The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2046-9012.htm

Valuing prior learning

Designing an ICT artifact to assess professional competences through text mining

Florian Fahrenbach
Vienna University of Economics and Business, Vienna, Austria

Kate Revoredo
Graduate Program in Informatics, UFRJ, Rio de Janeiro, Brazil, and

Flavia Maria Santoro
Department of Computer Science, UERJ, Rio de Janeiro, Brazil

Received 3 May 2019
Revised 5 August 2019
Accepted 8 August 2019

Abstract

Purpose – This paper aims to introduce an information and communication technology (ICT) artifact that uses text mining to support the innovative and standardized assessment of professional competences within the validation of prior learning (VPL). Assessment means comparing identified and documented professional competences against a standard or reference point. The designed artifact is evaluated by matching a set of curriculum vitae (CV) scraped from LinkedIn against a comprehensive model of professional competence.

Design/methodology/approach – A design science approach informed the development and evaluation of the ICT artifact presented in this paper.

Findings – A proof of concept shows that the ICT artifact can support assessors within the validation of prior learning procedure. Rather the output of such an ICT artifact can be used to structure documentation in the validation process.

Research limitations/implications – Evaluating the artifact shows that ICT support to assess documented learning outcomes is a promising endeavor but remains a challenge. Further research should work on standardized ways to document professional competences, ICT artifacts capture the semantic content of documents, and refine ontologies of theoretical models of professional competences.

Practical implications – Text mining methods to assess professional competences rely on large bodies of textual data, and thus a thoroughly built and large portfolio is necessary as input for this ICT artifact.

Originality/value – Following the recent call of European policymakers to develop standardized and ICT-based approaches for the assessment of professional competences, an ICT artifact that supports the automatized assessment of professional competences within the validation of prior learning is designed and evaluated.

Keywords Assessment, Proof of concept, Text mining, Design science, Professional competences, Validation of prior learning

Paper type Technical paper

The first author was partially funded by the EU H2020 program under MSCA-RISE agreement 645751 (RISE BPM) and FFG Austrian Research Promotion Agency (project number: 866270). The second author was partially funded by Österreichische Akademie der Wissenschaften. We are very grateful for Marcos José Moura's help with the data analysis.

European Journal of Training and Development
Vol. 44 No. 2/3, 2020
pp. 209-235
Emerald Publishing Limited
2046-9012
DOI 10.1108/EJTD-05-2019-0070

1. Introduction

The validation of prior learning (VPL) is the process of "assessing and recognizing a wide range of skills and competences which people develop through their lives and in different contexts, for example through education, work and leisure activities" (Bjørnåvold, 2000, p. 216). The European Union supports the validation of prior learning by introducing the Lifelong Learning Strategy (EU, 2006), the European Qualification Framework (EQF) (EU, 2017) and the recommendations on the validation of prior learning (EU, 2012). Having viable and efficient approaches for the assessment of professional competences within the validation of prior learning could help to lower the number of unemployed, increase labor market mobility and facilitate social cohesion within the European Union.

While policy frameworks for the assessment of professional competences within VPL are in place in most of the European countries, providing *specific* methods and approaches for the assessment proves to be a challenge for policy-making (EU, 2012, 2017) and scientific research (Bohne *et al.*, 2017; Brockmann *et al.*, 2009). In VPL, assessment is the phase in which a person's learning outcomes (i.e. professional competences) are "compared against specific reference points and/or standards" (Cedefop, 2015, p. 18). A standard or reference point is a document that describes which learning outcomes people have to obtain to be qualified on a certain level, e.g. a document that shows what a student should be able to do after finishing training or education.

As we lack innovative approaches to support the assessment of professional competences (Cedefop, p. 20), the European Union calls to develop standardized and information and communication technology (ICT)-based approaches for the assessment of professional competences within the VPL (Cedefop, 2017). Currently, the VPL procedures remain a labor-intensive manual task. The assessment of competences within the VPL has to be done by qualified assessors, who need to be trained to guide individuals through the validation process (Diedrich, 2013). Consequently, it takes weeks or even months to conduct a validation procedure before individuals can show their qualifications to employers. Our research question is: *How to automatize the assessment of professional competences within the VPL?* Our research objective is to introduce an ICT artifact that supports the assessment of professional competences within the validation of prior learning by matching a documentation of professional competences with a given standard, a predefined theoretical model of professional competences.

In this paper, we draw on a design science research (Hevner *et al.*, 2018; Hevner *et al.*, 2004; Gregor and Hevner, 2013) approach to develop an artifact (i.e. an algorithm) that uses text mining to match a repository of curriculum vitae (CV) with a given theoretical model of professional competences. The designed artifact allows us to compare each CV individually with the predefined theoretical model. We refer to the activities of this artifact as "competence mining." This proof of concept shows that such an artifact may support the assessment of professional competences within the VPL by assigning documented competences to a standard or reference point. Practically, we introduce an artifact that can be applied to automatically match textual data (e.g. portfolios or CV) to a standard or reference point (e.g. a theoretical model or qualification standard according to EQF). Based on previously *identified* (i.e. made explicit or spoken out) and *documented* (i.e. written down) evidence, the artifact is able to *assess* professional competences (i.e. compare them against a standard or reference point). This artifact may help assessors within the VPL procedure as it can give a hint about the candidate's competence profile, thus making the VPL procedure less time-consuming and tedious (Han and Lee, 2016). Theoretically, we add to the debate revolving the standardization of VPL procedures. We find that standardizing VPL to a

certain extent may diminish the negative effects that the VPL procedures can bring about (Diedrich, 2013).

The remainder of the paper is structured as follows. Section 2 consists of a literature review that introduces related theoretical and practical approaches for the ICT-supported assessment of professional competences. Section 3 more closely describes the design science approach (Hevner *et al.*, 2018; Hevner *et al.*, 2004; Gregor and Hevner, 2013). In section 4, we describe the designed artifact. In section 5, we present a proof of concept of the artifact by matching a repository of CV to a theoretical model of professional competence. In section 6, we discuss the findings and show how they relate back to the research question and objectives. In section 7, we outline potential limitations of the artifact, point out further research endeavors and conclude.

2. Literature review

2.1 The assessment of professional competences within the validation of prior learning

A person acquires professional competences mainly through experiences and learning that can be formal, non-formal or informal. Formal learning, occurring in an "organized and structured context (formal education, in-company training, etc.) is designated as learning" (Bjørnåvold, 2000, p. 204) and comparably easy to assess because licenses or degrees are awarded that explicitly specify the learning outcomes. Differently, non-formal and informal learning outcomes are partly tacit (Polanyi, 1966). Non-formal learning, "planned activities that are not explicitly designated as learning, but which contain an important learning element" (Bjørnåvold, 2000, p. 204), is considerably harder to assess as documentation may have different grades of trustworthiness. Informal learning or experiential learning that can "be understood as accidental learning" (Bjørnåvold, 2000, p. 204) is even more situated in the environment (Lave and Wenger, 2011) and occurs in day-to-day activities related to work, family or leisure, including language learning or parenting and more challenging to assess. For example, what a person is able to do is comparably easy to assess based on a university degree, comparably harder to assess based on certifications of massive open online courses or courses from tertiary education and even harder from learning that the person is not aware of.

To validate formal, non-formal and informal learning, tacit knowledge and competences must be made explicit and documented in a social process (Nonaka, 1994; Nonaka *et al.*, 2000) – the VPL. Consequently, the VPL usually consists of four phases: identification, documentation, assessment and recognition of prior learning (Cedefop, 2015). First, a qualified assessor supports individuals in identifying previously acquired knowledge, skills and competences from different contexts using reflection (Schön, 1983) and dialogue (Bohm, 2012) with the aim that individuals become increasingly aware of prior achievements. The "discovery and increased awareness of own capabilities is a valuable outcome of the process" (Cedefop, 2015, p. 18). Second, documenting learning outcomes or stocktaking requires people to provide evidence through "building" of a portfolio that tends to include a CV and a career history of the individual, with documents and/or work samples that attest to their learning achievements (Cedefop, 2015, p. 18). Individuals have to approach authorities, peers or former supervisors who are willing to provide evidence of the identified learning (e.g. certificates, licenses, proof of voluntary work). Third, assessment is the phase in which "an individual's learning outcomes are compared against specific reference points and/or standards" (Cedefop, 2015, p. 18). Standards or reference points (Bohlinger, 2017) are set by companies or professional associations and assessment methods range from written, oral or practical tests/examinations to portfolios. Fourth, recognition is the certification of previously assessed learning through the award of a qualification by an authority

(Cedefop, 2015, p. 18). The *identification* and *documentation* of professional competences is crucial for their subsequent assessment (Annen, 2013). Starting from the premise that professional competences have been previously *identified* and *documented*, this paper only deals with the assessment phase in the VPL.

We identify two – well documented – main challenges in the assessment of professional competences. First, a validity challenge (Stenlund, 2010): does the artifact assess what it promises to do? A person documenting competence in business and management subjects should show these competences in the relevant dimensions of the assessment. We propose a comprehensive model of professional competences as a standard or point of reference in Section 4. This model is – based on the Occupational Information Network (O*Net; Peterson *et al.*, 2001) – able to assess professional competences in all relevant domains (and is not limited to a certain profession). The second challenge is to determine the level of acquired competence by assigning a numerical value to the content dimension (Anderson and Krathwohl, 2001; Dreyfus and Dreyfus, 1987). In other words, how can we determine the level of competence, based on a thorough documentation of prior learning? This challenge refers to determining whether a person is a beginner, intermediate, advanced or expert in a certain field. We refer to established taxonomies of competence development and descriptions of the complexity of learning outcomes (Anderson and Krathwohl, 2001; Bloom *et al.*, 1956; Dreyfus and Dreyfus, 1987; Krathwohl, 2002) to determine the documented level of competence.

2.2 The assessment of competences using text mining

Extracting professional competences via content analysis from documents such as job advertisements or CV has a long tradition. We can observe this within the academic literature but also more practical fields[1]. We can distinguish between approaches that depart from the occupational side and use job advertisements to examine competence requirements for a specific occupation (Müller *et al.*, 2014; Gallivan *et al.*, 2004; Aken *et al.*, 2010; Todd *et al.*, 1995) and approaches that depart from the analysis of individual CV (Darabi *et al.*, 2018; Gorbacheva *et al.*, 2015; Han and Lee, 2016; Lichtnow *et al.*, 2008; Patel *et al.*, 2017; Valdez-Almada *et al.*, 2017). While extracting competence *requirements* must depart from the occupational side, the assessment of professional competences must begin with the individual CV.

In recent years, text mining is often used to assess large amount of textual data. Text mining is a form of data mining (Romero and Ventura, 2010; Sachin and Vijay, 2012) and is often used in educational settings. In this context, it is referred to as educational data mining (Romero and Ventura, 2010). It comprises a set of methods to analyze unstructured data such as texts or narrations. Text mining techniques "[. . .] allow to automatically extract implicit, previously unknown, and potentially useful knowledge from large amounts of unstructured textual data in a scalable and repeatable way" (Debortoli *et al.*, 2016, p. 556). In this regard, text mining helps to foster knowledge discovery because very large amounts of data can be analyzed simultaneously (Kobayashi *et al.*, 2018). Text mining usually follows the common steps of other data mining techniques, namely, pre-processing, data mining and post-processing (Debortoli *et al.*, 2016; Kobayashi *et al.*, 2018; Romero and Ventura, 2010).

Concerning the assessment of competence requirements, text mining was used in several studies. Darabi *et al.* (2018) use text mining to identify skills and qualifications which employers search for in engineering fields by comparing job postings to the O*Net. Debortoli *et al.* (2014) use latent semantic analysis (LDA) to develop a competency taxonomy of business intelligence and big data jobs based on job advertisements. Karakatsanis *et al.* (2017) use latent semantic indexing to match job postings on the Web with descriptors from

the O*Net. They aim at identifying the most in-demand occupations on the job market. Kobayashi *et al.* (2018) aim at introducing organizational researchers with the fundamental logic underpinning text mining and use topic modeling in a job analysis case study.

We consider a work as related if the approach uses text mining methods to assess individual CV. Table I summarizes these works. While there is a considerable amount of work in the field, the application of text mining procedures on large amount of CV remains, with notable exceptions (Darabi *et al.*, 2018; Gorbacheva *et al.*, 2015; Han and Lee, 2016; Lichtnow *et al.*, 2008; Patel *et al.*, 2017; Valdez-Almada *et al.*, 2017) scarce. These works aim at extracting competences in specific directions, such as engineering education (Darabi *et al.*, 2018), business process management (Gorbacheva *et al.*, 2015), construction work (Han and Lee, 2016), computer science (Lichtnow *et al.*, 2008), computer science and engineering majors (Patel *et al.*, 2017) and software engineering (Valdez-Almada *et al.*, 2017). Thus, extraction of competences is limited to a certain field. We address this limitation by designing an artifact which is, because of the comprehensiveness of its underlying model, able to assess individual competences of several professions.

3. Method

Our approach draws on a design science paradigm (Gregor and Hevner, 2013; Hevner *et al.*, 2004; Peffers *et al.*, 2007) to guide the development of the artifact. While the natural and social sciences aim to understand reality, "design science attempts to create things that serve human purposes" (Simon, 1996, p. 55). Design science comprises the creation (Section 4) and evaluation (Section 5) of an "innovative, purposeful artifact for a specified, currently unresolved problem domain" (Hevner *et al.*, 2004, p. 82). With the artifacts utility as an ultimate goal in mind, it addresses research challenges through the "*building* and *evaluation* of artifacts designed to meet the identified [. . .] need" (Hevner *et al.*, 2004, pp. 79-80). An artifact is "a thing that has, or can be transformed into, a material existence as an artificially made object (e.g. model, instantiation) or process (e.g. method, software)" (Gregor and Hevner, 2013). The design science research process includes six steps: problem identification and motivation; definition of the

Approach	Field	Goal	Method
Darabi *et al.* (2018)	Engineering education	Identify skills and qualifications employers search for in stem fields by comparing job postings to the O*Net	Text mining (NLTK)
Gorbacheva *et al.* (2015)	Business process management	Offering a gender perspective on business process management competences	Text mining (latent semantic analysis)
Han and Lee (2016)	Human resource allocation	Analyze CV to allocate positions in construction projects	Text mining (KNIME)
Lichtnow *et al.* (2008)	Knowledge management	Analyze CV to identify areas of expertise and build yellow pages	Text mining
Patel *et al.* (2017)	Big data computing	CaPaR: introducing a recommendation system for career paths; using text mining to scan resumes and profiles to identify key skills	Text mining
Valdez-Almada *et al.* (2017)	Software engineering	Analyzing CV to identify knowledge profiles for software engineering positions	Text mining (Stanford CoreNLP)

Table I.
Related work in the field

objectives for a solution; design and development; demonstration; evaluation; and communication (Peffers *et al.*, 2007, p. 46). Methodological rigor is achieved by "appropriately applying existing foundations and methodologies" (Hevner *et al.*, 2004, p. 80) in design science. Subsequently, we describe the designed artifact and evaluate the artifact on a set of CV.

4. Artifact description

The foundation of the artifact to be applied is a comprehensive model of professional competences (see Table AI in Appendix). It merges the normative European competence perspective (Cheetham and Chivers, 1996; Le Deist and Winterton, 2005; Mulder *et al.*, 2007) which focuses on *what a person is able to do* with the descriptive content model of the O*Net that provides a comprehensive and detailed taxonomy of occupational descriptors (Peterson *et al.*, 2001). The underlying model contains 4 general competence dimensions and 32 sub-competences (Fahrenbach *et al.*, 2019). To create the dictionary (see Table AII in Appendix), we characterized each of the 32 sub-competences of the underlying model with the descriptors in version 22.2 of the O*Net content model[2]. In total, the dictionary contains 1,255 descriptors for the 32 sub-competences (average: 39.2 descriptors per sub-competence; minimum: 8; maximum: 213). In case new competences or skills arise (e.g. programming languages), they can be updated in the dictionary.

The designed artifact for the assessment of professional competences receives a documentation of learning such as a repository of CV (in principle, it could receive any textual documentation of learning outcomes) and a dictionary of competences (which serves as a standard or reference point for the assessment) and returns a match between them. It has two main activities. In the first activity, the set of CV is processed to generate a bag of word representation for each of the CV. Natural Language Processing (Bird *et al.*, 2009) is used for tokenization (i.e. splitting up sentences into words), removal of stop words (i.e. removal of words that do not meaningfully contribute such as "and" or "or"), lemmatization (i.e. removal on inflectional word endings and return of the dictionary form) and the subsequent collection of relevant words. The second activity receives the bag of relevant word representations for each CV and the dictionary of competences and performs a match of these sets. In this regard, we base our analysis on the classical vector space model (Salton *et al.*, 1975) in which documents (such as CV or standards) are represented as vectors of terms. By matching, we mean that the artifact creates a term–document matrix. A collection of documents is then represented in such a term–document matrix which contains the number of occurrences each term appears in each document (Manning *et al.*, 2008). In other words, the artifact counts the number of coincidence words for each CV, each competence and its sub-competences. The counts are returned as the output of the designed artifact. With the number of matches, it is possible to conduct further analysis and we propose an optional third step. In this step, a thorough statistical analysis can be conducted. It is, for example, possible to rank CV based on a particular competence or to provide an overall description of a CV among all competences. An overview of the designed artifact is given in Figure 1. In sum, we provided an overview on the designed artifact in this section.

5. Evaluation

In this section, we evaluate the designed artifact on a set of CV gathered from LinkedIn. Section 5.1 describes the data collection and the data set used. Section 5.2 outlines the processing of CV, including data preparation and pre-processing. Section 5.3 outlines the application of the designed artifact. In Section 5.4, we analyze the results of applying the designed artifact on an aggregated level.

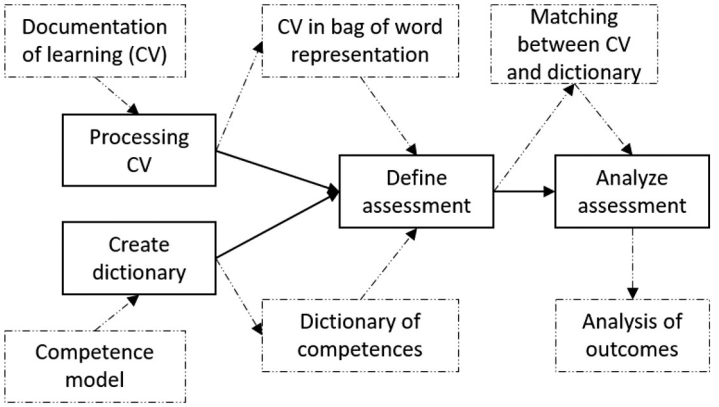


Figure 1.
Activities of the designed artifact

5.1 Data collection

We used an openly available data set[3] from a blog post with 1,445 URLs to CV from LinkedIn as primary data source. LinkedIn is a social media platform on which users can create an online portfolio and headhunters or companies can search through these for recruiting purposes (Bastian *et al.*, 2014). LinkedIn is increasingly used for employee selection and hiring (Roulin and Levashina, 2018) but also for research purposes. Although its use is hotly debated, first studies indicate good psychometric properties and validity of information reported on LinkedIn (Roulin and Levashina, 2018). We decided to use CV from this data set, as it is openly available on the Web and we could avoid possible biases, such as selection bias in data collection (Heckman, 1979). To scrape LinkedIn CV, we used an openly available webscraper[4]. The original data set provides reference to 1,488 individuals; however, there are only 1,445 URL to LinkedIn CV reported, of which two entries were duplicates. Furthermore, we could not access 8 URLs from the 1,443 links because of deleted accounts or updated privacy settings. In total, we scraped 1,435 CV from the original data set. All CV were stored in JSON arrays and saved on local hard drives. The scraping of CV took place between August 10, 2018 and August 27, 2018.

Each scraped CV is organized in a similar way, consisting of six general categories. "General information" includes the name, company, school and a short description or statement of purpose, "jobs" include the names of companies, job titles and job descriptions, "schools" include name of schools and degrees, "details" include personal websites and social media accounts, "skills" include self-assessed skills and endorsements from externals and "allskills" include a list of all reported skills separated by a comma (self-assessed and endorsed). A demographic overview of the scraped profiles is given in Table AIII in the Appendix. The demographic characteristics point at a skewed distribution with regard to gender, ethnicity and place of education. All of the 1,435 individuals work in 192 venture capital firms, either as an associate, principal or partner. Venture capital organizations "raise money from individuals and institutions for investment in early-stage businesses that offer high potential but high risk" (Sahlman, 1990, p. 473). According to literature, successful CEOs in venture capital firms rely on a set of characteristics, which can be summarized by two factors. Factor 1 is described by "general management ability" and factor 2 by "communication and interpersonal skills with a focus on execution and resoluteness" (Kaplan *et al.*, 2012, p. 1005).

5.2 Processing curriculum vitae

The 1,435 scraped CV from LinkedIn, stored in the JSON file, served as input for the designed artifact, which has as first activity the processing of CV. We used Jupyter to convert the CV in JSON format to python objects for further processing and text mining. For preparing and pre-processing of the CV, we followed common text mining procedures (Debortoli *et al.*, 2016; Kobayashi *et al.*, 2018). For the text mining itself, we relied on the python library nltk (Bird *et al.*, 2009). We identified the relevant stop words in English from this library (Rajaraman and Ullman, 2011) as well. For natural language pre-processing, we lemmatized the words (Debortoli *et al.*, 2016). To do so, we imported the Word Net Lemmatizer library to extract the non-inflected (canonical or lemma) form of each word (Miller, 1995). We also applied tokenization, which allows to split up documents into sentences and sentences into words (Debortoli *et al.*, 2016). In sum, we followed common text mining procedures to remove words that create noise in the data set.

5.3 Define assessment

As outlined above, the assessment of competences entails to compare previously *identified* and *documented* competences against a standard or point of reference (Cedefop, 2015; Bjørnåvold, 2000). To do so, the designed artifact counts occurrences of matching words between the repository of CV and the dictionary. The artifact evaluates each word in each of the CV against each word in the dictionary and saves the result in a vector. If there is a match between the CV and the dictionary, the result is stored as "1" in the vector, otherwise as "0." Based on these vectors, we summed up all matches per CV and sub-competence. This activity resulted in a data set with 1,435 rows, indicating the CV and 32 columns indicating the matches for each sub-competence. The subsequent analysis of the artifact is based on the already aggregated data on the level of 32 sub-competences. Using the LinkedIn URL and an ID, we can track each individual in the original and resulting data set.

5.4 Analyze assessment

Applying the artifact resulted in a data set with 67,522 matches between the 1,435 CV and the dictionary in total. The average number of matches per sub-competence is 2,110 (min: 18; max: 12,485; median: 926; SD: 2925). Table AIV in the Appendix shows the number of matches per sub-competence.

To get a better overview regarding which sub-competence matched frequently with the CV, the upper part of Figure 2 shows the ordered frequencies (y-axis) of matches per sub-competence dimension (x-axis). The upper part of Figure 2 also shows that CV matched to mostly four different sub-competences [MC9 (business management) accounted for 18.5 per cent, PC3 (suitability based on interests) for 13.1 per cent, DC1 (domain knowledge) for 12.1 per cent and MC5 (performing complex technical activities) for 9.3 per cent of all matches]. As a result, 4 out of 32 sub-competence dimensions account for 60 per cent of the matches.

There are several explanations for these results, which are outlined below. First, individuals working as venture capitalists seem to rely on a comparable set of competences, mainly related to business management (as indicated by the prevalence of MC9). In the dictionary, MC9 (business management) was described with terms such as "business; business and management; business administration; accounting; human resource management; HRM; material resource management; organizations; organization; sales; marketing; sales and marketing; economics; office information; enterprise resource planning; organizing systems; economics; administration and management; strategic planning; resource allocation; human resource modelling; resource allocation; [. . .]." We interpret the frequency of matches with MC9 as closely related to the factor "general management

ability." In this regard, our findings are in line with previous research (Kaplan *et al.*, 2012). Second, the large number of matches in PC3 (suitability based on interests) can be explained theoretically. The underlying theory of occupational interests by Holland (1997) defines, among others, "enterprising interests" which are described by entrepreneurial activities and interest in management. In this regard, the dictionary described PC3 with terms such as "entrepreneur; realistic; pragmatic; social; artistic; enterprise; convention; conventional; hands-on problems; investigation; investigate; problem-solving; thinking; design patterns; teaching; service; entrepreneurship; project management; leadership; business; risk taking; routines; procedures; [. . .]." Third, the large amount of matches in DC1 (domain knowledge) can be explained by the breadth and depth of the entry in the dictionary (213 descriptors). DC1 describes domain-specific knowledge and includes a wide variety of cross-occupational knowledge and school subjects such as "computers and electronics; engineering and technology; biology; psychology; arts and humanities; [. . .]," which explains the number of matches in this domain. Fourth, MC5 (performing complex technical activities) is strongly related to perform skilled activities in technical fields. MC5 is described in the dictionary with "technical activities; skilled activities; coordinated movements; movements; coordination; computers; computer; PC; software; hardware; tools; computer systems; programming; computer programming; data entry; process information; Coding; Code; functions; electronics; [. . .]." Individuals working as venture capitalists seem to have considerable technical experience (given the high number of matches in MC5). This finding can be explained by 27 per cent of individuals with an engineering degree in the original data set. The upper part of Figure 2 also indicates that other competence dimensions match considerably less. For SC8 (conflict management), the artifact returned only 18 matches (0.03 per cent).

The lower part of Figure 2 indicates the number of matches on the *x*-axis and the number of CV on the *y*-axis (also in Table AV of the Appendix). Figure 2 shows that a large amount of CV only match to a modest number with the dictionary (117 CV do not match at all, 262 CV show one to nine matches with the dictionary, and only a small number of CV show considerable matches with the dictionary). The average number of matches per CV is 47 (median: 37; SD: 43). This finding can be explained with the fact that many individuals provide only very few information about themselves on their LinkedIn CV (Gorbacheva *et al.*, 2015; Roulin and Levashina, 2018). However, the automatized assessment of professional competences relies on a large repository of documents and textual data (Han and Lee, 2016). These can be written and narrative statements of purpose, a detailed

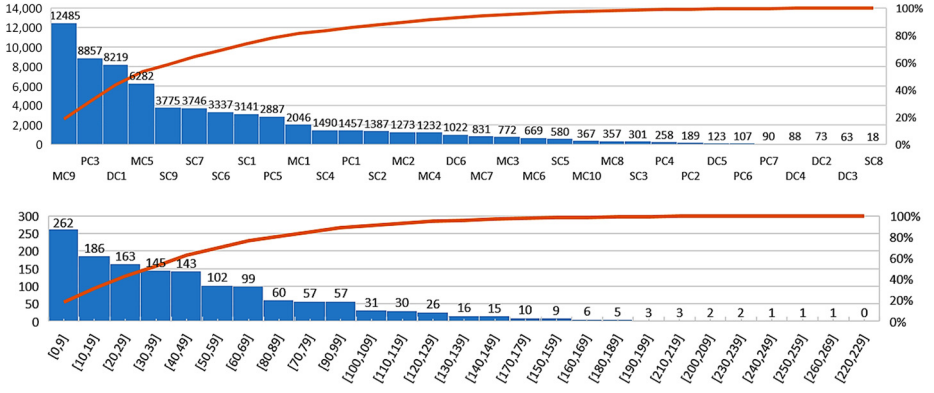


Figure 2.
Upper part shows the number of matches per sub-competence dimension. Lower part shows the number of matches per CV

description of previous work activities or any other textual document. Thus, the (very few) CV with the most matches have an extensive statement of purpose uploaded to their LinkedIn CV.

6. Interpretation and application of the artifact

As we set out to answer the research question *How to automatize the assessment of professional competences within the VPL²*, this section interprets the findings and outlines possible areas of application with two examples from the data set. We argue that a viable answer to the research question and objective can be the designed artifact. Starting from the premise that a person *identifies* his/her competences and *documents* them thoroughly in a (guided) self-assessment, the designed artifact is able to match the documents to a predefined standard (the comprehensive competence model).

While we analyzed results on the level of the whole data set in the last section, we take a look at two individual competence profiles in this section. On an individual level, the designed artifact results in a distinct competence profile, such as the green field in Figure 3(a)-(c). In Figure 3(b)-(d), the red line indicates a standard, against which the individual competence profile is assessed (in this case, the standard is of illustrative nature).

To demonstrate the application of the assessment for individual CV, and to assess the level of competence, we refer to common taxonomies which suggest six levels of competence (Anderson and Krathwohl, 2001; Dreyfus and Dreyfus, 1987) in which 1 represents a beginner and 6 represents an expert. To align the results to these taxonomies, we normalized the data set to a scale from 0 to 6 by using the following formula:

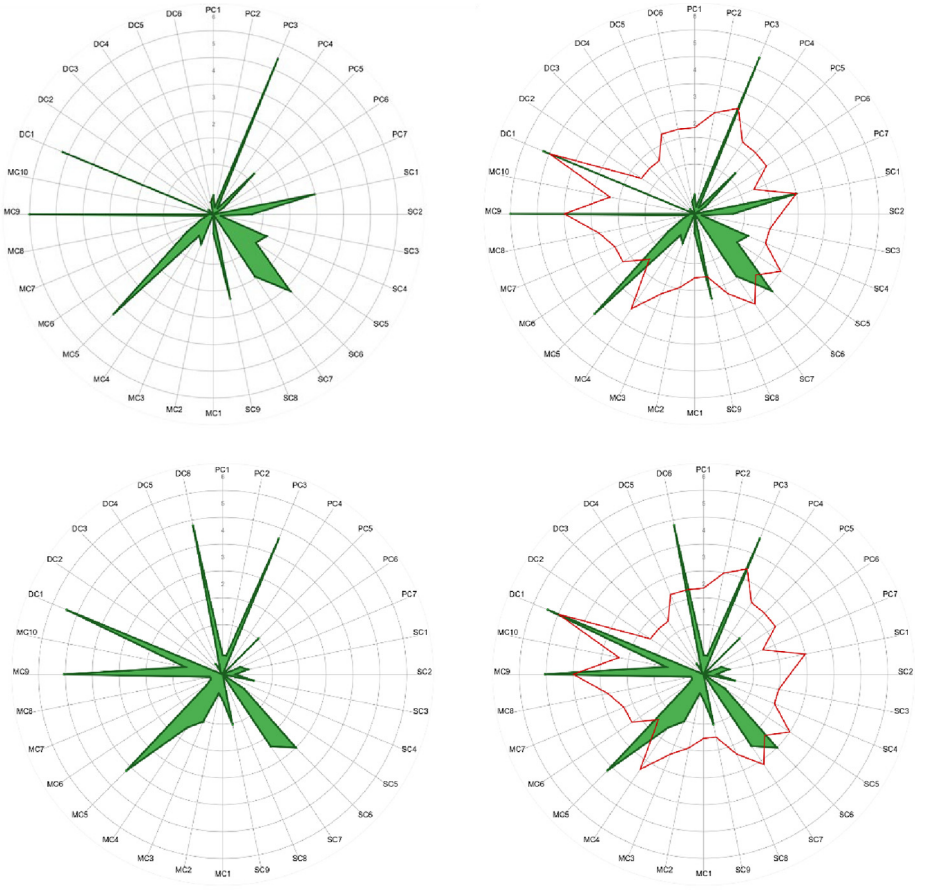
$$x\_new = \frac{x - x\_min}{x\_max - x\_min}$$

As the data in Figure 3 is normalized to a scale from 0 to 6, we have to know the overall number of matches to interpret the distribution of assessed professional competences. Figure 3 shows the analysis of two individuals with the most matches in the data set. Figure 3(a)-(b) shows an individual with 258 matches in total (min: 0; max: 36 matches). Figure 3(c)-(d) shows a different individual but the same standard as in Figure 3(b). The CV of this individual showed 261 matches in total (min: 0; max: 36 matches).

We introduce three different areas of application for the designed artifact. First, as outlined above, the artifact can be used for the assessment of competences when professional associations set standards for occupational fields on a certain level, such as within the EQF (EU, 2017). Second, the artifact can be used in organizations within human resources allocation or hiring decisions, when searching for a single best individual. For example, an organization defines competence requirements (Campion *et al.*, 2011) (see Figure 3, red lines) and an individual applies with a certain competence profile (see Figure 3, green field). Using the designed artifact, it is possible to select a single best individual for a given standard of competence to point at learning fields in which the individual has to acquire additional competences to fit to the organization's competence requirement. Third, within human resource development, organizations can use the artifact to assess the competences of their employees and tailor specific learning interventions accordingly, based on the gap between the competence profile and a previously set standard (Swanson, 2001).

7. Limitations and conclusion

In this paper, we designed and proposed an artifact to assess professional competences of individuals to value their prior learning. The artifact applies a text mining algorithm to



Notes: Upper left: (a) the green field shows the competence profile of individual 1; upper right: (b) shows the same profile with a standard or point of reference (red line); lower left: (c) the green field shows the competence profile of individual 2. Lower right: (d) shows the same profile with a standard or point of reference (red line)

Figure 3.
Two individuals with their specific distribution of assessed competences normalized to a scale from 0 to 6

make the assessment of professional competences more efficient and less tedious. The designed artifact can be a part in the VPL procedure. Subsequently, we present limitations and further research.

First, limitations concern the data set we used. All individuals in our data set work for venture capitalist firms within the USA. In this regard, the demographic and professional variety within the data set is limited as can be seen in Table AIII of the Appendix. Further research should apply the designed artifact to textual data from different professions and countries. Further research should, nevertheless, test the designed artifact with several documents of one person outlining the competences in different areas of professional and personal life. Also, many of the CV did not produce a match or produce only a very small number of matches between the repository and the competence model. In this regard, we

support the call to use long and descriptive or narrative resumes as repository for text mining methods (Han and Lee, 2016). Long, descriptive and narrative CV may also support the assessment of competences in VPL.

Second, limitations concern the designed artifact. The artifact should be only used for already *identified* and *documented* competences. It has to be pointed out that the artifact does not validate competences automatically, rather it may help in organizing documented professional competences for a reviewer or external assessor. In this regard, the resulting competence profiles may serve as a heuristic for further dialogue between an assessor and a candidate and can be a basis for a thorough psychological assessment. In this regard, the designed artifact is also not a behavioral assessment. To assess behavioral competences, i.e. whether a person is really able to perform a certain occupation, further behavioral simulations have to be conducted (Epstein, 2002). In other words, if we are to find out whether a person is really able to bake, i.e. possesses the necessary experience and tacit knowledge to do so, automatized assessments will only be of little help (Ribeiro and Collins, 2007). In this regard, the occurrences of matches between a body of documents and a standard may serve as an approximation toward competence and should be interpreted as first impression. Furthermore, the normalization of results to a scale from 0 to 6 may distort the results to some extent as the highest number of matches automatically gets assigned the value 6 and the lowest number a value near to 0. Further research should test the designed artifact with different procedures of normalization. Even though we build a dictionary based on a comprehensive model of professional competences (Peterson *et al.*, 2001), further research should aim to use an even more detailed model of professional competence as a standard.

Notes

1. Available at: https://patents.google.com/patent/US20150127567A1/en
2. Available at: www.onetcenter.org/dictionary/22.2/excel/content_model_reference.html
3. Available at: https://blog.usejournal.com/where-did-you-go-to-school-bde54d846188
4. Available at: https://phantombuster.com

References

Aken, A., Litecky, C., Ahmad, A. and Nelson, J. (2010), "Mining for computing jobs", *IEEE Software*, Vol. 27 No. 1, pp. 78-85.

Anderson, L. and Krathwohl, D. (2001), *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Abridged ed., Longman, New York, NY.

Annen, S. (2013), "Recognising non-formal and informal learning: typology and comparison of selected European approaches", *Literacy Information and Computer Education Journal (LICEJ)*, Vol. 4 No. 1, pp. 928-937.

Bastian, M., Hayes, M., Vaughan, W., Shah, S., Skomoroch, P., Kim, H. and Lloyd, C. (2014), "Linkedin skills: large-scale topic extraction and inference", *Proceedings of the 8th ACM Conference on Recommender systems*, ACM, pp. 1-8.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Sebastopol.

Bjørnåvold, J. (2000), "Making learning visible: identification, assessment and recognition of non-formal learning in Europe", CEDEFOP reference document, Office for Official Publications of the European Communities, Luxembourg.

Bloom, B.S., Krathwohl, D. and Masia, B.S. (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals: Cognitive Domain*, Longman, New York, NY.

Bohlinger, S. (2017), "Comparing recognition of prior learning (rpl) across countries", in Mulder, M. (Ed.), *Competence-Based Vocational and Professional Education*, Technical and Vocational Education and Training, Springer International Publishing, Cham, pp. 589-606.

Bohm, D. (2012), "On dialogue", *Routledge Classics*, 2nd ed., Taylor and Francis, Hoboken.

Bohne, C., Eicker, F. and Haseloff, G. (2017), "Competence-based vocational education and training (vet)", *European Journal of Training and Development*, Vol. 41 No. 1, pp. 28-38.

Brockmann, M., Clarke, L. and Winch, C. (2009), "Competence and competency in the eqf and in european vet systems", *Journal of European Industrial Training*, Vol. 33 Nos 8/9, pp. 787-799.

Campion, M.A., Fink, A.A., Ruggeberg, B.J., Carr, L., Phillips, G.M. and Odman, R.B. (2011), "Doing competencies well: best practices in competency modeling", *Personnel Psychology*, Vol. 64 No. 1, pp. 225-262.

Cedefop (2015), *European Guidelines for Validating Non-Formal and Informal Learning*, 2nd ed., Office for Official Publications of the European Communities, Luxembourg.

Cedefop (2017), "European inventory on validation of non-formal and informal learning – 2016 update-: synthesis report", Luxembourg.

Cheetham, G. and Chivers, G. (1996), "Towards a holistic model of professional competence", *Journal of European Industrial Training*, Vol. 20 No. 5, pp. 20-30.

Darabi, H., Karim, F., Harford, S., Douzali, E. and Nelson, P. (2018), "Detecting current job market skills and requirements through text mining", in *2018 ASEE Annual Conference and Exposition, Conference Proceedings*.

Debortoli, S., Müller, O. and Vom Brocke, J. (2014), "Comparing business intelligence and big data skills", *Business and Information Systems Engineering*, Vol. 6 No. 5, pp. 289-300.

Debortoli, S., Müller, O., Junglas, I.A. and Vom Brocke, J. (2016), "Text mining for information systems researchers: an annotated topic modeling tutorial", *Communications of the Association for Information Systems*, Vol. 39, pp. 555-582.

Diedrich, A. (2013), "Translating validation of prior learning in practice", *International Journal of Lifelong Education*, Vol. 32 No. 4, pp. 548-570.

Dreyfus, H. and Dreyfus, S. (1987), "From socrates to expert systems: the limits of calculative rationality", *Bulletin of the American Academy of Arts and Sciences*, Vol. 40 No. 4, pp. 15-31.

Epstein, R.M. (2002), "Defining and assessing professional competence", *JAMA*, Vol. 287 No. 2, pp. 226.

EU (2006), "Council decision of 15 november 2006 on establishing an action programme in the field of lifelong learning", *Official Journal of the European Union*.

EU (2012), "Council recommendation of 20 december 2012 on the validation of non-formal and informal learning", *Official Journal of the European Union*.

EU (2017), "Council recommendation of 22 may 2017 on the european qualifications framework for lifelong learning", *Official Journal of the European Union*.

Fahrenbach, F., Kaiser, A. and Schnider, A. (2019), "A competence perspective on the occupational information network (O*NET)", in Bui, T. (Ed.), *52th Hawaii International Conference on Systems Science*, pp. 5651-5660.

Gallivan, M.J., Truex, D. and Kvasny, L. (2004), "Changing patterns in it skill sets 1988-2003: a content analysis of classified advertising", *ACM SIGMIS Database*, Vol. 35 No. 3, pp. 64-87.

Gorbacheva, E., Stein, A., Schmiedel, T. and Müller, O. (2015), "A gender perspective on business process management competences offered on professional online social networks", in ECIS 2015 Completed Research Papers.

Gregor, S. and Hevner, A.R. (2013), "Positioning and presenting design science research for maximum impact", *MIS Quarterly*, Vol. 37 No. 2, pp. 337-355.

Han, S. and Lee, G. (2016), "A preliminary study on text mining based human resource allocation in a construction project", *Proceedings of the 33rd International Symposium on Automation and Robotics in Construction (ISARC)*.

Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, Vol. 47 No. 1, pp. 153-161.

Hevner, A.R., Vom Brocke, J. and Maedche, A. (2018), "Roles of digital innovation in design science research", *Business and Information Systems Engineering*, Vol. 6 No. 1, pp. 39.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004), "Design science in information systems research", *MIS Quarterly*, Vol. 28 No. 1, pp. 75-105.

Holland, J.L. (1997), *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*, 3rd ed., Psychological Assessment Resources, Odessa.

Kaplan, S., Klebanov, M. and Sorensen, M. (2012), "Which ceo characteristics and abilities matter?", *The Journal of Finance*, Vol. 67 No. 3, pp. 973-1007.

Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M.A., Aung, Z. and Woon, W.L. (2017), "Data mining approach to monitoring the requirements of the job market: a case study", *Information Systems*, Vol. 65, pp. 1-6.

Kobayashi, V.B., Mol, S.T., Berkers, H.A., Kismihók, G. and Den Hartog, D.N. (2018), "Text mining in organizational research", *Organizational Research Methods*, Vol. 21 No. 3, pp. 733-765.

Krathwohl, D.R. (2002), "A revision of Bloom's taxonomy: an overview", *Theory into Practice*, Vol. 41 No. 4, pp. 212-218.

Lave, J. and Wenger, E. (2011), *Situated Learning: Legitimate Peripheral Participation*, Cambridge Univ. Press, Cambridge.

Le Deist, F.D. and Winterton, J. (2005), "What is competence?", *Human Resource Development International*, Vol. 8 No. 1, pp. 27-46.

Lichtnow, D., Loh, S., Carlos, L., Junior, R. and Piltcher, G. (2008), "Using text mining on curricula vitae for building yellow pages", Working paper.

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

Miller, G.A. (1995), "Wordnet: a lexical database for English", *Communications of the ACM*, Vol. 38 No. 11, pp. 39-41.

Mulder, M., Weigel, T. and Collins, K. (2007), "The concept of competence in the development of vocational education and training in selected eu member states: a critical analysis", *Journal of Vocational Education and Training*, Vol. 59 No. 1, pp. 67-88.

Müller, O., Schmiedel, T., Gorbacheva, E. and Vom Brocke, J. (2014), "Towards a typology of business process management professionals: identifying patterns of competences through latent semantic analysis", *Enterprise Information Systems*, Vol. 10 No. 1, pp. 50-80.

Nonaka, I. (1994), "A dynamic theory of organizational knowledge creation", *Organization Science*, Vol. 5 No. 1, pp. 14-37.

Nonaka, I., Toyama, R. and Konno, N. (2000), "Seci, ba and leadership: a unified model of dynamic knowledge creation", *Long Range Planning*, Vol. 33 No. 1, pp. 5-34.

Patel, B., Kakuste, V. and Eirinaki, M. (2017), "Capar: a career path recommendation framework", in *3rd IEEE International Conference on Big Data Computing Service and Applications*, pp. 23-30.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007), "A design science research methodology for information systems research", *Journal of Management Information Systems*, Vol. 24 No. 3, pp. 45-77.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., Fleishman, E.A., Levin, K.Y. and Dye, D. M. (2001), "Understanding work using the occupational information network (o*net): implications for practice and research", *Personnel Psychology*, Vol. 54 No. 2, pp. 451-492.

Polanyi, M. (1966), *The Tacit Dimension*, University of Chicago Press, Chicago.

Rajaraman, A. and Ullman, J.D. (2011), "Data mining", in *Mining of Massive Datasets*, Cambridge University Press, pp. 1-17.

Ribeiro, R. and Collins, H. (2007), "The bread-making machine: tacit knowledge and two types of action", *Organization Studies*, Vol. 28 No. 9, pp. 1417-1433.

Romero, C. and Ventura, S. (2010), "Educational data mining: a review of the state of the art", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 40 No. 6, pp. 601-618.

Roulin, N. and Levashina, J. (2018), "Linkedin as a new selection method: psychometric properties and assessment approach", *Personnel Psychology*, Vol. 72 No. 2, pp. 187-211.

Sachin, R.B. and Vijay, M.S. (2012), "A survey and future vision of data mining in educational field", *in 2nd International Conference on Advanced Computing and Communication Technologies*, IEEE, pp. 96-100.

Sahlman, W. (1990), "The structure and governance of venture-capital organizations", *Journal of Financial Economics*, Vol. 27 No. 2, pp. 473-521.

Salton, G., Wong, A. and Yang, C.-S. (1975), "A vector space model for automatic indexing", *Communications of the Acm*, Vol. 18 No. 11, pp. 613-620.

Schön, D.A. (1983), *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York, NY.

Simon, H.A. (1996), *The Sciences of the Artificial*, 3rd ed., MIT Press, Cambridge, Mass.

Stenlund, T. (2010), "Assessment of prior learning in higher education: a review from a validity perspective", *Assessment and Evaluation in Higher Education*, Vol. 35 No. 7, pp. 783-797.

Swanson, R.A. (2001), "Human resource development and its underlying theory", *Human Resource Development International*, Vol. 4 No. 3, pp. 299-312.

Todd, P.A., McKeen, J.D. and Gallupe, R.B. (1995), "The evolution of is job skills: a content analysis of is job advertisements from 1970 to 1990", *MIS Quarterly*, Vol. 19 No. 1, pp. 1-27.

Valdez-Almada, R., Rodriguez-Elias, O., Rose-Gomez, C., Velazquez-Mendoza, M. and Gonzalez-Lopez, S. (2017), "Natural language processing and text mining to identify knowledge profiles for software engineering positions: generating knowledge profiles from resumes", in *5th International Conference in Software Engineering Research and Innovation*, pp. 97-106.

Appendix

Table AI.
The comprehensive competence model includes personal competence (PC1 – PC7), social competence (SC1 – SC9), method competence (MC1 – MC10) and domain competence (DC1 – DC6) which served as the starting point to create the dictionary

ID	Name	The person is able to . . . at his/her workplace.
PC1	Socialization through education or culture	Use his/her education and cultural background to perform appropriate
PC2	Suitability based on personality characteristics	Perform based on his/her personality characteristics
PC3	Suitability based on interests	Reflect on his/her professional interests and match these to the demands
PC4	Achievement motivation	Reflect on his/her key strengths and use them
PC5	Management of values	Reflect on his/her values and on organizational values
PC6	Setting and pursuing goals	Eet goals and pursue them
PC7	Act practically intelligent	Use his/her common sense
SC1	Sense of social appropriateness	Act in a socially appropriate way
SC2	Communication and interaction	Communicate and interact with others in a goal-oriented and appropriate way
SC3	Active and passive feedback	Give feedback to others and receive feedback from others
SC4	Empathy	Act in a friendly, cooperative and empathic way with others
SC5	Ability to form and maintain relationships	Support others and to build strong relationships with others
SC6	Occupational roles	Negotiate about the own role in the occupation
SC7	Leadership and social influence	Exert influence in social systems and to lead others
SC8	Conflict management	Solve conflicts constructively
SC9	Advice and development	Advice others and be responsible for their professional development
MC1	Socio-technical systems	Understand, monitor and improve socio-technical systems
MC2	Resource management	Manage his/her and organizational time and finances
MC3	Human resources systems and practices	Ensure that an organization has fitting employees to meet their organizational goals
MC4	Solving complex problems	Solve new, ill-defined and complex problems in the real world
MC5	Performing complex technical activities	Perform skilled activities using coordinated movements
MC6	Operate and use machines and technical systems	Use his/her developed capacities to design, Set-up, operate and correct malfunctions in Machines and technical systems
MC7	Digital communication	Appropriately use different methods and ways of digital communication
MC8	Manage knowledge and information	Identify and manage knowledge and information
MC9	Business management	Apply knowledge of principles and facts related to business management
MC10	Administrative work	Perform routine operations like administration, staffing or controlling
DC1	Domain knowledge	Use domain-specific knowledge to perform
DC2	Work settings	Work in different physical environments
DC3	Environmental conditions	Withstand extreme environmental conditions
DC4	Handling of dangerous conditions	Handle different dangerous or hazardous conditions
DC5	Physical and cognitive requirements	Handle the physical and cognitive requirements
DC6	Work conditions	Work under different and changing conditions

Table AII. Dictionary based on the model of professional competences that served as a standard or reference for the CV (total: 1,255 descriptors)

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
PC	Personal competence	Personal competence describes the "willingness and ability, as an individual personality, to understand, analyse and judge the development chances, requirements and limitations in the family, job and public life, to develop one's own skills as well as to decide on and develop life plans. It includes personal characteristics like independence, critical abilities, self-confidence, reliability, responsibility and awareness of duty, as well as professional and ethical values" (Le Deist and Winterton, 2005, p. 38)		
PC1	Socialization through education or culture	The person is able to use his/her education and cultural background to perform appropriate at his/her workplace	Learn; education; educate; school; pedagogy; reading; writing; listening; conversation; speaking; mathematics; mathematic; problem-solving; knowledge-acquisition; science; thinking; think; critic; critical-thinking; decision-making; logic; culture; socialization	23
PC2	Suitability based on personality characteristics	The person is able to perform at his/her workplace based on his/her personality characteristics	pressure; stress; criticism; setbacks; setback; work-related problems; maturity; flexibility; flexible; poise; self-control; emotion-control; anger; aggression; calmness; stress-tolerance; openness to change; commitment; dependability; carefulness; trustworthiness; trust; accountability; detail-orientation; attention; reliability; responsibility; dependability; fulfilling-obligations; carefulness; honesty; integrity; conscientiousness; conscientious; extraverted; extroverted; emotional stability; emotion	38
PC3	Suitability based on interests	The person is able to reflect on his/her professional interests and match these to the demands at the workplace	Interests; entrepreneur; realistic; pragmatic; social; artistic; enterprise; convention; conventional; hands-on problems; investigation; investigate; problem-solving; thinking; design patterns; teaching; service; entrepreneurship; project management; leadership; business; risk taking; routines; procedures	24
PC4	Achievement motivation	The person is able to reflect on his/her key strengths and use them at the workplace	Achievement; motivation; persistence; accomplishment; initiative; accomplish; ability utilization; goal-setting; competence; competences; competencies; competency; achievement orientation; achievement goals; mastering tasks; effort; obstacles; challenges; challenge; responsibility; responsibilities	21

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
PC5	Management of values	The person is able to reflect on his/her values and on organizational values	Values; value; creativity; creative; responsible; responsibility; autonomy; idea; ideas; decision; decision; autonomy; supervision; supervise; recognition; advancement; leadership; prestige; prestigious; authority; recognition; social status; status; advancement; opportunities; recognize; organizational values; tradition; traditional; stable; stability; innovation; innovate; collaboration; collaborate; opportunity recognition; taking chances; guiding principle; excellence; high standard; high standards; openness; honesty; honest; transparence; transparency; flexibility; adapting to change; adaption; fairness; justice; just; precision; detail-oriented; stability; getting things done; well-being; caring; innovation; openness to change; openness to ideas; aggressiveness; customer value; valuing customer	64
PC6	Setting and pursuing goals	The person is able to set goals and pursue them at the workplace	Goal; goals; explicit goals; smart goals; goal characteristic; goal setting; goal attaining; quantification; goal pursuing; feedback; goal feedback; specific goals; specificity	13
PC7	Act practically intelligent	The person is able to use his/her common sense at the workplace	Pragmatic; pragmatism; practical intelligent; practical intelligence; idea generation; generate ideas; creativity; alternative solutions; work-related problems; common sense; logic; work related issues; analytical thinking; analytic thinking	14
SC	Social competence	Social competence describes the "willingness and ability to experience and shape relationships, to identify and understand benefits and tensions, and to interact with others in a rational and conscientious way, including the development of social responsibility and solidarity" (Le Deist and Winterton, 2005, p. 38)		
SC1	Sense of social appropriateness	The person is able to act in a social appropriate way at the workplace	Shape relationships; relationship; social; competence; social competence; responsibility; solidarity; perceptiveness; coordination; adjustment; persuasion; persuade; negotiate; negotiation; reconciliation; instructing; teaching; helping; help; social orientation; service orientation; service; civil service; social	24
SC2	Communication and interaction	The person is able to communicate and interact with others in a goal-oriented and appropriate way at the workplace	Communication; communicate; interact; interaction; interpretation; meaning; translation; explanation; explaining; supervisor; supervisors; peers; subordinates; peer; email; e-mail; telephone; phone; public relations; pr; relationships; relationship; constructive; assistance; medical attention; emotional support; care; selling; sell; merchandise; goods; complaints; complaint; grievance; conflicts; conflict; negotiating; negotiation; dispute; disputes; restaurant; store	42

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
SC3	Active and passive feedback	The person is able to give feedback to others and receive feedback from others at the workplace	Feedback; active feedback; passive feedback; supervisor; co-worker; performance; monitoring; monitor	8
SC4	Empathy	The person is able to act in a friendly, cooperative and empathic way with others at the workplace	Empathy; pleasantness; sympathy; interpersonality; interpersonal; easy going; cooperation; good-natured attitude; concern; sensitive; helpful; understanding; concern for others; social orientation; social; personal connection; human interaction; interaction; relationship; relationships; responsibility; health; safety	23
SC5	Ability to form and maintain relationships	The person is able to support others and to build strong relationships with others at the workplace	Relationship; service; non-competitive environment; co-workers; moral values; social service; relation; pressure; freedom; morality	10
SC6	Occupational roles	The person is able to negotiate about the own role in the occupation at the workplace	Roles; role conflict; role overload; conflict; conflicts; demands; requests; groups; supervisor; supervision; role negotiability; negotiation; overload; demand; assignment; adequate resource; role relationship; teamwork; team; group-work; customers; customer; coordination; leadership; lead; coordinate	26
SC7	Leadership and social influence	The person is able to exert influence in social systems and to lead others at the workplace	Leadership; influence; impact; social influence; design; crafting; craft; supervisor; take charge; be in charge; supervisory leadership; supervisor; leadership; friendly; supportive; support; goal setting; planning work; planning tasks; schedule; plan; assign tasks; assignment; vision; group vision; organizational vision; vision development; problem-solving; difficulties; support; human relations; relations; company policies; fairness; fair treatment; leader	36
SC8	Conflict management	The person is able to solve conflicts constructively at the workplace	conflict; argument; argumentation; conflictual contact; deescalation; unpleasant; angry; discourteous; aggressive; aggression; violence; conflict resolution; resolution; compromise; consens	15
SC9	Advice and development	The person is able to advice others and be responsible for their professional development at the workplace	Cooperation; sensitivity; easy-going; good-natured; concern; understanding; helpful; helpfulness; sensitive; connection; social orientation; interpersonal relationships; relationship; development; counselling; training; human interaction; responsibility for others; responsible; responsibility; outcomes; results; health; safety; apprenticeship; apprentice; mentoring; mentor	28
MC	Method competence	Method competence arises "from the implementation of transversal strategies and processes of invention and problem-solving" (Le Deist and Winterton, 2005, p. 36). Transversal strategies are cross-functional and span a variety of occupations		

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
MC1	Socio-technical systems	The person is able to understand, monitor and improve socio-technical systems at the workplace	System; technical system; social system; socio-technical system; visioning; understanding; improvement; socio-technical systems; requirement analysis; systems perception; system changes; consequences; identification of consequences; change in operations; key causes; cost and benefits; judgment; decision-making; system performance; system evaluation; analysis; evaluation; measures; indicators; performance; performance improvement	26
MC2	Resource management	The person is able to manage his/her and organizational time and finances	Resources; allocate resources; resource management; time management; finance; financial resources; money; accounting; expenditures; material resources; material; equipment; facilities; facility; personnel; motivating employees; motivating people; motivation; personnel management; personnel resources; directing people; directing; developing people; developing; personnel selection	25
MC3	Human resources systems and practices	The person is able to ensure that an organization has fitting employees to meet their organizational goals	Human resources; HR; human resource systems; human resource practices; policies; recruitment; selection; recruitment and selection; hiring; hiring decisions; promotion; promoting; personnel decisions; personnel; vacancy; recruiting plans; job interview; recruitment operations; assessment center; assessment; job selection; training; human resource development; human resource methods; training methods; formal training; informal training; training programs; content of training; skill training; technical training; sponsored training; compensation; reward; reward system; non-monetary benefits; performance; knowledge; skills; seniority; team performance; job attributes; organizational performance; compensation; pensions; insurance; paid leave; awards; bonuses	49
MC4	Solving complex problems	The person is able to solve new, ill-defined and complex problems in the real world	Problem; problem identification; identify information; information identification; complex problems; complexity; problem solving; information gathering; essential information; find information; information organization; organize information; classify information; classification; synthesis; information synthesis; knowledge synthesis; reorganization; idea generation; creativity; idea evaluation; idea implementation; implementation planning; solution appraisal; appraisal; solution; problem observation; problem evaluation; outcome evaluation; lessons learned; reasoning; reason; decision-making; judgment; analyzing information; evaluating results; best solution; solve problems;	80

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
			creative thinking; thinking creatively; design thinking; developing ideas; designing ideas; cerating ideas; developing applications; applications; application; designing applications; designing application; creating application; creating applications; developing ideas; designing ideas; designing relationships; creating ideas; developing systems; system development; system design; product design; artistic; knowledge use; knowledge application; update knowledge; use knowledge; relevant knowledge; up-to-date; technical knowledge; scheduling; scheduling work; time management; schedule activities; organizing work; work organization; organization; work planning; accomplish; accomplishment; prioritization; prioritize; goal development	
MC5	Performing complex technical activities	The person is able to perform skilled activities using coordinated movements	Technical activities; skilled activities; coordinated movements; movements; coordination; computers; computer; pc; software; hardware; tools; computer systems; programming; computer programming; data entry; process information; coding; code; functions; drafting; laying out; specifying technical devices; technical documentation; technical instructions; technical drawings; drawings; specifications; fabrication; construction; fabricated; constructed; assembled; assembly; modification; modified; maintained; maintain; maintenance; usage; used; service; repairing; repair; adjust; adjusting; testing machines; testing; test; machine; devices; device; moving parts; mechanical principles; principles; reparation; maintaining; mechanical equipment; electronic; electronics; electronic maintenance; servicing; repairing; calibrating; fine-tuning; tuning; testing; machines; machine; device; devices; equipment; operation; electronic principles; electric principles; documenting; recording; information; document information; record information; enter information; entering information; transcribe information; transcribe text; record information; store information; information storage; maintain information; written information; electronic information; magnetic information	90
MC6	Operate and use machines and technical systems	The person is able to use his/her developed capacities to design, set-up, operate and correct malfunctions in machines and technical systems	Technical skills; technical machines; design machines; set up machines; operate machines; correct machines; malfunctions; technological systems; operations analysis; operations management; product requirements; customer needs; design requirements; design needs; generate technology; adapt	68

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
			technology; adapt equipment; needs; technology design; equipment selection; tool use; installing equipment; equipment; installation; wiring; program machines; wiring machines; specifications; design specifications; programming; program; operation monitoring; watch gauges; watch dials; control operations; control equipment; control systems; quality management; quality evaluation; product evaluation; product inspection; inspection; routines; routine; maintenance; equipment; equipment maintenance; troubleshooting; error determination; error; troubleshoot; reparation; repairing; repair; repair machines; repair systems; test conducting; conductor; inspections; inspecting products; inspecting services; testing products; testing services; evaluate quality; evaluate performance; analyse quality; control quality; quality control	
MC7	Digital communication	The person is able to appropriately use different methods and ways of digital communication	Communication; public speaking; telephone; electronic mail; email; e-mail; letters; memos; face-to-face; discussions; teams; individuals; digitalization; ICT	14
MC8	Manage knowledge and information	The person is able to identify and manage knowledge and information at the workplace	Knowledge management; information management; information input; job-related information; receiving information; getting information; observing; receiving information; obtaining information; monitor processes; process monitoring; reviewing information; monitor materials; monitor surroundings; monitor events; monitor environment; detect problems; assess problems; information interpretation; information identification; information evaluation; job-relevant information; estimating size; estimating distances; estimating quantities; size; distance; quantity; determining time; determining costs; resources; time cost; materials; work performance; product characteristics; event characteristics; information characteristics; data science; information processing; data processing; judging things; judging quality; judging people: assessing value; assessing importance; assessing quality; assessing people; processing information; compiling; coding; compiler; code; categorizing; calculating; tabulating; auditing; audition; calculation; verification; verifying information; verifying data; evaluate information; compliance; laws; regulations; standards; norms; norm compliance; law compliance; regulations compliance; analyzing data; analyzing information; identifying principles; principle identification; identify reasons; identify facts; identify informations; reasoning; decision-making; judgment	79

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
MC9	Business management	The person is able to apply knowledge of principles and facts related to business management at the workplace	Business; business and management; business administration; accounting; human resource management; HRM; material resource management; organizations; organization; sales; marketing; sales and marketing; economics; office information; enterprise resource planning; organizing systems; economics; administration and management; strategic planning; resource allocation; human resource modelling; resource allocation; leadership; leadership technique; production methods; coordination of people; coordination; coordination of resources; clerical; clerical procedures; administrative procedures; word processing; managing files; managing records; stenography; transcription; designing forms; office procedures; office; economics and accounting; financial market; finance; financial data; reporting financial data; analysis of financial data; banking; showing products; promoting products; selling products; selling services; promoting services; marketing; marketing strategy; tactics; sales and marketing; product demonstration; sales technique; sales control systems; customer service; personal service; customer needs; needs assessment; quality standards; service quality; customer satisfaction; personnel recruitment; personnel selection; recruitment; selection; training; compensation; benefits; negotiation; labor relations; personnel information systems; personnel	76
MC10	Administrative work	Persons are able to perform routine operations such as administration, staffing or controlling at the workplace	Administrative activities; processing paperwork; paperwork; information files; maintaining files; recruiting; interviews; selection; hiring; promoting employees; staffing; monitoring; controlling; control; overseeing; spending	16
DC	Domain competence	Domain competence describes the "willingness and ability, on the basis of subject-specific knowledge and skills, to carry out tasks and solve problems and to judge the results in a way that is goal-oriented, appropriate, methodological and independent. General cognitive competence . . . the ability to think and act in an insightful and problem-solving way" (Le Deist and Winterton, 2005, p. 38)		

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
DC1	Domain knowledge	The person is able to use domain-specific knowledge to perform at the workplace	Production; manufacture; agriculture; agricultural goods; goods storage; goods processing; manufacturing; raw materials; production process; quality control; costs; goods distribution; planting; growing; harvesting; harvest; plant; grow; animal; consumption; food production; engineering; technology; computers; electronics; circuit boards; processors; processor chips; electronic equipment; computer hardware; software; applications; programming; practical application; design techniques; design tools; design principles; technical plans; blueprints; drawings; models; construction; building; materials; house construction; building construction; building; house; highway; highways; roads; road; mechanical knowledge; physics; biology; mathematics; geography; algebra; arithmetic; calculus; statistics; statistic; fluid dynamics; material dynamics; atmospheric dynamics; mechanical structures; electrical structures; atomic structures; subatomic structures; fluid processes; mechanical process; electrical process; atomic process; chemical composition; substance structure; chemical process; chemical transformation; chemicals; chemical interactions; chemistry; danger signs; disposal methods; plant; animal; organism; tissues; cells; human behavior; psychology; individual differences; ability; personality; interests; learning; motivation; psychological research methods; research methods; psychological assessment; disorder treatment; behavioral disorder; affective disorder; sociology; group behavior; group dynamics; societal trends; human migration; migration; ethnicity; ethnic; culture; history; origin; land; sea; air masses; health service; health service; health; diagnosis; curing; preventing disease; disease prevention; physical health; mental health; well-being; preserving health; improving health; medicine; dentistry; diagnosis; injuries; diseases; deformities; symptoms; treatment alternatives; drug properties; drugs; preventive medicine; preventive health care; rehabilitation; physical dysfunctions; mental dysfunctions; career counselling counselling; guidance; education; training; curriculum; training design; training instruction; training effects; arts and humanities; arts; humanities; human thought; English language; English; composition; grammar; rules of composition; foreign language; pronunciation; music; dance; visual arts; drama; sculpture; fine arts; history; archeology; historical events;	213

(*continued*)

Table AII.

ID	Name	Description	Descriptors of the dictionary	No. of descriptors
			civilizations; cultures; philosophy; theology; religion; religions; values; ethics; customs; practices; human culture; regulations; property; injury; damage; public conduct; legislation; political process; law; public safety; security operation; data protection; property protection; legal codes; laws; court procedures; precedents; government regulations; executive orders; agency; rules; democratic process; political process; delivering information; telecommunication; transmission; broadcasting; switching; control; media production; communication; dissemination; transportation	
DC2	Work settings	The person is able to work in different physical environments	Physical surroundings; work setting; indoors; indoor; environmentally controlled; warehouse; heat; cold; outdoors; exposed; weather; working outdoors; exposure; under cover; open vehicle; tractor; car	16
DC3	Environmental conditions	The person is able to withstand extreme environmental conditions at the workplace	Environmental condition; environment; extreme condition; extreme environment; physical proximity; sounds; noise; distraction; uncomfortable; temperatures; hot; cold; very hot; very cold; brightness; lightning; lightning condition; contaminants; pollutants; gases; dust; odors; cramped workspace; cramped; awkward position; vibration; jackhammer; whole body	28
DC4	Handling of dangerous conditions	The person is able to handle different dangerous or hazardous conditions at the workplace	Hazard; hazardous condition; frequency; exposure; radiation; disease; infection; high places; high place; hazardous equipment; burns; cuts; bites; stings; injured; injury; serious; outcome; work attire; attire; dress; protective equipment; safety shoes; glasses; gloves; hard hats; life jackets; breathing apparatus; safety harness; full protection suits; radiation protection	31
DC5	Physical and cognitive requirements	The person is able to handle the physical and cognitive requirements at the workplace	Physical position; body position; sitting; standing; climbing ladders; ladders; scaffolds; poles; walking; running; kneeling; crouching; stooping; crawling; keeping balance; regaining balance; balance; twisting body; bending body; handle; feel objects; feel tools; control tools; repetition	24
DC6	Work conditions	The person is able to work under different and changing conditions	Activity; compensation; independence; security; variety; work condition; busy; independence; variety; compensation; security; working conditions	11

Table AIII. Demographic description of the scraped CV which served as the repository for text mining (consolidated from primary data source)

Demographic characteristic	Men (*N* and %)	Women (*N* and %)	*N* (total and %)
Ethnicity Caucasian	831 (71)	164 (61)	995 (69)
Ethnicity African-American	31 (3)	10 (4)	41 (3)
Ethnicity Hispanic	16 (1)	7 (3)	23 (2)
Ethnicity Asian	289 (25)	86 (32)	375 (26)
Role of associate	220 (19)	118 (44)	338 (24)
Role of principal	168 (14)	58 (21)	226 (16)
Role of partner	777 (67)	92 (33)	869 (60)
Engineering degree	336 (28)	56 (20)	392 (27)
Educated at Harvard or Stanford	446 (38)	128 (48)	574 (40)
Total	1,165 (81)	268 (19)	1,435 (100)

Table AIV. Number of matches per sub-competence

Matches per sub-competence dimension	PC (*N* and %)	SC (*N* and %)	MC (*N* and %)	DC (*N* and %)
1	1,457 (2.12)	3,141 (4.65)	6 (3.03)	8,219 (12.17)
2	189 (0.28)	1,387 (2.05)	1,273 (1.89)	73 (0.11)
3	8,857 (13.12)	301 (0.45)	772 (1.14)	63 (0.09)
4	258 (0.38)	1,490 (2.21)	1,232 (1.82)	88 (0.13)
5	2,887 (4.28)	580 (0.86)	6,282 (9.30)	123 (0.18)
6	107 (0.16)	3,337 (4.94)	669 (0.99)	1,022 (1.51)
7	90 (0.13)	3,746 (5.55)	831 (1.23)	
8		18 (0.03)	357 (0.53)	
9		3,775 (5.59)	12,485 (18.49)	
10			367 (0.54)	
Total 67,522 (100%)	13,845 (20.50)	17,775 (26.32)	26,314 (38.97)	9,588 (14.20)

Range	*N*	(%)
0	117	8.15
[0,9]	262	18.25
[10,19]	186	12.96
[20,29]	163	11.35
[30,39]	145	10.10
[40,49]	143	9.96
[50,59]	102	7.10
[60,69]	99	6.89
[70,79]	57	3.97
[80,89]	60	4.18
[90,99]	57	3.97
[100,109]	31	2.16
[110,119]	30	2.09
[120,129]	26	1.81
[130,139]	16	1.11
[140,149]	15	1.04
[150,159]	9	0.62
[160,169]	6	0.41
[170,179]	10	0.69
[180,189]	5	0.34
[190,199]	3	0.20
[200,209]	2	0.13
[210,219]	3	0.20
[220,229]	0	0
[230,239]	2	0.13
[240,249]	1	0.06
[250,259]	1	0.06
[260,269]	1	0.06

Table AV.
Number of matches per CV

Corresponding author
Florian Fahrenbach can be contacted at: florian.fahrenbach@gmail.com