Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

A Tool for Visualizing Skill Requirements in ICT Job Advertisements

Harri Hämäläinen, Jouni Ikonen, Jari Porras

Abstract: *Many students do not necessarily have a clear idea what types of skills are typically required by industry for different job positions and how the skills relate to each other. Job advertisements are a good source for finding out these skills and analyzing them. In this paper we introduce a tool that is used to search and analyze skills introduced in job advertisements. The tool examines the frequency of different skills in advertisements and their relations. This information is used to recognize the skills typically required for employment and to visualize their relations through a web-based user interface.*

Key words: *Knowledge-based requirements, hard skills, soft skills, morphological analysis.*

INTRODUCTION

Professional objectives and expectations of students are often based more on the job titles and possibly the primary technical (hard) skills than on the knowledge and full set of comprehensive skills. They do not necessarily understand the importance of the skills they have learned and will learn during their education, but consider only the value of the degree. Although the educational institutes aim to offer courses and study paths that create sufficient know-how and basis for learning and adapting technologies, students still might have difficulties to divide the industrial requirements into more specific skills or even to recognize them, especially concerning the soft skills. In addition, the relations between different skills are not necessarily obvious. However, the skills that are learnt and the skills that companies require play the essential role when getting employed.

Job advertisements are a good and real time source for collecting and examining the apparent requirements that companies, and thus whole the industry, set to the applicants. Since advertisements typically list sets of skills, collecting and finding out the skills and their relations to each other can provide good help for students even before seeking job is topical. In addition, this information is usable for teachers as well when mapping how extensively the content of courses and teaching meet the requirements of industry over the topic in question.

Finding information about technology trends and the most significant technical skills in labour market is presumably doable using existing tools and services such as Google Trends. Instead, it is more challenging to realize which secondary skills, both hard and soft, appear in pursuance of searched primary skills and if some skills are strongly related to certain job titles. By analyzing job advertisements the relations of these skills can be determined and the current situation and development in labour market can also be taken into an account.

Job advertisements on ICT industry are nowadays being published in various web-based employment sites. Because of the previously listed reasons and to support the decision making when students are building their personal study and development plans, we have implemented a web-based tool called *JobSkillSearcher* to collect, identify, analyze and present the information from Finnish ICT labour market. The analyzed information is shared for our students to support their decision making and consciousness of demands.

Our motivation on implementing the system is to provide the students a tool that they can use for multiple purposes. First of all, in the beginning of their studies they are able to get an overall picture of what kinds of job positions exist and which skills are demanded for these positions at labour market. In later phases the information can be examined more

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

deeply; which types of skills are strongly related to each other. In teaching and motivating the students, teachers can have an option to prove the importance of course contents by showing the connection of technologies in question in larger scale which is often challenging for students.

This paper is organized as follows. In chapter 2 we discuss about the research that has been completed on ICT job advertisements in past. Chapter 3 represents the implemented tool and the foundations for its design. Chapter 4 discusses about the preliminary results we have collected and analyzed until now. In the last chapter we conclude this paper and describe the future steps.

RELATED WORK

The content of job advertisements in ICT industry has been analyzed over the past decades. According to Surakka [9] the research of the industry requirements for skills has traditionally been completed by interviewing different industry representatives or by manually investigating the job advertisements from newspapers or professional magazines. However, interviewing or analyzing the advertisements manually requires a considerable amount of resources. On the other hand, quality of manual analysis completed by skilled persons is presumably more reliable than completing it by computers when using rather simple algorithms.

Another challenge on the manual investigation of advertisements is the growing number of existing technical skills. Based on numerous researches the mean of listed skills in an advertisement has had a rising trend already since the 70's. Todd et al. [11] reported that the number of technical phrases in job advertisements for programmer positions increased from the mean of 2.2 in 1970 to 4.2 in 1990. Surakka [9] has reported that the number of desired technical skills in general ICT advertisements has also been increasing during the past two decades from 3.57 in 1990 to 7.66 in 2004. However, the comparison of individual researches of different research groups is difficult since the used terms have not been published in each of the cases and the words that have been recognized may have changed over the years.

In addition to the increased demand of technical skills, numerous researches, such as [12], [3], and [4], have revealed the importance of non-technical (soft) skills. For example interpersonal skills play an important role also among ICT professionals who are regularly interacting with colleagues and customers. This also reflects to the job advertisements and also to the recruitment process where the role of soft skills seems to have become more important than earlier. Therefore the ability of analyzing and even somehow evidencing personal soft skills might be advantageous for a job seeker.

Internet-era for its part has changed the way how companies publicly search for employees: the number of advertisements in newspapers has decreased while the web-based services have gained popularity. Due to this the structure and content of advertisements have also changed. Formerly when the space and number of words in an advertisement was limited because of expenses, advertisements were concise and focused only on the core competences. This change increases the amount of manual work to be even more than earlier.

Since the web-based services have become more important data source for job seekers than newspapers, it is reasonable to take advantage of the digital material while analyzing the advertisements of present time. In longitudinal researches where trends are analyzed there is a need for accessing and going through the advertisements over a longer period of time to find out the changes. This is one of the reasons why the web-based employment services that typically provide only the non-expired advertisements

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

have previously not been used. However, after collecting the advertisements from the services constantly this limitation will be eliminated allowing longitudinal analysis for the data as well.

Automated systems for parsing the skills from job advertisements are not widespread and our academic search provided only few systems that can by any means be compared to ours. Litecky et al. [6] have developed a software tool for parsing and extracting the pre-defined set of 239 skill terms from the advertisements of three US-origin employment sites. Using this tool they have extracted and analyzed more than 200.000 job advertisements that require a degree on computing science programs. However, the results that are collected are used for research and analysis purposes and no public interface for outside users is provided. Loth et al. [7] have developed a prototype for indexing the contents of online job advertisements using latent semantic analysis.

IMPLEMENTATION OF THE APPLICATION

To collect information about job opportunities in Finland we have implemented an application called *JobSkillSearcher* that collects and analyzes ICT job advertisements from various employment sites and provides a user interface to access this information. Our primary purpose is to expose the relations between skills and terms for the users by providing a web based tool to access the analyzed information, not researching and trying to analyze the information and comprehensive meaning of advertisements too deeply.

Separate modules are used to download job advertisements from four different Finnish employment websites. Advertisements are retrieved daily from RSS-feeds if available, or parsed from HTML-files using document object model to find the meaningful content of advertisements. The retrieved advertisements are selected based on the classification of advertisements in the service. The system database contains a list of terms that are used to recognize the terms when the advertisements are being processed. Indexed terms are divided into classes which are later used when accessing the content and statistics of advertisements. These classes include e.g. several types of technical skills, soft skills, job titles, educational requirements and locations of workplaces. The architecture of the application is presented in Figure 1.

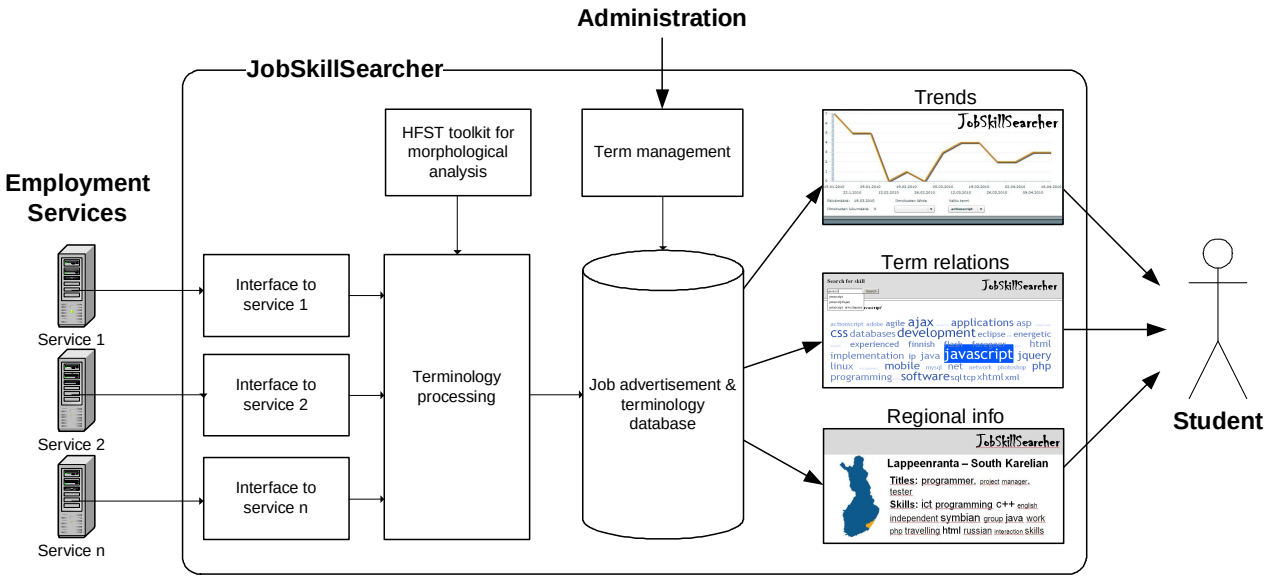

Figure 1: Architecture of the *JobSkillSearcher*.

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

The challenge of using an automatic tool for advertisement analysis from semantic approach has been recognized and considered as a limiting factor. Since the application uses Finnish employment websites where most of the advertisements are written in Finnish, some additional features had to be included for the word processing. In Finnish language verbs, nouns and adjectives may all be conjugated thus morphological analysis is needed to find the basic forms for the words in the advertisement to ease the automatic recognition against the defined terms in the database. We are using the Helsinki Finite-State Transducer toolkit (HFST) [5] based on SFST [8] for the analysis of the terms and the list of Finnish words provided by [2].

Indexed list of existing terms are created of each advertisement. This generates the possibility of searching for any terms or skills that are potentially related to each other. When new meaningful terms are recognized and classified, they will later be represented for users in their search results. The set of recognized terms has so far been maintained by service administrator. We have to consider carefully whether there is an opportunity to exploit the users on identifying and categorizing new terms.

RESULTS

There are several ways how students can put the application into use. First of all, they can use the tool to search for existence of a single term in advertisements. The tool can provide historical data about how many occurrences the term has within the selected interval at a time. When examining technical skills within a longer period, user can see if the trend is ascending or descending. As well, user can create combined searches using two or more terms – skill sets – having individual weights in which case the results are ordered based on the given weights.

Another type of search is based on the terms which appear in association with the searched term. This is useful later, when student has already some level of an idea about the potential type of job he is preparing himself/herself for. These positions require some primary skills – such as management skills or ability to program using a certain language – but also some secondary skills that are often mentioned in association with the selected primary skill. To separate the different types of terms from each other to give a better picture of requirements based on users' personal needs, each related term is presented based on our classification. A screenshot of search results for term "programming" provided by *JobSkillSearcher* based on the collected data and classification in our database is presented in Figure 2. When completing the search using a different keyword, such as "consultant" or "customer relationship management", the terms and required skills look understandably completely different.

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

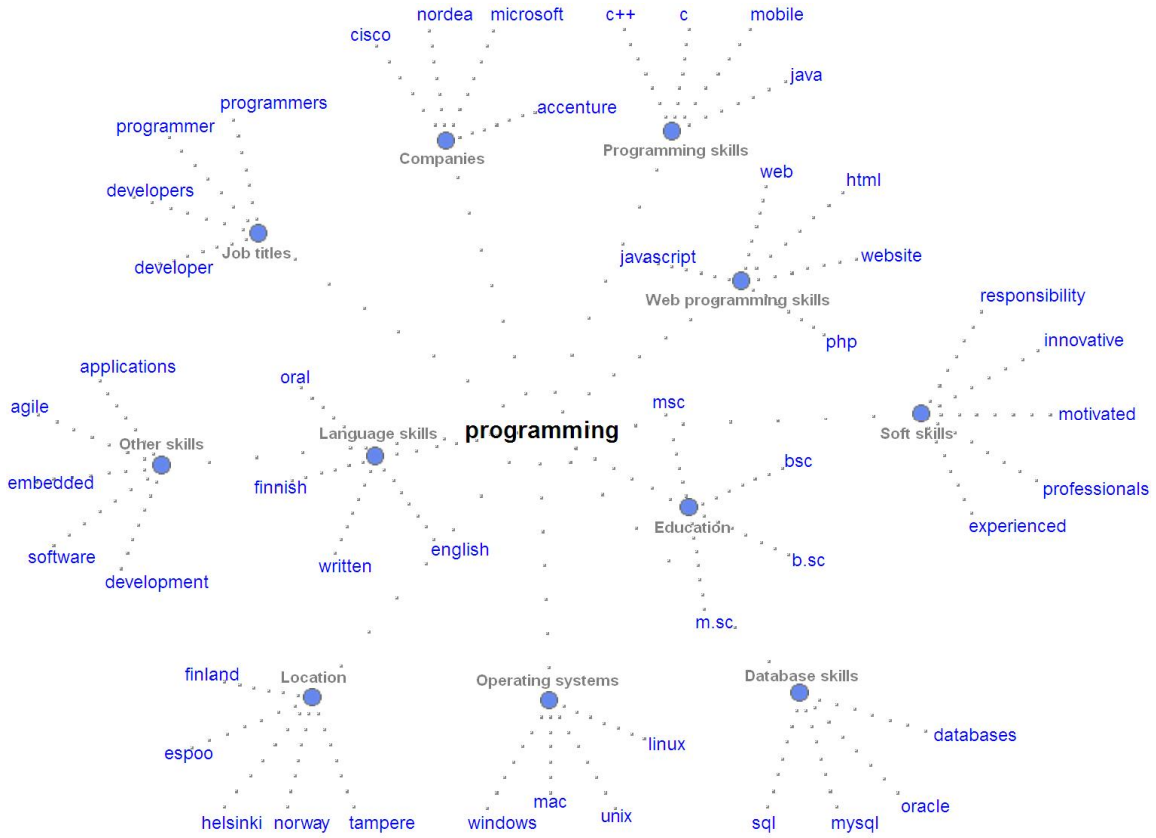

Figure 2: The most relevant terms related to term "programming" classified in groups by term type.

Sorting the announcement based on other terms such company name or location of the workplace provides a chance to examine the advertisements with more specific limitations. One can study e.g. whether some regional skill requirements can be found, whether there are differences within the soft skill requirements between different companies or which types of skills different job titles typically might require.

At this moment we have more than 3.800 advertisements in our database since March 2011. The collected advertisements consist of more than 370.000 words out of which 48.000 being different. Most of these words are useless from the point of view of skill analysis. The number of manually selected terms related to the skills that are being recognized and classified from the advertisements is currently more than 2.300, including synonyms and various conjugations of recognized terms. In addition to this, the database contains pre-defined terms that are used to identify the job title mainly based on research completed by the Federation of the Finnish Information Industries [10], names of all the localities in Finland and the provinces to recognize the location of workplace, and a bunch of ICT company names operating in Finland to find out their requirements. The structure and the list of terms have however in many respects been manually updated ever since the system was taken into use.

Based on the preliminary results, existence of relations between different hard skills is quite easy to prove. However, the number of relations even between different technical terms is notable. At the moment with our vocabulary we are able to recognize 10,2 terms in average from an advertisements which indicates that the vocabulary of terms and parsing methods still have to be improved. To be able to provide meaningful trends from a longer period, we are going to need years of time to collect more advertisements.

Regarding to Hyvärinen [1] different sizes of companies typically have different types of needs; whereas smaller companies need extensive employees that are capable of

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

doing many types of work, larger companies are often looking for specialists in more narrow area. Therefore, defining a detailed and comprehensive skill set that should be occupied for all and any type of IT-professional is impossible and there might not necessarily be significant associations between different skills for making too universal assumptions based on those. The most important issue when applying for a single job is to answer and present the skills that are related to the applied job and requested in the advertisement. For this reason *JobSkillSearcher* should not be considered as an improvement for creating personal job applications from this approach but instead as a tool used to support learning to give better idea of the potential skills that should be learnt.

CONCLUSIONS AND FUTURE WORK

In this paper we have introduced an application for identifying and analyzing meaningful terms from job advertisements and presenting the information in different ways for the users. The results are promising: Even a relatively small amount of advertisements have shown us that there are numbers of hard skills that relate to each other and different job titles from advertisement to another. Being able to find the terms related to soft skills is more challenging and their relations to each other is more complicated.

The next step is to evaluate the quality of the system by inspecting a sufficient number of advertisements manually to find out the skills and terms that are being listed there. These results will then be compared to the results that the system is able to provide. The results can be evaluated in two ways. Firstly, the evaluation will consider, how extensive list of skills we have in the system. Secondly, the demanded skills in the advertisements are not necessarily represented as a single term, but the demand of a skill might be presented in a lot more complicated way. These results will hopefully reveal us the need for further development of the application itself.

After collecting adequate number of advertisements, we are going to search the relations of the skills, their weights with each other more deeply and their dependencies to job titles. As well, we are interested about the balance between hard and soft skill requirements and if we are also able to show some regional emphasis on job descriptions.

Universities and other educational institutes having a goal to educate skilled students who answer to the needs of industry and employ can benefit from the information provided by our application. Although the primary sources for current and especially upcoming trends are others, the evolving trends as well as evanescent technologies in industry can also be observed from the information that is being collected. We believe that this information is useful for students when they set their learning goals and select the courses. The information can also be used when analyzing our teaching and whether the methods being used answer to the demands of the ICT industry in Finland.

REFERENCES

[1] Hyvärinen, J, Saranen Consulting. Interview. 24.11.2009. Espoo.

[2] Kotimaisten kielten tutkimuskeskuksen nykysuomen sanalista. Accessed 21.1.2011. Available at http://kaino.kotus.fi/sanat/nykysuomi/.

[3] Lee, D.M.S, E.M. Trauth, D. Farwell. Critical Skills and Knowledge Requirements of I/S Professionals: A Joint Academic/Industry Investigation, MIS Quarterly, 1995, vol. 19 (3), pp. 313-340.

[4] Lethbridge, T.C. What Knowledge Is Important to a Software Professional? IEEE Computer, 2000, vol. 33 (5), pp. 44-50.

Preprint. Use following information for referencing: Harri Hämäläinen, Jouni Ikonen and Jari Porras: A Tool for Visualizing Skill Requirements in ICT Job Advertisements. 7th E-learning Conference, e-Learning'11 (E-Learning and the Knowledge Society), Bucharest, Romania, August 25-26th, 2011, pp. 254-259, ISBN 978-606-505-459-2.

[5] Lindén, K., M. Silfverberg, T. Pirinen. HFST Tools for Morphology – An Efficient Open-Source Package for Construction of Morphological Analyzers. Communications in Computer and Information Science, 2009, Volume 41, pp. 28-47.

[6] Litecky, C., A. Aken, A. Ahmad, H.J. Nelson. Mining for Computing Jobs. IEEE Software, Jan./Feb. 2010, vol. 27 (1), pp. 78-85.

[7] Loth, R., D. Battistelli, F. Chaumartin, H.D. Mazancourt, J. Minel, A. Vinckx. Linguistic information extraction for job ads (SIRE project). In Proceedings of RIAO. 2010, Paris, France, pp. 222-224.

[8] Schmid, Helmut. A Programming Language for Finite State Transducers, Proceedings of the 5th International Workshop on Finite State Methods in Natural Language Processing (FSMNLP 2005), Helsinki, Finland.

[9] Surakka, S. Analysis of Technical Skills in Job Advertisements Targeted at Software Developers. Informatics in Education, 2005, vol. 4 (1), pp. 101–122.

[10] Tietoalojen liitto ry. IT-alan tehtävänimikkeistö. 11.3.2002. Accessed 11.11.2009, no longer available. http://www.ek.fi/tietoalojen_liitto/suomi/ tietoa_toimialasta/IT-alan_tehtavanimikkeisto.php.

[11] Todd, P.A., J.D. McKeen and R.B. Gallupe (1995). The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990. MIS Quarterly, vol. 19 (1), 1–27.

[12] Wade M.R., M. Parent. Relationships Between Job Skills and Performance: A Study of Webmasters, Journal of Management Information Systems, 2002, vol. 18 (3), pp. 71-96.

ABOUT THE AUTHORS

Harri Hämäläinen – M.Sc, Lappeenranta University of Technology, +358 5 621 2828, harri.hamalainen@lut.fi.

Assoc. Prof. Jouni Ikonen – DSc. (Tech.), Lappeenranta University of Technology, +358 50 555 0665, jouni.ikonen@lut.fi.

Prof. Jari Porras – DSc. (Tech.), Lappeenranta University of Technology, +358 400 555 427, jari.porras@lut.fi.