IJCTA, 9(35), 2016, pp. 37-48

Decision Support System for Talent Management: Text Mining for Competency Assessment

Nikita Nikitinsky* Roman Teplov Alexey Voronets*** and Alexey Nesterenko******

Abstract : In spite of extensive works addressing the Human Capital Management area in companies the existing solutions still do not enable fully automated competence identification and assessment. Often, companies still have to rely on pen and paper tests which assume high degree of subjectivity in evaluation. In the following paper we describe the model and the prototype of a Decision Support System for Talent Management (DSSTM) The system is based on text mining approach and enables automated competence assessment. We outline concepts of modules of the system, and give the general picture of the approach we utilized to create the system. Overall, the results of application of the system in the case company demonstrated the increased efficiency of HR management but at the same time revealed certain limitations of the system, in particular in the assessment of "soft", non-technical competences.

Keywords : Automated competence assessment; talent management; text mining; LSA.

1. INTRODUCTION

Automated regulation of human resources allocation and estimation of employees' experience, competencies and productivity gave a broad area for the Human Capital Management (HCM) software, such as SAP SuccessFactors [1] or Oracle Taleo [2]. High technological products are actively used in eGovernment field. In the nearest future there is even a possibility of decision support systems for administrative institutions or gamification in healthcare supporting governmental healthcare policy [7].

Different approaches for competence assessment have been discussed in academic literature. Thus, Berio and Harzallah [4] in their earlier work discuss the model to manage four essential processes: identification, assessment, acquisition and usage of the competences. García-Barriocanal et al. [6] developed a generic ontological model which should facilitate the quantitative competences analysis. Work of Rauffet et al. [13] propose methodology for assessment of organizational capabilities.

Although a complicated multimodal system gives an opportunity to examine the object from several angles simultaneously, the set of tools is still limited and human-biased. The problem is, therefore, to create the system, which enables managers and team leaders to overcome the limitations of traditional competence evaluation methods (such as 360 degree-feedback or ordinal professional skills tests) and automate the process of competence identification and assessment.

In this paper, we describe a Decision Support System for Talent Management (DSSTM), which is based on text mining approach and enables automated competence assessment, outline concepts of

* NAUMEN 47 Varshavskoye Highway, Moscow 115230, Russian Federation
** Lappeenranta University of Technology, P.O. Box 20, Lappeenranta 53851, Finland
*** Tsentr Razrabotki, LLC, 20/1 Pozharova Street, Sevastopol 299008 , Russian Federation
**** Independent researcher Corresponding author

modules of the system, and give the general picture of the approach we utilized to create the system. The goal of this paper is to present and discuss the model and prototype of the system we developed (in the paper, we will call it DSSTM), which is currently being tested in a software development company.

The rest of this paper is organized as follows. Section 2 briefly outlines the motivation for DSSTM, section 3 outlines the methods of competence assessment, sections 4 and 5 discuss qualification assessment and professional interest discovery respectively, section 6 describes additional modules of the system, section 7 provides the results of practical application in the case company and section 8 concludes with the final remarks.

2. MOTIVATION FOR DSSTM

In this chapter, we outline the motivation for the system. Figure 1 displays the general high-level structure of the DSSTM.

The system is designed to cope with various Talent Management tasks, such as:

- Assisting the C-level executives and managers in search for the most skilful employees or group of employees.
- Assisting HR managers in employees competences evaluation
- Enabling the employees to track the development of their competences and identify new acquired competences
- Facilitating the communication between departments by automatically connecting employees from different departments with similar professional interests and comparable competences
- Facilitating knowledge diffusion within a company by recommending content created by one employee to other employees based on their skills and positions

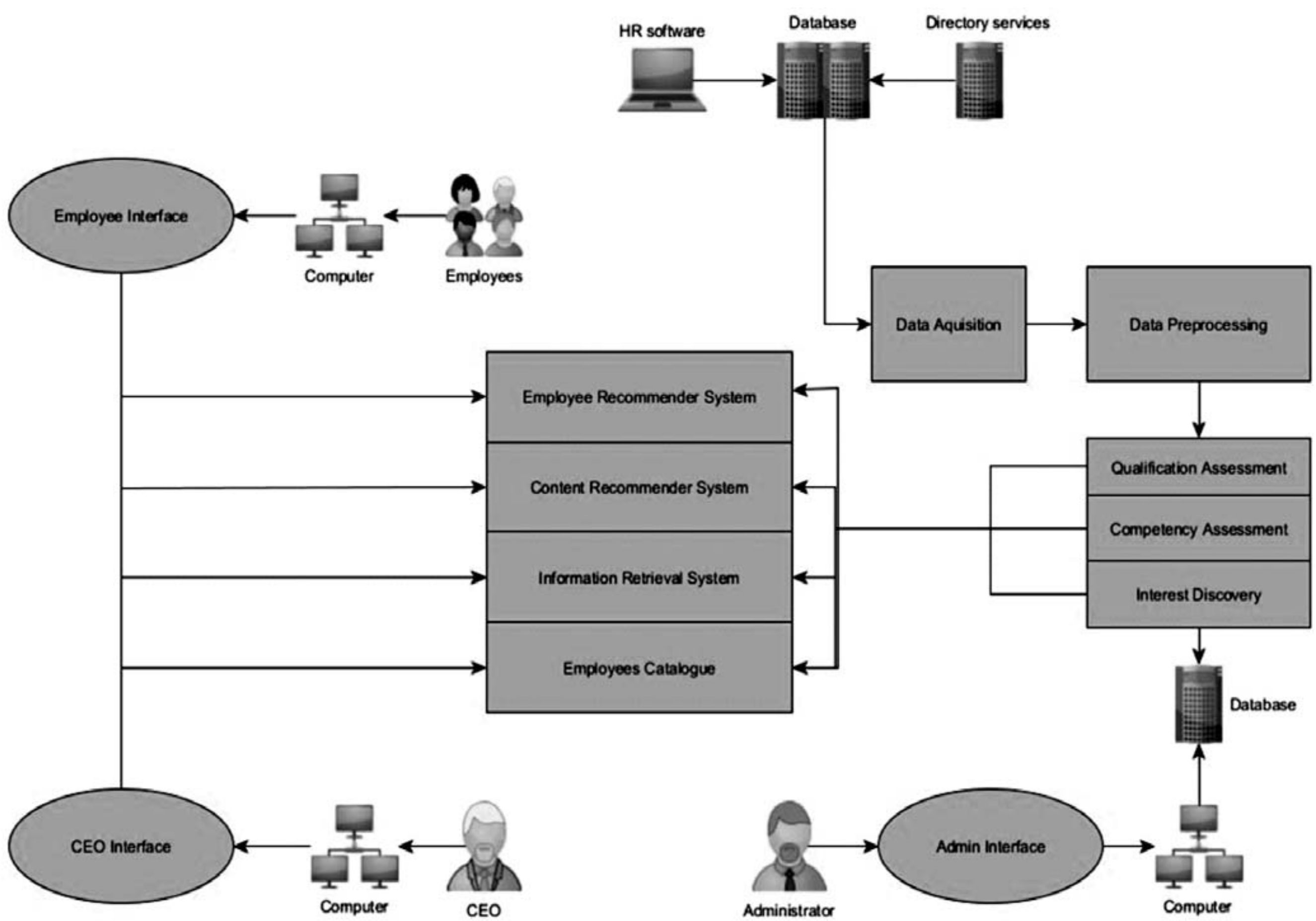

Figure 1: The schema of DSSTM as described in the paper

2.1. Data

Currently, for competence assessment the system combines the information from three large sources: employee HR profile, all the text documents produced by employees (*e.g.* scientific publications, work reports) and results of professional skill tests and other traditional competence assessments methods.

To obtain this data, DSSTM uses connectors to the most popular HR software systems, databases and directory services

As DSSTM focuses on text document analysis, it obtains texts of documents produced by employees. Either, the system takes into account all the metadata derived from the documents, including:

- Document type (reports, R&D documents, etc.)
- Authorship and co-authorship
- Evaluation the document provided by colleagues

When it comes to the employee profile, DSSTM uses the following basic employee attributes:

- Overall work experience
- Work experience in the company
- Information about the achieved KPI
- Current education level and the information about all further education courses
- Current department and the position in the company

Besides, the system also employs the results of professional skill tests passed by employee.Therefore, DSSTM uses various data types that enables the identification and assessment of employees' competences with higher precision and less biased by human influence.

2.2. Preprocessing

Whereas some data (*e.g.* information from the employee profile) demands relatively simple transformation (*e.g.* ranking, averaging, normalization), pre-processing of documents is a more complex task. Apart from tokenization, stop-words removal and morphological analysis of the contents (*i.e.* part-of-speech tagging, lemmatization etc.), we employ word embeddings and latent semantic analysis (LSA) for text classification and key terms extraction.

Text classification

Text classification allows DSSTM to define document subject areas such as scientific areas (physics, chemistry, etc.) and then compare vectorised document of certain subject area to a benchmark document in order to define its quality.

The method is based on a well-known technique named word2vec.Word2vec [10] is a neural network model used for word embeddings analysis. The assumption behind the model is that words located in the similar contexts tend to have semantic closeness (i.e. similar meanings). The model supports two architectures: continuous bag-of-words (CBOW) and continuous skip-gram. In DSSTM, we utilized skip-gram approach as it have been found more suitable for working with texts containing less frequent words.

The algorithm for text classification in DSSTM is following:

- The word 2 vec model is trained with any given corpus of texts related to required topics
- For each topic from the corpus the text is converted into the sum of word vectors from the word-2vec model
- Similarly, the input text (the text, which need to be classified) is projected into word2vec space and sum of word vectors for the input text is returned

- The obtained input text vector is then compared with every topic vector (cosine similarity is applied)
- Topics with cosine similarity more than threshold (0.8 by default) are assigned to the input text. If too many topics demonstrate the similarity above the threshold, then only n topics with the highest cosine are considered (n can be assigned by user, by default $n = 5$)

Thus, we obtain classified texts and for some of them (say, some types of reports) we may try to estimate their quality by comparing to benchmark document of certain type and subject area. The information about topics and document quality is then used both for competence assessment and for content recommendation system.

LSA semantic space construction

LSA is a natural language processing technique, which analyses relationships between a set of documents and the terms they contain [9]. The assumption behind the algorithm is very similar to that of word2vec. Currently, researchers actively use LSA to solve various text analysis tasks [5; 15]. In this technique, a weighted term-document matrix is constructed, where rows represent unique words, and columns represent documents. The matrix is built using well-known mathematical technique called singular value decomposition (SVD). SVD allows approximation of initial matrix with the lower rank matrix increasing the lemma's weight and reducing noise. However, the dimensionality reduction is a trade-off between the richness of semantic groups included and the level of noise.

A singular value decomposition of some $m \times n$ matrix M is a factorization of the form:$M = U\Sigma V^*$, where M is $m \times n$ matrix whose entries come from some field K, U is $m \times m$ matrix, Σ is $m \times n$ diagonal matrix with non-negative real numbers on the diagonal and V^* is an $n \times n$ unitary matrix over K.

For the weighting method we apply LogEntropy. The method has proved its efficiency in variety practical applications [8; 11]. We build the LSA semantic space over all the documents obtained by DSSTM. The created semantic space is later used for key terms extraction and clustering.

Key term extraction

For key term extraction we apply a combination of LSA approach with rule-based approach.

For every text document, we extract two types of key terms – local and global. Local key terms consist of lemmas contained in the analysed document whereas global key terms include lemmas from all corpus of documents. Local key terms are intended to describe the document itself while the global ones should describe the subject area of the document.

The algorithm for key term extraction is the following :

- We select candidates for key terms from the document with preliminary defined rules.
- To obtain local key terms, we estimate similarity between each candidate-term vector and document vector in LSA space (cosine similarity).
- For global key terms, we estimate similarity between the document vector and all lemmas in the semantic space
- Both for global and local key terms we select top-*n* terms (*n* can be assigned by user, by default $n = 30$) with cosine similarity above certain threshold.

The extracted key terms are then used for competence assessment and content recommendation system.

3 COMPETENCY ASSESSMENT

The core functionality of DSSTM is competency assessment. In DSSTM, a competence is a combination of a skill applied to a certain domain. For instance, the skill "technical writing" applied to a domain of Physics is a competence "technical writing in Physics".

Competence evaluation consists of two steps:

- I dentification of competence. In this step, the system checks the presence of certain competence for an employee.
- Evaluation of competence. In this step, the competence is being evaluated with help of modifiers.

To make sure that all the input parameters are normally distributed and fitted to a given scale, we apply scaling and normalization to the parameters.

3.1. Scaling and normalization

All the input parameters are normalized and scaled to the given scale in order of standardization of the parameters. It is worth noting, that every competence modifier in DSSTM is evaluated relative to the average score for this competence modifier in a department or a whole company. This is done in order to quickly detect the best and worst performers for every competence and help CEOs to easily find them.

Thus, each modifier is compared with the average value for the department or for the whole organization (+/– std deviation). If the values are equal then the modifier has no impact on the basic sore. If the modifier value differs, the basic element can be adjusted according to the formula:

$$\mathrm{D}(b_i) = \begin{cases} 1 + \dfrac{b_i - \min(\mathrm{All})}{\max(\mathrm{All}) - \min(\mathrm{All})} \text{if } b_i > \mathrm{mean(All)} + \mathrm{std(All)} \\ \qquad\qquad\qquad\qquad \text{if } b_i \in \mathrm{M} \\ 1 - \dfrac{b_i - \min(\mathrm{All})}{\max(\mathrm{All}) - \min(\mathrm{All})} \text{if } b_i < \mathrm{mean(All)} - \mathrm{std(All)} \end{cases} \tag{1}$$

where $\mathrm{D}(b_i)$–adjusted modifier element, b_i – normalized parameter, All = $\{bo_1, bo_2, \ldots, bo_n\}$ – set of all values for the parameter, M = $\{bo_1, bo_2 \ldots bo_n | \, bo_i \in \{\text{mean (All) std} \pm \text{(All)}\}$ – set of all parameter values within standard deviation, mean(All) – average value for the parameter, std(All) – standard deviation of the parameter, min (All)–the min value of the parameter, max (All)–the max value of the parameter

3.2. Identification of competence

DSSTM uses a rule-based approach for competence identification. Based on possible features, obtained from profile, text documents or professional skills test results, DSSTM user may create rules to identify the presence of competence for an employee.

The example of such rule, which may be created for identification of a competence "Scientific research in Biology" is presented below:

- Check for presence of documents of specific type (*e.g*. scientific paper, thesis, dissertation etc.) and specific subject area (*i.e*. Biology)
- Eevaluate the documents' relevance to the competence key terms and benchmark documents (*i.e*. compare the semantic similarity of competence key terms and benchmark documents to the employee documents of selected types in LSA semantic space)
- Calculate the amount of documents with semantic similarity above given threshold'
- If the amount of such documents is more than zero, then the employee has this competence

The result of the competence evaluation must be a number in order to be used in the process of competence evaluation.

3.3. Evaluation of competence

The general formula for competence assessment thus may be expressed as:

$$\mathrm{rate_{comp}} = \mathrm{B_{comp}} * \left(\frac{1}{3}(\mathrm{B} + \mathrm{HR} + \mathrm{TXT})\right) \tag{2}$$

Where $rate_{comp}$ – competence score, B_{comp} – basic score, B-basic parameters modifier score, HR-HR modifier score and TXT-text modifier score.

For the convenience, we scale the resulting competence score to the conventional scale (say, 1–5, where 1 means junior level of competence and 5 – expert compared to other employees with the same competence). Therefore, the results from the formula (2) need to be adjusted accordingly:

$$rate_{comp} = \begin{cases} \text{Max Scale if } rate_{comp} > \text{Max Scale} \\ rate_{comp} \text{ if } rate_{comp} \in \text{Scale} \\ \text{Min Scale if } rate_{comp} < \text{Min Scale} \end{cases} \quad (3)$$

Where, $rate_{comp}$ – competence score, Max Scale – the maximum value for the scale applied (in our model Max Scale = 5), MinScale-the minimum value for the scale applied (in our model Min Scale = 1) and Scale-the scale value within the range

As the result, an employee gets his competence level evaluated and scaled to a conventional scale. For example, “technical writing in Physics – 3”.

It can be seen that formula of competence assessment (2) is composite and consists of basic score and several modifier scores – B, HR and TXT. In the following sections we will discuss them in more details.

3.4. Basic score

First, we calculate so called basic score, which reflects how evident the identified competence is. The basics core is calculated asratio of the result of competence identification for certain employee to maximum result of competence identification in a department or whole organization

This is done in order to define how certain employee is compared to the best employee in a department or company with this competence.

$$B_{comp} = \frac{CI_{res}}{CI_{max}} \quad (4)$$

Where B_{comp} – basic score, CI_{res} – the result of competence identification for certain employee, CI_{max} – maximum result of competence identification in a department or whole organization

3.5. Modifier score

Modifier can increase or decrease the basic sore and consists of three elements:

- Basic parameters modifier (work experience, courses participation, etc.)
- HR modifiers (represents the results of qualification tests)
- Text modifier (represents the quality of documents produced by employee)

Basic parameters modifier takes into account the quantitative features from the employee profile, such as overall work experience, work experience in the company, amount of KPI achieved etc. Qualitative features can be converted into categorical or numeric with conversion rules. For example, education level can be assessed with 1-5 scale, where 1 relates to secondary education and 5 to doctoral degree. Overall, the formula for the modifier enables including various amount of parameters (b_1, b_2, …, b_n,…, b_i) depending on the company needs:

$$B = gl * (\textstyle\sum_{i=1}^{n} (D(b_i) * \text{imp})) \quad (5)$$

Where $D(b_i)$ – adjusted modifier element, imp – the weight for each element (by default equal 1), gl – global weight for the element (by default equal $1/n$)

HR modifier is based on traditional evaluation methods such as *e.g.* 360 degrees, various professional tests, surveys. The modifier employs scores earned by employee in such tests. In particular, the modifier calculation method considers following:

- Average/median test score.
- Stability of tests results (for this parameter Shannon information entropy is used: $St = -\sum_{i=1}^{n} p(est_i) \log_2 p(est_i)$, the employee should pass at least 5 professional tests (est_1, est_2, …, est_n, …, est_i).
- Frequency of completing tests: $Fq = \frac{nt}{ny}$, where nt-number of tests passed, ny-the required time period (6 months by default)
- Relative success compared to colleagues at the same position in the department or the whole organization

After scalingall the features with (1) HR modifier can be calculated as:

$$HR = gl * \left(\sum_{i=1}^{n} (D(hr_i) * \mathrm{imp})\right) \tag{6}$$

where, $D(hr_i)$ – adjusted modifier element, imp – the weight for each element (by default equals to 1), gl – global weight for the element (by default equals to $1/n$)

Text modifier evaluates the indirect quality of text documents.

The modifier currently includes following components:

- Document type
- Average length of the document in words.
- Flesch-Kincaid readability index. The index is rescaled to avoid negative values: $fl_i = fl_i + |\min(Fl_m)|$. where Fl_m – the set of index values for all employees. In other words, the absolute value of minimal index within the organization is added to the calculated index for each employee.
- SMOG readability index. The assumption of using readability indices is that more experienced employees tend to create more readable texts.
- Lexicon uniqueness (number of unique lemmas in relation to the overall amount of words in the text): $LD = |U|/|A|$, where U- is the number of unique words in the text after lemmatization, A-is the overall amount of words in the text after lemmatization. The assumption is that more experienced employees tend to use more diverse lexicon.
- Amount of co-authors and their input. To assess co-authorship for each text generated by employee the array of co-authors is evaluated (based on the competences of each co-author). In the similar way the input of all co-authors (for all texts generated by the given employee) is assessed.
- Text evaluation provided by colleagues

After scaling all the features with (1), text modifier can be calculated as follows:

$$TXT = gl * \left(\sum_{i=1}^{n} (D(txt_i) * \mathrm{imp})\right) \tag{7}$$

where, $D(txt_i)$ – adjusted modifier element, imp – the weight for each element (by default equal 1), gl – global weight for the element (by default equal $1/n$)

The competence assessment algorithm is highly customizable and allows adding more rules and parameters into the formulas. However, high adjustability may lead to relatively long and difficult initial tuning for every competence.

4. QUALIFICATION ASSESSMENT

Evaluation of separate competences cannot tell the DSSTM user, how qualified each employee is in general and in comparison to others with same competences and same position in a company. The following methodology enables evaluation of the overall level of qualification. The method provides two alternatives of assessment: average qualification and the qualification with reference to job requirements.

Average qualification evaluates the current qualification level of employee. The index represent the average of present competences:

$$Q = \frac{1}{n}\sum_{i=1}^{n} \text{rate}_i \quad (8)$$

where, Q-the level of employee qualification, rate_i is the current level of certain employee competence.

Qualification with reference to (further) job requirements assesses the employee’s qualification required for promotion to the next position within the company structure. This parameter compares the current level of employee’s qualification with the required level for specific position, which has to be defined in advance. In essence, this type of qualification is computed as intersection of the competences and other parameters of an employee (including data from personal profile, and the results of traditional competence assessments). The resulting qualification is expressed as a number between 0 and 100, where 0 is total unsuitability for the promotion for the next position and 100 is full suitability for promotion.

5. PROFESSIONAL INTEREST DISCOVERY

Professional interest discovery is intended to detect the subject areas, which are most interesting for an employee. This information will be might be useful *e.g.* for team leader and contribute to employee motivation by assigning tasks which relate to employees’ areas of interest.

Currently, DSSTM manages discovery of global and local interests. Global interests are computed as top-n most frequent subject areas from the employee’s text documents (see chapter 2.2.1).

To extract local interests, we need to conduct some more data processing:

- Create personal employee’s semantic LSA subspace (as a part of the overall organization term-document semantic space - for the method of creation of semantic space for organization documents see chapter 2.2.2). The created individual semantic subspace is then clustered with clustering algorithm. Currently, we employ Clustering by Committee (CBC) algorithm for this [12]. We preferred this algorithm to others as it has been created specifically to cluster text data and can partially handle homonymy.
- To determine the optimal number of clusters we typically use Silhouette coefficient [14]. Silhouette coefficient compares the average distance from element to element within a cluster with the average distance to elements in other clusters, assigning highest scores to the number of clusters, where objects are densely distributed within the clusters, while the clusters are located far from each other.
- And, finally, we extract key terms for each detected cluster. To do this, we apply method described in the chapter 2.2.3about it.

As the result, we obtain a list of key terms for every local professional interest, which is human-interpretable. Either, we sort the discovered local interests according to their significance (based on occurrence in texts) and relate them to global ones to let the user of DSSTM understand that, for example, certain employee is interested in specific branch of physics. However, it may be a challenging task to identify the name of a very specific professional interest.

Both global and local professional interests are later used for content recommendation system and employee recommendation system as features.

6. APPLICATION OF ASSESSED COMPETENCES

The assessed competences, qualification and professional interests are used to create a personal DSSTM profile for every employee, perform search for employees and let the employees with similar competences communicate and share knowledge. Furthermore, DSSTM computes competence statistics for a company, which can be used for analysis purposes – for example, to find out crucial competences, which are weakly represented in the company.

6.1. Information Retrieval System

DSSTM the system currently provides search for:

- Employees based on specified parameters of competence, qualification and/or professional interests
- Documents based on specified parameters such as document type, subject area etc.
- Task teams with specified set of competences (and certain level), qualifications and/or professional interests.
- Task teams with unknown set of competences. This search is based on text analysis – DSSTM user provides project description and DSSTM detects the competences required to cope with it. To detect the competences, text classification and competence assessment methods are applied (see chapters2.2.1and 3for these methods).

Currently, search system is based on open-source tools, such as Apache Lucene. Apart from traditional full-text search, DSSTM employs semantic search within LSA semantic space and word2vec model, which allows the user retrieve results that are more relevant.

6.2. Content recommendation system

The purpose of the system is to share knowledge between employees by recommending content created by one employee to other employees.

The system employs two approaches:

- Content filtration approach (for employees who produced sufficient amount of documents). This approach consists of projection of employee's documents vectors to the LSA semantic space and ranking them.
- Rule-based method (for employees who produced small number of documents).

Content filtration method recommends the content produced by the employee according to the following algorithm:

- Text document is pre-processed and vectorised
- The vectorised document is iteratively compared to all other employee's personal semantic LSA subspaces (*i.e*. to every cluster in that subspace, see chapter 5for details) and cosine similarity is computed.
- If the document is semantically close to the employee's content, (cosine similarity is above threshold, which is 0.8 by default), then it undergoes post-filtration, which is based on the employee's preferences

If the document passes post-filtration successfully, then the content recommendation system recommends it to the employee

The main advantage of content filtration approach is decent accuracy of the recommendations. However, this approach cannot be applied to an employee, who does not have a sufficient amount of documents (*e.g*. if he or she has recently joined the organization). Therefore, for such cases in order to avoid the "cold start" problem the rule-based approach is applied.

The rule-based approach uses as an input data from the personal employee's profile, including the information about competences and qualification (if there are some). The assumption behind this approach is that the employees with similar parameters from personal profile tend to have common interests. Thus, DSSTM selects top-n employees, whose profiles are most similar to the given employee, and recommends the content produced by those employees.

6.3. Other modules

Based on the approach described in chapter 6.2, DSSTM recommends employees to each other, assuming that the employees with close profiles (including competences and qualification) and semantically close documents tend to have common tasks and interests and may help each other in their daily work.

DSSTM also automatically creates an employees catalogue, which groups employees with close professional characteristics based on all the data about the employees in the system. To create the catalogue, OPTICS clustering algorithm is used, as it copes with arbitrary shaped clusters and selects the best number of clusters by itself [3]. This catalogue is intended to help employers with organization of task team in a bit different way than task-team search.

7. CASE STUDY AND DISCUSSION

The prototype version of the system has been functioning for three months in the software developing company in Russia. The preliminary results are summarized in the Table 1.

Table 1

The results of practical implementation of the system

Controlled Parameter	*Average in Russia (based on hh.ru statistics*)*	*In the case company, 3 months before launching DSSTM*	*In the case company, 3 months after launching DSSTM*	*Results*
Average time to find an employee with required competencies within the company	No data available; depends on the size and structure a of company	2-3 hrs. (in average)	0.7-1.5 hrs.	decreased by 35%
Average monthly outflow of employees	Varies too much according to the domain	4 persons per 100 employees	3 persons per 100 employees	decreased by ~20%
Average monthly expenses for employees' competency assessment by HR department	62 hrs.	58 hrs.	54 hrs.	decreased by 6%
Key productivity indexes average fulfillment during the first three months of work	60-75%	70-80%	75-83%	Increased by 4-5%

Note: * hh.ru – Russian largest HR web-portal.

The practical implementation of the system demonstrated promising results but also revealed certain limitations of the model. First, currently it can detect and estimate mostly technical competencies. Core and soft skills can hardly be unambiguously formalized and estimated with the text analysis. Currently, the model considers only some aspects of employee's performance, which cannot cover all the employees' activities. Moreover, the model works well only for those employees, who produce at least some text (*e.g.* programmers, analysts, marketing managers etc.). In addition, it may be a challenging task to create detection conditions for some competences.

Despite the number of limitations of the current prototype version, we consider usage automated Competence Analysis prospective for Decision Support and Talent Management. The potential of Deep Learning in forecasting the career development of an employee in a new environment might become a strategy for the further product development. So far, we lean towards turning to the several potential scenarios of the future DSSTM development:

- Profound competence assessment (cross-subject competences, flexibility in cognitive styles and other)
- Forecasting employees productivity (modelling productivity in different tasks and environments based on the previous statistics),
- Assessing professional development potential (based on comprehension of cross-subject issues, fundamental knowledge, flexibility in perception of new information, potential to go deeply in specific subjects),
- Research teamwork productivity, correct roles and tasks assignment

The successful deployment of DSSTM, requires the access to the documents produced by employees is a, thus limiting its effectiveness for managing employees dealing with manual labour. The key area of application therefore is management of R&D, and managerial workforce.

8. CONCLUSIONS AND FURTHER RESEARCH

In this study, we introduced the model for competence and qualification assessment, the Decision Support System for Talent Management based on this model and the current results of testing the prototype in a case company.

Key features of both the model and the DSSTM are extensive use of text mining and high automation of the process of competence assessment in order to improve precision of assessment and make it less biased. The model is highly customizable and allows the users to add their own features into the model. DSSTM provides various modules to help C-level executives and managers solve different HRM tasks. The most interesting additional modules are information retrieval module, which helps to find employees and create task-teams with specified competencies and content and employee recommendation system module, which helps to preserve and share knowledge within the company.

Despite some certain limitations, the prototype version of DSSTM showed promising results after three months testing in the case company.

In our future research, we plan to cope with the current limitations of the DSSTM according to the results of testing in a case company and improve the overall performance of the model. In particular, we draw following perspective directions for further works:

- Experiment with larger amount of different features to the model, e.g. more text features
- Improve the quality of recommendation systems and cluster analysis by introducing semi-supervised clustering to the model
- Experiment with elements of fuzzy logic in order to be able to let the DSSTM work with estimation of soft and core competencies
- Conduct research on applying some aspects from psychology and sociology to the system

We do also believe that the presented approach will not only facilitate work of HR managers and team leaders but can be also applied to other domains opening new possibilities for both researchers and practitioners.

9. ACKNOWLEDGEMENT

This work was supported by the Ministry of Education and Science of the Russian Federation [grant number RFMEFI57914X0091]

10. REFERENCES

1. SAP SuccessFactors. (n. d.). Date Views: 23.10.2016 https://www.successfactors.com/
2. Talent Management for the Enterprise (n. d.). Date Views: 23.10.2016 http://www.oracle.com/us/products/applications/taleo/enterprise/overview/index.html
3. Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. In ACM Sigmod Record, 28(2), 49-60
4. Berio, G., & Harzallah, M. (2007). Towards an integrating architecture for competence management. Computers in Industry, 58(2), 199-209.
5. Chen, Y., Tsai, F. S., & Chan, K. L. (2008). Machine learning techniques for business blog search and mining. Expert Systems with Applications, 35(3), 581-590.
6. García-Barriocanal, E., Sicilia, M. A., & Sánchez-Alonso, S. (2012). Computing with competencies: Modelling organizational capacities. Expert systems with applications, 39(16), 12310-12318.
7. Kachurina, P., Kolomiytsev, S., Bespalov, A., Barchuk, A., & Pantyukhov, P. (2015, November). The Synergy of Gamification and Mathematical Modelling in eHealthcare. In Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (pp. 116-122). ACM.
8. Korzycki, M., & Korczyński, W. (2015). Does topic modelling reflect semantic prototypes? In New Research in Multimedia and Internet Systems. Springer International Publishing, pp. 113-122.
9. Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). Handbook of latent semantic analysis. Lawrence Erlbaum Associates.
10. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Proceedings of workshop at ICLR
11. Naud, A., & Usui, S. (2008). Exploration of a collection of documents in neuroscience and extraction of topics by clustering. Neural Networks, 21(8), 1205-1211.
12. Pantel, P. A. (2003). Clustering by committee (Doctoral dissertation, University of Alberta).
13. Rauffet, P., Da Cunha, C., & Bernard, A. (2014). A dynamic methodology and associated tools to assess organizational capabilities. Computers in Industry, 65(1), 158-174.
14. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.
15. Wang, J., Peng, J., & Liu, O. (2015). A classification approach for less popular webpages based on latent semantic analysis and rough set model. Expert Systems with Applications, 42(1), 642-648.