A data mining approach for creating a job position in the system for evaluating competencies

Ondrej Pektor¹, Bogdan Walek¹* and Ivo Martiník²

¹ Department of Informatics and Computers, University of Ostrava, Ostrava, Czech Republic
² Department of Computer Science, VSB-Technical University of Ostrava, Ostrava, Czech Republic

*bogdan.walek@osu.cz

Abstract. This paper focuses on a data mining approach for automated retrieval of job position competencies based on a given job title and keywords that represent important competencies or concepts that are associated with a given position. The main aim is to retrieve and process the content of relevant job vacancies on the selected job portal. The output is a list of relevant words found along with their occurrence in crawled job vacancies that indicate important competencies that are required for a given job position. The HR manager then obtains a list of relevant competencies for the selected job position and the selected competencies can be added to the profile of job position in the system. Additionally, for the HR manager, the output is also a proposal to remove competencies from the job position profile, because they are not relevant (not found in searched job vacancies). These outputs are then a support tool for HR managers to create a list of all the competencies of a given job position and to store them in the job position profile. The implemented approach is verified on a specific example and the results are presented.

1. Introduction

Nowadays, the area of recruiting suitable and competent staff to job positions is a very current issue. For all companies, it is crucial to find workers who will not only be loyal and reliable but will, above all, have all the required competencies required for the job.

In companies, it is emphasised that job positions are occupied by competent employees. Competencies are defined as knowledge, skills and attributes that distinguish an expert from a layman. Another definition is that competency is a construction which helps define the level of skill and knowledge [1][2]. A person competent to take up a particular job must meet the given competencies to a minimum. This information is called the competency model. It can be successfully used to define which employee skills are crucial for the job position [3][4].

For instance, based on the evaluation of the job applicant competencies, the overall job position suitability can be determined.

Evaluating the employee's competencies can take place in various situations. For instance, in decision-making and recruitment of job applicants or regular evaluation of existing staff. The assessment of each competency can take the form of a language expression, years of work experience or percentage.

In the case of the number of years of practice, this rating can be converted to percentages according to
the given reference value (maximum, minimum, etc.) and in the case of a language expression, the conversion to percent is easy if the number and order of language expressions are fixed. This paper focuses on a data mining approach for automated retrieval of job position competencies based on a given job title and keywords that represent important competencies or concepts that are associated with a given position. The main aim is to retrieve and process the content of relevant job advertisements on the selected job portal. The output is a list of relevant words found along with their occurrence in crawled job advertisements that indicate important competencies that are required for a given job position. The HR manager then obtains a list of relevant competencies for the selected job position and the selected competencies can add to the profile of job position in the system. Additionally, for the HR manager, the output is also a proposal to remove competencies from the job position profile, because they are not relevant (were not found in searched job vacancies). These outputs are then a support tool for HR managers to create a list of all the competencies of a given job position and to store them in the job position profile.

2 Data mining approach

The structure of the proposed data mining approach is shown in the following figure:

![Figure 1. Proposed data mining approach.](image)
Particular steps and functionalities of the proposed approach are described in more detail in the following subchapters.

2.1 Creating the name of the job position and keywords to determine the competencies
In the first step, the name of the job position is created. Creating the right name of the job position is very important because this name will be used to search job advertisements on the job portal. Job vacancies on the job portal will be searched based on the job title you have entered. In addition, it is important to include keywords in the system that indicate important competencies or concepts related to the given job position. An example may be an "Administrative assistant" job position. Keywords for this position can be: work with PC, Office, Word, Excel, etc.

2.2 Loading data source - job portal
In the next step, a data source is entered into the system by the administrator as a URL to the job portal. One or more data sources can be entered into the system, each represents the selected job portal.

2.3 Loading job position data and processing
Based on the job position title and the given data source, an automated search of all relevant advertisements is performed. Each advertisement is typically represented by an HTML page that contains its description under the advertisement name. The data mining module then looks for the relevant competencies in advertisement description. Each advertisement is then categorised and divided into words and phrases. Then the competencies are extracted and the list of relevant competencies is created.

The data mining algorithm is described below:

1. Creating a request to display relevant job vacancies on the job portal using the job title
2. Analysis of HTML response - gather links to individual job vacancies and find out the number of pages of all job vacancies
3. Sending requests for each HTML page and gather links to individual job vacancies
4. Submit a request for each of the collected links to job vacancies
   a. Analysis of the HTML code of the job vacancy - analysis of the job vacancy structure and detection if the job vacancy is in the Czech language - if not, the job vacancy is excluded
   b. Searching for the keywords entered in the job vacancy, if no keyword is found, the job vacancy is excluded
   c. The job vacancy description is divided into individual words
      i. Every word is processed and system detects if the word is on the list of excluded words and if it consists only of the letters, if not, the word is skipped
      ii. Identifying competencies that are already added to the work position and that contain the word being processed (these competencies will not be proposed to be deleted)
      iii. If statistics for that job vacancy do not include this word, the word is added if the word contains statistics, it increases the number of occurrences
      iv. Searching for a database of competencies that are not assigned to a job position and comparing it with a word being processed. If a match is found, the competency is added to the list of existing competencies recommended to add to the given job position (if the competency added is already added, its frequency counter will be increased)
      v. Comparing the word with the competency register from the database. If a match is found, the word is added to the new competencies, if it has already been added, its occurrence counter
5. Suggested keywords are extracted from the statistics of all job vacancies if their occurrence is higher than 2

2.4 Creating a list of competencies
Based on the data from the previous step, the system shows the outputs directly to the user (typically HR manager). And the user can work immediately with the found competencies.
We distinguish two basic types of outputs:
- List of found competencies that occurred in the system database and in the found job vacancies, but are not part of the currently searched position in the system
- List of new proposed competencies - a list of all new competencies and keywords that a user can add to their own modelled job position along with the number of occurrences of a given competency

2.5 Proposal to add new competencies and remove the competencies which are not found
Based on the previous step, the user can decide which newly-proposed competencies will be inserted into the modelled position and removed from the modelled position. The advantage is that the data mining approach proposed by us is in the form of a module directly integrated into the system for evaluating competencies. The user can drag and drop the competencies into a modelled position and assign the importance for later evaluation. A comprehensive system for evaluating competencies that includes the data mining approach proposed in this article, is more described in these publications [5] [6].

2.6 Modelling the job position
In the final step, the user creates the final list of all relevant competencies. Then the user saves the modelled job position and the job position is ready for later evaluation.

3 Verification
For verification, a part of the web system has been created and contains all the necessary components. The web system is built on the .NET Core platform, so it is a multiplatform. As a data warehouse, the PostgreSQL database management system is used and the Object-Relational Entity Framework is used to access it. The Frontend system runs on ASP.Net Core MVC using responsive framework Bootstrap.

The fuzzystring library, authored by Kevin D. Jones, was used to compare the words of the ad and the Competence Register. [7].

To validate the proposed approach, a job was created in the system called Programmer. For this position, the following competencies were added to the system:
- Analytical reasoning
- Technical ability
- Working with computer
- Communicativeness
- Compliance with corporate culture

Then, there is competency dictionary created in the system, which contains a list of pre-created competencies that can be used for insertion into different job positions. The HR manager can modify and complement this competency dictionary.

The HR manager then entered the job position "Programmer" and the keywords that are shown in Table 1 to identify other appropriate competencies in the datamining module.
Table 1. Keywords used for job position "Programmer"

| Keyword   |
|-----------|
| CNC       |
| Java      |
| NET       |
| Postgres  |

In the next step, the request for searching job vacancies with the title "Programmer" and the inserted keywords is created. The Datamining Module then automatically searched the www.jobs.com work portal and excluded irrelevant job vacancies. The statistics on the number of job vacancies used and excluded are shown in Table 2.

Table 2. The number of used and excluded job vacancies (job vacancies in a non-Czech language or the absence of any of the keywords)

| Type               | Number |
|--------------------|--------|
| Excluded job vacancies | 57     |
| Used job vacancies   | 63     |
| Σ                  | 120    |

Then, the module suggests competencies that exist in the system database for other job positions and also occur in found job vacancies. Table 3 shows selected proposed competencies.

Table 3. Proposed competencies existing in the database (not selected for the current position) and found in job vacancies

| Competency name | Number of occurrences |
|-----------------|-----------------------|
| PostgreSQL      | 68                    |
| Oracle           | 28                    |
| Office           | 26                    |

In the next step, the module proposes to include the competences found in the job position, which are in the competency dictionary, but are not filled in the system at any job position. Table 4 shows the most numerous competencies.

Table 4. Competencies that are listed in the competency dictionary but are not in the database for any job positions and were found in job positions (10 most numerous)

| Competency name | Number of occurrences |
|-----------------|-----------------------|
| java            | 112                   |
| data            | 36                    |
Then, the module shows a list of competencies that are inserted to the current job position in the system but were not found in job vacancies. These competencies are shown in Table 5. The HR manager can decide about removing these competencies from the current job position.

Table 5. A list of competencies that are inserted in the current position but not found in the job vacancies

| Competency name               |   |
|-------------------------------|---|
| Working with computer         |   |
| Communicativeness             |   |

The, the module shows the statistics of the most common keywords that were found in job vacancies. This keyword list can be beneficial to HR manager because it displays the most used keywords in relevant job vacancies for the current job position.

Table 6. The most used keywords found in job vacancies

| Competency name | Number of occurrences |
|-----------------|-----------------------|
| team            | 40                    |
| work            | 39                    |
| holidays        | 33                    |
| development     | 32                    |
| ambience        | 32                    |
| java            | 24                    |
| sql             | 21                    |
| spring          | 13                    |
| documentation   | 12                    |
| hibernate       | 10                    |

4 Conclusion
The paper proposes a data mining approach for creating a job position in the system for evaluating competencies. The main aim of the data mining approach is to retrieve and process the content of relevant job advertisements on the selected job portal. The output of the approach is a list of relevant words found along with their occurrence in crawled job vacancies that indicate important
competencies that are required for a given job position. The HR manager then obtains a list of relevant competencies for the selected job position and the selected competencies can add to the profile of job position in the system. Additionally, for the HR manager, the output is also a proposal to remove competencies from the job position profile, because they are not relevant (not found in searched job vacancies). The proposed approach was verified on a specific example and the results were presented.

Acknowledgments
This work was supported within the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme “Research and Development for Innovations” funded by Structural Funds of the European Union and state budget of the Czech Republic. The work was created during the completion of a Student Grant SGS02/UVAFM/2018 with student participation, supported by the Czech Ministry of Education, Youth and Sports.

References
[1] J. S. Shippmann, et al., "The practice of competency modelling", Personnel psychology. 2000, Volume 53, Issue 3, pp. 703-740.
[2] M. Harzallah, F. Vernadat, "IT-based competency modelling and management: from theory to practice in enterprise engineering and operations", Computers in industry, 2002, Volume 48, Issue 2, pp. 157-179.
[3] M. A. Campion, et al., "Doing competencies well: Best practices in competency modelling" Personnel Psychology, 2011, Volume 64, Issue 1, pp. 225-262.
[4] E. García-Barrioconal, M. A. Sicilia, S. Sánchez-Alonso, "Computing with competencies: Modelling organizational capacities. Expert Systems with Applications", 2012, Volume 39, Issue 16, pp. 12310-12318.
[5] B. Walek, O. Pektor, R. Farana, “Proposal of the Web Application for Selection of Suitable Job Applicants Using Expert System”, Software Engineering Perspectives and Application in Intelligent Systems. Springer, Cham, 2016, pp. 363-373.
[6] O. Pektor, B. Walek, I. Martinik, “Proposal of a system for evaluating competencies”, Proceedings of the 2018 19th International Carpathian Control Conference, Szilvasvarad, 2018, pp. 502 – 507.
[7] K. D. Jones, “Fuzzystring library”, https://github.com/kdjones/fuzzystring