethiopia. Before joining the University of Siegen, he was Lecturer, ICT Director and System Administrator in Haramaya University, Ethiopia. He also worked on research on Statistical Machine Translation at German Research Center for Artificial Intelligence (DFKI) in Saarland University. His research interests include machine translation, knowledge management, data mining and machine learning, information retrieval and application of ICT in various problem domains. Sisay Adugna Chala is the corresponding author and can be contacted at: sisay.chala@uni-siegen.de

Fazel Ansari obtained PhD in Computer Science from the University of Siegen, Germany, in the specialization of knowledge management and knowledge-based systems. From 2013 to 2016 he was part of the Supervisory Board of the FP7 Marie Curie Initial Training Network (ITN) Eduworks. Fazel is currently working as Assistant Professor at the Institute of Management Science, Vienna University of Technology, Austria. The focus of his interdisciplinary research activities and his publications is placed on Industry 4.0, knowledge-based maintenance and human-centered cyber physical production systems.

Madjid Fathi is Professor and Chair of Institute of Knowledge-Based Systems and Knowledge Management, Department of Electrical Engineering and Computer Science, University of Siegen, Germany. Professor Fathi received his MSc degree in Informatics and PhD in Mechanical Engineering, both from the University of Dortmund. He obtained his Habilitation degree in Informatics and Automation from the University of Ilmenau. Professor Fathi has published several journal articles, book chapters, peer-reviewed conference papers, textbooks and edited books in the aforementioned areas of interest. His research interests include knowledge management applications in medicine and engineering, computational intelligence and knowledge discovery from text (KDT).

Kea Tijdens is Research Coordinator at AIAS/University of Amsterdam and Professor of Women's Work at the Department of Sociology, Erasmus University Rotterdam. She studied Sociology and Psychology at Groningen University, and obtained her PhD in Sociology in 1989. She is the scientific coordinator WageIndicator and a member of Webdatanet, INGRID and EDUWORKS. For WageIndicator she developed the web survey on work and wages, and performed the systems design for global databases on minimum wages, labor law and collective agreements. Her publications focus on women's work, wage setting processes, measurement of occupations and bias related to volunteer web surveys.