International Journal of Engineering Education Vol. 33, No. 4, pp. 1361–1368, 2017
Printed in Great Britain
0949-149X/91 $3.00+0.00

Analysis of Software Engineering Industry Needs and Trends: Implications for Education*

FATIH GURCAN and CEMAL KOSE
Karadeniz Technical University, Department of Computer Engineering, Faculty of Engineering, 61080, Trabzon, Turkey.
E-mail: fgurcan@ktu.edu.tr, ckose@ktu.edu.tr

In modern day software development environments, analysis and understanding of the emerging industry needs is of strategic importance for a more effective software engineering (SE) education that is innovative and responsive to changing industry needs. Considering the demand for well-trained software engineers in the near future, an empirical study was performed on SE job postings in order to identify the emerging needs and trends in the software industry. The methodology of this study was based on semantic topic analysis implemented by latent Dirichlet allocation (LDA), a probabilistic generative approach for topic modeling. The findings of the study indicated that, the software industry has a wide spectrum in terms of professional roles, responsibilities (in-demand skills) and combinations of programming languages. Each of the professional roles is profoundly based on specific skill sets that reflect the dynamics of the software industry. Also, the topics discovered by LDA highlighted a broad range of the characteristics of the SE, such as contemporary trends, demands, skills, tools, platforms, methodologies, and technologies that indicate the level of progress in this dynamic field. In light of these findings, an innovative academic curriculum for SE education can be designed consistent with the emerging needs and trends in the software industry. In this regard, the findings can provide valuable implications for the industry, academia, and SE community to close the gap between the industry needs and the current SE education.

Keywords: software engineering education; software engineer skills; software industry needs; topic modeling; latent Dirichlet allocation

1. Introduction

The 21st century has a significant progress in the information technology (IT). Therefore, the 21st century is defined as the information age. Consistent with the developments in the IT, it has experienced a major evolution in software development technologies. At the present time, the wide spectrum of software applications is used effectively in every phase of daily life. From this perspective, software engineering (SE) plays a crucial role in this lifecycle as an engineering discipline based on the application of engineering practices to software development process [1–3]. In the new world order led by the IT, understanding of emerging needs and trends in dynamic software industry is a strategically key factor for the SE discipline in order to keep pace with industrial modernizations [3, 4]. The software industry has the dynamic, entrepreneurial, and collaborative working environments in which all processes are based on the cognitive labor force, and so human resources are used effectively [2, 3]. In these working environments, as the leadings actors, software engineers are expected to have a wide spectrum of roles, responsibilities, and skills frequently changing. Besides, contemporary software development process requires to use of the different combinations of programming languages [5]. For this reason, the software engineers should always keep their knowledge and skills up to date [6]. As the software industry advances, new career opportunities are opening up every day for the software engineers. Online employment platforms (web sites) are intensively used by employees and employers in order to provide the interactions between them [5]. The volume and variety of shared information in these platforms are ever-increasing in recent times due to this intensive usage.

Numerous SE jobs are published every day on the platforms. From this perspective, the SE job postings can be seen as an indicator of the industry needs and trends in this field [5–7]. Therefore, the studies based on analysis of the SE jobs and determination of the needs and trends may provide valuable contributions for the engineers, instructors, and companies in the SE field.

Given this background, numerous studies were performed for the determination of professional qualifications required for IT workforce by analyzing online job postings [5–8]. The studies based on analysis of the job postings can reveal the up-to-date industry needs and trends related to the SE field. Given this background, the education of software engineers consistent with industrial needs and trends is seen as an essential open question in terms of future of the SE discipline [2]. In this regard, the gap between software industry needs and academic preparation were discussed in a number of studies [2–9], and various approaches were proposed in order to close this gap by developing new learning models [10] and methodologies [11], investigating the industry needs and trends

* Accepted 10 February 2017.

[12, 13] analyzing educational requirements for software engineers [14], determining professional roles and responsibilities (in-demand skills) for software engineers [5–8], and leveraging SE practices [1]. In most of the above-mentioned studies, the traditional content analysis techniques (without sematic analysis or probabilistic topic modeling approaches) were used to discover industry needs and trends. In this dynamic framework, more research is needed to analysis and interpretation of these emerging needs and trends [2]. In particular, the supplementary studies based on generative models, semantic analysis, and topic modeling will contribute to SE research and practice in approved manner.

In this study, we investigated the emerging roles, responsibilities, trends, and demands for software engineers by analyzing the texts of the SE job postings. The background of this study was based five focal points: (1) identification of the professional roles and responsibilities of software engineers, (2) determination of the most in-demand combinations of the programming languages used in today's software development environments, (3) identification of the educational requirements for software engineers, (4) detection of trending topics at a high-granularity level in the SE jobs, and finally in light of these findings, (5) providing of valuable contributions and insights for the design of an innovative SE curriculum consistent with the emerging trends and demands in the software industry. Based on this purpose and scope, an empirical topic analysis was implemented on SE jobs using a generative topic-modeling approach called as latent Dirichlet allocation (LDA) [15]. In this analysis performed by LDA, the 30 latent topics were discovered at optimal level and these topics have enabled us to carry out the qualitative and quantitative evaluations about the SE trends. The findings of the study demonstrated that today's software engineers are expected to undertake the wide spectrum of roles and responsibilities. From this point of view, the software engineers are characterized by the roles and different combinations of the responsibilities (in-demand skills). The topics discovered by LDA highlighted a broad range of the characteristics of the SE, such as contemporary trends, demands, skills, tools, platforms, and technologies that reflect the level of progress in this dynamic field.

Our LDA-based topic analysis can provide valuable contributions to better understanding of the changing nature of the SE trends. The findings of this study can be helpful for (1) software engineers to evaluate and update their individual capabilities, (2) software corporations to select and employ the qualified software engineers, (3) educational institutions to design SE programs and core curriculum consistent with emerging needs and trends, and (4) students interested in SE to design their future careers. The rest of the paper is organized as follows. The research methodology and research data are included in Section 2. The results are shown in Section 3. The findings are discussed in Section 4. Finally, the conclusions, limitations and future work are given in Section 5.

2. Research methodology

The research methodology of this study was based on semantic topic analysis of the SE job postings using LDA-based topic modeling, a quantitative approach to analyze qualitative data [15, 16]. The methodology was designed according to the focal points of the study and consisted of a number of sequential phases as shown in Fig. 1. Initially, the job postings were collected and the dataset was created. Next, the data preprocessing steps were implemented to dimensional reduction and to increase the success of the analysis. In order to perform the numerical analysis on the dataset, the document-term matrix (DTM) were created. After this process, the semantic analysis was performed by implementing LDA-based topic modeling on the DTM to discover latent topics. Finally, the results of the analysis were presented and empirical findings were discussed in light of related studies. In the subsequent section, each phase of the methodology is described in more detail.

2.1 Data collection and preprocessing

Although there are numerous tech-focused job boards, considering the purpose of the study,

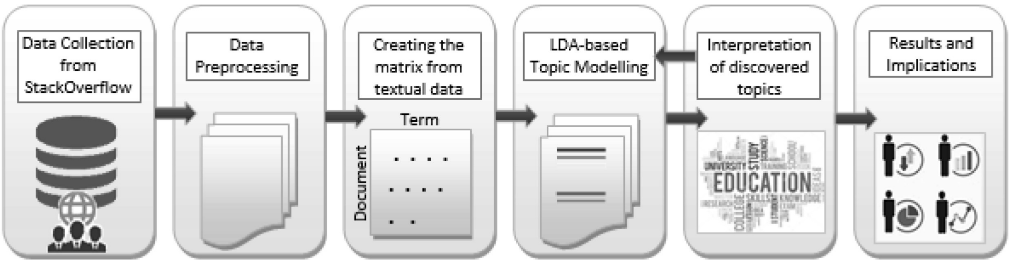

Fig. 1. An overview of the research methodology.

Stack Overflow Careers [17] was selected as the data source. In this selection, two main criteria were taken into account for the study. The first criterion was that the board selected as a data source should be only related to SE field. The second criterion was that the board includes the jobs from different countries. Besides, Stack Overflow [18] is a popular question-answer sharing platform and an intensive interaction platform used by software engineers [13]. For this reason, the job postings on this board are followed and discussed by the engineers in a comprehensive manner. In this context, our dataset consisted of 2533 unique SE job postings. Time period of the data was six months, from January 2016 to June 2016. In the data set, a typical job posting contained various information such as roles, responsibilities, location, major skills, requirements, and job description.

After the data collection phase, the data preprocessing was performed on the textual dataset. Data preprocessing is an essential process used to increase the quality of the analysis in text mining and information retrieval [19]. In this context, the preprocessing performed in this study consisted of four steps implemented consecutively. That is, tokenization, lowercase conversion, deleting special characters and stop words (is, a, the, for etc.). On the other hand, no stemming process was implemented in order to avoid any loss of sense for the reason that the textual context of the empirical data consisted of numerous technical terms. In the last step prior to the topic analysis, the textual data was converted into DTM in order to perform numerical topic analysis on the data. The DTM is a two-dimensional mathematical matrix that describes the frequency of words that occur in a collection of documents. In the DTM, rows represent the documents in the collection and columns represent the words in a document. The DTM is commonly used as the primary input in the text mining process [20]. After the preprocessing steps described above, the dataset was represented by the lower dimensional DTM.

2.2 Topic analysis and interpretation

Text documents consist of latent semantic structures, which are called "topics". In topic analysis, each text document is represented by a combination of topics and each topic is represented by frequently co-occurring words having a probability distribution [15, 16]. In this study, latent Dirichlet allocation (LDA) [15], a generative topic modeling approach, was used to discover emerging needs and trends in the software industry. LDA-based topic modeling is effectively used for the semantic analysis of document collections in text mining. Learning in LDA model is unsupervised, so millions of textual documents can be analyzed in a short time [16, 21]. For these reasons, LDA is the most suitable method for the detection of trending topics in our empirical dataset.

In our experiments, we used the LDA implementation of the MALLET [22] open source software that is used for statistical natural language processing and the topic modeling. MALLET uses Gibbs sampling algorithm [23] for parameter estimation. We implemented the MALLET with 1500 iterations of Gibbs sampling for each experiment. The number of topics is ranging between 15 and 50 to achieve an optimal setting [21]. The desired inferences were achieved when the number of topics was set to 30.

Bigram Topic Model: Essentially, LDA uses "bag of words" assumption that does not take into account the order of words [16]. However, the order and proximity of words is significant for specific semantic analysis. In this sense, bigram topic model is considered as an implementation of LDA by incorporating the order of words [24]. The bigram topic model is generally used to uncover semantic relations between words. For example, "web developer" words are processed as a bigram or two individual unigrams, "web" and "developer". The meaning of "web" is different from the meaning of "web developer". In this model, each word is evaluated with the characteristics of the previous word, so the model has increased the semantic accuracy of the topic modeling [24]. In this study, trigram word distributions were used as bigram topic model in order to identification of the triple combinations of programming languages. Bigram model can be implemented to the data using parameter of "--keep-sequence-bigrams TRUE" in the MALLET.

3. Results

3.1 The roles and responsibilities of software engineers

As a result of the analysis, the 20 main roles were identified for software engineers. The roles are sorted in descending order of their frequency of occurrences and presented in Table 1. According to the findings, the highest in-demand role was "Software Engineer" (12.4%), followed by "Mobile Developer" (10.6%), and "Frontend Developer" (9.0%). The roles, that appear with low-frequency (the frequency was smaller than 1%) or misspelled were assigned to their nearest group. We considered that the findings are descriptive for the roles and responsibilities of software engineers presented in the table. Therefore, there is no need to additional descriptive information about the role definitions. The roles and responsibilities offer a descriptive information about the areas of expertise in the SE field.

Table 1. Distribution of the roles and responsibilities

Role	Responsibilities (in-demand skills)	Rate	%
Software Engineer	java, python, c#, javascript, c++	315	12.4%
Mobile Developer	android, ios, mobile, unity, objective-c	267	10.6%
Frontend Developer	javascript, html5, css, java, angularjs	227	9.0%
Software Developer	java, c#, c++, javascript, python	179	7.1%
Full Stack Engineer	java, python, javascript, c++, c#	159	6.3%
Data Engineer	sql, mysql, oracle, hadoop, java	148	5.8%
Backend Developer	php, java, mysql, python, nodejs	143	5.6%
DevOps Engineer	linux, aws, devops, puppet, java	128	5.1%
Java Developer	java, spring, sql, hadoop, javascript	113	4.4%
Cloud Systems Engineer	cloud, linux, java, c#, aws	103	4.0%
Web Developer	html, javascript, css, angularjs, jquery	94	3.7%
Ruby on Rails Developer	ruby on rails, ruby, python, backbonejs, nodejs	91	3.6%
System Engineer	linux, windows, python, ruby, java	88	3.5%
UI/UX Developer	ui, javascript, ux, html, css	83	3.3%
JavaScript Developer	javascript, html, css, angularjs, nodejs	81	3.2%
Python Developer	python, django, postgresql, php, jquery	76	3.0%
C++ Developer	c++, c#, .net, sql, c	71	2.8%
Quality Assurance Engineer	qa, sql, testing, java, c++	67	2.6%
Application Developer	java, c++, .net, ruby, linux	53	2.1%
.Net Developer	.net, c#, java, asp.net, oop	46	1.8%
Total		**2533**	**100%**

As for the responsibilities of software engineers, the 476 different skills were identified in various areas of expertise for software engineers. Given the frequency of the in-demand skills, the top five most required skills were identified. Sorting of the top skills is as follows: Java (21%), JavaScript (18%), Python (12%), Html (11%), C++ (8%). The findings indicated that the programming languages are the core competencies of the SE field, and also the dominance of scripting programming languages is remarkable. For clarity, the responsibilities were clustered according to the roles and presented in Table 1. The skills are sorted in descending order of their frequency of occurrences for each role. For example, the required skills for software engineers are sorted in such as "java", "python", "c#", "javascript" and "c++". That means that "java" is the first most important skill, "python" is the second most important skill and "c#" is the third most important skill for the software engineers.

3.2 The most in-demand combinations of the programming languages

Based on the findings on the in-demand skills given previous section, the most in-demand triple combinations of the programming languages were identified via trigram topic model. These combinations are sorted in descending order of their frequency of occurrences and presented in Table 2. According to the findings, the highest in-demand triple combination was "html, css, javascript" (12.4%), followed by "html, javascript, sql" (6.88%), and ".net, asp.net, c#" (6.61%). As understood from the table, knowledge of only one programming language is not enough for the software engineers. Today's software development environments require the knowledge of multiple programming languages together.

3.3 The educational requirements for software engineers

In this phase, the dataset was analyzed to determine the educational requirements for software engineers. The obtained results were associated with the roles that previously achieved. With this way, the frequency of educational requirements was calculated for each of the roles and demonstrated in Table 3. According to the findings, quality assurance engineer has the highest rate (76%), whereas ruby on rails developer has the lowest rate (21%) in

Table 2. The top 20 most in-demand triple combinations of programming languages

The Combination	Rate
html, css, javascript	12.4%
html, javascript, sql	6.88%
.net, asp.net, c#	6.61%
.net, asp.net, sql	6.50%
javascript, jquery, html	5.76%
java, javascript, sql	5.70%
java, javascript, html	5.68%
java, javascript, css	5.20%
html, java, css	4.86%
.net, asp.net, javascript	4.81%
python, java, c++	4.43%
java, html, sql	4.09%
java, javascript, jquery	3.67%
java, javascript, xml	3.64%
python, java, sql	3.59%
ruby, java, python	3.59%
javascript, java, spring	3.22%
c#, c++, java	3.19%
python, java, javascript	3.14%
.net, asp.net, jquery	3.07%
Total	**100%**

Table 3. The educational requirements for the roles

Role	Rate
Quality Assurance Engineer	76%
UI/UX Developer	64%
Software Engineer	63%
Cloud Systems Engineer	60%
Java Developer	59%
Data Engineer	52%
C++ Developer	51%
Mobile Developer	47%
DevOps Engineer	44%
Python Developer	41%
Software Developer	41%
System Engineer	40%
.Net Developer	39%
Backend Developer	36%
JavaScript Developer	34%
Application Developer	32%
Full Stack Engineer	31%
Frontend Developer	29%
Web Developer	27%
Ruby on Rails Developer	21%
Total	**45%**

terms of educational requirements. In total, the educational requirements were demanded in 45% of all the SE jobs. This finding indicated a considerable gap (55%) between the software industry and academia. Educational requirements for the software engineers typically contained a bachelor's or a master's degree in computer science, information science, engineering, information systems or other related majors. A degree in computer science or related fields was preferred in the majority of job postings. According to the topic analysis outcomes that presented in the following section, the most common descriptive keywords related to this topic are listed as follows: computer, science, degree, related, engineering, field, ms, and bachelor.

3.4 *Trending topics in the SE field*

Determination of trending topics in the SE field was an empirical process and it required semantic topic analysis on different levels. To this end, we implemented LDA-based topic modeling approach to discover the trending topics in this field. Thus, the 30 most trending topics were discovered as a result of the analysis. The discovered topics and related keywords are presented in Table 4.

The topic names were manually assigned to briefly describe each topic. The top eight descriptive keywords related to the topics are determined and sorted in descending order of frequency for each topic. According to the topics, the areas of expertise in SE covered a wide variety of up-to-date skill sets. Some of these included; distributed systems, real-time processing, cloud-based development, open source development and scripting languages. The discovered topics also highlighted the many emerging trends such as the programming languages,

Table 4. The top 30 topics discovered by LDA

Topic Name	Top LDA words	Rate
Communication Skill	skills communication written english ability verbal excellent oral	15.9%
Educational Requirements	computer science degree related engineering field ms bachelor	11.0%
Database Technologies	database sql git mysql nosql relational oracle sqlserver	8.1%
Scripting Languages	java python languages php ruby language programming scripting	4.6%
JavaScript Frameworks	javascript html css jquery frontend angularjs ajax bootstrap	4.3%
Web Services	web service api rest json based backend xml restful technology	3.8%
Cloud Management	tool management cloud process automation aws build	3.6%
Mobile Development	android mobile app ios developer application platform objective	3.6%
System Administration	system linux required windows admin operating unix command	3.3%
Software Testing	testing tools integration test automation build unit tdd quality	3.3%
Teamwork Skills	work team ability problem solving strong motivated player sense	3.1%
Object Oriented Programming	experience object oriented programming java professional year role	2.9%
Code Writing Quality	code writing quality develop automated maintaining test clean	2.8%
Multi-platform Development	multi multiple platform ios web android mobile development	2.7%
Open-source Development	open source system online development software driven github	2.3%
Networking Security	problems solve security network system complex issues http	2.3%
JavaScript Libraries	javascript js nodejs modern libraries similar spring angularjs django	2.2%
Quality Assurance	technical quality engineer support QA training level optimize	2.1%
Project Management	project process strong managing github drive team good personal	2.1%
Model-View-Controller	technology working including mvc c# large scale .net c++	2.0%
Business Solutions	business solution product designing communicate internal customer	1.8%
Configuration Management	customer product puppet chef configuration support operations	1.8%
Distributed Systems	distributed scalable cloud open system high building performance	1.7%
Big Data Processing	big data learning algorithms processing record hadoop structures	1.5%
Real-Time Processing	real time process data project stream analytics make	1.3%
Software Designs	software development patterns practices principles concepts skills	1.3%
User Interface Design	design user interface coding responsive analysis complex create	1.3%
Agile Development Model	development agile test continuous integration driven practices scrum	1.3%
Cloud-based Development	cloud developer java python service relevant practical desirable	1.2%
Software Development Cycle	development software lifecycle method enterprise process cycle	0.9%
Total		**100%**

skills, tools, platforms, competencies and technologies that indicate priorities in this ever-growing software industry.

4. Discussion

Our analysis revealed the vocational qualifications, combinations of programming languages, educational requirements, and trending topics demanded in the dynamic SE field. The findings of this analysis were presented in detail in the previous section. In this section, the findings are discussed in light of related studies.

The first finding was that a significant evolution was observed in the professional roles of software engineers. These roles have an increasing diversity over time, considering the study conducted by Litecky et al. [5]. In particular, expanding coverage of mobile and web applications have increased the diversity of the professional roles. Most of the roles such as Frontend Developer, Full Stack Engineer and Cloud System Engineer, etc. have emerged recently. A notable outcome is that many programming languages are directly demanded as the roles. For example, Python developer, Java developer, JavaScript developer, Ruby on Rails developer, and C++ developer etc. Another notable outcome is that the developer title is often preferred, instead of programmer title that was widely used previously, as indicated in the study conducted by Chen et al. [4]. Developer, engineer, and administrator titles are used in defining professional roles of software engineers.

The second finding was that the in-demand skills for software engineers have a wide range of diversity, as stated in other studies [5, 13]. As a result of the analysis, the 476 different skills were extracted. The range of the skills illustrates the boundaries of software industry. This means that the software engineers should have various combinations of the skills in today's progressive software market. Due to the developments in the software industry, the in-demand skills especially programming languages and their specific combinations are ever-changing over time, and being an integral part of the SE field. Despite increasing diversity in programming languages, the dominance of Java in the last decade is remarkable. Java is followed by JavaScript and Python, respectively. The widespread use of JavaScript libraries and Html5 increases the diversity of the web developer skills. The importance of multi-platform applications on the platform of web, mobile, social, and cloud are increasing day by day as the emerging platforms. In this regard, the multi-platform applications will provide new carrier opportunities for the software engineers in the future [13].

The third finding demonstrated the educational requirements for professional roles of software engineers. The educational requirements varied according to the roles as outlined in the results section. In total, the requirements were demanded in the 45% of the software jobs. This rate indicated that there is a considerable gap (%55) between software industry and academia, as discussed in several studies previously [2, 6–9, 25]. The academic institutions have a crucial responsibility to close this gap. In light of the findings of similar studies, the SE curricula can be modernized by taking into account the emerging needs. In this regard, these and similar implications can provide valuable insights for training of software engineers according to up-to-date industry needs [2, 7, 9, 25, 26].

The fourth finding revealed that the discovered topics reflect the main themes and trends in the SE field. These topics had many conceptual and interpretable inferences. One of these was the SE field has an important evolution between emerging technologies and declining technologies. The topics outlined a wide range of skills, areas of expertise, working environments in this field. Besides many non-technical skills, interpersonal skills, personal skills, and organizational skills were observed as well as technical skills in the topics. The wide coverage of the topics was emphasized in a similar study based on the analysis of big data jobs using Latent Semantic Analysis method [27]. In the previous studies that rely on the analysis of the job postings, keyword indexing approaches were frequently used as a content analysis method [7]. Therefore, supplementary studies based on probabilistic topic modeling are needed to assess the effectiveness of our methodology.

5. Conclusions

In this study, our main objective is to analyze the SE industry needs and trends, and to reveal the implications for education in this dynamic field. To this end, we conducted an empirical analysis to provide valuable insights and contributions to SE education. The methodology of this study is based on semantic topic analysis of the SE job postings using LDA model, a probabilistic topic modeling approach, which used to discover the latent semantic patterns called as topics in order to identify emerging needs and trends in the dynamic software industry. In this context, the findings of this study were: (1) As leading actors, software engineers have a wide spectrum of professional roles and responsibilities in the software industry. (2) Today's software development environments require the effective usage of specific combinations of the programming languages. (3) In terms of educational requirements, the software engineers are desired to have at least a bachelor's degree in approximately half of all software jobs, and this finding underlined the notable gap between the software industry and academia. (4) The topics discovered via LDA revealed the required qualifications for software engineers as well as the emerging needs and trends in dynamic SE field.

At individual level, our findings can be helpful for software engineers to evaluate and update their own skills, the instructors to train qualified software engineers, and the students to plan their future careers in this field. At institutional level, the findings may provide guidance to software companies in selection of qualified software engineers, and academic institutions in meeting the need for well-trained workforce for software industry. Considering the findings of this study, an innovative academic curriculum for SE education can be designed consistent with the industry needs and trends. Furthermore, the research methodology can be used for semantic content analysis on different dataset such as forums, online communities, blogs, social networks, etc. In summary, the findings of this study may provide valuable contributions into the SE field.

As in all research, this study was constrained by several limitations. The findings were based on snapshots of the SE trends covering the six-months period from January 2016 to June 2016. The analysis was limited only by English job postings. The methodology could not be applied to multi-languages due to the nature of the analysis performed in the study. This study can serve as a basis for future studies in the multi-languages. Finally, our study was based on empirical analysis using the combination of many parameters. For this reason, selection of the optimal parameters is a main task that is still an open problem for LDA. In particular, determination of the optimal number of topics required several experimental processes. Thus, further confirmatory research is needed to validate and refine our results. Our methodology can be improved with new supportive approaches in order to perform context-sensitive semantic analysis in more efficient levels. In future study, we plan to extend our methodology using different topic modeling techniques, and implementing on learning analytics data to perform a more precise evaluation of SE education.

References

1. K. A. Cary, The software enterprise: Practicing best practices in software engineering education, *International Journal of Engineering Education*, **24**(4), 2008, p. 705.
2. A. M. Moreno, M. I. Sanchez-Segura, F. Medina-Dominguez,and L. Carvajal, Balancing SE education and industrial needs, *Journal of systems and software*, **85**(7), 2012, pp. 1607–1620.
3. A. Mishra and D. Mishra, Industry Oriented Advanced Software Engineering Education Curriculum, *Croatian Journal of Education*, **14**(3), 2012, pp. 595–624.
4. Y. Chen, R. Dios, A. Mili, L. Wu and K. Wang, An empirical study of programming language trends, *Software, IEEE*, **22**(3), 2005, pp. 72–79.
5. C. Litecky, A. Aken, A. Ahmad and H. J. Nelson, Mining for computing jobs, *Software, IEEE*, **27**(1), 2010, pp. 78–85.
6. B. Prabhakar, C. R. Litecky and K. Arnett, IT skills in a tough job market, *Communications of the ACM*, **48**(10), 2005, pp. 91–94.
7. D. Smith and A. Ali, Analyzing Computer Programming Job Trend Using Web Data Mining, *Issues in Informing Science and Information Technology*, **11**, 2014, pp. 203–214.
8. Y. Kim, J. Hsu and M. Stern, An update on the IS/IT skills gap, *Journal of Information Systems Education*, **17**(4), 2006, p. 395.
9. I. Aldmour, A New Computer Engineering Curriculum Based on Technology Expansion to Address the Needs of Developing Communities, *International Journal of Engineering Education*, **30**(6), 2014, pp. 1590–1601.
10. S. Chandrasekaran, A. Stojcevski, G. Littlefair and M. Joordens, Project-oriented design-based learning: aligning students' views with industry needs, *International journal of engineering education*, **29**(5), 2013, pp. 1109–1118.
11. H. Artail, A methodology for combining development and research in teaching undergraduate SE, *International Journal of Engineering Education*, **24**(3), 2008, pp. 567–580.
12. T. Abraham, C. Beath, C. Bullen, K. Gallagher, T. Goles, K. Kaiser and J. Simon, IT workforce trends: Implications for IS programs, *Communications of the Association for Information Systems*, **17**(1), 2006, p. 50.
13. A. Barua, S. W. Thomas and A. E. Hassan, What are developers talking about? an analysis of topics and trends in stack overflow, *Empirical SE*, **19**(3), 2014, pp. 619–654.
14. P. A. Laplante, *What every engineer should know about SE*. CRC Press, 2007.
15. D. M. Blei, A. Y. Ng and M. I. Jordan, Latent dirichlet allocation, *The Journal of Machine Learning Research,* **3**, 2003, pp. 993–1022.
16. D. M. Blei, Probabilistic topic models, *Communications of the ACM*, **55**(4), 2012, pp. 77–84.
17. Stack Overflow Careers, http://careers.stackoverflow.com, Accessed 28 July 2016.
18. Stack Overflow, http://stackoverflow.com, Accessed 21 July 2016.
19. M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*, John Wiley & Sons, 2011.
20. M. A. Pathak, *Beginning Data Science with R*, Springer, 2014.
21. T. L. Griffiths and M. Steyvers, Finding scientific topics, *Proceedings of the National Academy of Sciences,* **101**(suppl. 1), 2004, pp. 5228–5235.
22. A. McCallum, (2002) MALLET: a machine learning for language toolkit, http://mallet.cs.umass.edu, Accessed 29 Jun 2016
23. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* **6**, 1984, pp. 721–741.
24. H. M. Wallach, Topic modeling: beyond bag-of-words, In *Proceedings of the 23rd International Conference on Machine Learning*, ACM, 2006, pp. 977–984.
25. G. V. B. Subrahmanyam, A dynamic framework for software engineering education curriculum to reduce the gap between the software organizations and software educational institutions, In *2009 22nd Conference on Software Engineering Education and Training*, IEEE, 2009, pp. 248–254.
26. N. E. Gibbs and R. E. Fairley (Eds), *Software Engineering Education: The Educational Needs of the Software Community*, Springer Science & Business Media, 2012.
27. S. Debortoli, O. Müller and J. vom Brocke, Comparing Business Intelligence and Big Data Skills, *Business & Information Systems Engineering*, **6**(5), 2014, pp. 289–300.

Fatih Gurcan is an instructor in Department of Informatics, and PhD student in Department of Computer Engineering at Karadeniz Technical University, Trabzon, Turkey. He received BS degree in Department of Statistics and Computer Sciences, and MS degree in Department of Computer Engineering from Karadeniz Technical University, Trabzon, Turkey. His research interests include trend analysis, sentiment analysis, statistical topic modeling, engineering education, big data analytics, and text mining.

Cemal Kose is a Professor and Head of the Computer Engineering Department at Karadeniz Technical University, Trabzon, Turkey. He received BS, and MS degrees in the Department of Electrical and Electronic Engineering from Karadeniz Technical University, Trabzon, Turkey. He received the PhD degree in the Department of Computer Science from the University of Bristol, Bristol, UK. His research interests include text mining, information retrieval, topic analysis, image processing, pattern recognition, parallel computers, and computer graphics.