Social Network Analysis and Mining (2019) 9:43
https://doi.org/10.1007/s13278-019-0583-9

ORIGINAL ARTICLE

Building a soft skill taxonomy from job openings

Imane Khaouja[1,2] · Ghita Mezzour[1] · Kathleen M. Carley[3] · Ismail Kassou[2]

Received: 26 October 2018 / Revised: 13 July 2019 / Accepted: 17 July 2019 / Published online: 7 August 2019

Abstract
Soft skills are crucial for candidates in the job market, and analyzing these skills listed in job ads can help in identifying the most important soft skills required by recruiters. This analysis can beneft from building a taxonomy to extract soft skills. However, most prior work is primarily focused on building hard skill taxonomies. Unfortunately, methodologies for building hard skill taxonomies do not work well for soft skills, due to the wide variety of terminologies used to list soft skills in job ads. Moreover, prior work has mainly focused on extracting soft skills from job ads using a simple keyword search, which can fail to detect the diferent forms in which soft skills are listed in job ads. In this paper, we develop TaxoSoft, a methodol
ogy for building a soft skill taxonomy that uses DBpedia and Word2Vec in order to fnd terms related to diferent soft skills. TaxoSoft also uses social network analysis to build a hierarchy of terms. We use this method to build soft skill taxonomies in both English and French. We evaluate TaxoSoft on a sample of job ads and fnd that it achieves an *F*-score of 0.84, while taxonomies developed in prior work achieve an *F*-score of only 0.54. We then use the proposed methodology to analyze soft skills listed in job ads in order to fnd the skills most required in the American and Moroccan job markets. Our fndings can ofer insights to universities about the top soft skills requested in the job market.

1 Introduction

The term 'soft skills' refers to a broad set of skills, behav iors, attitudes and personal qualities that allow people to adapt efectively to their environment, to work well with others, to perform well, and to achieve their goals. These skills are broadly applicable, and complement 'hard' or aca demic skills. Soft skills are of paramount importance for the development of human capital and workforce success, and most employers seek candidates with specifc soft skills in order to guarantee the success of projects and to ensure that the organization will thrive. A growing evidence base shows that the value of soft skills rivals that of hard skills in terms of predicting employment and earnings, among other outcomes (Balcar 2014; Kautz et al. 2014). Unfortunately, a shortage of such skills has been noted by many employ ers around the world, who report that candidates lack the soft skills needed to fll the available positions (Hurrell 2016; Manpower 2017). Many universities, job seekers and employability programs are now starting to place more emphasis on soft skills (Cornali 2018; Hillmer et al. 2007; Maitra and Gopalram 2016; Smith et al. 2015), and candi
dates can beneft from an understanding of which soft skills are sought by employers. Employers specify the attributes needed by potential candidates in job advertisements, and an analysis of the soft skills listed in job ads can help toward an understanding of the need for such skills. This analysis can beneft from the construction of a soft skill taxonomy to help identify the relevant skills.

Most prior work analyzing soft skills in job ads has used simple keyword searches and generic synonyms. Moreover, the existing taxonomies for the identifcation of skills are primarily built to identify hard skills. Unfortunately, using the same methodology for soft skills works poorly, because recruiters use diferent terms referring to the same skill (e.g., 'teamwork' and 'collaboration'). For example, a soft skill can be cited as 'writing reports,' which implies that the candidate should have good written communication. Moreover, unlike hard skills, which are mainly expressed using concept words, soft skills are varied in term of the parts of speech used: they may be nouns, adjectives, adverbs, verbs or even phrases. For example, 'collaboration' may be

 Imane Khaouja
imane.khaouja@uir.ac.ma

[1] Universite Internationale de Rabat, TicLab, Rabat, Morocco [2] Universite Mohammed V de Rabat, Rabat, Morocco

[3] Institute for Software Research, Carnegie Mellon University, Pittsburgh, PA 15213, USA

expressed using an adjective such as ‘collaborative’ or the

adverb ‘collaboratively’ or via the phrase ‘to collaborate with coworkers.’

In this paper, we develop TaxoSoft, a methodology for building a soft skill taxonomy that helps in extracting soft skills from job ads. First, we collect around a hundred thou sand of job ads in French and around three hundred thou sand in English. We identify an initial set of soft skills and extract terms related to these soft skills from DBpedia and Word2vec; then, we keep the common terms between these two sets of related terms. This intersection provides a set of related terms that better refects how soft skills are listed in job ads than either DBpedia or Word2vec alone. We then represent these related terms as a network and use network centrality measures to build a hierarchy of terms, generat ing a taxonomy of soft skills. We use our methodology to build a taxonomy of soft skills in both English and French. When used to identify soft skills listed in English job ads on a random sample of job ads, our taxonomy achieves an *F*-score of 0.84, compared to an *F*-score of 0.54 for taxono mies developed in prior work. Finally, we use our methodol ogy to identify the most requested soft skills in the Moroc can and American job markets in order to highlight their importance. This study allows us to shed light on soft skills in the Moroccan job market, and one interesting fnding is that most jobs require at least one soft skill and the most required soft skills are communication and teamwork. Using this methodology, we can also identify soft skills in larger collections, and can cluster these soft skills based on their co-occurrence in job ads in order to identify groups of soft skill that appear together in job ads.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 gives the problem statement and the motivation behind this work. Section 4 describes the tools used in TaxoSoft. In Sect. 5, we give an overview of TaxoSoft before explaining each component separately. In Sect. 6, we describe the tagging process of soft skills in job ads using our taxonomy. Finally, in Sect. 7, we evaluate TaxoSoft by performing a comparison with existing bases.

2 Related work

Related work includes the development of skill bases, assess ing the soft skills of students and analyzing the required soft skills in job ads. Prior work in the generation of skill bases is mainly focused on hard skills.

Various skill bases and extraction systems exist (Bas tian et al. 2014; Javed et al. 2017; Kivimäki et al. 2013; Malherbe and Aufaure 2016; Zhao et al. 2015). Kivimäki et al. (2013) uses the LinkedIn skills taxonomy along with the spreading activation algorithm applied on the Wikipe dia hyperlink graph to extract both inferred and explicitly stated skills from the job ad, while the LinkedIn Skills

system proposed by Bastian et al. (2014), uses a data-driven approach to build a skills folksonomy. Javed et al. (2017) and Zhao et al. (2015) examined the skill sections of resumes and Wikipedia categories to defne and develop a taxonomy of professional skills. They also used Word2vec to disam biguate skills, for instance where the term ‘BI’ could refer to the Bank of Indonesia or to business intelligence. Malherbe and Aufaure (2016) built a skill base that relied on page redirection in Wikipedia, where skills were frst extracted from candidate profles found in various professional social media, and page redirections in Wikipedia were added to these skills as aliases or alternative labels. These bases are focused primarily on hard skills, while the methodology works poorly with soft skills, which are not fully extracted from job postings. Only a proportion of soft skills that can be found in job ads is extracted. One exception is ESCO,[1] which is a project that categorized skills, occupations and other relevant competencies in diferent languages; however, a very costly manual design was required to create the ESCO taxonomies where many stakeholders were involved in the development of such base (De Smedt et al. 2015). ESCO does contain a section called transversal skills, where some soft skills are listed with their alternative labels. Relatively few skill bases have been developed in languages other than English; these include ESCO that was developed in diferent languages and the work in Malherbe and Aufaure (2016) that was developed in French and English.

Several prior studies have analyzed job ads in order to extract job market requirements in term of soft skills; how ever, these skills were extracted manually, using a simple keyword search or based on generic synonyms (Calanca et al. 2018; Daneva et al. 2017; Fernandez-Sanz 2010; Yanaze and Lopes 2014; Florea and Stray 2018; Matturro et al. 2015; Matturro 2013). Recent papers (Brooks et al. 2018; Gardiner et al. 2018) confrm the importance of soft skills in new types of jobs, such as those working with Big Data.

Some papers have also investigated how to assess and evaluate soft skills among students or professionals (Blake and Gutierrez 2011; Joseph et al. 2010; Monasor et al. 2014; Zaharim et al. 2012). Blake et al. assessed professionalism, which is a combination of soft skills such as autonomy and commitment, using LSA over student writing texts (Blake and Gutierrez 2011). Monasor et al. created a model that simulated and measured the soft skills of workers in a soft ware company (Monasor et al. 2014). The authors proposed a simulation model that trains learners to recognize commu nication problems by reproducing diferent scenarios where a virtual guide corrects the leaner mistakes and assesses the learners progress.

[1] https://ec.europa.eu/esco.

Table 1 Extracted terms from DBpedia in English using a public SPARQL endpoint

Term Page redirections Hyperlinks Backward hyperlinks

Active listening Active listening Confict resolution, cooperation, paraphrase, coun seling, artifcial intelligence, communication, etc.

Oral communication, verbal communication, interactivity, active listening, coaxial cable, com merce, conversation, pharmaconomist, etc.

Pseudolistening, serious play, small talk, listening, social skills, etc.

Communication Communication Interpersonal communication, telecommunication, pathogenicity, signal, etc.

Table 2 Extracted terms from DBpedia in French using a public SPARQL endpoint

Term Page redirections Hyperlinks Backward hyperlinks

Flexibilité Flexible, fexibilite Anatomie, souplesse, physique, économie, etc.

Persuasion Active listening Convaincre, foi, hypnose, logique, manipulation, etc.

organisation, hydrogel, compliance, etc.

Argumentation, communication, sophiste, persuasif, soumission, etc.

Prior work has typically processed soft skills in the same way as hard skills; however, they need a diferent type of processing, since they are not technical words or concept terms and can appear within a job ad in the form of adjec tives, verbs, adverbs or even phrases.

3 Problem statement and motivation

Since soft skills are becoming increasingly important in order to fnd a job, we aim to automatically identify the soft skills required by the job market in order to better refect these requirements for students and universities. Our fnd ings will help raise awareness among graduates and uni versities to better develop soft skills. According to a recent study conducted by PayScale in the USA in 2016,[2] 36% of managers surveyed suggest that recent graduates[3] are def cient in teamwork-related and interpersonal competencies (Lacerenza et al. 2018).

Our main focus is on automatically identifying soft skills listed in job ads based on their diferent terminology. How ever, current bases that are used to extract skills from job openings are unable to automatically identify these soft skills. The skill bases found in most recent prior work were built using page redirections found in Wikipedia or DBpe dia. However, page redirections give little information about synonyms when dealing with soft skills. Pages redirection terms are URL redirection terms (see Tables 1 and 2), Wu and Weld (2010) give some detailed example such as the USA is redirected to the article on the 'United States.'

[2] https://www.payscale.com/data-packages/job-skills.

[3] https://www.payscale.com/data-packages/job-skills.

Consequently, we use internal hyperlinks terms to bet ter refect these soft skills and their diferent terminology (see Tables 1 and 2). An internal hyperlink or wikilink is a link from one Wikipedia page to another; in other words, hyperlinks are the names of other Wikipedia pages. We can access these hyperlinks terms by querying Dbpedia. While querying DBpedia, we extract the titles of the internal hyper links or wikilinks (using SPARQL). For instance, from the wikilink 'https://en.wikipedia.org/wiki/Nonverbal_communication' The query will return 'Nonverbal communication.' Then, we use Word2vec, trained on job ads, to flter these hyperlinks terms. Section 5 describes our methodology in more detail.

Lexical databases such as WordNet (Fellbaum

1998) or Roget's Thesaurus (Roget 1911) encode relations between words, such as synonymy and hypernymy (Gabrilovich and Markovitch 2007) A traditional thesaurus also provides only synonymous words; however, DBpedia gives both syno
nyms and related words, where we can fnd that the term 'writing' is related to 'grammar.' More specifcally, such resources contain few proper names and domain-specifc technical terms. Furthermore, these resources have strong lexical orientation and mainly contain information about individual words but little world knowledge in general about *n*-grams (Gabrilovich and Markovitch 2007). Also DBpe
dia structures the Wikipedia content, which is the largest encyclopedia in existence, into structured knowledge so that semantic web techniques can be applied to it. Therefore, we choose to use DBpedia to extract related words of soft skills. Then, we use Word2vec, Word2vec produces word vectors that maps semantic relationships between words (Mikolov et al. 2013) given an input text, to flter those related terms and select the most relevant. Word2vec gener ates a word vector representation where similar words tend to be close to each other.

4 Background

The background to our approach consists of methods for extracting terms related to a given word using DBpedia and Word2vec. Here, we discuss DBpedia and Word2vec that will be used in our methodology.

4.1 DBpedia

The DBpedia project focuses on converting Wikipedia con tent into structured knowledge so that semantic web tech niques can be applied to it. Each concept has a category, a description, a type, aliases corresponding to URL redi rections, and internal hyperlinks between pages (Auer et al. 2007).

The DBpedia data set can be accessed online via a SPARQL query endpoint and as Linked Data. More specif cally, we use a public SPARQL endpoint over the DBpedia data set in this paper.[4],[5]

The work in Wu and Weld (2010) used Wikipedia redi rection pages and backward links to automatically construct sets of synonyms (see Tables 1 and 2). These tables show the extracted hyperlinks and page redirections from DBpedia in English and French, respectively. More specifcally, these links refer to other pages in Wikipedia.

4.2 Word2vec

The Word2vec model learns vector representations of words from a large corpus. These vectors, called word embeddings, are produced using two model architectures: the continuous bag of words model (CBOW), or the continuous skip-gram model (Mikolov et al. 2013). These embeddings are the out
put of the model. The vectors produced in this way map semantically similar words to nearby points in word embed dings. Word embeddings are a learned representation for text where words that occur in the same context tend to have a similar representation. To generate Word2vec semantic rep resentations, we used the Gensim Python library[6] (Rehurek and Sojka 2010).

Hence, given a word, we can extract related terms by computing the cosine distance to the other words within the word embeddings. This means that we can evaluate the degree of similarity between two words by evaluating their cosine distance. Semantically, related words have a high cosine distance. Tables 3 and 4 show related words with their cosine distance extracted using Word2vec trained on job ads in English and French, respectively. (We trained

[4] http://dbpedia.org/sparql.

[5] http://fr.dbpedia.org/sparql.

[6] https://radimrehurek.com/gensim/models/word2vec.html.

Word2vec for each language apart.) The quality of the word embeddings strongly depends on the size of the input text, and larger corpora give better results (Altszyler et al. 2017; Lai et al. 2015). Also, training on an in-domain corpus can signifcantly improve the quality of word embeddings for a specifc task.

On the other hand, generic embeddings are very broad and diverse. For instance, Word2vec trained on Wikipedia articles gives good results, but these not specifc to a single domain (see Fig. 1). Figure 1 shows words related to 'com
munication,' using Word2vec trained on Wikipedia articles.

5 Methodology

5.1 Overview

Soft skills are hard to measure consistently from job ads, since they appear in a range of diferent expressions written by diferent individuals. To tackle this problem, we propose to use a combination of Word2vec and DBpedia in order to extract alternative labels for a given soft skill. As shown in Fig. 2, the proposed approach can be decomposed into the following fve steps:

Step 1 *Data collection* We collect online job ads written in French and in English

Step 2 *Data preprocessing* We establish an initial set of soft skills by extracting them from the collected job ads. Since our initial set is limited to a small number of skills, we add those found in prior work (Daneva et al. 2017; Yanaze and Lopes 2014).

Step 3 *Related terms using Word2vec* We extract related words using Word2vec (see Tables 3 and 4), where Word 2vec is trained on our collected job ads.

Step 4 *Related terms from DBpedia* We extract hyperlink terms of the initial set of soft skills by querying DBpedia (see Tables 1 and 2).

Step 5 *Alternative labels* We then perform the intersec tion of these two extracted sets (see Tables 7 and 8) to give alternative labels for soft skills. Finally, the hierarchy between skills is defned using the centrality measure, where we represent soft skills in a network.

5.2 Taxonomy generation

5.2.1 Data

In order to generate our soft skill taxonomy, we collect job postings from diferent job boards. The job ads are gathered from diferent sources, specifcally French and English job ad websites, and the numbers of job ads collected are shown in Table 5. These websites include

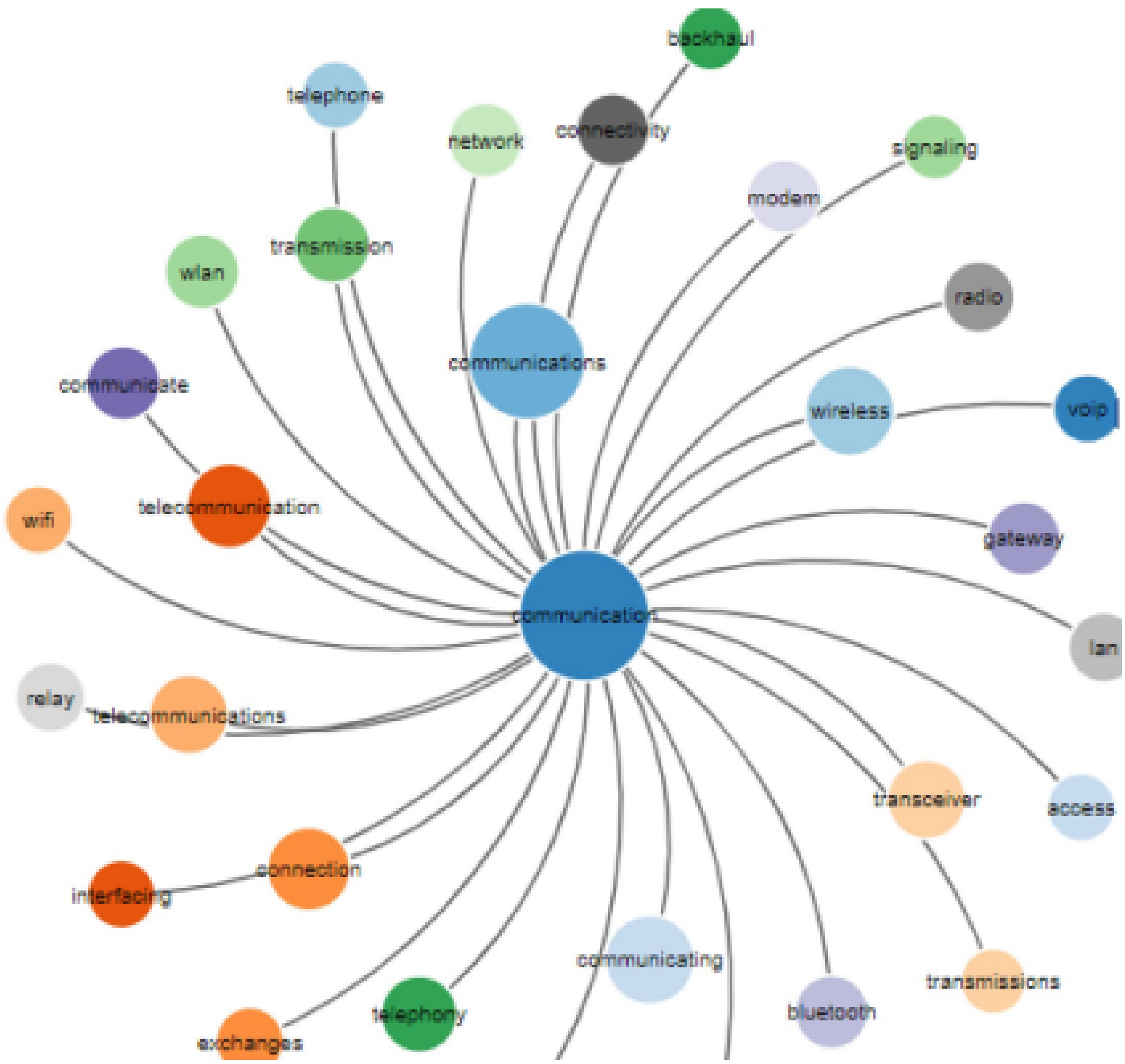


Fig. 1 Top 30 analogous words extracted from Word2vec trained on Wikipedia articles (www.wordsimilarity.com)

Table 3 Extracted terms from Word2vec in English

Term	Related terms from Word2vec
Communication	(Interpersonal communication, 0.63), (time management, 0.61), (organizational, 0.49), (oral communication, 0.41), (verbal com munication, 0.41), (interaction, 0.40), (active listening, 0.31) ...
Active listening	(efective, 0.51), (listening, 0.49), (criti cal thinking, 0.49), (lead infuence, 0.49), (diplomacy, 0.45), (confict resolution, 0.42), (communication, 0.31) ...

CareerBuilder, Apec, Keljob, Rekrute and Emploi.ma (see Table 6). These primary websites were identifed through prior work (Javed et al. 2017; Malherbe and Aufaure 2016) and through a job aggregator in order to select the most popular job boards. We scraped the abovementioned web sites and other miscellaneous French and English websites until we collected a sufcient number of job ads to vali date our work. We also scraped English job ads found on

Table 4 Extracted terms from Word2vec in French

Term	Related terms from Word2vec
Fléxibilité	(Adaptabilité, 0.63), (écoute, 0.60), (réactiv ité, 0.60), (anticipation, 0.59), (evemplarité, 0.59), (amabilité, 0.58), (soupless, 0.56) ...
persuasion	(Conviction, 0.85), (proposition, 0.62), (pug nacité, 0.62), (persérance, 0.60), (négocia tion, 0.53), (diplomate, 0.52), (propositions, 0.52) ...

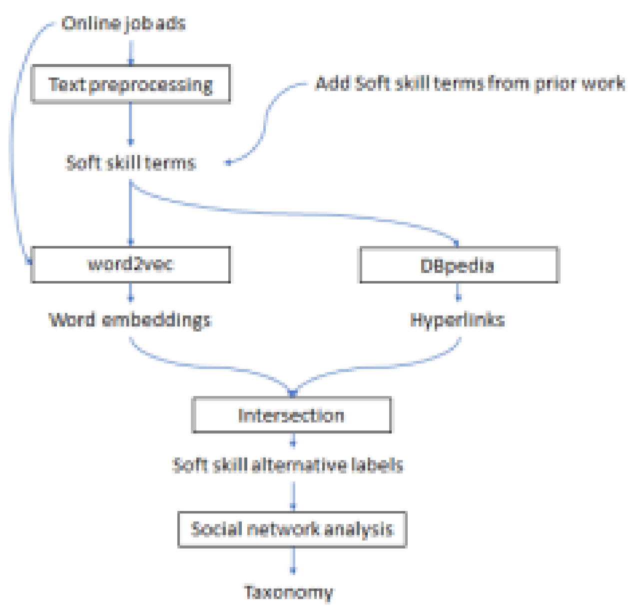


Fig. 2 Methodology overview

CareerBuilder, and used the Kaggle data set[7] in order to increase the amount of data. The Kaggle data set contains approximately 360,000 job ads. Also, the amount of data scraped from CareerBuilder is 20,000 ads.

We scraped all job ofers in those websites without dis tinction by developing scrapers in Python for each website that extract the text from HTML tags.

The number of English job ads is greater than those in French, as no prior database existed for French ads. Although our work focused on analyzing ads written in French, we collected those written in both languages, as English is the most common language used in prior work and allows us to test and apply TaxoSoft more widely.

5.2.2 Data preprocessing

First, we defne the n-gram words using DBpedia. More specifcally, we defne n-gram by querying DBpedia. To do this, we remove articles from the text, such as 'the' in Eng lish and 'la/le/les' in French. This step allows us to avoid defning proper nouns as n-grams, such as 'The Team', or 'The Ofce.' (In Wikipedia, 'The Team' refers to a radio network and 'The Ofce' to a sitcom.) This step is important in order to defne n-gram and avoid defning proper nouns as n-grams.

Following this, we extract the most common 2- and 3-gram phrases cited in our corpora of job ads that occurred at least 10 times in our corpora; this threshold was set after

Fig. 3 Tagging process

Table 5 Number of job ads

Language of text	Number of job ofers
French	160,000
English	380,000

Table 6 Number of job ads extracted from each website

Website	Number of job ofers
CareerBuilder	20 000
Apec	23 500
Keljob	38 500
Rekrute	7000
Emploi.ma	9000

examining the common 2-gram and 3-gram and in order to reduce complexity. We validate the *n*-grams using DBpedia: if we fnd a page associated with the *n*-gram within DBpe dia, then we keep it; otherwise, we discard it. For instance, for the *n*-gram ‘active listening,’ we can fnd the associ ated page ‘http://dbpedia.org/page/Active_listening’. Then, we separate these *n*-grams by an underscore in order to be considered as single words by Word2vec. (‘Active listening’ will become ‘active_listening.’)

On the other hand, we split the description into sentences using punctuation since our job ads lack a defned struc ture. These sentences are classifed into two categories, a description of the enterprise and a description of the job (its requirements and duties), using a naive Bayes text classifer trained on a manually labeled sample (see Fig. 3) containing

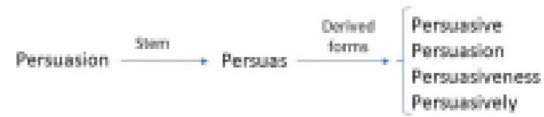


Fig. 4 Example of adding derived forms

[7] https://www.kaggle.com/c/job-salary-prediction/data.

Table 7 Final terms after intersection in English

Soft skill	Final related words
Communication	Interpersonal communication, oral communication, verbal communication, interaction, active listening
Active listening	Listening, communication, confict resolution, social skills

100 job ads per language. We implement this step in order to set apart the job description and the enterprise descrip tion as our job ads are unstructured. In fact, this step can be skipped in the case the job ads are structured, especially where the job description and the enterprise are separated in two diferent sections.

We use the **job description** to extract the frst set of soft skills. We query DBpedia and select words having the terms ‘feeling,’ ‘attitude’ or ‘ability’ in their categories or in their hyperlinks in DBpedia. We choose those words because we found that soft skills have these words in common (see Table 15).

We then validate the resulting list manually (see Table 15 in the appendix). However, the collected soft skills (of which there are 20) may represent only a fraction of the possible skills, and we therefore enrich the list by adding terms found in the literature (Daneva et al. 2017; Yanaze and Lopes 2014) (around 20 soft skills).

5.2.3 Selection of soft skills alternative labels

In order to select the most relevant terms related to a given soft skill, we use the steps in Algorithm 1. In fact, Algo rithm 1 takes as input the soft skill term from the list of soft skills, found in Sect. 5.2.2, and a threshold fxed to 0.3 in our case. This threshold was set empirically, where we compared values from 0 to 0.5 with a step of 0.05 and took the one that produced the best alternative labels. We have observed that this threshold gave the best alternative labels and the most representative synonyms of the soft skills when compared to the alternative labels produced by other values.

Table 8 Final terms after intersection in French

Soft skill	Final related words
Flexibilité	Souplesse, fexibilité
Persuasion	Preuve, éloquence, argumentation, négociation

In line 1 of Algorithm 1, we extract the hyperlinks from DBpedia of the term S, which is a soft skill from our list (see Sect. 5.2.4 for more details). RTD contains the related terms extracted from DBpedia. In other words, RTD contains the hyperlink term extracted from DBpedia. Then, in line 5, we add the derived forms of each returned hyperlink term from RTD using a stemmer to RTDstem (Sect. 5.2.6 explains this step in more detail), an illustration is given in Fig. 4. In line 7, we compute the cosine similarity using Word2vec between the input soft skill term S and each ele ment of RTDstem. We train the Word2vec model using our job ads, as explained more in detail in Sect. 5.2.5. In line 8, only words with a score greater than the threshold are kept. This step represents the intersection between DBpedia terms and related terms from Word2vec (see Tables 7 and 8). These tables show the results of such intersection. Finally, in line 12 of Algorithm 1, we obtain alternative labels for the soft skill in our input list. Tables 13 and 14 in the Appendix contain more examples of our taxonomy. We apply this algorithm to all the soft skills found in our list, as described in Sect. 5.2.2.

Table 9 Final terms after intersection and adding their derived forms in English

Soft skill	Final related words	Final related words with their derived forms
Teamwork	Team, team player, team building, coopera tion, collaboration, synergy	Team, teams, teaming, teamed, teamwork, teamworks, teamworker, teamworking, team player, team building, collaboration, collaborative, collaborate, collabo ratively, cooperate, cooperatively, cooperation, cooperative, synergis, synergi, synergy, synergies
Persuasion	Persuading, persuasive, persuade, persuasion	Persuasion, persuasively, persuade, persuading, persuades, persuade, persuasive, persuasiveness

Table 10 Final terms after intersection and adding their derived forms in French

Soft skill	Final related words	Final related words with their derived forms
Flexibilité	Souplesse, fexibilité	Souplesse, souplesses, fexibilité
Persuasion	Preuve, éloquence, argumentation, négociation	Persuadé, preuves, négocier, persuasion, arguments,

5.2.4 Extraction of related words from DBpedia

To extract related words for a given skill, we extract a subset of DBpedia data that match our skills in French and Eng lish. For each skill, we extract internal direct and backward hyperlinks (see Tables 1 and 2). This extracted set contains related words for the input skill. Although hyperlinks con tain terms that are related and similar to the input term, they may be very distant from the context of soft skills, as shown in Tables 1 and 2, and refning is therefore necessary.

5.2.5 Extraction of related words using Word2vec

We use our job ads to train the Word2vec model, and n-gram words are separated by an underscore. We set min-count to 10 in order to better control the level of noise. Since our sample is small, the output is noisy if we use the default value of the model. (The default value for min-count is 1.) The minimum count defnes the minimum number of occur rences required for a word to be included in the word vectors. Only words that were cited more than 10 in the corpora will be considered by Word2vec model; this step reduces noise by removing misspelled words and words that are sparsely mentioned. We tested several hyperparameters such as 1, 5, 10, 15. We have observed a signifcant improvement in the semantic relevancy of the trained vectors with the value 10. Also, we increase the vector size to 200 from its default value of 100 as done in Zhao et al. (2015). For each selected term, we extract related terms using Word2vec (see Tables 3 and 4). Since more distant words are usually less related to the current word than those close to it, we set a threshold for the returned words, and only consider those with a score greater than the fxed threshold. In our methodology, this persuasif, persuadant, persuader, négociateur, éloquent, argumenter, preuve, éloquence, persuasive, argumentation, négociation

threshold was fxed at 0.3. This threshold value was selected empirically.

5.2.6 Stemming process

Soft skills are not always referred to in the same manner in job ads, and can be expressed as adjectives, adverbs or verbs. Hence, we enrich the related words by adding word inflections such as adjectives, verbs and adverbs. We add these inflections using a stemmer, which is run on all words in the collected ads in order to obtain the stems and their inflections. More specifically, we stem all words found in the sentences in the preprocessed job description. Next, words with the same stem are clustered into separate groups in with the number of their occur rences in the job ads. Then, inflections with the highest occurrence (the top 4) are added to the term (see Tables 9 and 10). Although stemmers do not provide information about parts of speech classes, we consider the most com monly occurring inflections to contain these classes. Also, the n-grams are not considered in this stemming process.

5.3 Hierarchical taxonomy

After defning our taxonomy, we explore the relationships between skills by mapping the connections between skills in a network representation. In TaxoSoft, we consider soft skills as nodes, and build the hierarchy by exploring the relationships between nodes. An edge between two skills is settled when one skill belongs to the set of related terms of the other; for example, the term ‘communication’ belongs to the group of terms related to active listening (see Table 6).

Table 11 Soft skill taxonomy

Parent	Child
Communication	Verbal_communication, Interpersonal_communication, active_listening
Written_communication	Writing
Persuasion	Presentation, infuence
Accountability and integrity	
Social skills	Emotional intelligence
Mentorship	Mentor, coaching, mentorship, mentor ing
Critical_thinking	Learning, argument
Creativity	Curiosity
Decision_making	Decision
Enthusiasm	Eagerness, passion

Several prior studies have used centrality measures to defne the latent hierarchy between nodes (Boldi and Monti 2016; Heymann and Garcia-Molina 2006), and Benz et al. (2010) used betweenness centrality. To build the hierarchy

between soft skills, we compute the betweenness centrality for each node, where the node with the highest centrality represents the parent. In each component, the node with the highest betweenness centrality is considered the parent for the direct nodes connected to it. In Fig. 5, ‘communication’ has the highest centrality in the component; this means that ‘communication’ is the parent, and ‘oral communication,’ ‘verbal communication,’ ‘interpersonal Communication’ and ‘active listening’ represent the children. However, if all nodes have a centrality of zero, this means that they repre sent the same skill, such as ‘integrity’ and ‘accountability,’ as shown in Fig. 5 and Table 11.

6 Application

6.1 Soft skill tagging

Before implementing the tagging process, we split the job ad into sentences, where each sentence is classifed into two categories, a job description and an enterprise description,

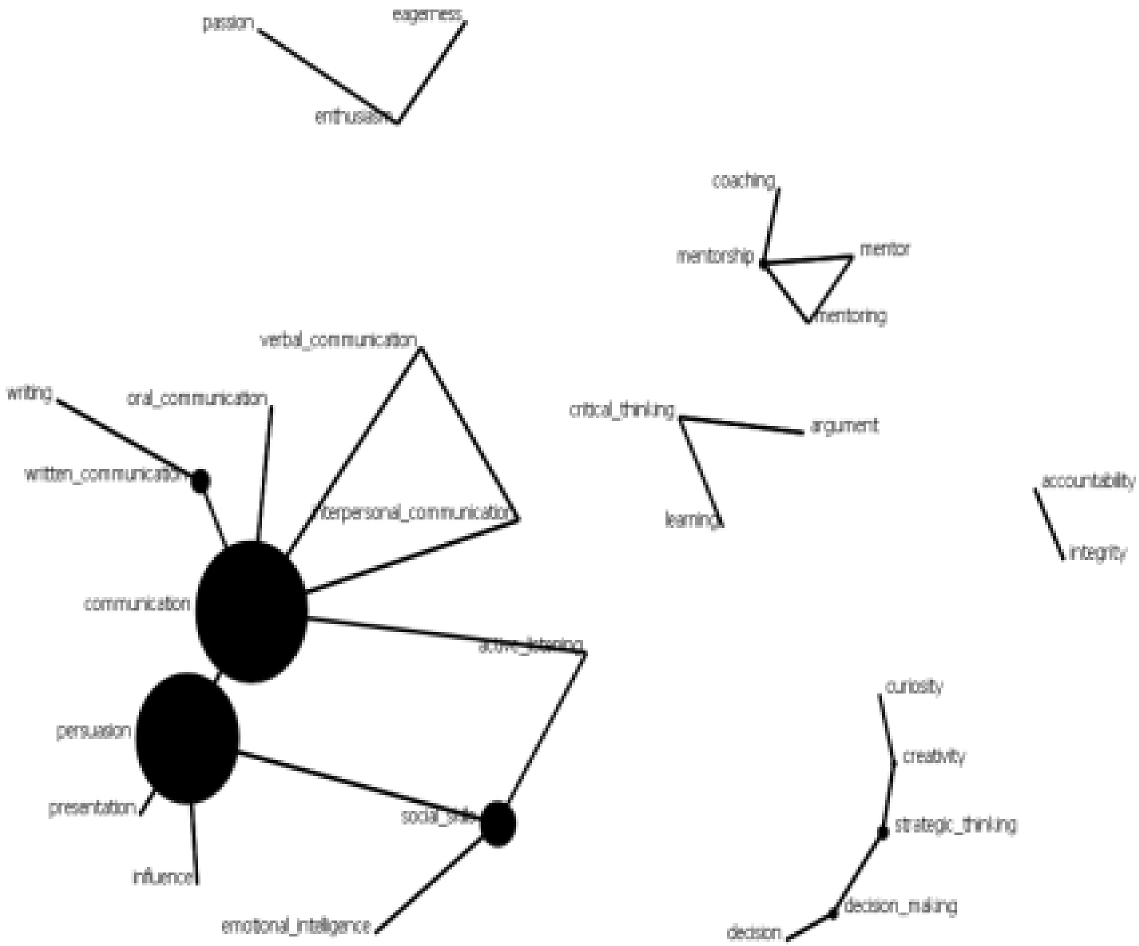


Fig. 5 Example of hierarchical taxonomy

Restaurant-General-Manager
What we're looking for:
A friendly, enthusiastic attitude Someone who knows the ins-and-outs of running a business (P&L management, food safety, sales-forecasting, staffing... you get the gist) Someone who can oversee the preparation of safe and delicious food and maintain a clean, neat and organized restau rant
A creative approach to marketing (fundraisers, community-engagement, etc.)
Someone that loves to develop and lead a team
The ability to deliver a great guest experience
A team player who can jump in where needed
Previous restaurant experience
The ability to communicate in the primary lan guage(s) of the work location

Extracted soft skills:
Enthusiasm, Creativity, Hospitality, Team work, Communication

Fig. 6 Extracted soft skills from a job ad

which contain some words that can distort the tagging (see Fig. 3). Then, tagging is carried out based on an exact match between the job description and our taxonomy, as depicted in Fig. 6. This fgure shows an example of job ad and the extracted soft skills, where the underlined words are iden tifed using the alternative labels found in our taxonomy.

6.2 Case study

6.2.1 Soft skills according to years of higher education

The aim of this taxonomy is to identify soft skills within job ads in order to better understand the needs of the job market. We extract the required soft skills from a job ad, as shown in Fig. 6. However, to get better insight into job market needs, we gather job ads from various Moroccan websites between February 2017 and August 2018. These websites were iden tifed using a job aggregator. Before measuring soft skills, we preprocessed our job ads by performing deduplication using a shingling algorithm (Manku et al. 2007). The same preprocessing was performed for English job ads, which were gathered from CareerBuilder over a month (Febru ary 2018) in a single crawl. We divided the Moroccan job ads according to the level of higher education required. The academic degrees awarded upon completion of fve years of higher education in Morocco are known as Master's and Engineering degrees. Universities and vocational schools also ofer degrees upon completion of two to three years of training. Some job ads in the Moroccan job market did not require higher education, and we also examined this group.

We split the English job ads according to the level of degree required. Figures 7 and 8 show that the number of job ads in which soft skills were cited increases with the number

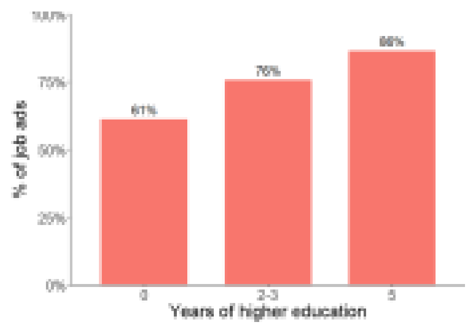


Fig. 7 French job ads from Moroccan websites containing soft skills

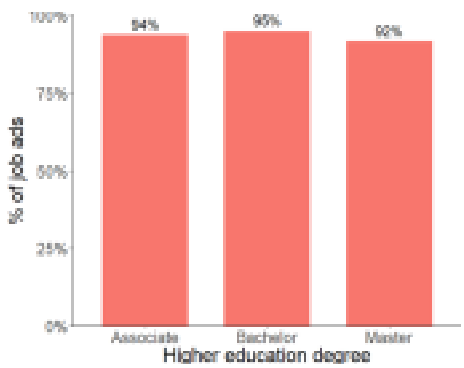


Fig. 8 English job ads From CareerBuilder containing soft skills.

The tree degrees require more than 90% which indicates the importance of soft skills in the American job market

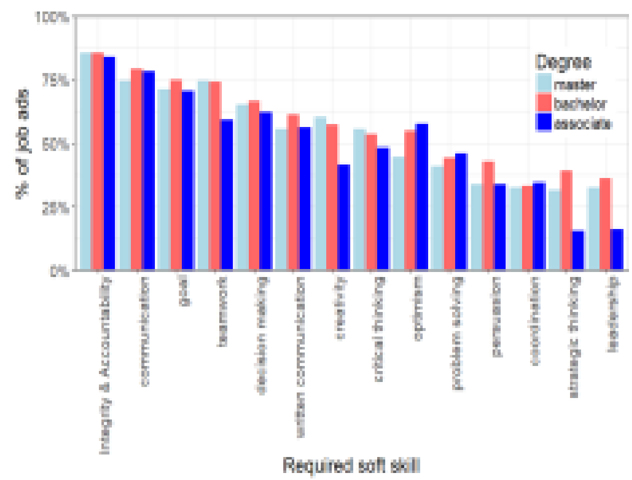


Fig. 9 Top Soft skills in the American job market

of years of higher education required. This confrms that soft skills are important to employers and can make a diference in the recruitment process, probably because higher posi tions involve more responsibility. Thus, universities should emphasize soft skills in their curricula. From Fig. 9, we can see that the soft skills required in job ads are almost identical,

Social Network Analysis and Mining (2019) 9:43

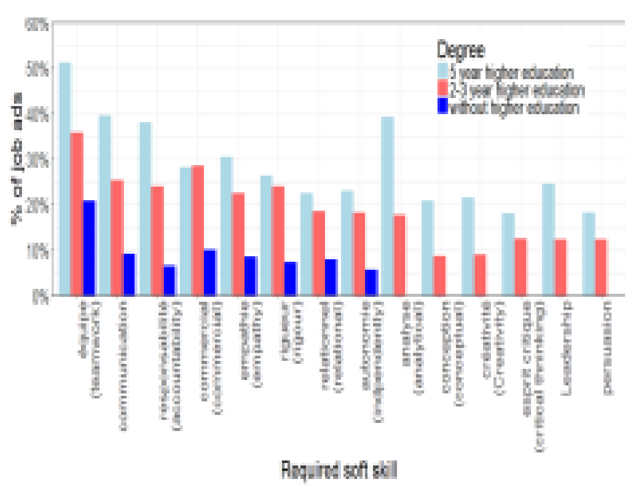


Fig. 10 Top Soft skills in the Moroccan job market written in French

although they are requested more often for jobs requiring a higher degree, such as a Bachelor's or a Master's. This means that soft skills are more often needed in jobs requiring more higher education and involving more responsibility.

We can see from Fig. 10 that the Moroccan job ads requiring an Engineers or Master's degree focus primarily on teamwork, communication and analytical skills. Teamwork is important because working on projects requires working with other people. Communication is also important since meetings, presentations, reports and even emails require solid communication skills. Analytical skills are needed to identify problems before they occur, to determine the root causes of problems, to create and test prototypes and to evaluate test results. We can also see that job ads for people with up to three years of higher education require higher levels of commercial skill, mainly because most of these are commercial positions. Our results also demonstrate that adverts for people without no qualifcations focus more on motivation, which refers to the willingness of an employee to work and to meet the company's goals.

In the American job market (see Fig. 9), adverts requir ing a Bachelor's or Master's degree require more integrity, accountability, communication and teamwork. According to a recent study conducted by PayScale in the USA in 2016,[8] 36% of managers surveyed feels that recent graduates are defcient in teamwork-related and interpersonal competen cies (Lacerenza et al. 2018).[9] Accountability and integrity are some of the most important attributes of individuals in the workplace. Accountability involves being responsible or answerable for ones actions, while integrity describes an individual who demonstrates sound moral and ethical

[8] https://www.payscale.com/data-packages/job-skills. [9] https://www.payscale.com/data-packages/job-skills/methodology.

principles at work (e.g., honesty). These values are the foun dation on which coworkers build relationships, trust and efective interpersonal relationships. Jobs requiring associ ate degree holders also require optimism (60%); a positive attitude fuels more positive results.

6.2.2 Soft skills according to sectors

We also analyze the soft skills required in diferent sectors in Morocco. Figure 11 shows the top ten soft skills required in banking, IT ofshoring, the automotive industry, education, justice, medicine and call centers. We can clearly see from Fig. 11 that the requirements for soft skills vary across sec tors, with each having certain specifc needs.

In banking, the soft skills required are teamwork, com mercial skills, communication and analytical skills, while in education, teaching and communication skills are the most requested. In medicine, teamwork and accountability are the most often mentioned. In justice, candidates need to work in teams while showing independence, rigor and analytical skills are also important, while in call centers, candidates need strong motivation, commercial skills and communication.

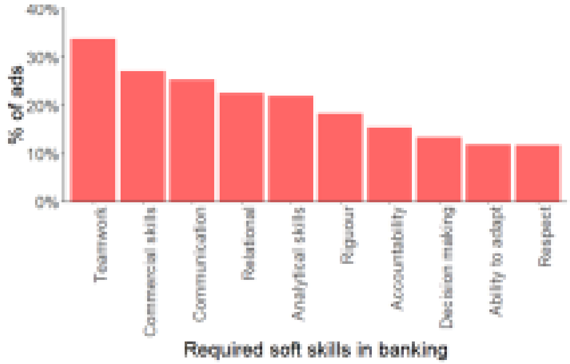
6.2.3 Co-occurring soft skills

In order to detect co-occurring soft skills, we build a co occurrence graph, in which soft skills are represented as nodes and an edge between two nodes indicates the co-occur rence of these terms, with a weight proportional to its fre quency. We start by building a matrix in which the columns represent soft skills and the rows represent job ads. Each cell indicates whether or not a soft skill (or its alternative label) is cited. When we have constructed this matrix, we fold it in order to obtain a matrix of the co-occurrence soft skills. We then use Louvain's community detection algorithm to discover groups of soft skills that co-occur within the job ads. Community detection has been used in prior work to detect co-occurring terms and keywords (Grineva et al. 2009). From the clusters in Fig. 12, we can see that groups of soft skills appear together. The fgure shows that analytical skills, ability to learn, self-confdence, curiosity, decision making, critical thinking, proposition and problem resolu tion co-occur, indicating that students should develop these skills in order to succeed in their future careers.

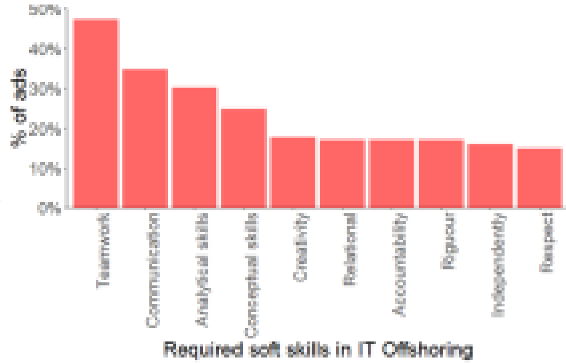
7 Evaluation

7.1 Evaluation metrics

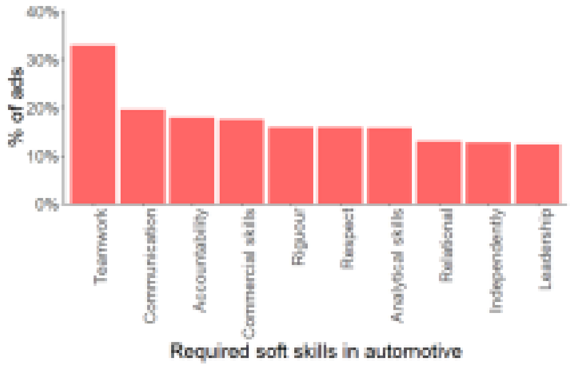
In order to evaluate the accuracy of TaxoSoft, we compute three metrics: precision, recall and *F*-score. These metrics



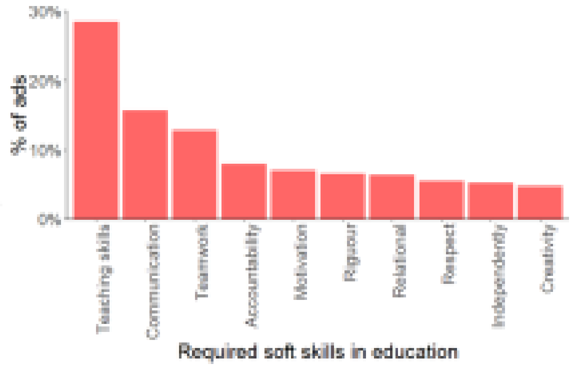
(a) Top soft skills in banking



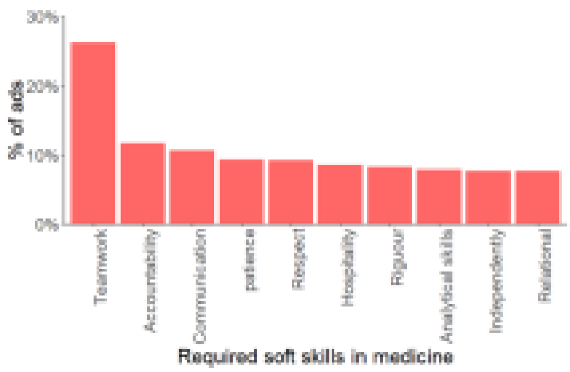
(b) Top soft skills in IT offshoring



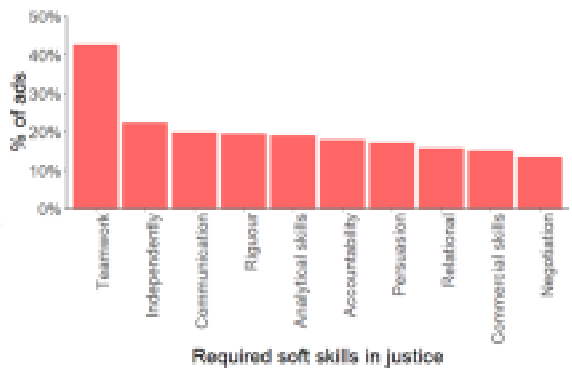
(c) Top soft skills in automotive



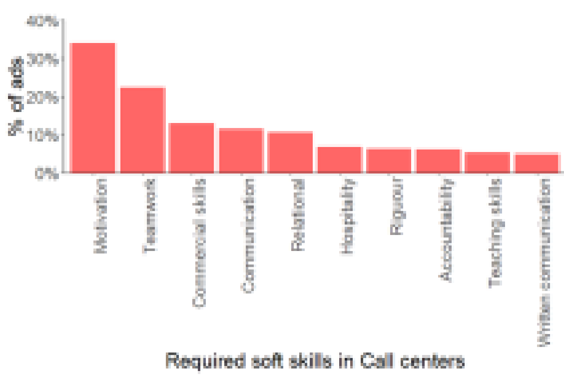
(d) Top soft skills in education



(e) Top soft skills in medicine



(f) Top soft skills in justice



(g) Top soft skills in Call centers

Fig. 11 Top soft skills in diferent sectors

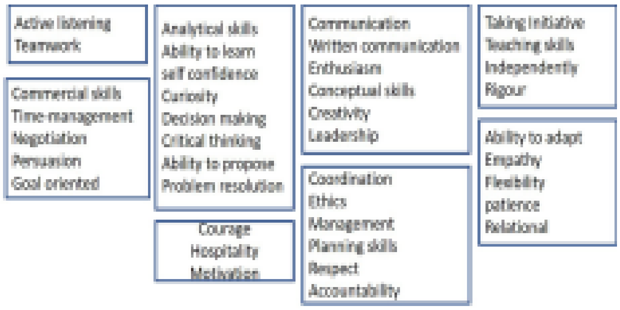


Fig. 12 Clusters of Soft skills in Morocco

Table 12 Comparison between the existing bases

Language	Taxonomy	Precision (%)	Recall (%)	*F*-score (%)
English	Esco	85	10	17
	Knowledge base	94	41	54
	TaxoSoft	82	89	84
French	Esco	100	1	2
	Knowledge base	90	42	57
	TaxoSoft	81	80	80

have been used in prior work (Javed et al. 2017; Malherbe and Aufaure 2016).

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$F\text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where True Positives = Skills extracted correctly from the job ad, False Positives = Skills extracted and do not exist within the job ad, False negatives = Skills not extracted and exist within the job ad

7.2 Comparison

The predefned metrics are evaluated on a random sample of 100 job ads, in the same way as in Malherbe and Aufaure (2016). We extract soft skills from job ads using TaxoSoft. More specifcally, we use the soft skills and their alternative labels to extract soft skills (see Tables 13 and 14 in Appen dix). These alternative labels were found using TaxoSoft. We also extract soft skills from job ads using ESCO and the knowledge base used in Malherbe and Aufaure (2016). In ESCO, we use only the transversal skills, which are soft skills. We replicated the methodology described in Mal herbe and Aufaure (2016) as far as possible, in which page redirections (see Table 1) are considered to be aliases or

synonyms to soft skills. Since we could not access the data used in their paper, we used the soft skills set described in Sect. 5.2.2. The database generated from the work in Mal herbe and Aufaure (2016) is called the knowledge base. In order to avoid bias in our evaluation, we manually extracted soft skills from the random sample, and then we compared the extracted soft skills with the soft skills identifed by the knowledge base and TaxoSoft.

The results in Table 12 show that ESCO has low recall, which is as expected, since the transversal skills section in ESCO has little alternative labels for each skill. However, the knowledge base also has low recall, because pages redi rections give little information about soft skills alternative labels. TaxoSoft has the highest recall and *F*-score for both languages, where a skill is detected in its diferent wordings within a job ad. However, TaxoSoft has a lower value in French than English, which is partly due to the size of the collection of ads passed to Word2vec in the input. Over all, TaxoSoft has much higher *F*-score than ESCO and the knowledge base. This improvement is essentially due to the fact that we tailored our methodology to extract soft skills. TaxoSoft could be improved by the use of a larger collec tion of job ads. We can also train Word2vec using sentences that contain soft skill requirements, which would give better results.

8 Limitations and future work

Our methodology was evaluated on a relatively small set of job ads, since the manual labeling of data takes a consider able time. In the future, we intend to evaluate it on a larger sample using mechanical Turks. We also intend to measure soft skills required by diferent occupations, in order to con struct a chart of soft skills for each occupation apart. To do this, we need to categorize the job ads into diferent clusters based on the function of their occupation. We plan to apply our methodology to other languages, since any language for which DBpedia is rich enough and can be considered. We also intend to add a stemmer for bigrams, in order to better match the bigrams found in the text with the soft skills, and 'problem solving' could be linked to 'resolve problems.' A larger sample would help to improve the extraction of soft skills, and Word2vec trained on sentences that contain soft skills would improve our results.

9 Conclusion

Soft skills are gaining in importance in the job market; they are highly sought after by employers, as they facilitate the success of projects and allow organizations to thrive. We present a methodology to generate a soft skill taxonomy in

French and English, which uses a combination of DBpedia and word embeddings. This combination provides a set of related terms that refect how soft skills are listed in job ads, and achieves better results than DBpedia or word embed dings alone. We then build a soft skills hierarchy by rep resenting our taxonomy as a network and using centrality measures. To evaluate our methodology, we use our meth odology to extract soft skills from a random sample of 100 job ads written in both French and English. We found that our methodology could extract more soft skills than other knowledge bases, and it outperformed existing bases by 30% in terms of its *F*-score. This improvement is essentially due to the fact that we tailored our methodology for soft skills extraction. We then used our taxonomy to identify impor tant soft skills in both the Moroccan and American job mar kets. To do this, we gathered a signifcant number of job ads from various websites identifed in prior work and using a job aggregator. Our results indicate that the need for soft skills increases with the number of years of higher education specifed in the ad (up to Master's level). This is probably due to the fact that higher positions involve more respon sibility. Thus, universities should emphasize soft skills in their curricula.

Moreover, while some soft skills such as teamwork are common in both the USA and Morocco and at diferent edu cational levels, others are specifc to a particular country, education level or sector. In Morocco, jobs for Engineers and Master's degree holders focus primarily on teamwork, communication and analysis skills. We also note that jobs requiring up to three years of higher education need more commercial skills, mainly because most of these jobs are commercial positions. Our results also show that advertise ments for people without no qualifcations focus more on motivation, a quality that demonstrates the willingness of an employee to work and meet the company goals, and respect, which helps ensure a good working environment.

Furthermore, soft skills vary across sectors. In Morocco, IT and banking jobs require teamwork, communication and analytical skills, while in education, teaching skills are the skills most needed in order to efectively convey and explain academic subject matter, and to support student learning. Teaching skills may be considered as hard skills in educa tion; however, it is considered as soft skills in other sectors (Nolinske and Millis 1999).

However, call centers primarily require motivation. Accountability is required in various sectors, and especially in medicine, education and the automotive industry. This analysis helped us identify the most important soft skills at each educational level and in each sector. Our results show that universities should enhance teamwork among their stu dents, in order to help them develop their teamwork and communication skills. Moreover, students pursuing a Mas ter's degree should develop their analytical skills; these can be improved by encouraging students to draw conclusions from texts, images, charts or case studies. In the American job market, jobs for Bachelor's and Master's degree holders primarily need integrity, accountability, communication and teamwork. Accountability and integrity are some of the most important values required of individuals in the workplace. Accountability involves being responsible or answerable for an action, while integrity describes an individual who dem onstrates sound moral and ethical principles at work (e.g., honesty). These values are the foundation on which cow orkers build relationships, trust and efective interpersonal relationships. Jobs for associate degree holders require opti mism, since a positive attitude fuels more positive results. We also fnd clusters of soft skills in Morocco; these clusters give insights into soft skills that go together.

Finally, this methodology can be replicated in order to generate this taxonomy in other languages and to examine the evolution of soft skills in job ads. Our taxonomy can help universities, job seekers and employability programs to identify the most important soft skills.

Acknowledgements This work is supported in part by the United States Agency for International Development (USAID) under grant AID-OAAA-11-00012 and by a Google Africa PhD fellowship. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the ofcial poli cies, either expressed or implied of USAID or Google. The authors would like to thank Mehdi Zakroum and Ibtissam Makdoun for useful comments and discussion.

Appendix

See Tables 13, 14 and 15.

Table 13 Soft skill taxonomy in English

Soft skill	Alternative labels
Accountability	Responsibilities, responsibility, transparently, responsible, accountability, transparency, transparent, transparencies, accountable, account, accountabilities
Active listening	Listens, listen, communicator, social skills, listening, confict resolution, communications, listener, communicate, communication, active listening
Adaptive	Adaptability, adaptive, stressed, adapt, adaptable, stressful, stress, stresses, adapting
Analytical	Synthesis, analytically, analytical, analytic, analytical skills, analytics
Argumentation	Interpreting, interpret, debate, opinion, logically, arguments, logical, debated, opinionated, interpreted, debating, argument, logic, opinions, debates, interpretation, logics, argumentation
Coaching	Coaching, coaches, performance management, coached, sprinting, coach, sprints, sprint
Commitment	Commitment, commitments, committed, commit
Communication	Interactive, communicator, communication skills, understandable, interpersonal, communications, understanding, interpersonally, communicate, communicating, interact, communication, understand, interpersonals, interacting, open communication, interaction, understands, interpersonal communication,
Conceptual	Paradigm, conceptualize, conceptually, primes, pragmative, conceptual, concepts, pragmatism, concepting, primed, conceptualization, conceptualism, pragmatically, paradigms, pragmatic, conception, concept, priming, prime,
Confict management	Confict management, counselling, counsels, counsel, counseling
Coordination	Coordinating, coordination, coordinated, coordinator, coordinate
Creativity	Original, geniuses, musical, creative thinking, originally, insights, imagine, ingenuous, creatives, innovate, imaginable, origination, insightful, thinking outside the box, creatively, innovation, music, genius, originality, creative, creativity, musics, graphic design, intellect, innovative, musicals, imagination, innovations, insight, insightfully, ingenuity, imaginative
Critical thinking	Rationalization, intellectuals, reasonable, rationally, thought, critic, humanism, reasonably, critically, thoughts, rationality, reason, critical thinking, intellectually, critics, criticality, relevance, rational, critical, relevancy, foresight, reasoning, thoughtfully, advanced level, thoughtful, intellectual, reasons, relevent, relevant, criticism
Curiosity	inquisitional, curious, curiosity, curiousity, interested, inquisitive, interesting, inquisitiveness, interest, inquisition, interests
Decision	Discern, decision, decisions, decisive, judgmental, discerning, discernment, discerned, judgment, judgments, decisioning, discernible, judgmentability
Decision making	Data quality, estimation, action, costs, actioned, actionable, advocacy, decision making, intuitive, estimate, intuit, data governance, memorandums, estimator, estimating, pros, business intelligence, costings, costing, cost, intui tively, actions, estimates, memorandum, intuition
Detail	Details, detail, detailing, detailed
Diverse	Diversify, diversifes, multiculturalism, diverse, diversion, diversions, multicultural, diversifed, multicultured, multiculture, diversity, diversifying
Eagerness	Persistence, persist, persistently, eager, eagerly, persistent, eagerness
Emotional intelligence	Motivating, empathi, motivate, emotional intelligence, motivated, empathy, motivation Enthusiasm Enthusing, emotional, spirited, enthuse, enthusiastic, spirit, spiritability, enthused, enthusiastically, enthusiasma bility, enjoyment, emotionally, enthuses, enjoying, enjoys, enthusiasm, emotions, spirits, enthusiast, enjoyable, enthusiasts, emotion, enjoy
Ethic	Ethics, respect, respected, ethically, respectful, respective, ethical, ethic
Flexibility	Flexibly, fexibilities, fexibility, fexible

Goal	Purpose, profts, planned, milestoned, action plan, purposeful, milestones, desire, milestone, goals, objects, desired, object, plans, primary objective, purposely, objective, goaled, proftability, objectives, mission statement, proft, goal, proftable, plan, desireable, planning, purposes, desirable
Hospitality	Hotels, bartenders, hotel manager, entertaining, hospitality, bartend, guests, entertain, hotel, hospitality industry, hospitable, revenue management, entertainment, guest, bartender, entertainments, bartending
Impartiality	Impartially, objectivity, objective, objects, object, objectives, impartial, impartiality
Infuence	Infuentially, persuasively, infuencer, infuences, infuencing, infuence, persuasive, infuential, persuasion, persuasiveness
Initiative	Initiator, initiative, initial, initiatives, initially
Integrity	Confdentiality, trusted, integrated, credibly, trust, accuracies, trusts, integrity, data integrity, trusting, credibility, confdential, honesty, confdentialities, accuracy, confdentially, credible
Interpersonal communication	people skills, public speaking, interpersonal communication

Table 13 (continued)

Soft skill	Alternative labels
Kindness	Warming, kindness, kinds, warms, patience, warm, compassion, kindly, warmly, kind
Leadership	Leadership qualities, leadership skills, facilitate, facilitates, leadership, leaderships, leadership development, facilitating, business acumen, leadership position, facilitation, succession planning
Mentoring	Onboarded, onboard, mentored, mentor, mentoring, onboarding, mentors, onboardings, mentorship, management development
Motivated	Motivating, motivate, motivation, motivated
Negotiation	Negotiable, negotiation skills, arbitral, procuring, procured, arbitrations, procurement, mediate, mediating, negotiations, arbitrate, negotiating, negotiate, mediator, procure, diplomacy, negotiator, mediation, arbitration, negotiation
Optimism	Optimist, positive attitude, optimistic, optimism, personality, person
Oral communication	Oral communication, orals, oralism, oral, orally
Passion	Passion, passionate, passions, passione, passionately
Persuasion	Persuasion, persuasively, persuade, persuading, persuades, persuader, persuasive, persuasiveness
Presentation	Presenting, audienceability, audiences, exhibition, present, audienc, presentations, presentation, exhibits, exhibi tions, audience, sales presentation, exhibit
Problem solving	Problem analysis, probleme, problem, troubleshooter, troubleshooting, decomposition, troubleshoots, problem solving, troubleshoot, problems, problem management
Self-confdence	Self-esteem, confdently, confdent, confdant, self-confdence, self-confdent, confdence
Self-organized	Self-organized
Social skills	Assertiveness, social skills, assertively, interpersonal skills, assert, assertive, life skills
Speaking	Speak, speaking, speech, speaks, speeches
Strategic thinking	Strategizing, creatives, strategies, creatively, strategic thinking, strategize, strategy, creative, strategically, creativ ity, strategie, business planning, strategys, strategic planning, strategic
Teamwork	Team, teams, synergies, teaming, collaboratively, cooperate, teamworks, teamworker, collaboration, collaborative, cooperatively, collaborate, team player, cooperation, teamwork, synergis, cooperative, synergy, team building, teamworking, teamed, synergi
Time management	Punctuality, task, tasked, tasking, punctually, time management, punctual, project planning, tasks
Trustworthy	Trustworthiness, trustworthy
Verbal communication	Verbal communication
Writing	Authoring, proofread, translated, publications, social media, record, recording, proofreading, journalism, gram mar, journaling, journals, spelling, authorities, spelled, records, spell, journal, publicity, public, translate, writers, write, writing, texting, writer, writes, author, publication, recorded, grammarability, proofreader, spells, translation, proofreads, white papers, grammars, texts, translating, authority, text, writings
Written communication	Written communication

Table 14 Soft skill taxonomy in French

Soft skill	Alternative labels
Adaptation	Résilient, adaptée, adaptation, résiliente, adapter, résilience
Analyse (analytical)	Analyser, analyses, mathématiques, synthése, mathématique, analyse, synthése, synthéses
Apprentissage (ability to learn)	Apprentissages, changer, autoform, changement, changements, apprendrez, autoformation, appren dra, apprentissage, apprendre, autoformer
Autonomie (autonomy)	Autonome, autonomes, autonomique, autonomie
Commercial	Commerciale, commercant, courtiers, commerciaux, commercial, courtier, commercante, commerce
Communication	Communiquer, animer, communiquant, communiquez, animation, communication orales, commu nication, communication orale, communications, communication verbale, animations, dialoguer, dialogues, dialogue, oral, orales, orale
Conception (conceptual)	Idée, créationnisme, création, concepts, idées, créations, conception, concept
Confance (confdence)	Confances, coopérateurs, sujet, coopérer, optimiste, sujets, confance, coopération, optimisme, crédibilité, optima
Coordination	Coordinateur, coordination, coordinatrice,
Courage	Courages, endurant, endurance, courage, courageux,
Créativité (creativity)	Imaginatif, personnalité, créatifs, créativité, efcace, leadership, innovant, designer, designs, créa tives, innovation, innovants, leaderships, personnalités, efcacité, créative, artistiques, design, imagination, créatif, efcaces, imaginer, artistique, humour, artiste
Curiosité (curiosity)	Curiosité
Décision (decision)	Décisif, risque, cadrage, décision, décidabilité, décider, risques, décide, décidé, cadrages, risqué, décisions
Écoute active (active listening)	Ecoute activement, écoute active, reformuler, reformulation, ecoute active
Écrit (writing)	Écrits, écrit, écrite
Empathie (empathy)	Compréhensions, compréhension, empathe, empathique, assertivité, empathie, empath, assertif, sourires, assertive, empathies, compréhensif, sourire
Enthousiasme (enthousiasm)	Joie, enthousiast, enthousiasmant, enthousiasme, enthousiasmé, enthousiastes, enthousiaste, enthousiasmes
Équipe (teamwork)	Équipe, collaborateur, equipe, collaboration, equipes, équipes, collaborateurs
Esprit critique (critical thinking)	Jugerez, convictions, jugement, raison, logiques, juger, raisonnable, logique, raisonnement, esprit critique, conviction, raisons
Éthique (ethic)	Devoirs, dignité, normés, déontologie, éthique, normes, norme, éthiques, devoir, confdential, valeur, valeurs, politique, confdentialit, déontologique, confdentialité, déontologiques, objectivité
Flexibilité (fexibility)	Souplesse, souplesses, fexibilité
Gestion du stress (stress management)	Gestion du stress
Gestion du temps (time management)	Gestion de temps

Hospitalité (hospitality) Accueillir, hospitalité, accueillant, accueil
Initiative Initiative, initiatives, initiateurs
Leadership Leader, chéf, chef, chefs, lead, leads
Manager Manager, manage, gestionnaires, management, gestionnaire
Motivation Motivant, énergique, motivé, énergies, énergie, motivation
Négociation (negotiation) Négociation, techniques de vente, négocier, diplomate, techniques de ventes, diplomatie, négocia teur, diplomates, négociations
Pédagogie (teaching skills) Pédagogie, intégrant, pédagogiques, psychologique, enseigne, intégration, pédagogue, pédagogique, enseignes, orienter, intégrer, pédagogisme,enseignement, pédagogues, enseignant, orientation, orienté
Patience Patiente, patients, patience, attentes, attente, attention, attentif, patient Persuasion Persuadé, preuves, négocier, persuasion, arguments, persuasif, persuadant, persuader, négociateur, éloquent, argumenter, preuve, éloquence, persuasive, argumentation, négociation
Planifcation Calendriers, optimisation, gestion de projets, gestion de projet, budgets, prévisionnels, aléas, project, budget, calendrier, aléa, budgétés, projects, stratégique, plan, plane, prévisionnelle, stratégie, plani fcation, optimiser, marchandises, optimisations, planning, planifcateur, ressources, stratégiques, prévisionnel, projection, plannings, planifcations
Prise de décision (decision making) Prise de décision
Proposition Propositions, proposition

Soft skill Alternative labels

Table 14 (continued)

Résolution de probléme (problem solving) Problémes, problem, résolution, résolution de problémes, problématique, résolument, résolutions, problématiques, résolution de probléme, probléme
Résultat (results) Résultats, résultat
Relationnel Relationnelles, relationnel, relationnelle,
Respect Respect, respectant, humilité, respecter, politesse, conscience,
Responsabilité (responsability) Responsabilités, responsabilites, responsable, responsabilité, responsabilities, responsabilite Rigueur (riguor) Rigueureux, rigoureuses, rigueure, rigoureux, rigueur, rigoureuse Stress Pression, stresse, pressions, stressant, stress

Table 15 List of soft skills

	Soft skills extracted from Dbpedia	Soft skills added from prior work
English	Time_management, ethical, fexible, independ ent, analytical_skills, creative, hospital ity, fexibility, communication, decisions, coaching, commitment, coordinate, detail, goal, infuence, initiate, interpersonal, oral, passion, strategic, speaking, teamwork	Accountability, active listening, adaptive, argu mentation, conceptual, confict management, critical thinking, curiosity, diverse, eagerness, emotional intelligence, enthusiasm, impartial ity, integrity, kindness,mentoring, motivated, presentation, leadership, problem solving, self-confdence, optimism, self-organized, social skills, trustworthy, writing communic tion
French	Manager, proposition, négociation, apprendre, commercial,esprit_critique, rigoureux,équipe, persuasion,autonome, enthousiasme, patience, éthique, résultat, relations,prise_de_ décision,écrit, fexibilité, responsable	Analyse, adaptation, conception, confance, coordination, courage, curiosité, écoute active, empathie, gestion du stress, gestion du temps, hospitalité, leadership, pédagogie, planifcation, résolution de probléme

References

Altszyler E, Sigman M, Ribeiro S, Slezak DF (2017) Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database. Conscious Cogn 56:178–187. https://

doi.org/10.1016/j.concog.2017.09.004
Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (2007) DBpedia: a nucleus for a web of open data. In: The semantic web, Lecture notes in computer science. Springer, Berlin, pp 722–735, https://doi.org/10.1007/978-3-540-76298-0_52
Balcar J (2014) Soft skills and their wage returns: overview of empirical literature. Rev Econ Perspect 14(1):3–15. https://doi. org/10.2478/revecp-2014-0001
Bastian M, Hayes M, Vaughan W, Shah S, Skomoroch P, Kim H, Uryasev S, Lloyd C (2014) LinkedIn skills: large-scale topic extraction and inference. In: Proceedings of the 8th ACM con ference on recommender systems, ACM, New York, NY, USA, RecSys'14, pp 1–8. https://doi.org/10.1145/2645710.2645729
Benz D, Hotho A, Stumme G, Stützer S (2010) Semantics made by you and me: self-emerging ontologies can capture the diversity of shared knowledge. In: Proceedings of the 2nd web science confer ence (WebSci10)
Blake R, Gutierrez O (2011) A semantic analysis approach for assess ing professionalism using free-form text entered online. Com put Hum Behav 27(6):2249–2262. https://doi.org/10.1016/j. chb.2011.07.004
Boldi P, Monti C (2016) Cleansing Wikipedia categories using centrality. ACM Press, New York, pp 969–974. https://doi. org/10.1145/2872518.2891111
Brooks NG, Greer TH, Morris SA (2018) Information systems secu rity job advertisement analysis Skills review and implications for information systems curriculum. J Educ Bus 93(5):213–221. https ://doi.org/10.1080/08832323.2018.1446893
Calanca F, Sayfullina L, Minkus L, Wagner C, Malmi E (2018) Responsible team players wanted an analysis of soft skill require ments in job advertisements. arXiv:181007781
Cornali F (2018) Training and developing soft skills in higher edu cation. In: 4th international conference on higher education advances (HEAD'18), Editorial Universitat Politecnica de Valen cia, pp 961–967
Daneva M, Wang C, Hoener P (2017) What the job market wants from requirements engineers? An empirical analysis of online job ads from the Netherlands. In: 2017 ACM/IEEE international sym posium on empirical software engineering and measurement (ESEM), pp 448–453. https://doi.org/10.1109/ESEM.2017.60
De Smedt J, le Vrang M, Papantoniou A (2015) ESCO: towards a semantic web for the European labor market. In: LDOW@ WWW Fellbaum C (1998) A semantic network of english: the mother of all wordnets. Comput Human 32(2):209–220. https://doi. org/10.1023/A:1001181927857
Fernandez-Sanz L (2010) Analysis of non technical skills for ICT profles. In: 5th Iberian conference on information systems and technologies, pp 1–5
Florea R, Stray V (2018) Software tester, we want to hire you an analy sis of the demand for soft skills. In: Garbajosa J, Wang X, Aguiar A (eds) Agile processes in software engineering and extreme programming. Lecture notes in business information processing. Springer, Berlin, pp 54–67

Gabrilovich E, Markovitch S (2007) Computing semantic related ness using Wikipedia-based explicit semantic analysis. In: Pro ceedings of the 20th international joint conference on artifcal intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, IJCAI'07, pp 1606–1611. http://dl.acm.org/citation. cfm?id=1625275.1625535. Event-place: Hyderabad, India
Gardiner A, Aasheim C, Rutner P, Williams S (2018) Skill require ments in big data a content analysis of job advertisements. J Comput Inf Syst 58(4):374–384. https://doi.org/10.1080/08874 417.2017.1289354
Grineva M, Grinev M, Lizorkin D (2009) Extracting key terms from noisy and multitheme documents. In: Proceedings of the 18th international conference on World wide web—WWW'09, ACM Press, Madrid, Spain, p 661. https://doi.org/10.1145/15267 09.1526798
Heymann P, Garcia-Molina H (2006) Collaborative creation of commu nal hierarchical taxonomies in social tagging systems. Technical report, Stanford
Hillmer G, Fink C, Foradori M, Gall M, Kilian D, Sparer W (2007) Social and soft skills training concept in engineering education. Innovations 2007: world innovations in engineering education and research, International network for engineering education and research, pp 355–366
Hurrell SA (2016) Rethinking the soft skills defcit blame game: employers, skills withdrawal and the reporting of soft skills gaps. Hum Relat 69(3):605–628
Javed F, Hoang P, Mahoney T, McNair M (2017) Large-scale occu pational skills normalization for online recruitment. https://aaai. org/ocs/index.php/IAAI/IAAI17/paper/view/14922
Joseph D, Ang S, Chang RHL, Slaughter SA (2010) Practical intel ligence in IT: assessing soft skills of IT professionals. Commun ACM 53(2):149–154. https://doi.org/10.1145/1646353.1646391
Kautz T, Heckman JJ, Diris R, Weel Bt, Borghans L (2014) Foster ing and measuring skills: improving cognitive and non-cogni tive skills to promote lifetime success. Working Paper 20749, National Bureau of Economic Research. https://doi.org/10.3386/ w20749
Kivimäki I, Panchenko A, Dessy A, Verdegem D, Francq P, Bersini H, Saerens M (2013) A graph-based approach to skill extraction from text. In: Proceedings of TextGraphs-8 graph-based methods for natural language processing, pp 79–87
Lacerenza CN, Marlow SL, Tannenbaum SI, Salas E (2018) Team development interventions: evidence-based approaches for improving teamwork. Am Psychol 73(4):517
Lai S, Liu K, Xu L, Zhao J (2015) How to generate a good word embedding? arXiv:150705523
Maitra S, Gopalram K (2016) Ethics and soft skill assessment tool for program outcome attainment: a case study. In: 2016 IEEE 4th international conference on MOOCs, innovation and technol ogy in education (MITE), pp 317–324. https://doi.org/10.1109/ MITE.2016.069
Malherbe E, Aufaure MA (2016) Bridge the terminology gap between recruiters and candidates: a multilingual skills base built from social media and linked data. In: 2016 IEEE/ACM international conference on advances in social networks analysis and min ing (ASONAM), pp 583–590. https://doi.org/10.1109/ASONA M.2016.7752295
Manku GS, Jain A, Das Sarma A (2007) Detecting near-duplicates for web crawling. In: Proceedings of the 16th international conference

on World Wide Web, ACM, New York, NY, USA, WWW'07, pp 141–150. https://doi.org/10.1145/1242572.1242592
Manpower (2017) Talent shortage 2016–2017 | ManpowerGroup. http://manpowergroup.com/talent-shortage-2016
Matturro G (2013) Soft skills in software engineering A study of

its demand by software companies in Uruguay. In: 2013 6th inter national workshop on cooperative and human aspects of software engineering (CHASE), pp 133–136. https://doi.org/10.1109/ CHASE.2013.6614749

Matturro G, Raschetti F, Fontàn C (2015) Soft skills in software devel opment teams a survey of the points of view of team leaders and team members. In: 2015 IEEE/ACM 8th international workshop on cooperative and human aspects of software engineering, pp 101–104. https://doi.org/10.1109/CHASE.2015.30

Mikolov T, Chen K, Corrado G, Dean J (2013) Efcient estimation of word representations in vector space. arXiv:13013781

Monasor MJ, Noll J, Vizcaìno A, Piattini M, Beecham S (2014) Walk before you run: using heuristic evaluation to assess a training tool prototype. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering, ACM, New York, NY, USA, EASE'14, pp 41:1–41:10. https://doi.org/10.1145/2601248.2601271. Event-place: London, England, United Kingdom

Nolinske T, Millis B (1999) Cooperative learning as an approach to pedagogy. Am J Occup Ther 53(1):31–40

Rehurek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: In Proceedings of the LREC 2010 work shop on new challenges for NLP Frameworks, Citeseer

Roget PM (1911) Roget's thesaurus of English words and phrases. TY Crowell Company, Philadelphia

Smith SP, Hickmott D, Bille R, Burd E, Southgate E, Stephens L (2015) Improving undergraduate soft skills using m-learning and serious games. In: 2015 IEEE international conference on teaching, assessment, and learning for engineering (TALE), pp 230–235. https://doi.org/10.1109/TALE.2015.7386049

Wu F, Weld DS (2010) Open information extraction using Wikipedia. In: Proceedings of the 48th annual meeting of the association for computational linguistics, Association for Computational Linguis tics, pp 118–127

Yanaze LKH, Lopes RdD (2014) Transversal competencies of electri cal and computing engineers considering market demand. In: 2014 IEEE frontiers in education conference (FIE) proceedings, pp 1–4. https://doi.org/10.1109/FIE.2014.7044169

Zaharim A, Ahmad I, Yusof YM, Omar MZ, Basri H (2012) Evaluat ing the soft skills performed by applicants of malaysian engineers. Procedia Soc Behav Sci 60:522–528. https://doi.org/10.1016/j. sbspro.2012.09.417

Zhao M, Javed F, Jacob F, McNair M (2015) SKILL: a system for skill identification and normalization. In: Proceedings of the 29th AAAI conference on artificial intelligence, AAAI Press, AAAI'15, pp 4012–4017. http://dl.acm.org/citation.cfm?id=2888116.2888273. Event-place: Austin, Texas