

Using Kibana and ElasticSearch for the Recommendation of Job Offers to Students

Antonio Robles-Gómez¹, Salvador Ros¹, Antonio Martínez-Gómez², Roberto Hernández¹, Llanos Tobarra¹, Rafael Pastor¹, Agustín C. Caminero¹, and Jesús Cano²

¹ Control and Communications System Department, ETSI Informática
Universidad Nacional de Educación a Distancia (UNED), Spain
{arobles, sros, roberto, llanos, rpastor, accaminero, jcano}@scc.uned.es

² Student, ETSI Informática
Universidad Nacional de Educación a Distancia (UNED), Spain
martinezgamez.antonio@gmail.com

Abstract. This paper proposes the use of Kibana and ElasticSearch for the efficient recommendation of job offers to students, in order to allow them to improve their professional career. Both Kibana and ElasticSearch tools are integrated in a self-developed web-system at UNED. By using Kibana, users can get a graphical representation of the job situation for a time period, which could indicate them what is the knowledge most wanted by companies. Also the system can let users search for information using keywords within the job offers stored in the system. This is intended to provide users with a picture of the job market, which allows them to focus their training on the most demanded professional profiles by employers, or to find hiring trends. For storing all the collected information, ElasticSearch has been deployed in order to perform agile queries from not-structured information, like job descriptions.

Keywords: Academic Analytics (AA); Kibana/ElasticSearch; Visualization, Education; Professional Career.

1 Introduction

In recent years, the production of digital information has raised in the context of employment. One of the key features of this growth is the existence of multiple massive sources of information, including social networks (such as InfoJobs [4] or Monster [8]) and Internet services, among others. The use of social networking in education as a motivating tool and its use to determine the impact of education on labor market integration is not a new issue. For instance, LinkedIn is employed in [1]. A key topic associated with the volume of data is the semantic and structural heterogeneity of such data (structured, semi-structured...). This fact adds additional complexity to the extraction, processing, and visualization of data, since classical methods cannot be employed.

This variety of heterogeneous pieces of information gives birth to the Big Data paradigm, which has a great potential to revolutionize different dimensions of society.

However, there are still research areas to explore to obtain a higher throughput. In the field of education, the Learning Analytics (LA) area covers this paradigm.

Our current work is focused on analyzing the students' learning to improve it or to make recommendations, but not in the students' learning process itself. More specifically, our attention is focused at the field of Academic Analytics (AA) in an institutional level, in order to help our University to analyze students' learning profiles and academic performance. It is very important for an Institution or University to give their students the possibility to check the job offers available depending of the level of knowledge. Also, the administrative staff could carry out this task in order to analyze the requirements of companies. This can be a starting point for the Institution to perform additional studies in order to improve the employability of students.

Therefore, this paper proposes a self-developed web-system at UNED based on Kibana and ElasticSearch. It can obtain information from several sources of job offers, such as Infojobs and Monster [8], and make an educational counseling for the successful employment of students. This is a relevant society challenge, since this fact increases the students' possibilities of graduation and professional development. Kibana [7] allows us to graphically represent a selection of relevant indicators for job offers. Users in the system can play an active role, since they get a graphical representation of the job situation with relevant information about the job offers, and are also able to search information of interest by using keywords. This is intended to provide users with a picture of the job market to adapt their professional career to the profiles most demanded by employers or find hiring trends, niche specialization, etc.

From the back-end side, job offers are gathered by using both web-scraping techniques and/or by means of APIs, depending of the availability of information in the sources of information [9]. To manage no-structured or semi-structured information, ElasticSearch [3] is being broadly employed in the context of Big Data paradigm, for agile visual queries, like job descriptions. In the LASI 2016 conference, a workshop about Kibana and ElasticSearch was presented. More information can be found in [10].

The structure of this paper is as follows: Section II depicts the data extraction and processing procedures. After that, Section III focuses on the data visualization phase for the recommendation of job offers to students. Finally, Section IV discusses our conclusions and suggests guidelines for future work.

2 Data Extraction and Processing

The information hosted on the web is normally unstructured, or at most semi-structured –in the case that access mechanisms have been provided to developers. Some websites take advantage of APIs (Access Programming Interfaces). These APIs operate as a bridge between the users' browser and the REST services and they allow getting JSON responses to source's invocations by using the field selectors. For instance, the LinkedIn API has many restrictions which limit its use. For this reason, this data collection method is not the most appropriate in some cases, as stated in [2]. Depending on the user's permissions, several professional on-line networks allow getting information, such as the actual job position, job requirements, summary description, and so on. This have been the method employed to obtain job offers in the

Infojobs professional network for the current work. In the case of the Monster professional network, a set of scrapping techniques have been employed [9]. Currently, our system supports both professional social networks. It is intended to be extended with additional social networks in a nearby future, such as LinkedIn, Twitter, and so on. The state of the art of the data extraction and processing of information can be found in [2] and [9].

As for our web-system, the administrator can manage the extraction and processing of the different sources of information, for instance, starting, stopping or modifying parameters such as the time between captures. Job offers will be stored in ElasticSearch for a more advanced management of job offers.

Once data is retrieved from its source, our self-developed web-system carefully preprocesses information to guarantee the quality of the data collection and its later aggregation [6]. Once a job offer is extracted from the corresponding source of information, this is cleaned and adapted to a new ontology defined in [9] for the educational counselling based on professional indicators. Depending on the professional social network, the gathered information can have a different level of quality. We filter all data to keep common information in our dataset, independent from the specific source of information, in order to have common indicators for all the job offers. The UNESCO's ICSED [5] education degree levels classification presents a revision of the ISCED 1997 levels of education classification. These levels of education are employed for matching job offers to students' preferences.

On the other hand, we use a relational database to save the data provided by web crawling tools in a structured way. It has also been considered appropriate the storage of data in a powerful server of searches, in order to experiment with flexibility which functionalities and performance this type of product presents when storing and performing searches by text. Within this category we can mention Apache Solr, ElasticSearch, or MongoDB. In these kinds of databases, the information can be stored in the form of JSON documents, which are indexed so that access to information is much faster than with relational databases. These are open-source, and can be used for the most common programming languages by using REST services. They also offer additional tools that give value and complete the initial product, among others, data visualization and analysis tools. The most powerful and agile database is ElasticSearch, after studying all of them. ElasticSearch is based on Lucene, so it can be seen as a powerful search-engine for texts, which can be deployed easily in a distributed way. Additionally, it can handle RESTful services in an efficient way.

3 Data Visualization

The final decision of using ElasticSearch compared to any other competitor was also due to the need to implement a system that offered different graphs that allowed users to interpret the labor market. This is where Kibana [7], another product of the company Elastic takes part. Kibana eases the analysis and visualization of data thanks to its intuitive interface, which also offers the power to share or insert control panels. Fig. 1 shows an example of job offers stored in ElasticSearch through the Discover tab of Kibana.

Kibana only queries data held in Elasticsearch, which acts as both the data store and the analytics engine. Fig. 2 show a visualization panel, a Dashboard, which has been fully-integrated from Kibana as a frame in our self-developed web-system. This panel is totally flexible and adaptable from the Kibana interfaces, as shown in Fig. 3. When any of the graphics of the panel in use is modified in Kibana, this is automatically updated in the associated frame integrated in our web-system.

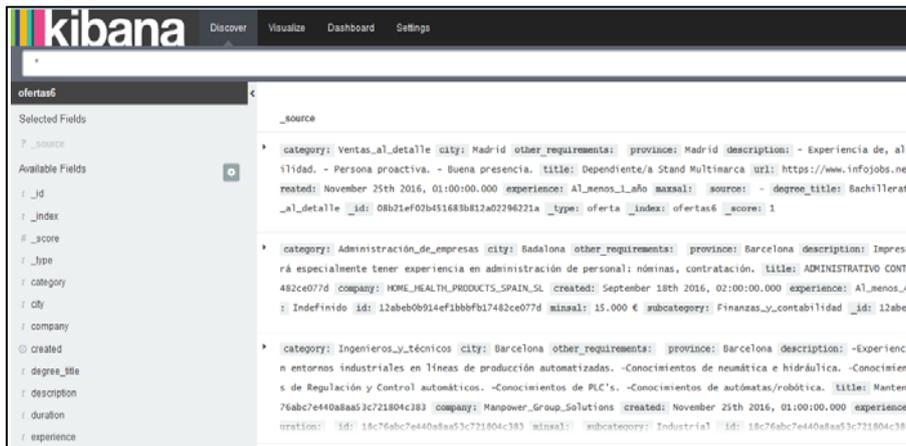


Fig. 1. Job offers stored in Elasticsearch by using Kibana (Discover interface) [7].

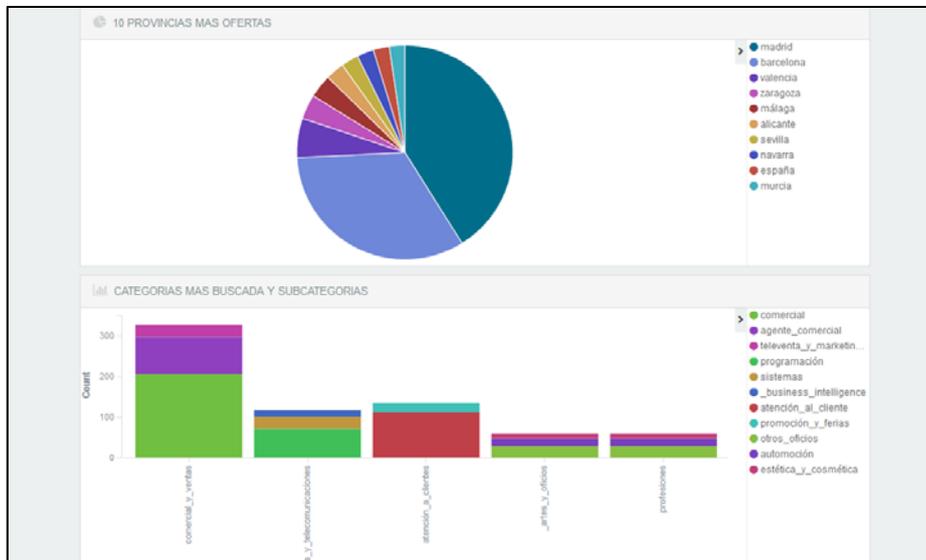


Fig. 2. Visual dashboard integrated in our web-system.

As an example, Fig. 2 and Fig. 3 show a set of analytics for job offers collected from the different professional networks that our expert system supports. In both figures, we can observe the ten provinces with more job offers. In this case, Madrid and Barcelona are the cities with more job offers. The second graphic is related to number of categories and subcategories of job offers. The “commercial_y_ventas” category is the one with more job offers and, as the most popular subcategory, we have “comercial”. In Fig. 3, we can observe additional graphics, and other relevant information, such as the amount of job offers located in the system, and the date of the last offer captured by the system.



Fig. 3. Visual dashboard located in the Kibana website.

Finally, users can perform free-text searches by employing the powerful ElasticSearch engine from our application, so looking for relevant offers that could be interesting from his/her point of view. In Fig. 4, an example is shown. In this case, the user has performed a search using the “Informática” word, and our system has searched and presented all the related job offers located in the ElasticSearch database.

Oferas

ANALISTAS Y PROGRAMADORES: TESTING

<https://www.infojobs.net/madrid-barcelona-zaragoza/analistas-programadores-testing/of-45d037e38348319da228116a4a1a97>

Experiencia mínima de 3 años en cualquiera de los entornos: BARCELONA: JAVA J2EE, PL/SQL, HOST, HTML5, Liferay, QView, TESTING, BIG DATA, ASP, NET, C++, TIBCO. Todos con INGLÉS. VALENCIA: JAVA J2EE, PL/SQL. Todos con INGLÉS. ZARAGOZA: HOST, JAVA J2EE, C#, NET, C++ MADRID: HOST, HTML5, JAVA J2EE, BIG DATA, .NET, C++, PL/SQL.

ANALISTAS Y PROGRAMADORES: FRONTEND ANGULAR

<http://www.infojobs.net/barcelona/analistas-programadores-frontend-angular/of-ec0716f5354d52ab1be13bb40f6130>

Desarrolladores (AF, AT, AP) en alguno de los siguientes entornos : JAVA J2EE, PL/SQL, C# NET, HTML5, Liferay, C++, COBOL/CICS/DB2.

Analista Programador_Java con Microservicios

http://www.infojobs.net/alicante-atacantanalista-programador_java-con-microservicios/of-1b77d1190034a41bedb1638e1662f56

GfI, compañía internacional de Servicios Informáticos, especializada en el desarrollo de proyectos tecnológicos, busca Analistas Programadores Java/J2EE con Spring, para su Centro de Servicios en Alicante. El candidato seleccionado trabajará en un proyecto basado en Microservicios, implementando Java/J2EE con distintas versiones de Spring como Spring-Data o Spring-Boot con funciones tanto técnicas como de análisis, integrándose a equipos altamente cualificados Java/J2EE, en los que podrá adquirir y compartir experiencias y conocimientos. Ofrecemos: - Contrato indefinido - Posibilidad de evolución en la compañía - Remuneración competitiva, con una evaluación y revisión salarial anual - Formación en idiomas (inglés, francés) - Entorno adecuado para continuar desarrollando habilidades de programación y análisis Java/J2EE.

AIT- Project Manager apps móviles

<https://www.infojobs.net/barcelona/ait-project-manager-apps-moviles/of-4eed8dd5af4a66a27396cfc0fac931>

Experiencia en la gestión de proyectos de apps móviles. Inglés indispensable. Conocimientos de infraestructura Backend valorable. Conocimientos en Java, HTML y CSS valorable y en general lenguajes de programación iOS y Android muy valorable.

Analista programador/a Java

<http://www.infojobs.net/valencia/analista-programador-java/of-129b24cb58f4a818bc5e91a0fa3198>

Fig. 4. Free-text search during the registration process along the job offers stored in ElasticSearch. Example of searching the “Informática” word.

4 Conclusions

Learning does not take place in an isolated context, but occurs in a social, cultural and working atmosphere. For this reason, it is very relevant for our students the analysis and aggregation of external sources of information from current professional networks. This will help students to guide their education and, therefore, their professional career. For this purpose, several professional sources of information, such as Infojobs and Monster, are available on the Internet. For this reason, a large amount of data is currently being obtained from these websites in the field of Big Data, and more specifically, in the topic of Academic Analytics (AA), as stated above.

This work proposes the use of Kibana and ElasticSearch for the recommendation of job offers to students to improve their professional career. Both Kibana and ElasticSearch tools are integrated in a self-developed web-system. By using the Kibana interface, users can get a graphical representation of the job situation during a period of time, which could indicate them what the most demanded knowledge by companies is, and search relevant keywords that fit their necessities or preferences. This is intended to provide users with a picture of the job market that allows them to focus their training on the professional profiles most demanded by employers. To store the collected information, ElasticSearch has been deployed for its ability to perform agile

visual queries from not-structured information, like job descriptions. Both Kibana and ElasticSearch are fully-connected and integrated with our web system, used for linking job offers and users' preferences. The recommendation step matches the ontology proposed in [9] with the common selected indicators for job offers. In addition to this, users can perform searches with some criteria, such as location, category, and so on.

Acknowledgements.

Authors would like to acknowledge the support of the European research project ERC-2015-STG-679528 POSTDATA, and the local project (2014I/PPRO/031) from UNED and Banco Santander; and the Region of Madrid for the support of E-Madrid Network of Excellence (S2013-ICE2715). The authors also acknowledge the support of SNOLA, officially recognized Thematic Network of Excellence (TIN2015-71669-REDT) by the Spanish Ministry of Economy and Competitiveness.

References

1. Case, T., Gardiner, A., Rutner, P., Dyer, J.: A LinkedIn Analysis of Career Paths of Information Systems Alumni. *Journal of the Southern Association for Information Systems*, vol. 1, no. 1, Winter 2013
2. Dai, K., Nespereira, C. G., Vilas, A. F., Redondo, R. P. D.: Scraping and Clustering Techniques for the Characterization of LinkedIn Profiles. In *Proc. 4th International Conference on Information Technology Convergence & Services*, pp. 1–15, January 2015
3. ElasticSearch. <https://www.elastic.co/> (*accessed 2017-06-07*)
4. Infojobs. <https://www.infojobs.net/> (*accessed 2017-06-07*)
5. Institute for Statistics of the United Nations Educational, Scientific and Cultural Organization (UNESCO): "International Standard Classification of Education: ISCED 2011". <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf> (*accessed 2017-06-07*)
6. Jacikevicius, Z: Introduction to Data Quality, February 2016, <http://www.datasciencecentral.com/profiles/blogs/introduction-to-data-quality> (*accessed 2017-06-07*)
7. Kibana. <https://www.elastic.co/products/kibana> (*accessed 2017-06-07*)
8. Monster. <http://www.monster.es/> (*accessed 2017-06-07*)
9. Robles-Gómez, A., Ros, S., Martínez-Gámez, A., Hernández, R., Caminero, A. C., Tobarra, L., Pastor, R & Cano, J.: Defining a Novel Ontology for Educational Counselling based on Professional Indicators. In *Proc. Workshop on Applied and Practical Learning Analytics (WAPLA)*, pp. 1–8, September, 2016
10. Vozniuk, A., Rodríguez-Triana, M. J., & Gillet, D. CLEO - Workshop on Contextual Learning Analytics Enforcing data Ownership. Building Interactive Learning Dashboards with Logstash, Elasticsearch and Kibana (ELK). In *Learning Analytics Summer Institute 2016 (LASI 2016)*, Bilbao, Spain, June, 2016. <https://infoscience.epfl.ch/record/220334> (*accessed 2017-06-07*)