

Feeding employee analytics with knowledge graphs

Maria Cesaria Giordano^{1,3}, Pasquale Davide de Palma³,
Andrea Giovanni Nuzzolese², Paolo Ciancarini¹

¹DISI, Università di Bologna, Bologna, Italia

²STLab, ISTC-CNR, Roma, Italia

³HRCoffee srl, Molfetta, Italia

maria.giordano15@unibo.it, davide@hrcoffee.it,
andrea.nuzzolese@cnr.it, paolo.ciancarini@unibo.it

Abstract

In this paper we present an industrial research project lasting 36 months carried out at the Department of Computer Science and Engineering of the Alma Mater Studiorum University of Bologna within the XXXVII Cycle of doctorates. The project is a joint collaboration among companies and research institutions, involving the Italian National Research Council (CNR), Confindustria and the startup HRCoffee. The focus of this research is about the extensive adoption of knowledge extraction and engineering solutions for enabling advanced AI capabilities in the context of People Analytics. The objective is to investigate how to create semantic knowledge graphs from different sources and applications that are typically used for the management of human resources (HRs) processes inside companies. This is key for enabling a variety of tasks encompassing decision making, expertise retrieval, and HR development. For example, an HR department might be interested in determining career plans projections of employees with certain job families and seniority in order to improve employee retention and reduce turnover costs.

Node of the CINI-AIIS: Università degli Studi di Bologna

Unit: Dipartimento di Informatica - Scienza e Ingegneria (DISI)

1 Introduction

In recent years, according to “AI at Work”¹ survey carried out by Oracle on over 600 HR managers, AI is being used in various processes related to the world of human resources. In particular, respondents use this technology in several sectors, that is: (i) 40% in the sourcing, screening and interview phases; (ii) 29% in career development and training; (iii) 26% to improve the recruiting of candidates; (iv) 24% for perfor-

¹<https://www.oracle.com/human-capital-management/ai-at-work/> last visited on January 18th 2022.

mance management; and (v) 23% to provide recruiters with insights and predictive actions on the hiring process.

The use of AI in the recruiting process has improved the completion rate of the job application, decreased the screening time, automated the process of answering candidates’ doubts through the use of chatbots, and reduced time to hire and guaranteed feedback to each candidate. Thanks to the AI it was possible to carry out repetitive procedures quickly, in which the human contribution does not give added value, such as the sourcing phase, CV screening, the organization of physical interviews, the creation of an order of preferences (ranking) between CVs. According to [Jarrahi, 2018] all those situations demand the processing of masses of information at a speed beyond the cognitive capabilities of even the smartest human decision makers. The latter is known in literature as the *information horizon problem* [Gangemi e Presutti, 2010]. However, in recent years, AI with superior quantitative, computational, and analytical capabilities has surpassed humans in complex tasks.

In this scenario, our research aims specifically to bring new methods supported by AI in the field of People Analytics for improving the employee retention processes. The increased acquisition of tools and applications for managing HR processes is a breakthrough towards the creation of “data-driven” organisations, i.e. organisations that effectively and consistently use data in their decision-making process across all levels of the organisation. This, thanks to the knowledge of people, requires a detailed and constant monitoring in order to collaborate with other departments of a same company and develop new individual and business growth scenarios.

2 Background

In 2019 IBM presented the “Proactive Retention” program, applied to its employees in which, thanks to Watson’s artificial intelligence, it analysed the position, title and salary of employees, correlating them with information on the history of promotions and relationships with management. On the basis of these data, an algorithm calculated the probability that some professional will leave the company, thus enabling a timely intervention to meet his specific needs. This process of retaining and mobilising its talents has saved IBM \$130 million in hiring costs. This application in the IBM context encourages and motivates us to develop new methods supported by AI in employee retention processes at a historic

moment like this, where due to the COVID-19 pandemic, employees have had the opportunity to re-examine their lives and priorities. Furthermore a recent analysis by Microsoft² states that 40% of the global workforce is planning to resign within the year (i.e. 2021). Hence, companies have and will increasingly need to show employees in a way clear and transparent the career path. Recent directions in human resources management rely on the creation of digital twins [Tao *et al.*, 2018]. In the context of HR a digital twin is the virtual replica of a physical employee and. Digital twins are useful because they provide models and data that can be used to understand what could be the key elements for improving motivation and well-being of employees. In particular, within this project we want to exploit knowledge extraction techniques and semantic technologies with knowledge engineering methods to formalise the emerging knowledge from People Analytics in semantic knowledge graphs on data coming from payroll, travel, assessments, skills, training, performance related to job family, gender, age of service, level of study, job rotation. Thus, we aim at developing a method for the extracting relevant knowledge, obtained by combining data from different sources. By relevant knowledge we mean the formalisation of a semantic knowledge graph that can be exploited by intelligent applications in the task of tracing projections and trends of contextualised career plans. In this way the company will have the possibility to compare the current job family levels with the past ones, understand the existing gap and intervene promptly with refresher courses, company or family benefits, increases etc... avoiding the flight of talents and decreasing company turnover, very widespread in recent years following the limitations imposed by Covid-19.

3 Project

Before COVID-19, SMEs and the PA struggled the adoption of digital technologies for the decentralised management of their employees, preferring conventional face-to-face and centralised approaches. However, the COVID-19 pandemic has favoring digital transitions in many industries and in society as a whole. Companies have been forced in reorganising their business models by rapidly adopting digital solutions and advanced technology tools [Golinelli *et al.*, 2020]. Additionally, corporates have questioned themselves by looking for new tools and trying to bridge the cultural gap between responsibilities and objectives. Smart working, or rather homeworking, has made both public and private organisations more fragile and at the same time rich in “technology” and data. Today, the same realities are faced with big data that are difficult to interpret due to their size and heterogeneity. In particular, generating and managing corporate knowledge automatically, with particular regard to its human resources, requires, at least, to effectively address two problems, which some a[Gangemi e Presutti, 2010] in the literature have identified as:

- The information horizon problem: the number of data affecting the internal wealth of a company and the its

²<https://www.microsoft.com/en-us/worklab/work-trend-index/hybrid-work> last visited on January 18th 2022.

staff has no well-defined boundaries and grows to a rhythm of production that goes beyond cognitive abilities human beings in particular in medium-large companies;

- The problem known as knowledge soup: corporate knowledge can emerge through the systematic analysis of platforms, technologies, databases, documents that are heterogeneous in nature and semantics. Taken as a whole, such documents would reveal heterogeneous, perhaps complementary or inconsistent, knowledge that humans must reconcile and make sense of.

For the first time in history, business leaders can make decisions about their people based on in-depth data analysis rather than traditional methods of personal relationships, experience-based decisions and risk prevention. For this reason, People Analytics with the support of Artificial Intelligence acquires a strongly interdisciplinary characterization, as the production of knowledge starting from data on human resources Expertise requires the combination of behavioral sciences, psychology, sociology, economics of Work Applied mathematics Statistical learning, Optimization, Predictive analysis, Data management, Reporting, Data Visualization, Machine Learning.

The method we envision allows us to formalise and aggregate data from the various tools and sources adopted and used in HR processes into semantic knowledge graphs for easing the HR decision making. This will be decisive for the development of a new method of human capital management based on functional and immediate organizational models. For this reason, in our project we want to investigate and experimenting with semantic knowledge graphs (SKGs) for understanding and representing how information flows in organisation. The general assumption is that information flows through the social network of employees in the organisation and not through “official” hierarchical or matrix-shaped organisation charts. Therefore it follows logically that it would really benefit the HR function to understand and analyse this information flow, through social network analysis. Although there is no consensus on what a knowledge graph is [Hogan *et al.*, 2021], by SKG we mean a mathematical graph representing factual, conceptual, and procedural knowledge in the form of *triples* (subject, predicate, object) defining binary relationships (via predicates) between entities (i.e. subject and object). Creating SKGs from information flows inside an organisation requires to define a framework based on a novel knowledge extraction (KE) approach that integrates machine reading [Etzioni *et al.*, 2006] with social network analysis tools and with the key metrics for HR management.

Machine reading is critical, because it allows us to gather knowledge graphs directly from unstructured text, which is frequently the core data source in many real word scenarios. Machine reading enables the generation of structured knowledge from text. An example of state of the art machine reader is FRED [Gangemi *et al.*, 2017]. Then, we apply People Analytics by leveraging the knowledge graphs resulting by our KE component. The People Analytics based on KGs will be integrated on the platform designed and implemented by HRCoffee, which is a leading company in Italy in the filed of

People Analytics.

We want to extend the IBM case to other sources. Examples of sources we want to consider are (i) company databases and social networks, (ii) curricula, (iii) posts from social media, (iv) the web of data for the retrieval of advanced contextual data (e.g. data geospatial), etc.

In Figure 1 we illustrate the approach we apply by leverage knowledge graphs, that is:

- analysis of key metrics at the state of the art for gathering relevant insights on workers' trajectories from HR data
- analysis of data generated by focusing on several aspects to understand how the variety of skills to each other within companies.
- application of People Analytics in order to define career trajectories in a predictive (cf. Future Trend in Figure 1) and prescriptive perspective, thus exploiting knowledge extraction techniques as previously described.

4 Related work

In the context of this project we must consider the three macro-areas indicated in the following subsections.

4.1 Complexity of organizations

The theory of complexity in the business environment is the study of causes and factors that determine behavioral or partially inexplicable modes in a logic that is merely referable to the sums of the parts, connected between people and the reference environment [Padroni,]. In practice, the system seems to assume an unexpected behavior that is unrelated to that of its components. The "system" can be made up of various phenomena not only material but also of culture, politics, economics. [De Toni, 2011; De Toni e Pessot, 2021] compare classical management with complex management. In the classic, it predicts that the organization is simple, in a stable environment and in the foreseeable future based on projections of time series. Success is achieved in these conditions through balance and stability. The complex model is based instead on the idea that the organization is a complex adaptive system, in a turbulent environment and in the foreseeable future only in part thanks to the study of megatrends. In this situation, success derives from non-equilibrium and from change, like survival for complex adaptive systems, i.e. according to [Pascale *et al.*, 2000] *equilibrium is dead*.

4.2 Knowledge Extraction e Knowledge Engineering

Knowledge extraction technician from heterogeneous sources (cf. knowledge soup problem) and semantic technologies with knowledge engineering methods to formalize the emerging knowledge from People Analytics in knowledge graphs. Regarding knowledge extraction techniques, we want to deepen (i) both tools able to manage structured data sources (ex. CSV, relationship database) through machine learning and distributional approaches [Mikolov *et al.*, 2013], (ii) and tools in able to extract business knowledge from unstructured sources (ex. natural language). A promising solution for the analysis of the last point exposed is FRED [Gangemi *et al.*, 2017].

4.3 People Analytics

In an increasingly competitive workplace characterized by numerous changes, making decisions based on real data becomes increasingly important. For this reason, People Analytics today plays an increasingly important role in all companies by providing a strategy to understand what works or not in order to improve the company and make it grow. The vast majority of the data collected by HR managers is nothing more than raw, generic and decontextualized data, which alone cannot provide any concrete information.

5 Conclusion

After analyzing the current scenario and the future needs of companies, with industrial research we want to create a human resource management model based on the predictive and prescriptive People Strategy capable of adapting to all types of public and private organizations. For the recovery of the data and the development of the methodology, the companies that use the services of Hrcoffee will be involved in the research.

References

- [De Toni e Pessot, 2021] Alberto F De Toni e Elena Pessot. Investigating organisational learning to master project complexity: An embedded case study. *Journal of Business Research*, 129:541–554, 2021.
- [De Toni, 2011] Alberto F De Toni. Teoria della complessità e implicazioni manageriali: verso l'auto-organizzazione. *Sinergie Italian Journal of Management*, 10(81):77–96, 2011.
- [Etzioni *et al.*, 2006] Oren Etzioni, Michele Banko, e Michael J Cafarella. Machine reading. In *AAAI*, volume 6, pages 1517–1519, 2006.
- [Gangemi *et al.*, 2017] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, e Misael Mongiovi. Semantic web machine reading with fred. *Semantic Web*, 8(6):873–893, 2017.
- [Gangemi e Presutti, 2010] Aldo Gangemi e Valentina Presutti. Towards a pattern science for the semantic web. *Semantic Web*, 1(1, 2):61–68, 2010.
- [Golinelli *et al.*, 2020] Davide Golinelli, Erik Boetto, Gherardo Carullo, Andrea Giovanni Nuzzolese, Maria Paola Landini, e Maria Pia Fantini. Adoption of digital technologies in health care during the covid-19 pandemic: systematic review of early scientific literature. *Journal of medical Internet research*, 22(11):e22280, 2020.
- [Hogan *et al.*, 2021] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 12(2):1–257, 2021.
- [Jarrahi, 2018] Mohammad Hossein Jarrahi. Artificial intelligence and the future of work: Human-ai symbiosis

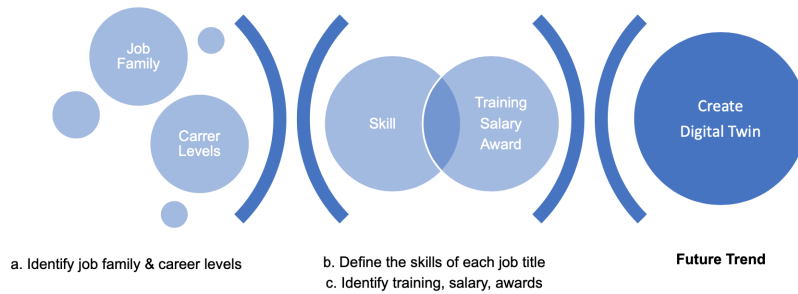


Figure 1: Research Approach

in organizational decision making. *Business Horizons*, 61(4):577–586, 2018.

[Manuti *et al.*, 2014] Amelia Manuti, Pasquale Davide De Palma, et al. *Why Human Capital is important for organizations: People come First*. Springer, 2014.

[Manuti e De Palma, 2017] Amelia Manuti e Pasquale Davide De Palma. *Digital HR: A Critical Management Approach to the Digitilization of Organizations*. Palgrave Macmillan, 2017.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, e Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, e Kilian Q. Weinberger, editors, *NIPS*, pages 3111–3119, 2013.

[Padroni,] Giovanni Padroni. *Aspetti della complessità e sensibilità «postmoderna» nelle dinamiche organizzative e del capitale umano*, volume 76.

[Pascale *et al.*, 2000] Richard T Pascale, Mark Millemann, e Linda Gioja. *Surfing the edge of chaos: The laws of nature and the new laws of business*. Currency, 2000.

[Tao *et al.*, 2018] Fei Tao, He Zhang, Ang Liu, e Andrew YC Nee. Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, 15(4):2405–2415, 2018.