International Journal of Information Management 50 (2020) 416–431

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

# Text mining of industry 4.0 job advertisements

Mirjana Pejic-Bach[a], Tine Bertoncel[b,*], Maja Meško[b], Živko Krstić[c]

[a] University of Zagreb, Faculty of Economics and Business, Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia
[b] University of Primorska, Faculty of Management, Cankarjeva ulica 5, 6000 Koper, Slovenia
[c] Atomic Intelligence d.o.o., Bencekovićeva 33, 10000 Zagreb, Croatia

A R T I C L E I N F O

*Keywords:*
Human resource management
Text mining
Job profiles
Big data analytics
Industry 4.0
Education
Smart factory

A B S T R A C T

Since changes in job characteristics in areas such as Industry 4.0 are rapid, fast tool for analysis of job advertisements is needed. Current knowledge about competencies required in Industry 4.0 is scarce. The goal of this paper is to develop a profile of Industry 4.0 job advertisements, using text mining on publicly available job advertisements, which are often used as a channel for collecting relevant information about the required knowledge and skills in rapid-changing industries. We searched website, which publishes job advertisements, related to Industry 4.0, and performed text mining analysis on the data collected from those job advertisements. Analysis of the job advertisements revealed that most of them were for full time entry; associate and mid-senior level management positions and mainly came from the United States and Germany. Text mining analysis resulted in two groups of job profiles. The first group of job profiles was focused solely on the knowledge related to Industry 4.0: cyberphysical systems and the Internet of things for robotized production; and smart production design and production control. The second group of job profiles was focused on more general knowledge areas, which are adapted to Industry 4.0: supply change management, customer satisfaction, and enterprise software. Topic mining was conducted on the extracted phrases generating various multidisciplinary job profiles. Higher educational institutions, human resources professionals, as well as experts that are already employed or aspire to be employed in Industry 4.0 organizations, would benefit from the results of our analysis.

## 1. Introduction

The fourth industrial revolution, also called Industry 4.0, has brought numerous changes in organizational processes, methods of work, and staff structure in organizations (Vogel-Heuser & Hess, 2016). Macurova, Ludvik, and Žwakova (2017) point out that organizations transitioning to Industry 4.0 will be at risk of having workers who are not skilled at preparing, implementing, and using technologies associated with Industry 4.0, since higher educational institutions lag behind in developing adequate educational programmes, both formal and informal. In the same time, Industry 4.0 could have a short-term negative impact on the decreased demand for the low-skilled workers (Weber (2016). A future challenge will be to restructure jobs and educational programmes (Kane, Palmer, Phillips, & Kiron, 2015) simultaneously. Therefore, it is crucial for manufacturing organizations, not only to prepare for the restructuring of their production processes (Ivanov, Dolgui, Sokolov, Werner, & Ivanova, 2016), but also to analyse job profiles prevalent in Industry 4.0 organizations, in order to determine the required employee competencies for such an organization (Imran & Kantola, 2018; Pecina & Sladek, 2017).

Due to the relevance of Industry 4.0, it has become of the upmost importance to develop a deep insight into the required knowledge and skills for this rapid changing area. Nevertheless, how we should gain up-to-date insight into the skills and competencies relevant for Industry 4.0 organizations? What is the best way, not only to gain insight but also to monitor trends in this area? In our paper, we shed some light on these questions using text mining approach to analyse job advertisements published on a website, for the purpose of uncovering required knowledge and skills of Industry 4.0.

Current research of the skills and knowledge needed by the Industry 4.0 is scarce. Several authors wrote about the job profiles in Industry 4.0, based on secondary sources, such as scientific papers (Hecklau, Galeitzke, Flachs, & Kohl, 2016; Pecina & Sladek, 2017). Pinzone et al. (2017) develop job profiles for Industry 4.0, based on the in-depth interviews and focus groups with managers working in Industry 4.0 organizations in Italy. However, these researches generate their conclusions mainly using the literature review analysis, but due to the fast changes in Industry 4.0 organizations, an approach that is more dynamic is needed. Since changes in job characteristics in areas such as Industry 4.0 are rapid, fast tool for analysis of job advertisements is

* Corresponding author.
*E-mail addresses:* mpejic@efzg.hr (M. Pejic-Bach), tine.bertoncel@gmail.com (T. Bertoncel), maja.mesko@fm-kp.si (M. Meško), zivko.krstic@live.com (Ž. Krstić).

https://doi.org/10.1016/j.ijinfomgt.2019.07.014
Received 30 December 2018; Received in revised form 21 July 2019; Accepted 23 July 2019
Available online 01 August 2019
0268-4012/ 

needed. Job advertisements are a relevant source of information about the required skills and knowledge, which can be used in order to provide fast insight into the changes in job profiles. Two main methods for analysing job advertisements are manual content analysis (e.g., Todd, McKeen, & Gallupe, 1995) and automated text analysis, often referred to as text mining (Amado, Cortez, Rita, & Moro, 2017). Text mining has relevant advantages compared to manual content analysis, such as less required time and human work needed for the analysis (Guo, Vargo, Pan, Ding, & Ishwar, 2016). In addition, with this paper we contribute to the growing body of research focusing to the utilization of social media for extracting knowledge relevant for gaining competitive advantage in various organizations (Kapoor et al., 2018; Shiau, Dwivedi, & Yang, 2017), as well as improving the efficiency of marketing efforts (AlAlwan, Rana, Dwivedi, & Algharabat, 2017; Dwivedi, Kapoor, & Chen, 2015). Text mining methods has been widely used for the purpose of analysing the information stored at social media websites, such as Facebook posts (Shiau, Dwivedi, & Lai, 2018).

Against on this background, the purpose of this paper is to clarify the competencies required for Industry 4.0 specialists, using a text mining analysis of job advertisement from LinkedIn, the leading website that publishes job advertisements (Skeels & Grudin, 2009). We have downloaded job advertisements published on LinkedIn, which contained on of the two key-terms: Industry 4.0 or Smart Factory. Downloaded data was cleaned and transformed, using lemmatization algorithm. The list of key words has been extracted, which were further classified into phrases and topics. Text mining has been conducted using the specialized text mining software WordStat Provalis ver. 8. Results of the text mining analysis were discussed in relation to the landscape of knowledge needed in Industry 4.0 organizations.

The rest of this paper is structured in the following manner. In the second section, we introduce the background for the research done in this study, by looking at concepts of Industry 4.0, research done on future job profiles in Industry 4.0, as well as the use of text mining techniques to study online job profiles. In the third section, the methodology for data gathering is outlined. The fourth section focuses on the results, which consist of descriptive analysis, text mining, phrase extraction, and topic mining. The fifth section presents the discussion of our results in the context of the theoretical and practical contributions. Finally, in the concluding section, we summarize the results and look at the limitations of the study and recommend further research.

## 2. Background

### 2.1. Industry 4.0 and the concept of smart factory

The concept of Industry 4.0 is based on the vision of future manufacturing, embedded in the notion of a smart factory, which could be considered as the next step in the evolution of factories (Ivanov et al., 2016; Lucke, Constantinescu, & Westkämper, 2008; Radziwon, Bilberg, Bogers, & Madsen, 2014; Wang, Wan, Li, & Zhang, 2016).

A smart factory describes a factory that utilizes the cyberphysical systems (CPS) for carrying out the production tasks, based on information gathered by sensors about the environment in which the production is occurring, as well as gathered information about the environment. CPS can be interlinked through the Internet of things (IoT), creating a flexible, agile and decentralized production system (Ghobakhloo & Azar, 2018; Lucke et al., 2008; Yoon, Shin, & Suh, 2012). Through AI or machine-learning, CPS devices have the potential of becoming self-organizing, self-learning and self-adapting, as well as capable of cooperating with other CPS devices through a broader IoT network (Stojmenovic, 2014; van Gerven, 2017; Wan et al., 2013). In addition, during the production process, the massive amounts of data are collected from smart objects and are transferred to the cloud, where advanced analytic techniques could be used the improvement of manufacturing performance (Flynn, Dance, & Schaefer, 2017; van Gerven, 2017).

Machine-learning algorithms, which enable machines to execute various functions autonomously, have an important role in the automation in Industry 4.0 organizations (Flynn et al., 2017; van Gerven, 2017). Although the complete automation is yet a far-reaching goal, important emerging concepts utilizing machine-learning algorithms are already being actualized in practice, mainly in regards to interoperability and mass customization (Gruber, 2013; van Gerven, 2017).

### 2.2. Future job profiles of industry 4.0

Because of these technological changes, Industry 4.0 brings changes in the knowledge and skills that employees should possess in smart factories (Jerman, Pejić Bach, & Bertoncelj, 2018; Morlock, Wienbruch, Leineweber, Kreimeier, & Kuhlenkoetter, 2016). Lorenz, Rüßmann, Strack, Lueth, and Bolle (2015) investigated 40 families of jobs in 23 industries and found that in the Industry 4.0 »more jobs will be gained than lost, but workers will require significantly different skills«. It is projected that by 2025 in Germany, as one of the leading countries in using advanced production technologies, Industry 4.0 technologies will decrease the number of assembly and production jobs by 610.000, while jobs in data science and information technology will increase by approximately 960.000 (Lorenz et al., 2015). Other authors provide the same conclusions. Achenhagen and Zeller (2011) expect that there will be an increase in machine operation, software maintenance, and hardware maintenance jobs in the future. A decrease will be seen in repetitive, routine and physically demanding jobs, while those requiring a higher level of education, flexible responses, problem solving and complexity will see an increase (Hecklau et al., 2016; Lorenz et al., 2015).

According to Lorenz et al. (2015), the number of interdisciplinary programs that would incorporate the optimal skill set for Industry 4.0 is lower than required. Although an increasing number of worldwide higher educational institutions are starting graduate or bachelor's programs focused on Industry 4.0, the number of these programmes is still too low. For example, several authors mention the lack of programmes in mechatronics engineering (Barger & Gilbert, 2018; Lorenz et al., 2015), which would incorporate the broad knowledge relevant for the automation, as well as an interdisciplinary skill set that would support the implementation of this knowledge, such as project management. These and similar new programs would be better able to prepare the student for work in smart factories (Barger & Gilbert, 2018; Kozák, Ružický, Štefanovič, & Schindler, 2018; Meek, Field, & Devasia, 2003).

The impact of Industry 4.0 on the job market is controversial. Weber (2016) discuss that technological changes in the past often triggered the fear of job loss. Expectations for the Industry 4.0 range from the optimistic ones, which envision an increase in employment and work conditions, to the pessimistic ones, which envision decrease of employment opportunities due to the replacement of human work with robots. Although both scenarios seem unrealistic, Weber (2016) points out that the previous industrial revolution had a strong impact on the job opportunities for low-skilled workers, which is likely to repeat in the Industry 4.0 era. Sommer (2015) expresses the same concern for SMEs, which will not be able to adapt as fast as needed.

Currently, the research on the skills and knowledge needed for Industry 4.0 has been scarce. Most of the research use secondary sources or qualitative research in order to shed some light to the competencies needed for Industry 4.0 organizations (Hecklau et al., 2016; Marnewick & Marnewick, 2019; Pecina & Sladek, 2017).

### 2.3. Usage of online job advertising for job profiling

Job advertisements are used in various areas as the important source of information about the changes in required knowledge and skills in a particular area, such as information systems (Todd et al., 1995), and project management (Ahsan, Ho, & Khan, 2013). Analysis of job profiles has been conducted using the two approaches: manual content analysis and text mining. Several years ago, the usage of text mining in the development of job profiles was still rare. For example, Harper (2012) point out that only 4% of the researches used text mining, among the analysed 70 research studies in the area of library and information systems. However, the relevance of text mining for the analysis of job advertisements in order to develop job profiles is increasing (Palshikar et al., 2019). Text mining techniques are already being used, to gain insight into job profiles, such as big-data in marketing (Amado et al., 2018), to classify online job advertisements (Boselli, Cesarini, Mercorio, & Mezzanzanica, 2018; Karakatsanis et al., 2016), match resources with job profiles (Gonzalez et al., 2012), as well as to predict future performance of employees (Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016).

Text mining has been used in numerous studies that analyse the job advertisements in order to extract the information useful for decision making. For example, text mining can be used in order to detect emerging talents in organizations, and to foster their employment or career advancement (Chamorro-Premuzic et al., 2016), to change the system of job classifications from the traditional to the one more adjusted to the market changes (Amado et al., 2018), or to make rapid decisions based on the observed changes at the job market (Mezzanzanica, 2017). Text mining uses algorithms, such as latent semantic analysis or average linkage, in order to extract meaningful information from the vast amount of textual data. In comparison with manual content analysis, text mining is less time consuming and less expensive (Guo et al., 2016).

Text mining involves various steps of pre-processing that are used for cleaning the text, using techniques such as tokenization, stop word removal, stemming, chunking, parts of speech tagging, which removes unnecessary words and letters from the text. After pre-processing, text mining algorithms extract words from the cleaned text of job advertisements. Extracted words are combined into phrases and topics, using various algorithms, such as cluster analysis (Kino, Kuroki, Machida, Furuya, & Takano, 2017), latent semantic analysis (Karakatsanis et al., 2016; Müller, Schmiedel, Gorbacheva, & vom Brocke, 2016), support vector machines, random forest, neural networks (Amato et al., 2015; Boselli et al., 2018), explicit rules and latent dirichlet allocation (Amato et al., 2015). Text mining has been used for varous purposes, such as investigating the consumer perceptions of hotels based on the online customer textual reveiws (Xu, Wang, Li, & Haghighi, 2017), utilization of social media posts (Facebook and Twitter) for conducting competitive analysis (He, Zha, & Li, 2013), usage of social media posts for product planning (Jeong, Yoon, & Lee, 2017), and big data analysis in financial sector (Pejić Bach, Krstić, Seljan, & Turulja, 2019).

Text mining research that analyses job advertisements could be consolidated into three groups. The first group of studies uses text mining for analysis of job advertisements in order to develop a new classification scheme of job advertisements and compare it with the standard classification system of occupations (e.g., Amato et al., 2015; Karakatsanis et al., 2016). For example, Mezzanzanica (2017) used text mining for investigating Italian job advertisements in the field of marketing in Italy. Authors used ESCO job classification taxonomy and machine-learning algorithm over 1.9 million of job advertisements, with the result of observed trends and dynamics in the evolution of the labour market, and identified several emerging potential occupations. The second group of researchers used text mining in order to improve the quality of job matching with the potential candidates, according to the ‘commute time, job location, job type, hourly rates, the skill set of candidate’ (Kino et al., 2017, p.1523). These researches take into account that suboptimal matching of candidates to job positions can severely cost organizations and that for this reason, higher quality-matching techniques, such as text mining are needed. Third group of research used text mining algorithms in order to develop job profiles for specific vertical areas, such as business process managment (Müller et al., 2016), knowledge management (Chang, Wang, & Hawamdeh, 2018) and big data (de Mauro, Greco, Grimaldi, & Ritala, 2017; Gardiner, Aasheim, Rutner, & Williams, 2018), or horizontal areas, such as lean managment (Kregel, Ogonek, & Matthies, 2019).

## 3. Methodology

Our methodology consists of two major steps: data extraction and machine-learning analysis, following the approach of Jimenez-Marquez, Gonzalez-Carrasco, Lopez-Cuadrado, and Ruiz-Mezcua (2019). In order to present a procedure that could be useful to other interested researchers, experts of human resource specialists, we provide the detailed description of the used data sources, and the used analytical approach (descriptive analysis and text mining).

### *3.1. Data source*

Job advertisements are the most important channel for attracting new employees. Traditionally, job advertisements have been published in printed newspapers, while in the last two decades; they are dominantly published online, either on specialised sites or on social media.

In our work, we use LinkedIn as a source for job advertisements for several reasons. First, LinkedIn has become one of the most prominent leaders in publishing job advertisements covering a broad range of organizations, countries, and job types (Bradbury, 2011). Second, their job advertisements have a semi-structured form, which is suitable for text mining analysis (Gassler, Zangerle, & Specht, 2014). Third, LinkedIn can be considered a suitable open source for sampling (Žmuk, 2017).

Selection of job advertisements has been conducted in the following manner. In order to extract relevant advertisements, we focused on job advertisements that contain one of the two phrases strongly related to the new trends emerging in manufacturing: “Industry 4.0” or “Smart factory”. The following Boolean word combination was used when searching for the relevant work (“Industry 4.0” OR “Smart factor*”).

The data for the text mining consisted of texts found in job advertisements published on LinkedIn, in a period of three months, between the 1st of April and 1st of July 2018. In this period, the total number of 2566 job advertisements that contained the word Industry 4.0 or Smart Factory were published on LinkedIn. However, text mining was only conducted on 1460 job advertisements that were in English. Although one of the possible limitations of our work is the usage of too short observational period, we consider our approach as relevant, following other researcher’s approach who also collected job advertisements during a period of several months (e.g. Boselli et al., 2018), or even collected job advertisements using a one-time research session (e.g. Müller et al., 2016).

Job advertisements have been traditionally examined using either content or text mining analysis, but as Guo et al. (2016) state, content analysis requires more time and more human work, which limits the number of advertisements that can be analysed, although more rich information about the advertisements can be gathered by human experts. On the other hand, text mining is appropriate for the big data sets, where unstructured automated analysis of relevant information should be conducted in order to extract useful information. Guo et al. (2016, p.2) implicate that ‘the criterion to evaluate whether a dataset is “big” is a function of the amount of time required for a human to make a decision on a given unit’, concluding that the dataset measured in thousands of items may be considered as “big”. According to this criterion, our data set could be treated as a big data set, and such a conclusion provides the additional argument for using text mining for analysis of job advertisements.

LinkedIn job advertisements are semi-structured. Job advertisements published by these organizations contain the following fields: company, Job title, Location, Job function, Employment type, Company industry, Seniority level, Job description. The structure of these fields is presented in Table 1. The research team added the following field:

**Table 1**
Structure of a typical LinkedIn job advertisement.
Source: Authors' work

| Field | Type of field | | Note |
|---|---|---|---|
| | Structured | Unstructured | |
| Company | | ✓ | Free style text |
| Job title | | ✓ | Free style text |
| Location | | ✓ | Free style text (city and country or only city) |
| Job function | ✓ | | Predefined classification for the job functions is allowed to use; maximum 3 job functions. |
| Employment type | ✓ | | Predefined classification is allowed with the following modalities: Full-time, Part-time, Contract, Temporary, Volunteer, Internship; only one can be chosen. |
| Company industry | ✓ | | Predefined classification for the company industry is allowed to use; maximum 3 industries. |
| Seniority level | ✓ | | Predefined classification is allowed with the following modalities: Internship, Entry level, Associate, Mid-Senior level, Director, Executive, Not Applicable; only one can be chosen. |
| Job description | | ✓ | Free style text |
| Skill words | ✓ | | Predefined classification for the skill words is allowed to use; maximum 10 skills. |

Language of the ad.

As is visible form the Table 1, most of these fields have the free style text input. For the purpose of our analysis, we have used the following fields:

- Location
- Employment type
- Seniority level
- Job description

The information about the language in which the job advertisement was published was generated manually. In some cases, job advertisements were published in more than one language. However, we only used in our analysis the advertisements published in English.

The field Location has been recoded so that it contains only the country in which the job is placed. This recoding has been conducted manually.

The field Employment type is a structured field, with the following predefined modalities: Full-time, Part-time, Contract, Temporary, Volunteer, Internship. Only one modality can be chosen. This field has been recoded into two categories: Full-time and Other, due to the small number of advertisements in the categories Part-time, Contract, Temporary, Volunteer, and Internship.

The filed Seniority level is a structured field, with the following predefined modalities: Internship, Entry level, Associate, Mid-Senior level, Director, and Executive. Only one modality can be chosen. Among 1460 English job advertisements, 199 job advertisements had an empty field, and these were coded manually by two researchers independently into one of the predefined modalities.

The field with the richest information is a Job description, which is unstructed and was the only field used for text mining, which will be described in the following subchapter.

Other fields (Job function, Company industry, and Skill words) had a large number of predefined modalities. In addition, these fields could have multiple values (3 job functions, 3 company industries, and 10 skill words could be selected for one job advertisement). For example, possible values for the field Company industry are Construction, Civil Engineering, Human Resources, Space Computer Software or Information Technology and Services, while for Job type, examples are Management, Information Technology, Project management, Manufacturing, and Customer Service. However, these fields were not used in the analysis, since a number of modalities is very high, which would result in a too large number of categories per job advertisement, which is not suitable for the analysis, due to high sparsity of the data. In addition, some of the modalities of Company Industry are too general, such as Services, or highly specific, such as Space Computer Software. The same conclusions refer to the fields of Job function and Skill words.

### 3.2. Data analysis

The collected data were analysed in two steps: descriptive analysis and text mining analysis.

Descriptive analysis was conducted in order to analyse the job advertisements according to the Location, Employment type, Seniority level, and language of the job advertisement. Additionally, we investigated the relationship between the seniority level and contract type, as well as the country of job placement, and language of the job advertisement. We aimed to investigate.

Electronic documents have become the primary means of storing and retrieving written communication. Data mining is focused on extracting useful knowledge (e.g., trends, patterns) from unstructured or semi-structured files, databases, XML files in order to create data-driven models, such as classification, regression analysis or clustering (de Mauro et al., 2017; Gandomi & Haider, 2015). Text mining is a particular type of data mining focused on the handling of unstructured or semi-structured text documents. Text mining can be used to identify research trends by identifying relevant words and relationships in order to categorize or draw conclusions.

The text mining approach was used to determine topics most often emerging in online job advertisements related to smart factories and Industry 4.0. We have conducted text mining using WordStat Provalis software ver. 8.0.

In the first step, text pre-processing has been conducted, using lemmatization and exclusion. Lemmatization pre-processes words so that plural words were transformed into singular words, and verbs were transformed into present tense versions from other tenses (e.g., past tense). Exclusion process is based on using the exclusion list. For our purpose, we used the exclusion list of standard English words provided with the Wordstat, which was expanded with additional words, which occur often in job advertisements, but are not relevant for our analysis (such as a word "to expect"), or are too general (such as a word "skill"). These additional words were selected by the researchers themselves, based on their knowledge of Industry 4.0 organizations.

In the second step, the most frequent words that occur in more than 10 job advertisements were extracted. Using this large process number of words was extracted. However, analysis based on words is often not useful, since phrases that consist of two or more words are richer in meaning. For example, the words "big" and "data" do not have any specific meaning in the context of smart factories, but the phrase "big data" is highly relevant. In total, 96 words were extracted.

In the third step, the most frequent phrases that consist of a maximum of 5 words and occur in more than 50 job advertisements were extracted. This approach resulted in 52 phrases that were used for the description of the required skills and knowledge for the Industry 4.0 jobs.

In the fourth step, the topic extraction has been conducted using cluster analysis, where the extracted phrases are put through an average-linkage hierarchical clustering algorithm, for creating a similarity matrix that sheds light on phrases (topics) that occur close to each other.

Distances between clusters are based on the Unweighted Pair Group Mean Averaging method, which calculates the averages of the distances between each phrase of one Cluster to every other phrase of a different Cluster.

$$d_{ab} = \frac{1}{kl}\sum_{i=1}^{k}\sum_{j=1}^{l} d(A_i, B_j) \tag{1}$$

Notation:

$A_1, A_2, \ldots, A_k$ = Observations from *Cluster* A

$B_1, B_2, \ldots, B_l$ = Observations from *Cluster* B

d (a,b) = Distance between a *Cluster* with an observation vector *a* and a *Cluster* with observation vector *b*

Jaccard's coefficient similarity measure was used to determine the association between phrases that occur close together. Phrases that are associated together are represented by a dendogram. This allowed for increased focus on the most meaningful and strongest associations among phrases. Branches within clusters can be rotated, as the distances are represented temporally and not linearly. This means that it is more important to look at which clusters connect to other clusters, instead of looking at the sequence of phrases within clusters.

## 4. Results

First, we present the results of the descriptive analysis. Second, we present the most frequent phrases extracted from the job advertisements related to Industry 4.0 and Smart factories. We provide cluster analysis on these phrases in order to develop the landscape of knowledge needed in Industry 4.0 organizations. Finally, we analyse extracted phrases and topics according to the seniority level of the job vacancy.

### 4.1. Descriptive analysis

Table 2 presents the seniority level and contract type of employment in job advertisements related to Industry 4.0 and Smart factories. More than 90% of all contract types were full time, for all levels of management, except for internships, which were full time in 56% of cases. The most job openings were available for entry level (32%), associate level (29%) or mid-senior level (27%), accounting for 88% of all the job advertisements.

Most of the job advertisements (48%) for Industry 4.0 come from Germany (30.48%) and the United States (17.46%) (Table 3). The United States, Great Britain, Singapore, Ireland, and South Africa are the only countries that had all of their advertisements in English (Table 3). Several other countries came close to having all of their advertisements in English, such as India (98%), Norway (96%) and Spain (91%). The least amount of English advertisements came from Korea (12%), Italy (17%), Russia (17%) and Germany (20%). The rest of the countries, Netherlands, China, Switzerland, Belgium, Sweden, Canada, France, Austria, the Czech Republic, and Poland, had between 21 and 90% of their advertisements in English.

**Table 2**
Seniority level and contract type of employment in job advertisements related to Industry 4.0 and Smart factories.
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

| Seniority level | Total | % | Contract type | | | |
|---|---|---|---|---|---|---|
| | | | Full-time | % | Other | % |
| Associate | 425 | 29% | 404 | 95% | 21 | 5% |
| Director | 72 | 5% | 71 | 99% | 1 | 1% |
| Entry level | 472 | 32% | 444 | 94% | 28 | 6% |
| Executive | 25 | 2% | 24 | 96% | 1 | 4% |
| Internship | 68 | 5% | 38 | 56% | 30 | 44% |
| Mid-Senior level | 398 | 27% | 385 | 97% | 13 | 3% |
| Total | 1460 | 100% | 1366 | 94% | 94 | 6% |

**Table 3**
Country of the establishment of the organization publishing job adds in job advertisements related to Industry 4.0 and Smart factories, according to the language of the add.
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

| Country | Total | % | The English language adds | % |
|---|---|---|---|---|
| Germany | 782 | 30,48% | 159 | 20% |
| United States | 448 | 17,46% | 448 | 100% |
| Netherlands | 183 | 7,13% | 82 | 45% |
| Great Britain | 124 | 4,83% | 124 | 100% |
| China | 91 | 3,55% | 80 | 88% |
| Italy | 83 | 3,23% | 14 | 17% |
| Switzerland | 79 | 3,08% | 33 | 42% |
| Belgium | 78 | 3,04% | 38 | 49% |
| Korea | 74 | 2,88% | 9 | 12% |
| Sweden | 59 | 2,30% | 33 | 56% |
| India | 51 | 1,99% | 50 | 98% |
| Canada | 42 | 1,64% | 32 | 76% |
| France | 40 | 1,56% | 22 | 55% |
| Singapore | 39 | 1,52% | 39 | 100% |
| Austria | 32 | 1,25% | 22 | 69% |
| Czech Republic | 30 | 1,17% | 11 | 37% |
| Poland | 28 | 1,09% | 22 | 79% |
| Russia | 24 | 0,94% | 4 | 17% |
| Norway | 24 | 0,94% | 23 | 96% |
| Spain | 22 | 0,86% | 20 | 91% |
| Ireland | 21 | 0,82% | 21 | 100% |
| South Africa | 20 | 0,78% | 20 | 100% |
| Other | 192 | 7,48% | 153 | 80% |
| Total | 2566 | 100,00% | 1460 | 57% |

### 4.2. Text mining analysis

Text mining generated numerous results related to the most frequent words, phrases, and topics. However, we shall present the most relevant results, which include the most frequent phrases and extracted topics.

Table 4 presents the most frequent phrases reflecting knowledge in job advertisements related to Industry 4.0 and Smart factories. The Column Term Frequency Inverse Document Frequency (TF*IDF) of Table 4 contains values of metrics for an estimation of the importance of a phrase in a collection of documents. The importance is based on the ratio between the frequency at which a phrase occurs in a document and the total number of phrases in the document (TF), along with the logarithm of the ratio between the total number of documents the phrase occurs in and the total number of documents present in the collection (*IDF). Within the context of this research, TD*IDF refers to documents consisting of job advertisements related to Industry 4.0. The TF-IDF metric is used to pre-process the collection of documents to exclude phrases that cannot be considered as key phrases.

The top ten most important phrases according to their frequency, are "supply chain", "project management"," machine-learning"," big data"," computer science"," internet of things"," software development"," digital manufacturing"," product development" and" business development".

Extracted phrases can be divided into two groups: the phrase describing general and specific knowledge. Most of the extracted phrases are related to specific knowledge, such as "supply chain management", "big data", or "lean manufacturing". However, there are several phrases that reflect general knowledge that is transferable across different

**Table 4**
Most frequent phrases advertisements reflecting knowledge in job advertisements related to Industry 4.0 and Smart factories (+50 cases).
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

| | FREQUENCY | NO. CASES | % CASES | LENGTH | TF • IDF |
|---|---|---|---|---|---|
| SUPPLY CHAIN | 794 | 284 | 11,12% | 2 | 757,4 |
| **PROJECT MANAGEMENT** | 471 | 332 | 13,00% | 2 | 417,3 |
| MACHINE-LEARNING | 312 | 170 | 6,66% | 2 | 367,2 |
| BIG DATA | 294 | 191 | 7,48% | 2 | 331,1 |
| COMPUTER SCIENCE | 293 | 265 | 10,38% | 2 | 291,3 |
| INTERNET OF THINGS | 288 | 227 | 8,89% | 3 | 302,7 |
| SOFTWARE DEVELOPMENT | 271 | 171 | 6,70% | 2 | 318,2 |
| DIGITAL MANUFACTURING | 240 | 108 | 4,23% | 2 | 329,7 |
| PRODUCT DEVELOPMENT | 239 | 158 | 6,19% | 2 | 288,8 |
| **BUSINESS DEVELOPMENT** | 203 | 133 | 5,21% | 2 | 260,5 |
| **DIGITAL TRANSFORMATION** | 187 | 142 | 5,56% | 2 | 234,7 |
| DATA ANALYTICS | 186 | 133 | 5,21% | 2 | 238,7 |
| **CONTINUOUS IMPROVEMENT** | 160 | 125 | 4,89% | 2 | 209,6 |
| ARTIFICIAL INTELLIGENCE | 153 | 139 | 5,44% | 2 | 193,4 |
| REAL TIME | 131 | 102 | 3,99% | 2 | 183,2 |
| **INNOVATIVE SOLUTIONS** | 127 | 117 | 4,58% | 2 | 170,1 |
| BUSINESS PROCESSES | 126 | 102 | 3,99% | 2 | 176,2 |
| CONTROL SYSTEMS | 126 | 85 | 3,33% | 2 | 186,2 |
| ADVANCED MANUFACTURING | 120 | 85 | 3,33% | 2 | 177,3 |
| ELECTRICAL ENGINEERING | 116 | 99 | 3,88% | 2 | 163,7 |
| SMART MANUFACTURING | 116 | 87 | 3,41% | 2 | 170,3 |
| SOFTWARE ENGINEERING | 116 | 83 | 3,25% | 2 | 172,6 |
| INDUSTRIAL ENGINEERING | 109 | 86 | 3,37% | 2 | 160,5 |
| LEAN MANUFACTURING | 98 | 80 | 3,13% | 2 | 147,4 |
| DATA MANAGEMENT | 95 | 70 | 2,74% | 2 | 148,4 |
| **LIFE CYCLE** | 86 | 75 | 2,94% | 2 | 131,8 |
| PRODUCT DESIGN | 86 | 65 | 2,55% | 2 | 137,1 |
| SUPPLY CHAIN MANAGEMENT | 85 | 64 | 2,51% | 3 | 136,1 |
| CLOUD COMPUTING | 83 | 71 | 2,78% | 2 | 129,1 |
| INDUSTRIAL AUTOMATION | 81 | 67 | 2,62% | 2 | 128,1 |
| FACTORY AUTOMATION | 81 | 53 | 2,08% | 2 | 136,3 |
| DATA SCIENCE | 80 | 58 | 2,27% | 2 | 131,5 |
| **DIGITAL TECHNOLOGIES** | 76 | 51 | 2,00% | 2 | 129,2 |
| **CUSTOMER SATISFACTION** | 74 | 72 | 2,82% | 2 | 114,7 |
| APPLICATION SOFTWARE | 73 | 50 | 1,96% | 2 | 124,7 |
| PRODUCT MANAGEMENT | 71 | 57 | 2,23% | 2 | 117,2 |
| **OPERATIONAL EXCELLENCE** | 68 | 63 | 2,47% | 2 | 109,3 |
| **ENTREPRENEURIAL SPIRIT** | 67 | 67 | 2,62% | 2 | 105,9 |
| ENGINEERING EXPERIENCE | 66 | 61 | 2,39% | 2 | 107 |
| AGILE DEVELOPMENT | 65 | 64 | 2,51% | 2 | 104,1 |
| DISCRETE MANUFACTURING | 65 | 63 | 2,47% | 2 | 104,5 |
| INDUSTRY EXPERIENCE | 65 | 55 | 2,15% | 2 | 108,3 |
| **QUALITY MANAGEMENT** | 65 | 53 | 2,08% | 2 | 109,4 |
| **DRIVE INNOVATION** | 63 | 63 | 2,47% | 2 | 101,3 |
| OBJECT ORIENTED | 63 | 56 | 2,19% | 2 | 104,5 |
| PRODUCT LIFECYCLE | 60 | 53 | 2,08% | 2 | 101 |
| PERSONALIZED PRODUCTS | 59 | 57 | 2,23% | 2 | 97,4 |
| DESIGN AND DEVELOPMENT | 58 | 57 | 2,23% | 3 | 95,8 |
| ENTERPRISE SOFTWARE | 58 | 55 | 2,15% | 2 | 96,7 |
| ADVANCED ANALYTICS | 58 | 51 | 2,00% | 2 | 98,6 |
| OPERATING SYSTEMS | 57 | 55 | 2,15% | 2 | 95 |
| ENGINEERING SERVICES | 53 | 53 | 2,08% | 2 | 89,2 |
| SAP DIGITAL | 53 | 51 | 2,00% | 2 | 90,1 |

Note: General knowledge phrases are outlined in bold.

business areas, such as “project management”, “innovative solutions”, and “customer satisfaction”. Some of the extracted phrases also overlap, such as “innovative solutions” and “drive innovation”, and “supply chain” and “supply chain management”.

### 4.3. Usage of phrases for weak-signal detection

Day and Shoemaker (2009) stress the need for organizations to scope, scan, interpret, probe and act on subtle information (i.e. weak signals) that is not easily accessible with traditional methods, which would in turn work as an early warning system, with which the organization can respond to opportunities or threats in their business environment. One method that is recommended for this purpose is advanced analytics, which can help an organization gain insight and explore hidden trends in the vast amount of unstructured, as well as structured data, gathered from various sources, such as cyberspace. Advanced analytic digital technologies, such as topic mining, have become an indispensable resource for managers, consultants, human resource professionals, and researchers, for gaining up-to-date information on the job market and required profession-specific skills and knowledge.

In order to detect the phrases that could be used as weak signals, we extracted the phrases that occur in not less than 10 cases, and not more than 30 cases. The results here are abundant since we extracted more than 900 phrases. Manual analysis of the phrases is conducted by the researches, and some examples of relevant phrases that indicate weak signals are presented here. These phrases refer to the emerging fields such as: “connectivity in harsh environments”, “image processing”,

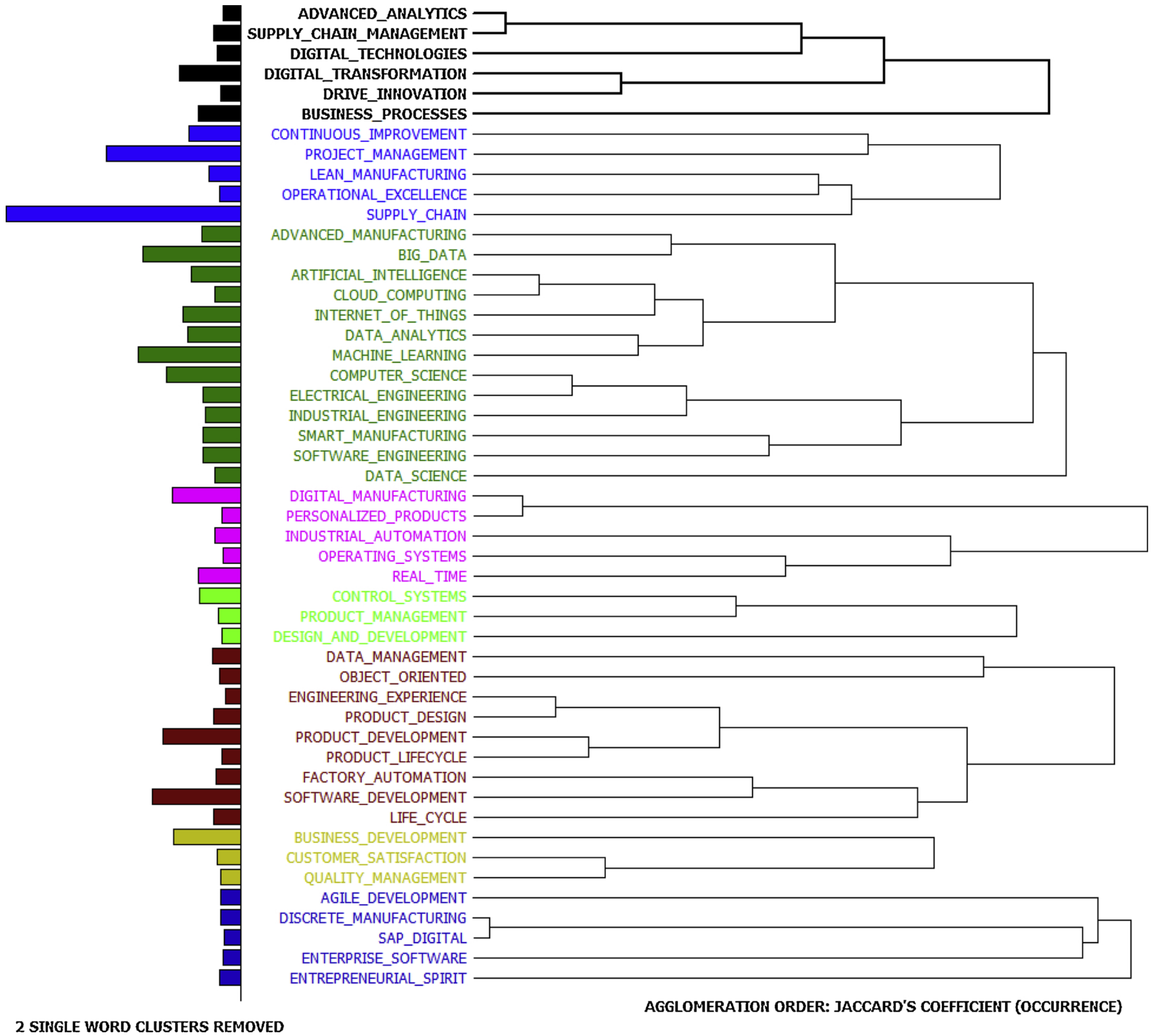



**Fig. 1.** Cluster analysis of phrases reflecting knowledge in job advertisements related to the Industry 4.0 and Smart factories (+50 cases).
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

"robotics and advanced materials", "production lifecycle management", and "ability to communicate complex ideas". Although these phrases occurred in a small number of advertisements, they provide examples of skills that could be highly relevant to the competitiveness of Industry 4.0 organizations.

### *4.4. Topic mining*

As previously stated, we used the Jaccard's coefficient similarity measure in order to determine the association between phrases extracted from job advertisements related to Industry 4.0 (Fig. 1). Results present the group of phrases that occur together in job advertisements. For example, project management and supply chain appear together in relation to continuous improvement, lean manufacturing, and operational excellence.

Fig. 2 presents the mapping of clusters, where the size of the circle reflects the frequency of the phrase in job advertisements, and grouping of circles reflect the strength of the relationship between the phrases in the clusters. As previously noted, two phrases stand out according to their frequency: supply chain and project management.

The cluster analysis identified eight topics regarding job advertisements related to Industry 4.0 and Smart factories, which are presented in Table 6. The second column presents the job profiles that were extracted using cluster analysis. We gave the titles to the job profiles in relation to the phrases grouped in the cluster (presented in the third column). The fourth column presents the distribution of job advertisements grouped in a cluster according to the seniority level. It can be noted that three job profiles have the highest ratio of mid-senior positions, four job profiles have the highest associate positions, and only one has the highest ratio of entry-level positions.

### *4.5. Cluster 1 - supply chain analyst*

The first cluster is described using specific knowledge phrases ("supply chain management", "business processes", and "advanced analytics") and several general knowledge phrases ("digital technology", "digital transformation", and "drive innovation") needed to support core process of supply chain management. The following phrases reflect the knowledge required for supply chain management from the Industry 4.0 standpoint: "advanced analytics", "digital transformation", "continuous improvement", and "lean manufacturing".

Advanced analytics, such as big data analytics, has a strong connection to supply chain management. Big data analytics can help organizations decrease costs and risks present in supply chains, for example, decreased operational costs resulting from the use of sensors and data analytics (Arunachalam, Kumar, & Kawalek, 2018; Govindan,

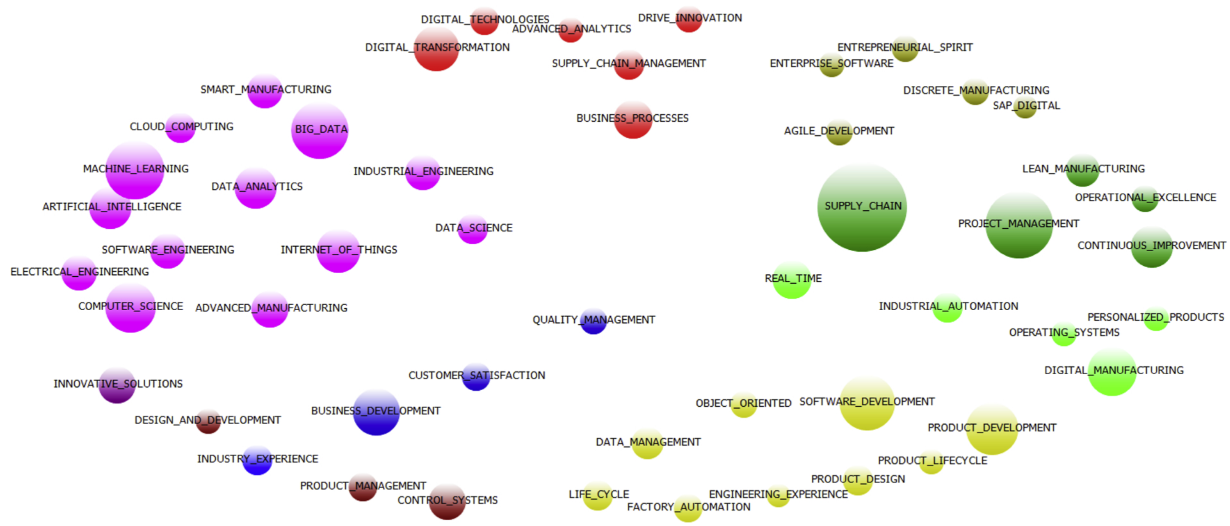



**Fig. 2.** Mapping of cluster analysis of phrases reflecting knowledge in job advertisements related to the Industry 4.0 and Smart factories (+50 cases).
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

Cheng, Mishra, & Shukla, 2018). There are several potential applications of big data in supply chain management, such as using data collected from RFID tracking devices attached to commodities, in order to optimize inventory policies. Within supply chain management, big data analytics can also help improve logistics/production planning and scheduling, as well as enhance forecasting and demand planning (Govindan et al., 2018). Big data analytics for supply chain management is also called supply chain analytics (SCA) (Arunachalam et al., 2018). The main sources of big data generation in SCA are enterprise information systems, sensors, and RFID devices coupled through IoT (Arunachalam et al., 2018).

### *4.6. Cluster 2 – supply chain engineer*

The second cluster combines the knowledge and skill relevant for the “supply” chain and “project management”, which both have a relationship with “lean manufacturing”, “continuous improvement” and “operational excellence”.

Lean manufacturing helps factories to organize production and processes in such a way as to achieve the shortest lead time and increase productivity by up to 25%, while keeping costs at a minimum and maintaining or increasing quality and customization of products (Brettel, Friederichsen, Keller, & Rosenberg, 2014; Kolberg & Zühlke, 2015; Küpper, Heidemann, Ströhle, Spindelndreier, & Knizek, 2017). In addition to reducing waste, continuous improvement of business processes and employee skills plays a key role in lean manufacturing and its corresponding potential for operational excellence. According to several experts from academia and industry, digitalization concepts, such as data acquisition, storage, and analytics have a slight to strong influence on continuous improvement (Hambach, Kümmel, & Metternich, 2017; Kolberg & Zühlke, 2015; Küpper et al., 2017).

### *4.7. Cluster 3 - cyberphysical systems (CPS) and the internet of things (IoT) for a robotized production engineer*

The third cluster is described using specific knowledge phrases supporting the notion of Industry 4.0: “advanced manufacturing “, “big data“, “artificial intelligence“, “cloud computing“, “internet of things“, “data analytics“, “machine-learning“, “computer science“, “electrical engineering“, “industrial engineering“, “smart manufacturing“, “software engineering“, and “data science“. Since these skills are central to the notion of Industry 4.0, the concepts related to this job profile will be described in details.

In regards to CPS and IoT, which are the backbone of Industry 4.0, job candidates for this job profile are be expected to have academic and work experience in electrical engineering, software engineering and/or industrial engineering. Skills from these three engineering areas are combined in the new area called mechatronics engineering (Barger & Gilbert, 2018; Benešová & Tupa, 2017; Fernández-Miranda, Marcos, Peraltam, & Aguayo, 2017; Kozák et al., 2018; Leminen, Rajahonka, & Westerlund, 2017; Meek et al., 2003; Sackley & Bester, 2016;).

IoT software knowledge is necessary for supporting numerous Industry 4.0 jobs. According to Taivalsaari and Mikkonen (2018), most IoT developers use open source code. Several open source technologies for cloud development are used in the following areas: security (HAProxy, NGINX), data acquisition (Apache Kafka), data analytics (Apache Storm/Spark, Apachi Hadoop), and monitoring (Graphite, Ichinga). Since current manufacturing networks do not have a far range across the factory, gateways are needed to pass the information to the cloud. However, recent advances in cellular radio technologies, such as NB-IoT, show potential for eliminating or decreasing the need for gateways (Taivalsaari & Mikkonen, 2018).

Low-end IoT devices within the context of cyber physical systems are actuators and sensors, for example, air quality sensors, thermostats, and automated door locks, used in various fields (e.g., Elhoseny et al., 2018). Low-end CPS devices are typically programmed with the programming languages C or C++, however other languages, such as assembly can also be used (Barr & Massa, 2006; Soulier, Li, & Williams, 2015; Taivalsaari & Mikkonen, 2018). Some low-end devices might have a real-time operating system (RTOS) sensors. In contrast with low-end devices, higher-end devices can have support for application development, dynamic programming, and full operating systems, for example, a Linux-based operating system. Such high-end devices are typically used to do advanced analytics by interacting with the cloud, with the help of web & mobile applications, as well as administrative and monitoring tools for CPS devices. They can be programmed with several programming languages, such as Python or JavaScript.

Big data analytics looks for patterns, trends and associations in big data sets, which in some cases can comprise of more than a petabyte of data (Govindan et al., 2018; Lorenz et al., 2015). In the Industry 4.0 organizations, such experts are required to understand and use advanced computational algorithms, such as machine-learning on data sets generated by the machines, e.g., IoT generated datasets.

### 4.8. Cluster 4 – digital manufacturing engineer

The fourth cluster contains the following specific knowledge phrases supporting the notion of Industry 4.0: “digital manufacturing“, “personalized products“, “industrial automation“, “operating systems“, and “real time“. This job profile needs to have the skills that support the organization of various digital manufacturing processes, such as mass customization.

The notion of personalized products is the prerequisite of mass customization. To cater the clients with mass customization, production control has to meet certain scheduling and machine operation needs. In these operations, the material flow needs to be monitored and analysed, particularly to increase synchronization of material flow and to conduct predictive maintenance (Foidl, Felderer, & Felderer, 2016; Sanders, Elangeswaran, & Wulfsberg, 2016; Zawadzki & Żywicki, 2016).

### 4.9. Cluster 5 – smart product designer

The fifth cluster contains the following specific knowledge phrases supporting the notion of Industry 4.0: “control systems“, “product management“, and “design and development“.

Smart product design can be delivered through various channels. For example, customers can use web applications in order to provide their desired product configurations. These configurations are then sent to Computer Aided Design software for 3D modelling and product lifecycle management, which makes all information about the product design available throughout the whole value chain (Schuh, Potente, Wesch-Potente, Weber, & Prote, 2014). These product designs can be prototyped and tested among various alternative solutions, through virtual reality or other forms of simulation, which is called the object oriented prototyping, with objects as a virtual model of specific products. Virtual reality can simulate operations, services, assembly and disassembly, workplace and product configurations, as well as visualize products for sales and marketing purposes (Flynn et al., 2017; Zawadzki & Żywicki, 2016). These above-mentioned approaches are referred to as mass customization, which has the goal to produce products customized according to customer preferences and simultaneously minimize the production time and the production costs.

### 4.10. Cluster 6 – ICT specialist for factory automation

The sixth cluster contains the following specific knowledge phrases supporting the notion of Industry 4.0: “data management”, “object oriented”, “engineering experience”, “product design”, “product development”, “product lifecycle”, “factory automation”, “software development”, and “life cycle”. This job profile is related to various aspects of factory automation.

Product design and production control are significantly changed in smart manufacturing, since the production system in smart production collects data through CPS devices, and distributes this knowledge through the IoT (Zawadzki & Żywicki, 2016). In addition, predictive maintenance can determine faults and anomalies in products and equipment and fix them on time for just-in-time delivery (Lee, Kao, & Yang, 2014; Sanders et al., 2016). The usage of RFID tags also contributes to the automation of production and personalization of products (Sanders et al., 2016).

### 4.11. Cluster 7 – customer satisfaction management

The seventh cluster contains the following general knowledge phrases supporting the notion of Industry 4.0: “business development“, “customer satisfaction“, and “quality management“. In the context of Industry 4.0 the general knowledge related to “business development, ““customer satisfaction, “and “quality management “describe a job profile that is in charge of aligning business processes with the needs of customers, resulting in increased loyalty and drive for the organization.

Various aspects of Inustry 4.0 contribute to customer satisfaction. On-time delivery and mass customization influence customer satisfaction, as high ration of products are tailored directly to the needs of the customer (Foidl et al., 2016; Sanders et al., 2016; Schuh et al., 2014; Zawadzki & Żywicki, 2016). Predictive maintenance and digital supply chain monitoring can drive business development, through its use of real-time analytic capabilities of monitoring up-to-minute changes in production, as well as up-to-minute changes in demand. Consequently, this increases supply chain visibility, inventory forecasting, and delivery reliability, which in turns enhanced quality management (Brettel et al., 2014; Elshendy & Fronzetti Colladon, 2017; Foidl et al., 2016; Kache & Seuring, 2015). With this respect, the goal of this job profile is to manage the big data analytics, CPS and mass customization in order to drive innovation in areas of organizational processes, integrated products, and service offerings, focused on customer satisfaction (Lee et al., 2014). However, the insight into the full text of the job advertisements for this profile indicates that the required knowledge and skills are not related to the technical issues, but to management issues. In line with that is that most of the jobs in this profile are for the mid-senior (41%) and director (7%) positions.

### 4.12. Cluster 8 - enterprise software specialist

The eight clusters contain the following specific knowledge phrases supporting the notion of Industry 4.0: “agile development”, “discrete manufacturing”, “SAP digital”, “enterprise software”, and “entrepreneurial spirit”.

In Industry 4.0 organizations, the relevance of enterprise software is increasing, since numerous specialised solutions are emerging. SAP digital is enterprise software for discrete manufacturing and Industry 4.0, which support discrete manufacturing and agile development. SAP digital suite, for example, contains SAP Leonardo, which through its cloud based analytics capabilities, provides an application that smart factories can use for predictive maintenance and supply chain management, for enhancing quality management and customer satisfaction. SAP digital provides manufacturing execution systems (MES) and enterprise resource planning (ERP) applications (SAP, 2017b, 2017b; SAP, 2018; SYSTEMA, 2018).

### 4.13. Distribution of phrases and clusters according to management level

Table 5 presents the most frequent phrases reflecting knowledge in job advertisements related with Industry 4.0 and Smart factories (+50 cases), which are to a greater or lesser degree, associated to different management levels of employment.

The first line of the table presents the distribution of job advertisements according to the management level for all the advertisements included in the analysis, with most of the job advertisements falling under the category of Associate (29%), Entry level (32%) and Mid-Senior level (27%) positions. A smaller number of job positions are for Director (5%) and Executive (2%) positions.

It can be noted that some phrases are more often related to Director and Executive positions than others are, such as “project management”, “digital transformation”, and “quality management”, which are general knowledge phrases. The phrases more often related to Associate position are in most of the cases reflecting specific knowledge, such as “big data”, “software development”, “product development”, “digital manufacturing”, “data analytics”, “artificial intelligence”; however, several general knowledge phrases also occur for Associate position: “digital transformation”, and “digital technologies”. The same conclusion can be made for the Mid-Senior level positions. Phrases “entrepreneurial spirit” and “innovative solutions” are also more often present for Associate and Mid-Senior level positions, indicating that the candidates should not also have specific knowledge of Industry 4.0 solutions, but instead should be able to put that knowledge in the context of new products and services, as well as inter-organizational innovations, such

**Table 5**
Most frequent phrases reflecting knowledge in job advertisements related to Industry 4.0 and Smart factories (+50 cases) according to the management level of employment.
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

| | Associate | Director | Entry level | Executive | Internship | Mid-Senior level | Total |
|---|---|---|---|---|---|---|---|
| TOTAL | 29% | 5% | 32% | 2% | 5% | 27% | 100% |
| SUPPLY CHAIN | 29% | 13% | 12% | 1% | 6% | 38% | 100% |
| PROJECT MANAGEMENT | 24% | 14% | 23% | 3% | 8% | 27% | 100% |
| MACHINE-LEARNING | 24% | 4% | 39% | 1% | 0% | 32% | 100% |
| BIG DATA | 37% | 3% | 32% | 2% | 1% | 26% | 100% |
| SOFTWARE DEVELOPMENT | 38% | 6% | 28% | 1% | 2% | 25% | 100% |
| PRODUCT DEVELOPMENT | 38% | 1% | 26% | 1% | 6% | 28% | 100% |
| COMPUTER SCIENCE | 30% | 4% | 38% | 1% | 2% | 25% | 100% |
| DIGITAL MANUFACTURING | 42% | 4% | 14% | 0% | 5% | 36% | 100% |
| BUSINESS DEVELOPMENT | 22% | 4% | 10% | 3% | 4% | 58% | 100% |
| DIGITAL TRANSFORMATION | 40% | 6% | 17% | 5% | 2% | 30% | 100% |
| INTERNET OF THINGS | 49% | 3% | 23% | 1% | 2% | 21% | 100% |
| DATA ANALYTICS | 33% | 3% | 34% | 3% | 7% | 20% | 100% |
| CONTINUOUS IMPROVEMENT | 28% | 11% | 32% | 0% | 2% | 26% | 100% |
| ARTIFICIAL INTELLIGENCE | 32% | 1% | 27% | 1% | 4% | 35% | 100% |
| BUSINESS PROCESSES | 29% | 8% | 13% | 5% | 2% | 43% | 100% |
| REAL TIME | 34% | 5% | 33% | 1% | 0% | 26% | 100% |
| INNOVATIVE SOLUTIONS | 32% | 1% | 15% | 3% | 3% | 46% | 100% |
| CONTROL SYSTEMS | 40% | 1% | 45% | 0% | 0% | 14% | 100% |
| ADVANCED MANUFACTURING | 44% | 13% | 25% | 0% | 2% | 16% | 100% |
| ELECTRICAL ENGINEERING | 30% | 2% | 33% | 0% | 10% | 24% | 100% |
| SMART MANUFACTURING | 29% | 5% | 34% | 7% | 2% | 23% | 100% |
| SOFTWARE ENGINEERING | 25% | 0% | 47% | 1% | 4% | 23% | 100% |
| INDUSTRIAL ENGINEERING | 29% | 5% | 28% | 2% | 16% | 20% | 100% |
| LEAN MANUFACTURING | 21% | 29% | 19% | 0% | 13% | 19% | 100% |
| DATA MANAGEMENT | 16% | 4% | 31% | 1% | 4% | 42% | 100% |
| LIFE CYCLE | 29% | 8% | 17% | 5% | 0% | 40% | 100% |
| PRODUCT DESIGN | 15% | 6% | 26% | 3% | 0% | 49% | 100% |
| SUPPLY CHAIN MANAGEMENT | 27% | 5% | 14% | 3% | 3% | 48% | 100% |
| CLOUD COMPUTING | 46% | 3% | 21% | 0% | 3% | 27% | 100% |
| INDUSTRIAL AUTOMATION | 27% | 0% | 43% | 0% | 0% | 30% | 100% |
| DATA SCIENCE | 31% | 2% | 43% | 5% | 5% | 14% | 100% |
| INDUSTRY EXPERIENCE | 34% | 4% | 21% | 3% | 3% | 34% | 100% |
| FACTORY AUTOMATION | 31% | 2% | 33% | 0% | 0% | 35% | 100% |
| DIGITAL TECHNOLOGIES | 35% | 2% | 20% | 0% | 2% | 41% | 100% |
| CUSTOMER SATISFACTION | 26% | 3% | 32% | 0% | 0% | 39% | 100% |
| PRODUCT MANAGEMENT | 25% | 7% | 26% | 0% | 18% | 25% | 100% |
| OPERATIONAL EXCELLENCE | 40% | 0% | 29% | 3% | 0% | 29% | 100% |
| ENTREPRENEURIAL SPIRIT | 34% | 4% | 12% | 1% | 6% | 42% | 100% |
| OBJECT ORIENTED | 28% | 3% | 41% | 2% | 2% | 24% | 100% |
| AGILE DEVELOPMENT | 28% | 0% | 34% | 0% | 0% | 38% | 100% |
| DISCRETE MANUFACTURING | 41% | 5% | 5% | 0% | 0% | 49% | 100% |
| QUALITY MANAGEMENT | 13% | 15% | 43% | 2% | 0% | 26% | 100% |
| DRIVE INNOVATION | 22% | 5% | 3% | 0% | 0% | 70% | 100% |
| DESIGN AND DEVELOPMENT | 31% | 2% | 39% | 0% | 2% | 27% | 100% |
| PRODUCT LIFECYCLE | 32% | 6% | 13% | 0% | 2% | 47% | 100% |
| PERSONALIZED PRODUCTS | 37% | 5% | 23% | 0% | 0% | 35% | 100% |
| ADVANCED ANALYTICS | 27% | 10% | 6% | 2% | 0% | 55% | 100% |
| ENTERPRISE SOFTWARE | 36% | 5% | 11% | 4% | 0% | 44% | 100% |
| OPERATING SYSTEMS | 47% | 0% | 36% | 0% | 2% | 15% | 100% |
| SAP DIGITAL | 33% | 2% | 2% | 0% | 0% | 63% | 100% |
| ENGINEERING EXPERIENCE | 29% | 4% | 17% | 0% | 0% | 50% | 100% |
| ENGINEERING SERVICES | 26% | 0% | 17% | 0% | 0% | 57% | 100% |

as process innovations.

Fig. 3 presents the distribution of job advertisements according to the management level within the cluster. It can be noted that in most of the jobs, the prevailing management level is mid-senior level (Cluster 1, Cluster 2, Cluster 7, and Cluster 8). In Cluster 3 and Cluster 5, the prevailing management levels are Entry Level. In Cluster 4 and Cluster 6, the prevailing management level is Associate. Directors are the most prevailed in Cluster 1 (7%), Cluster 2 (9%), and Cluster 7 (7%), and Executive in Cluster 1 (5%). It is worth to note that in these clusters, most of the general knowledge phrases are present.

## 5. Discussion

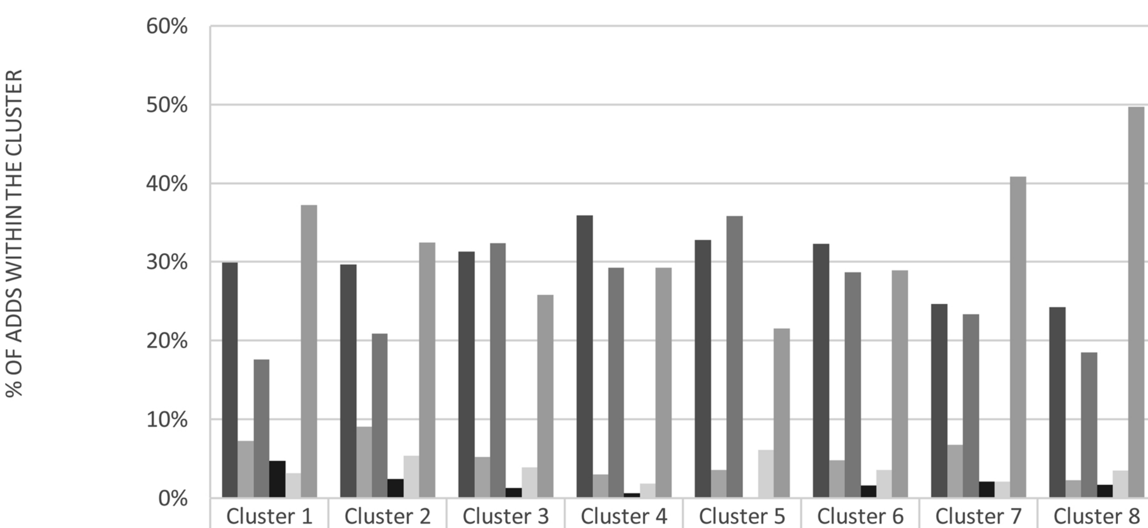
The objective of this work was to provide a detailed insight into the knowledge and skills relevant for the Industry 4.0 organizations. The first contribution of our work is in the area of Industry 4.0 relevant knowledge and skills. The previous work is based mostly on secondary sources and qualitative research. Pecina and Sladek (2017) used the literature review approach in order to analyse the impact of new technologies, such as robotics, virtual and augmented reality, automation, and 3D technologies, on engineering education. Marnewick and Marnewick (2019) also used the literature review approach in order to reveal the needed competencies of project managers and project team members in Industry 4.0 organizations. Hecklau et al. (2016) used a literature review on the challenges in Industry 4.0 organizations in order to develop four groups of competencies: technical, methodological, social, and personal. Only a few papers are using empirical research in order to develop Industry 4.0 competencies. For example,

**Table 6**
Job profiles extracted.
Source: Authors' work, based on job advertisements extracted from LinkedIn, 2018

| Cluster | Job profile | Phrases | Distribution of seniority level |
|---|---|---|---|
| C1 | Supply chain analyst | “advanced analytics“, “supply chain management“, “digital technology“, “digital transformation“, “drive innovation“, and “business processes. “ | Associate (30%), Director (7%), Entry level (18%), Executive (5%), Internship (3%), and **Mid-Senior (37%)** |
| C2 | Supply chain engineer | “continuous improvement“, “project management“, “lean manufacturing“, “operational excellence“, and “supply chain. “ | **Associate (30%)**, Director (9%), Entry level (21%), Executive (2%), Internship (5%), and Mid-Senior (26%) |
| C3 | Cyberphysical systems (CPS) and the Internet of things (IoT) for a robotized production engineer | “advanced manufacturing“, “big data“, “artificial intelligence“, “cloud computing“, “internet of things“, “data analytics“, “machine-learning“, “computer science“, “electrical engineering“, “industrial engineering“, “smart manufacturing“, “software engineering“, and “data science“ | **Associate (31%)**, Director (5%), Entry level (32%), Executive (1%), Internship (4%), and Mid-Senior (26%). |
| C4 | Digital manufacturing engineer | “digital manufacturing“, “personalized products“, “industrial automation“, “operating systems“, and “real time “ | **Associate (36%)**, Director (3%), Entry level (29%), Executive (1%), Internship (2%), and Mid-Senior (29%) |
| C5 | Smart product designer | “control systems“, “product management“, and “design and development“ | Associate (33%), Director (4%), **Entry level (36%)**, Executive (0%), Internship (6%), and Mid-Senior (22%) |
| C6 | ICT specialist for factory automation | “data management“, “object oriented“, “engineering experience“, “product design“, “product development“, “product lifecycle“, “factory automation“, “software development“, and “life cycle“ | **Associate (32%)**, Director (5%), Entry level (29%), Executive (2%), Internship (4%), and Mid-Senior (29%) |
| C7 | Customer satisfaction manager | “business development“, “customer satisfaction“, and “quality management“ | Associate (24%), Director (2%), Entry level (18%), Executive (2%), Internship (3%), and **Mid-Senior (50%)** |
| C8 | Enterprise software specialist | “agile development“, “discrete manufacturing“, “SAP digital“, “enterprise software“, and “entrepreneurial spirit“ | Associate (25%), Director (7%), Entry level (23%), Executive (2%), Internship (2%), and **Mid-Senior (41%)** |

Pinzone et al. (2017) used literature review, in-depth interviews and focus groups with managers from organizations operating in Industry 4.0, in order to in developing the job profiles relevant for these organizations. To the best of our knowledge, our study is the first to use machine-learning alogirthms, particularly hierarchical cluster analysis, to study job advertisements in regards to Industry 4.0 and smart manufacturing. As such, our study contributes to the growing field of research into Industry 4.0 and smart manufacturing, by shedding light into the most frequent phrases that occur in job advertisments, developing job profiles for Industry 4.0 organizations.

Pinzone et al. (2017) identified the five job profiles relevant for Industry 4.0: Operations Management, Supply Chain Management, Product-Service Innovation Management, Information technology-Operational technology (IT-OT) Integration Management, and Data Science Managment. Table 7 presents the overlapping of Pinzone et al. (2017) approach and job profiles extracted from the job advertisements




| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 | Cluster 8 |
|---|---|---|---|---|---|---|---|---|
| Associate | 30% | 30% | 31% | 36% | 33% | 32% | 25% | 24% |
| Director | 7% | 9% | 5% | 3% | 4% | 5% | 7% | 2% |
| Entry Level | 18% | 21% | 32% | 29% | 36% | 29% | 23% | 18% |
| Executive | 5% | 2% | 1% | 1% | 0% | 2% | 2% | 2% |
| Internship | 3% | 5% | 4% | 2% | 6% | 4% | 2% | 3% |
| Mid-Senior Level | 37% | 32% | 26% | 29% | 22% | 29% | 41% | 50% |

**Fig. 3.** Distribution of job advertisements according to management level within the clusters.
Source: Authors work, based on job advertisements extracted from LinkedIn, 2018

**Table 7**
Job profiles according to the main Industry 4.0 areas defined by Pinzone et al. (2017).
Source: Authors' work

| Job profiles | Industry 4.0 area according to Pinzone et al. (2017) | | | | |
|---|---|---|---|---|---|
| | Operations Management | Supply Chain Management | Product-Service Innovation Management | Data Science Management | IT-OT Integration Management |
| Supply Chain Analyst | | ✓ | | ✓ | |
| Supply Chain Engineer | ✓ | ✓ | | | |
| Cyberphysical Systems (CPS) and the Internet of Things (IoT) for a Robotized Production Engineer | | | | ✓ | ✓ |
| Digital Manufacturing Engineer | ✓ | | | | |
| Smart Product Designer | ✓ | | | | |
| ICT Specialist For Factory Automation | ✓ | | | | |
| Customer Satisfaction Manager | | | ✓ | | |
| Enterprise Software Specialist | ✓ | | | | |

in this study. Several differences are present. For example, our analysis divides the area of the supply chain into two job profiles: a supply chain analyst and supply chain engineer. In addition, these two job profiles combine the area of supply chain management, operation management, and data science management. This result indicates that current approach that is using literature review in order to develop job profiles for Industry 4.0 organizations is limited since it does not take into account multidiscliplinarity that is highly relevant for the organization of work in such organizations.

The second contribution refers to the multidiscliplinarity of the skills and knowledge required for the jobs in Industry 4.0 organizations. The analysis of the phrases for the job profile Cyberphysical Systems (CPS) and the Internet of Things (IoT) for Robotized Production Engineer indicates that professionals aiming to be employed in this area should have skills typical for three areas (industrial engineering, electrical engineering, and software engineering). Software engineers should have skills, which are more traditional in this area, such as managing gateways, cloud software, and databases, but also skills related to smart manufacturing, such as programming CPS devices. Industrial engineers should have competencies that are related to traditional production, such as optimization and machine maintenance, but from the perspective of augmented reality systems and human-robot collaborations. These specialists should also be able to use simulation modelling, which is typically a software engineer skill. Electrical engineering use software engineering skills for the conduction of typical tasks for the smart factory, such as IoT device management (Benešová & Tupa, 2017; Fernández-Miranda et al., 2017; Taivalsaari & Mikkonen, 2018). This overlapping is likely to results in a decrease in some of the traditional skills required of industrial engineers, due to the automation of production, mostly in regards to production planning, optimization of repetitive and simple tasks, traditional process control and manual parts assembly (Sackley & Bester, 2016).

Third, we extend the work of previous authors that used LinkedIn for their research. LinkedIn is an example of a big data ecosystem, which contains various information related to careers, such as professionals' profiles, organization profiles, networking groups, and job advertisements (Sumbaly, Kreps, & Shah, 2013). LinkedIn has been used in several research papers for developing new knowledge based on its users' profiles (e.g., Dai, Nespereira, Vilas, & Redondo, 2015; Dai, Vilas, & Redondo, 2017a; Dai, Vilas, & Redondo, 2017b). However, in their work, authors have been focused on user profiles, but not to job advertisements published over LinkedIn.

Fourth, we also contribute to the area of using text mining as a weak-signal that can provide an insight into emerging knowledge and skills. Yoon (2012) developed a text mining application that is used for the identification of weak signals based on keyword extraction. They used the application on the sample of Web news articles, which are related to solar cells; since the method revealed a convincing result, it is proposed to include it into long-term planning in order to support business experts in the identification of new trends based on weak signals. Although these methods are still in their early use, they show potential as a way to support decision-making, that is fact-based and rely less on intuition and its accompanying biases (Chamorro-Premuzic et al., 2016). Our contribution follows their work with providing the successful example of using text mining for detection of weak signals from the job advertisement analysis.

### 5.1. Implications for practice

Our research has important managerial implications for Industry 4.0 practitioners. Job advertisements are a relevant source of information that could be utilized by three groups of stakeholders: human resource professionals, higher educational institutions, and experts already employed or aiming to be employed in Industry 4.0 organizations. These stakeholders can benefit from the results of our research, but can also engage in developing their text mining project themselves.

The presented methodology in our study is described in details and is fairy simple, but useful, and it can be used by human resource practitioners to do their own text mining of relevant job postings. We propose three types of analysis: word frequency, phrase extraction, and topic extrations. Each of these anylysis provides a different type of useful information. The most frequent words and phrases can be used for the detection of the most often required skills of the jobs in Industry 4.0 organizations. On the other hand, the least freqent words and phrases can be used for the detection of emerging and rare skills, but which could be crucial for the competitiveness of Industry 4.0 organizations, i.e. those organizations which employ experts with rare skills may be more efficient in various tasks neded for the fast reactions at the market, such as data mining.

Practical implications of our research are primarily focused on the job of human resource professionals, who can benefit from our detection of prevalent job profiles required for Industry 4.0 organizations: Supply Chain Analyst, Supply Chain Engineer, Cyberphysical Systems (CPS) and the Internet of Things (IoT) for a Robotized Production Engineer, Digital Manufacturing Engineer, Smart Product Designer, ICT Specialist For Factory Automation, Customer Satisfaction Manager, and Enterprise Software Specialist. Using this job profile, human resource professionals in Industry 4.0 organizations can make more informed employment decisions. As previously proposed, they can also use the text mining software in order to develop a text mining project for their specific purpose, e.g., focusing to a narrower market (e.g., Germany or South Korea), or to a narrower type of job (Industry 4.0 jobs related to supply chain management). For that purpose, they need to acquire text mining skills which will become a necessity for human resource experts in Industry 4.0 and smart manufacturing, along with an understanding of the key concepts discussed in this research.

Higher educational institutions can also benefit from this research in the development of future programmes and courses. Our results indicated that there is a significant demand for experts with multidisciplinary skills in the three traditional areas (industrial engineering, electrical engineering, and software engineering). Job profile for the Cyberphysical Systems (CPS) and the Internet of Things (IoT) for Robotized Production Engineer combines these three fields. Therefore, new educational programmes should be designed by higher educational institutions, which would include such interdisciplinary knowledge and skills. However, these educational programmes should be both formal (e.g., Master studies), and informal (e.g., Life-long learning online courses).

Experts already employed or aiming to be employed in Industry 4.0 organizations can use the results of this research in order to check the level of their skills relevant for their current or targeted jobs. For example, graduates of industrial engineering may be based on the results of our research, pursue to gain additional skills related to software engineering, such as machine learning or IoT programming. For that purpose, they may participate in the life-long learning online courses, developed by the higher educational institutions, or engage in some of the massive open online courses (MooC) that has recently become a relevant source of education in various fields related to software engineering (Chang, 2016; Huang, Zhang, & Liu, 2017).

## 6. Conclusions

Industry 4.0 has already brought about a restructuring of jobs in manufacturing. For example, like automation of the manufacturing process increases, jobs in assembly and production will truly decrease, as control of assembly and production is taken over by highly educated data scientists and information technology specialists. Therefore, the fast tool that can track the changes in relevant knowledge and skills required in Industry 4.0 organizations is needed. Our study presents a detailed procedure of online job advertisements analysis for the Industry 4.0 jobs, by outlining the data structure, required manual transformation and text mining process (pre-processing, word and phrases extraction, and topic extraction). We have analysed more than 1400 current Industry 4.0 advertisements using the text mining approach, for extracting new job profiles. Extracted job profiles indicate that Industry 4.0 organizations are looking not for the traditionally educated industrial, software or electrical engineers, but for the experts with the multidisciplinary skills, which can help factories to create and improve low-end and high-end machine and device capabilities, using knowledge and skills in embedded and distributed system development. The study contributes to the existing knowledge in the Industry 4.0 by (i) proposing a new typology of job profiles in Industry 4.0; (ii) providing a practical guidance for human resources professionals on how to track changes in knowledge and skills needed in Industry 4.0; (3) providing a current information for higher educational institutions on how to update their curricula; and (4) presenting the usefulness of text mining approach for exploratory analysis of job advertisements.

### 6.1. Limitations

Our research has several limitations. First, the limitation stems from the selection of data source. The presented research has been conducted in a limited time-frame since the historical data of the job advertisements related to Industry 4.0 is not available. This limitation could be overcome by the periodical collection of job advertisements during a longer period, e.g., several years. In addition, we analysed the job advertisements published on one website (LinkedIn). Although LinkedIn is one of the leading websites for advertising jobs, the selection of this source increases the risk of bias. Instead of just using LinkedIn.com, our study, as well as future studies, could have used other job postings websitres, such as Monster, Dice, Indeed, glassdoor and careerbuilder (Radovilsky, Hegde, Acharya, & Uma, 2018). In addition, we analysed only the job advertisements published in English, which could be overcome in the future with the usage of automatic translation tools. Therefore, future research should include more websites that publish job advertisements, including those published in languages other than English.

Second limitation results from the usage of job advertisements as the source of the data for making conclusions about the knowledge and skills required in Industry 4.0 organizations. The text of job advertisement can be biased itself. For example, the text of the job advertisement could be written using the phrases that human resource specialists could consider as understandable by the potential job candidates, without consulting the domain experts in Industry 4.0. For example, Granville (2014) points out that management and human resources sometimes do not fully understand what data science is, or what skills are needed for big data analysis, resulting in job advertisements that are not specific enough or do not list relevant skills that are needed for the data scientists in the Industry 4.0 organization. Due to this, the results of our study could be biased in that they do not represent specific skills and experience expected from job seekers, but instead, represent a basic idea of what skill could be needed by organizations for their Industry 4.0 needs.

The third limitation arises from the usage of our methodological approach. Although the text mining has many advantages over the manual data analysis, the text mining process itself and the usage of its algorithms can also create bias. Bias can be the result of various decisions in different stages of text mining analysis. First, in our work, we pre-processed the initial list of extracted words with lemmatization approach, while it is also possible to use stemming. However, we have conducted the initial analysis with both approaches, and lemmatization approach resulted in richer information in relation to extracted phrases. Second, we decided that the cut-off for most frequent phrases in job advertisements would be 50, due to many phrases repeating themselves as synonyms, however. As a result, a number of the less frequent phrases were lost, which could have been of interest to the reader of this paper. Nonetheless, the authors believe that the most important frequent phrases were included in the study. Third, in our work we use Unweighted Arithmetic Average Clustering, while other approaches have also been proved useful for the purpose of topic extractions from job advertising, such as Latent Semantic Analysis (e.g., Müller et al., 2016). Fourth, results that are more comprehensive could stem from the combination of different methodologies. For example, by using a Delphi study with experts from various Industry 4.0 organizations and regions, this study could benefit from the relevant expert knowledge, and at the same time reduce bias.

### 6.2. Implications for research

Several opportunities for future research are opened by this study. First, phrases typical for Industry 4.0 relevant knowledge and skills are empirically extracted from online job advertisements, which could be considered as an initial dictionary of such skills. However, we analysed job advertisements only from one website during a limited period. Therefore, our dictionary provides a snapshot of the current state in this area. Several authors indicate that the development of data dictionaries based on the text mining of job advertisements would be a useful tool that would allow the tracking of changes in rapidly developing industries (Granville, 2014; Amado et al., 2018). In order to provide a sustainable tool with minimum bias, which could be used by human resources professionals and higher educational institutions, our approach should be repeated over several years, using more online sites that publish job advertisements, in various languages. Such an approach would allow that trends in Industry 4.0 jobs are identified in a timely manner.

Our approach did not allow complete automation. In our work, we used manual coding of countries, recoding of several fields (e.g., Employment type) and manual identification of languages. However, this approach is time consuming and allows the analysis of smaller samples of job advertisements, for which reasons such analysis are less likely to be repeated on a regular basis. In addition, manual filtering of the words and phrases increases the likelihood of bias in our research. Future research should aim towards the examination of various approaches towards these challenges (e.g., automatic language recognition, building a dictionary of general phrases in job advertisements, which are not specific for any industry).

We have identified knowledge and skills in the job advertisements that describe soft-skills and knowledge relevant for a specific area. Our results indicate that most of the job profiles are described using the soft-skills or general skills phrases (such as project management), besides the phrases typical for mechanical, software or electrical engineering. Soft skills were also more often detected in job profiles for which the higher seniority level of employment was indicated. This indicates the need for the development of the dictionary of soft skills relevant for a specific Industry 4.0 jobs. Such a dictionary could be developed by using further analysis of a broader database of job advertisements collected from various websites, using unsupervised text mining. The usage of the existing catalogue of soft skills and matching them with the descriptions in job advertisements, coupled with the qualitative research (case studies or Delphi study) could add the value to such a dictionary.

Finally, our approach could be used for detecting emerging jobs profiles in specific branches of Industry 4.0, such as the automotive or aviation industries.

## Declaration of Competing Interest

None.

## References

Achenhagen, C., & Zeller, B. (2011). *Zukünftige Qualifikationsanforderungen im "Internet der Dinge" in der industriellen Produktion.* (Accessed 8 December 2018) http://www.frequenz.net/uploads/tx_freqprojerg/Summary_indProd_final.pdf.

Ahsan, K., Ho, M., & Khan, S. (2013). Recruiting project managers: A comparative analysis of competencies and recruitment signals from job advertisements. *Project Management Journal, 44*(5), 36–54. https://doi.org/10.1002/pmj.21366.

AlAlwan, A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics, 34*(7), 1177–1190. https://doi.org/10.1016/j.tele.2017.05.008.

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics, 24*(1), 1–7. https://doi.org/10.1016/j.iedeen.2017.06.002.

Amato, F., Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., Moscato, V., et al. (2015). Challenge: Processing web texts for classifying job offers. *Semantic Computing (ICSC), 2015 IEEE International Conference on Semantic Computing,* 460–463.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capablities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review, 114*(6), 416–436. https://doi.org/10.1016/j.tre.2017.04.001.

Barger, M., & Gilbert, R. (2018). *Growth of 2-year programs for mechatronics. 2018 ASEE Southeastern Section Conference*. (Accessed 8 December 2018) http://www.aseese.org/proceedings/ASEE2018/papers2018/11.pdf.

Barr, M., & Massa, A. (2006). *Programming embedded systems with C and GNU development tools.* Sebstopol, CA: O'Reilly.

Benešová, A., & Tupa, J. (2017). Requirements for education and qualification of people in industry 4.0. *Procedia Manufacturing, 11*, 2195–2202. https://doi.org/10.1016/j.promfg.2017.07.366.

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2018). Classifying online job advertisements through machine learning. *Future Generation Computer Systems, 86*(9), 319–328. https://doi.org/10.1016/j.future.2018.03.035.

Bradbury, D. (2011). Data mining with LinkedIn. *Computer Fraud & Security, 2011*(10), 5–8. https://doi.org/10.1016/s1361-3723(11)70101-4.

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2014). How virtualization, decentralization and network building change the manufacturing landscape: An industry 4.0 perspective. *International Journal of Information and Communication Engineering, 8*(1), 37–44.

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world. *Industrial and Organizational Psychology, 9*(3), 621–640. https://doi.org/10.1017/iop.2016.6.

Chang, H. C., Wang, C. Y., & Hawamdeh, S. (2018). Emerging trends in data analytics and knowledge management job market: Extending KSA framework. *Journal of Knowledge Management*. https://doi.org/10.1108/JKM-02-2018-0088.

Chang, V. (2016). Review and discussion: E-learning for academia and industry. *International Journal of Information Management, 36*(3), 476–485.

Dai, K., Nespereira, C. G., Vilas, A. F., & Redondo, R. P. D. (2015). Scraping and clustering techniques for the characterization of LinkedIn profiles. *Proceedings of the Fourth International Conference on Information Technology Convergence and Services,* 1–15.

Dai, K., Vilas, A. F., & Redondo, R. P. D. (2017a). *A new MOOCs' recommendation framework based on LinkedIn data. Innovations in smart learning.* Singapore: Springer19–22.

Dai, K., Vilas, A. F., & Redondo, R. P. D. (2017b). The workforce analyzer: Group discovery among LinkedIn public profiles. *Journal of Ambient Intelligence and Humanized Computing,* 1–10.

Day, P. J. H., & Shoemaker, G. S. (2009). How to make sense of weak signals. *MIT Sloan Management Review, 50*(3), 81–89.

de Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2017). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing and Management, 54*(9), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004.

Dwivedi, Y. K., Kapoor, K. K., & Chen, H. (2015). Social media marketing and advertising. *The Marketing Review, 15*(3), 289–309. https://doi.org/10.1362/146934715X14441363377999.

Elhoseny, M., Abdelaziz, A., Salama, A. S., Riad, A. M., Muhammad, K., & Sangaiah, A. K. (2018). A hybrid model of internet of things and cloud computing to manage big data in health services applications. *Future generation computer systems, 86*, 1383–1394. https://doi.org/10.1016/j.future.2018.03.005.

Elshendy, M., & Fronzetti Colladon, A. (2017). Big data analysis of economic news: Hints to forecast macroeconomic indicators. *International Journal of Engineering Business Management, 9*, 1847979017720040. https://doi.org/10.1177/1847979017720040.

Fernández-Miranda, S. S., Marcos, M., Peraltam, M. E., & Aguayo, F. (2017). The challenge of integrating industry 4.0 in the degree of mechanical engineering. *Procedia Manufacturing, 13*(1229), 12436. https://doi.org/10.1016/j.promfg.2017.09.039.

Flynn, J., Dance, S., & Schaefer, D. (2017). ). Industry 4.0 and its potential impact on employment demographics in the UK. *Advances in transdisciplinary engineering, 6*, 239–244.

Foidl, H., & Felderer, M. (2016). Research challenges of industry 4.0 for quality management. In M. Felderer, F. Piazolo, W. Ortner, L. Brehm, & H. J. Hof (Vol. Eds.), *Innovations in Enterprise information systems management and engineering. ERP future 2015. Lecture notes in business information processing: Vol. 245*Cham, CH: Springer. https://doi.org/10.1007/978-3-319-32799-0_10.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gardiner, A., Aasheim, C., Rutner, P., & Williams, S. (2018). Skill requirements in big data: A content analysis of job advertisements. *Journal of Computer Information Systems, 58*(4), 374–384. https://doi.org/10.1080/08874417.2017.1289354.

Gassler, W., Zangerle, E., & Specht, G. (2014). Guided curation of semistructured data in collaboratively-built knowledge bases. *Future Generation Computer Systems, 31*, 111–119. https://doi.org/10.1016/j.future.2013.05.008.

Ghobakhloo, M., & Azar, A. (2018). Business excellence via advanced manufacturing technology and lean-agile manufacturing. *Journal of Manufacturing Technology Management, 29*(1), 2–24. https://doi.org/10.1108/JMTM-03-2017-0049.

Gonzalez, T., Santos, P., Orozco, F., Alcaraz, M., Zaldivar, V., de Obeso, A., et al. (2012). Adaptive employee profile classification for resource planning tool. *2012 Service Research and Innovation Institute Global Conference,* 544–553.

Govindan, K., Cheng, T. C. E., Mishra, N., & Shukla, N. (2018). Big data analytics and application for logistics and supply chain management. *Transportation Research Park E: Logistics and Transportation Review, 114*(6), 343–349. https://doi.org/10.1016/j.tre.2018.03.011.

Granville, V. (2014). *Developing analytic talent: Becoming a data scientist.* Indianapolis: John Wiley & Sons.

Gruber, F. E. (2013). Industry 4.0: A best practice project of the automotive industry. In G. L. Kovács, & D. Kochan (Eds.). *Digital product and process development systems, NEW PROLAMAT 2013 (36-40). Advances in communication technology, 411*Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-41329-2_5.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly, 93*(2), 332–359. https://doi.org/10.1177/1077699016639231.

Hambach, J., Kümmel, K., & Metternich, J. (2017). Development of a digital continuous improvement system for production. *Procedia CIRP, 63*, 330–335. https://doi.org/10.1016/j.procir.2017.03.086.

Harper, R. (2012). The collection and analysis of job advertisements: A review of research methodology. *Library and Information Research, 36*(112), 29–54. https://doi.org/10.29173/lirg499.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464–472. https://doi.org/10.1016/j.ijinfomgt.2015.12.007.

Hecklau, F., Galeitzke, M., Flachs, S., & Kohl, H. (2016). Holistic approach for human resource management in industry 4.0. *Procedia CIRP, 54*, 1–6. https://doi.org/10.1016/j.procir.2016.05.102.

Huang, L., Zhang, J., & Liu, Y. (2017). Antecedents of student MOOC revisit intention: Moderation effect of course difficulty. *International Journal of Information Management, 37*(2), 84–91. https://doi.org/10.1016/j.ijinfomgt.2016.12.002.

Imran, F., & Kantola, J. (2018). Review of industry 4.0 in the light of sociotechnical system theory and competence-based view: A future research agenda for the evolute approach. In J. Kantola, S. Nazir, & T. Barath (Eds.). *International Conference on applied human factors and ergonomics (118-128). Advances in human factors, business management and society, AHFE 2018. Advances in intelligent systems and computing, 783*Cham, CH: Springer. https://doi.org/10.1007/978-3-319-94709-9_12.

Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., & Ivanova, M. (2016). A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *International Journal of Production Research, 54*(2), 386–402. https://doi.org/10.1080/00207543.2014.999958.

Jeong, B., Yoon, J., & Lee, J. M. (2017). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2017.09.009 In press. Corrected Proof.

Jerman, A., Pejić Bach, M., & Bertoncelj, A. (2018). A bibliometric and topic analysis on future competences at smart factories. *Machines, 6*(3), 41. https://doi.org/10.3390/machines6030041.

Jimenez-Marquez, J. L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J. L., & Ruiz-Mezcua, B. (2019). Towards a big data framework for analyzing social media content. *International Journal of Information Management, 44*, 1–12. https://doi.org/10.1016/j.ijinfomgt.2018.09.003.

Kache, F., & Seuring, S. (2015). Challenges and opportunities of digital information of big data analytics and supply chain management. *International Journal of Operations & Production Management, 37*(1), 10–36. https://doi.org/10.1108/IJOPM-02-2015-0078.

Kane, G. C., Palmer, D., Phillips, A. N., & Kiron, D. (2015). Is your business ready for a digital future. *MIT Sloan Management Review, 56*(4), 37–44.

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers, 20*(3), 531–558. https://doi.org/10.1007/s10796-017-9810-y.

Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M. A., Aung, Z., et al. (2016). Data mining approach to monitoring the requirements of the job market: A case study. *Information Systems, 65*(4), 1–6. https://doi.org/10.1016/j.is.2016.10.009.

Kino, Y., Kuroki, H., Machida, T., Furuya, N., & Takano, K. (2017). Text analysis for job matching quality improvement. *Procedia Computer Science, 112*, 1523–1530. https://doi.org/10.1016/j.procs.2017.08.054.

Kolberg, D., & Zühlke, D. (2015). Lean automation enabled by industry 4.0 technologies. *International Federation of Automatic Control, 48*(3), 1870–1875. https://doi.org/10.1016/j.ifacol.2015.06.359.

Kozák, Š., Ružický, E., Štefanovič, J., & Schindler, F. (2018). *Research and education for industry 4.0: Present development. Proceedings of the 29th International Conference 2018 Cybernetics & Informatics (K&I)*1–8.

Kregel, I., Ogonek, N., & Matthies, B. (2019). Competency profiles for lean professionals–an international perspective. *International Journal of Productivity and Performance Management, 68*(2), 423–446. https://doi.org/10.1108/IJPPM-09-2017-0237.

Küpper, D., Heidemann, A., Ströhle, J., Spindelndreier, D., & Knizek, C. (2017). *When lean meets industry 4.0: The next level of operational excellence.* The Boston Consulting Group. (Accessed 23 December 2018) https://www.bcg.com/publications/2017/lean-meets-industry-4.0.aspx/.

Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia Cirp, 16*, 3–8. https://doi.org/10.1016/j.procir.2014.02.001.

Leminen, S., Rajahonka, M., & Westerlund, M. (2017). Actors in the emerging internet of things ecosystems. *International Journal of E-Services and Mobile Applications (IJESMA), 9*(1), 57–75. https://doi.org/10.4018/IJESMA.2017010104.

Lorenz, M., Rüßmann, M., Strack, R., Lueth, K. L., & Bolle, M. (2015). *Man and machine in industry 4.0.* The Boston Consulting Group. (Accessed 23 December 2018) https://www.bcg.com/publications/2015/technology-business-transformation-engineered-products-infrastructure-man-machine-industry-4.aspx/.

Lucke, D., Constantinescu, C., & Westkämper, E. (2008). Smart factory-a step towards the next generation of manufacturing. In M. Mitsuishi, K. Ueda, & F. Kimura (Eds.). *Manufacturing systems and technologies for the new frontier* (pp. 115–118). London: Springer. https://doi.org/10.1007/978-1-84800-267-8_23.

Macurova, P., Ludvik, L., & Žwakova, M. (2017). The driving factors, risks and barriers of the industry 4.0 concept. *Journal of Applied Economic Sciences, 12*(7), 2003–2011.

Marnewick, C., & Marnewick, A. L. (2019). The demands of industry 4.0 on project teams. *IEEE Transactions on Engineering Management. (Eerly Access),* 1–9. https://doi.org/10.1109/TEM.2019.2899350.

Meek, S., Field, S., & Devasia, S. (2003). Mechatronics education in the department of mechanical engineering at the university of Utah. *Mechatronics, 13*, 1–11. https://doi.org/10.1016/S0957-4158(01)00058-7.

Mezzanzanica, M. (2017). Italian web job vacancies for marketing-related professions. *Symphonya. Emerging Issues in Management, 3*(1), 110–124. https://doi.org/10.4468/2015.3.14mezzanzanica.

Morlock, F., Wienbruch, T., Leineweber, S., Kreimeier, D., & Kuhlenkoetter, B. (2016). Industry transformation of manufacturing companies – Maturity based migration to cyber-physical-manufacturing system. *ZWF Zeitschrift fuer wirtschaftlichen fabrikbetrieb, 111*(5), 306–309. https://doi.org/10.3139/104.111514.

Müller, O., Schmiedel, T., Gorbacheva, E., & vom Brocke, J. (2016). Towards a typology of business process management professionals: Identifying patterns of competences through latent semantic analysis. *Enterprise Information Systems, 10*(1), 50–80. https://doi.org/10.1080/17517575.2014.923514.

Palshikar, G. K., Srivastava, R., Pawar, S., Hingmire, S., Jain, A., Chourasia, S., et al. (2019). *Analytics-led talent acquisition for improving efficiency and effectiveness. Advances in analytics and applications*. Singapore: Springer141–160. https://doi.org/10.1007/978-981-13-1208-3_13.

Pecina, P., & Sladek, P. (2017). Fourth industrial revolution and technical education. L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.). *Proceeding of 11th International Conference on Technology, Education and Development (INTED),* 2089–2093. https://doi.org/10.21125/inted.2017.0621.

Pejić Bach, M., Krstić, Ž., Seljan, S., & Turulja, L. (2019). Text mining for big data analysis in financial sector: A literature review. *Sustainability, 11*(5), 1277. https://doi.org/10.3390/su11051277.

Pinzone, M., Fantini, P., Perini, S., Garavaglia, S., Taisch, M., & Miragliotta, G. (2017). Jobs and skills in industry 4.0: An exploratory research. *IFIP International Conference on Advances in Production Management Systems,* 282–288. https://doi.org/10.1007/978-3-319-66923-6_33.

Radovilsky, Z., Hegde, V., Acharya, A., & Uma, U. (2018). Skills requirements of business data analytics and data science jobs: A comparative analysis. *Journal of Supply Chain and Operations Management, 16*(1), 82–101.

Radziwon, A., Bilberg, A., Bogers, M., & Madsen, E. S. (2014). The smart factory: Exploring adaptive and flexible manufacturing solutions. *Procedia Engineering, 69*, 1184–1190. https://doi.org/10.1016/j.proeng.2014.03.108.

Sackley, S. M., & Bester, A. (2016). industrial engineering curriculum in industry 4.0: In a South African context. *South African Journal of Industrial Engineering, 27*(4), 101–114. https://doi.org/10.7166/27-4-1579.

Sanders, A., Elangeswaran, C., & Wulfsberg, J. (2016). Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing. *Journal of Industrial Engineering and Management, 9*(3), 811–833. https://doi.org/10.3926/jiem.1940.

SAP (2017b). *SAP Leonardo for Discrete Manufacturing, Service and Assets Option.* SAP. (Accessed 23 December 2018) https://www.sap.com/documents/2017/08/1871b631-d07c-0010-82c7-eda71af511fa.html/.

SAP (2017a). *SAP Leonardo for discrete manufacturing, logistics option.* (Accessed 23 December 2018) https://www.sap.com/documents/2017/08/8c9dba32-d07c-0010-82c7-eda71af511fa.html/.

SAP (2018). *SAP digital manufacturing insights | manufacturing process analytics.* (Accessed 23 December 2018) https://www.sap.com/products/digital-mfg-insights.html/.

Schuh, G., Potente, T., Wesch-Potente, C., Weber, A. R., & Prote, J. P. (2014). Collaboration mechanisms to increase productivity in the context of industrie 4.0. *Procedia CIRP, 19*, 51–56. https://doi.org/10.1016/j.procir.2014.05.016.

Shiau, W. L., Dwivedi, Y. K., & Yang, H. S. (2017). Co-citation and cluster analyses of extant literature on social networks. *International Journal of Information Management, 37*(5), 390–399. https://doi.org/10.1016/j.ijinfomgt.2017.04.007.

Shiau, W.-L., Dwivedi, Y. K., & Lai, H.-H. (2018). Examining the core knowledge on Facebook. *International Journal of Information Management, 43*, 52–63. https://doi.org/10.1016/j.ijinfomgt.2018.06.006.

Skeels, M. M., & Grudin, J. (2009). When social networks cross boundaries: A case study of workplace use of Facebook and Linkedin. *Proceedings of the ACM 2009 International Conference on Supporting Group Work,* 95–104. https://doi.org/10.1145/1531674.1531689.

Sommer, L. (2015). Industrial revolution-industry 4.0: Are German manufacturing SMEs the first victims of this revolution? *Journal of Industrial Engineering and Management, 8*(5), 1512–1532. https://doi.org/10.3926/jiem.1470.

Soulier, P., Li, D., & Williams, J. R. (2015). A survey of language-based approaches to cyber-physical and embedded system development. *Tsinghua Science and Technology, 20*(2), 130–141. https://doi.org/10.1109/TST.2015.7085626.

Stojmenovic, I. (2014). Machine-to-machine communications with in-network data aggregation, processing, and actuation for large-scale cyberphysical systems. *IEEE Internet of Things Journal, 1*(2), 122–128. https://doi.org/10.1109/JIOT.2014.2311693.

Sumbaly, R., Kreps, J., & Shah, S. (2013). The big data ecosystem at Linkedin. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data,* 1125–1134.

SYSTEMA (2018). *SAP manufacturing and SAP Leonardo – Industry 4.0.* SYSTEMA. Accessed 23 December 2018 https://www.systema.solutions/sap-manufacturing/.

Taivalsaari, A., & Mikkonen, T. (2018). *On the developments of IoT systems. Third International Conference on Fog and Mobile Edge Computing (FMEC)*.

Todd, P. A., McKeen, J. D., & Gallupe, R. B. (1995). The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990. *MIS quarterly,* 1–27.

van Gerven, M. (2017). Computational foundations of natural intelligence. *Frontiers in Computational Neuroscience, 11*(12), 7–30. https://dx.doi.org/10.3389/fncom.2017.00112.

Vogel-Heuser, B., & Hess, D. (2016). Guest editorial industry 4.0–prerequisites and visions. *IEEE Transactions on Automation Science and Engineering, 13*(2), 411–413. https://doi.org/10.1109/TASE.2016.2523639.

Wan, J., Yan, H., Liu, Q., Zhou, K., Lu, R., & Li, D. (2013). Enabling cyber-physical systems with machine-to-machine technologies. *International Journal of Ad Hoc and Ubiquitous Computing, 13*(3/4), 187–196. https://doi.org/10.1504/IJAHUC.2013.055454.

Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factory of industry 4.0: An outlook. *International Journal of Distributed Sensor Networks, 12*(1), 1–10. https://doi.org/10.1155/2016/3159805.

Weber, E. (2016). *Industry 4.0: Job-producer or employmentdestroyer?, Aktuelle Berichte, No. 2/2016.* Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Xu, X., Wang, X., Li, Y., & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of information management, 37*(6), 673–683. https://doi.org/10.1016/j.ijinfomgt.2017.06.004.

Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of web news. *Expert Systems with Applications, 39*(16), 12543–12550. https://doi.org/10.1016/j.eswa.2012.04.059.

Yoon, J. S., Shin, J. S., & Suh, S. H. (2012). A conceptual framework for the ubiquitous factory. *International Journal of Production Research, 50*(8), 2174–2189. https://doi.org/10.1080/00207543.2011.562563.

Zawadzki, P., & Żywicki, K. (2016). Smart product design and production control for effective mass customization in the industry 4.0 concept. *Management and Production Engineering Review, 7*(3), 105–112. https://doi.org/10.1515/mper-2016-0030.

Žmuk, B. (2017). Are publicly available online businesses lists appropriate to be used as sampling frames in Croatian business surveys? *Business systems research journal: international journal of the Society for Advancing Business & Information Technology (BIT), 8*(2), 26–39. https://doi.org/10.1515/bsrj-2017-0014.