Education Meets Knowledge Graphs for the Knowledge Management

Renato De Donato[1], Martina Garofalo[2], Delfina Malandrino[1], Maria Angela Pellegrino[1](✉), and Andrea Petta[1]

[1] Dipartimento di Informatica, Università di Salerno, Fisciano, Italy
rended83@gmail.com, {dmalandrino,mapellegrino}@unisa.it, andrpet@gmail.com
[2] ACT OR S.r.l., Rome, Italy
martina.garofalo@act-operationsresearch.com

Abstract. Data are a crucial source for informed decision making. However, their unrestrainable growth requires approaches and tools to learn how to query and identify data of interest. The problem we want to face is how to guide students in going beyond the passive inspection of results returned by a search engine, and in actively searching for the data that best answer their questions. Our final goal is to guide future citizens in actively creating their knowledge supported by data. In particular, we focus on Knowledge Graphs and how they can be used in the knowledge management process in the educational context. We present *ELODIE* and the *datalet* mechanism as a tool to support the knowledge management process. We will describe its features, and we will discuss how teachers and students could exploit it in the educational context.

Keywords: Knowledge management · Data literacy · Information retrieval · Data visualization · Knowledge Graphs · Semantic web · Semantic search

1 Introduction

Data are vertiginously rising [1] and because of their unrestrainable growth, "*data is the new oil, but if unrefined it cannot really be used*" [2]. Therefore, there is the necessity to learn how to read, work with, analyze, and argue with data [3], i.e., we should experience the *data literacy*. It is a crucial skill that future citizens have to acquire by 2030 [4] and it is strongly related to the ability of *searching* and *extracting knowledge* out of raw data. Knowledge is widely studied from several different perspectives, and there is no consensus on its definition. In this article, we consider *knowledge* as *information in action* [5]. It represents one of the phases of the *data* (or knowledge) *pyramid* [6] composed of *data*, *information* and *knowledge*. *Data* are observations or facts; the *information* is inferred from data in the process of answering questions such as *who*, *what*, *where*, *how many*,

Z. Kubincová et al. (Eds.): MIS4TEL 2020, AISC 1236, pp. 272–280, 2021.
https://doi.org/10.1007/978-3-030-52287-2_28

when and make data useful for actions or decisions; the *knowledge* is achieved by processing, organizing or structuring information. The process to reach the knowledge starting from data is referred to as *knowledge management* (KM).

The problem we want to face is how to engage students in the KM process actively and to make them leave the position of an indifferent spectator of retrieved results. Usually, users passively accept the list of first results returned by a search engine without further investigating the ones ranked as less important. Ironically, Elon Musk says that the "*safest place to hide a dead body is the second page of Google search results*" since most people stop on its first page. We desire to spur school pupils in assuming the *control* in searching on the Web, in *critically* choosing results of interest, and in *actively* creating their knowledge.

Search engines mainly query documents and look for the occurrence of the searched terms in those files. In this article, we propose a shift from querying documents to semantic searches, i.e., we focus on the exploitation of Knowledge Graph (KG) by the KM process in the educational context. A KG is a knowledge base modeled as a graph, and it combines linked data technologies and ontologies.

We have designed a guided workflow to query KGs and to visualize the retrieved results. While the component to query KG is named *ELODIE*, the results visualization component will be referred to as *datalet mechanism*. In this article, we present the combination of ELODIE and the datalet mechanism as a tool to support the KM process. In particular, it aims to guide students in querying KGs and in replying to questions formulated in natural language to retrieve information. Upon the retrieved results, students can make decisions and acquire knowledge. Our main contributions are the following:

– the proposal of a collection of educational use cases that can benefit from KGs by highlighting potentialities and required skills to query KGs;
– considerations and evidence that the KM process can take advantage of ELODIE and the datalet mechanism [7].

The structure of this article is the following: in Sect. 2, we consider the role of KM and KGs in the education; in Sect. 3, we analyze how the KM process can exploit KGs and in Sect. 4, we point out how ELODIE and the datalet mechanism can be used in the KM process; in Sect. 5, we discuss the potentialities and the required skills to exploit KGs in the educational context; finally, we will conclude with some final considerations and future directions.

2 Related Work

KM is a set of practices that leads to the use of data in decision-making [8]. Rodrigues and Pai [9] defined a list of crucial factors that must be taken into account to develop a suitable KM strategy for schools. Among others, they consider as a critical dimension the *technology and infrastructure*; the *acquisition and learning*, i.e., methods to improve the searching and learning of knowledge; the *dissemination and transfer* of gained knowledge with others. Therefore, students have to learn how to acquire knowledge and mainly how to disseminate it [10], supported by technological solutions. Our proposal, ELODIE and the datalet mechanism[1] [7], could represent a technological solution to support students in the KM process by enabling the knowledge acquisition by querying KGs. About the dissemination, ELODIE and the datalet mechanism are installed in a social platform to share and discuss acquired knowledge.

Our interest in bringing closer KGs and KM is justified by the observation that KGs can facilitate and enhance KM since they explicitly provide a structure to data, and this structure is instrumental in supporting semantic searches and answering more profound and complex questions [11]. KGs and KM are usually combined at the university level [12]. We propose to introduce this approach also in the early stage of the educational plan. This proposal is due to the observation that KGs are gradually applied to teaching and learning in the educational domain to spur students in thinking about entities and their logical relationships [13]. It can promote a more in-depth comprehension of how data are connected among them within a certain domain [13]. In the educational context, they are usually referred to as concept map [14]. It has been proved that concept maps can significantly improve students' learning achievement: it prompts constructive learning, reflective ability, and active interaction [15]. Thus, we assess that KGs can be positively perceived also by young students.

3 Knowledge Management by Knowledge Graphs

Advantages of Knowledge Graphs. An increasing interest is manifested over KG publication: the LOD Cloud [16] (a KG that collects most of the published KGs by academia and industry) counted 12 datasets in 2007 and currently contains 1,239 datasets. Despite the *quantitative* reason to exploit KGs, also the *provenance* is an important aspect: several virtuous institutions invested or are investing in publishing data in the linked format, such as, Europeana[2], Eurostat[3], ISTAT, Beni Culturali[4], the British Museum[5]. Furthermore, it is highly recommended to *interlink* KGs [17] and it implies the possibility to navigate from a KG to another. Because of the extensive range of heterogeneous information stored in KGs, for their easy navigation, thanks to their quantitative and qualitative properties, they could behave as a critical resource for KM.

Challenges Posed by Knowledge Graphs. First of all, it is not always immediately clear how to translate a natural language question in a query over a KG since data modeling can be domain-dependent or domain-agnostic. Therefore, it could be hard to conceptualise data to query and to guess the used terminology.

[1] In the article presented at CSCWD, the same tool is named SPLOD since it was presented as a component of SPOD. Here we consider the tool extrapolated from the surrounding platform, and we refer to it by ELODIE and the datalet mechanism.

[2] https://pro.europeana.eu/page/linked-open-data.

[3] https://ec.europa.eu/eurostat/web/nuts/linked-open-data.

[4] http://dati.culturaitalia.it/.

[5] https://old.datahub.io/dataset/british-museum-collection.

The second aspect is related to the available querying languages. SPARQL is the most common one even if it proves to be too challenging, mainly for lay users [18,19]. It requires technical skills in generic querying languages and in understanding the semantics of the supported operators by SPARQL. Therefore, there is an interest in developing tools able to implicitly compose queries by hiding the underlying complexity to open KGs also to lay users [20].

High-Level Knowledge Management Process. The KM process 1) starts from data that models facts, 2) reaches the information by querying data and collecting results, and 3) realises the knowledge by analysing and structuring achieved information in interpretable and shareable artefacts.

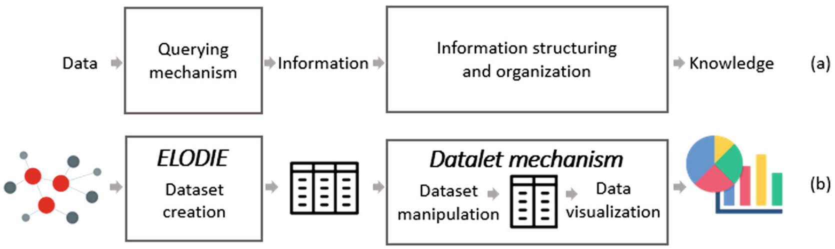

Fig. 1. In (a) we represent the high-level KM approach by KGs, while in (b) we report how we interpreted each phase in our tool. The KM starts from data (KG in our case); data must be queried to retrieve information (represented by a data table); information must be manipulated, and the knowledge can be gained. While the dataset creation phase is realized by ELODIE, the knowledge can be acquired by analyzing and making decisions upon the realized charts.

In Fig. 1a, we design the KM process by KGs, and we report how we interpreted each phase in our tool (Fig. 1b). KGs are the queried data. To move from data to information, we have to define a querying mechanism. As explained in the challenges posed by KGs, we aim to hide the underlying syntactical complexities while raising questions. Therefore, we designed a querying mechanism that leads to 1) the formulation of SELECT queries in natural language and 2) that organises results (i.e., the retrieved information) in data tables. We named this phase as *dataset creation*, and we address it by ELODIE. To move from information to knowledge, users need to process and organise retrieved results. Starting from the first year of their educational plan, children get used to reading tables, creating (simple) charts, performing (basic) manipulation, and interpreting results to reply questions [21]. Thus, we model the gained knowledge as shareable artifacts (e.g., charts) that can justify the decision made. To achieve knowledge, results can be manipulated to be compliant with visualization modes and, then, the desired chart can be realized by the datalet mechanism.

4 ELODIE and the Datalet Mechanism

ELODIE (pronounced elədē) is a guided workflow to extract a tabular representation of data queried by KGs, while the datalet mechanism is a scaffolded approach to visualize the retrieved results and create shareable artefacts. Their combination implements the trialogical learning approach that is articulated into three learning metaphors: knowledge acquisition, participation, and knowledge creation by reusable artifacts [22]. By ELODIE and the datalet mechanism, the user can experience an individual effort and can participate in social discussions. Furthermore, the tool enables the possibility to continue the query of another user by fulfilling the collaborative creation of shared artifacts [7].

Here, we propose ELODIE and the datalet mechanism as a tool for supporting the KM by KGs. While ELODIE realizes the transformation of data in information, the datalet creation leads to acquiring knowledge. We will now analyse each phase of the process as implemented in our tool and how it is related to the KM process. As use case, we consider the desire of a professor to create a library for his/her students and has to choose the best books to collect.

The tool can be accessed via registration in SPOD[6] or can be freely tested here[7]. The code is released on GitHub[8], while quick tutorials[9] are on YouTube.

Dataset Creation. In this phase, we want to move from data (i.e., KGs) to a tabular representation of retrieved results without facing SPARQL challenges. Therefore, we proposed a faceted search interface enhanced by a natural language query. ELODIE organises nodes (referred to as concepts) and links (referred to as predicates) of the KG by facets. Students can click on any provided option to formulate their queries. In Fig. 2, the interface of this phase is visible. Given the need to create a library for children, the professor is interested in retrieving books related to the Children's literature. Furthermore, he/she has to model all the desired parameters he/she wants to take into account in the book selection, for instance, the publication date, the length of the book, books rewarded by critics. By this step, users learn how to actively control retrieved data by choosing step by step, the option compliant with the desired goal. It goads to acquire *independence* and a complete *control* of the search process.

Dataset Manipulation *and* ***Data Visualization***. When the user is satisfied with the retrieved results, we can move on to the knowledge acquisition step, realized by the datalet mechanism. We split this phase in dataset manipulation and data visualization. First, the user has to manipulate the dataset to make it compliant with the desired visualization by aggregating, sorting, and filtering data. Finally, knowledge can be acquired. The creation of a shareable chart realises it. Back to our use case, we can opt for aggregating awards by authors

[6] http://spod.routetopa.eu/.
[7] https://deep.routetopa.eu/deep2/COMPONENTS/controllets/splod-visualization-controllet/demo.html.
[8] https://github.com/routetopa/deep2-components/tree/master/controllets/splod-controllet.
[9] https://youtu.be/e_o32GP-l1c.

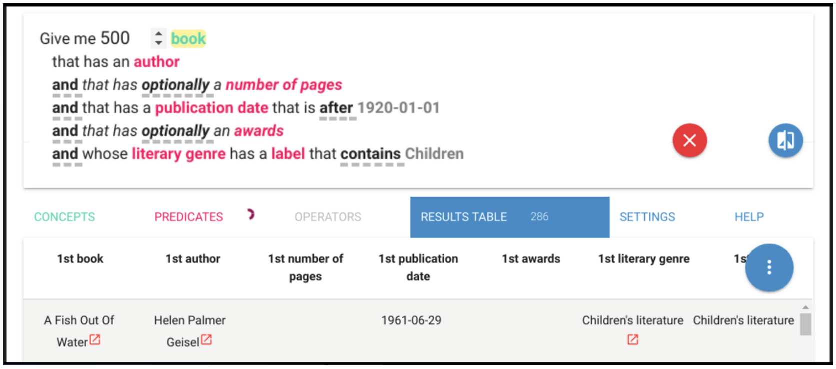

Fig. 2. It represents the dataset creation phase where users can move from data (e.g., books) to information (e.g., books matching the queried filters). By analyzing the retrieved results, the best author/book (i.e., knowledge) can be acquired.

and sort authors by counting the received prizes. We discover that Dr. Seuss (author of *The Cat in the Hat*), Dav Pilkey (cartoonist and author of *Captain Underpants*) and C.S. Lewis (author of the *Chronicles of Narnia*) are the preferred ones by critics. By considering the most recent works, we can cite *River Rose and the Magical Lullaby* written by Kelly Clarkson and *Diary of a Wimpy Kid* written by Jeff Kinney. Then, books of Dr. Seuss and Dav Pilkey are again at the top of the ranking. By considering the book-length, Dr. Seuss is at the top with 32 pages while C.S. Lewis wrote the most verbose stories. By these considerations, we acquire the awareness that Dav Pilkey cannot miss in our library. By this phase, the users learn how to acquire knowledge and how to represent it as a shareable artifact. Each performed analysis so far can be represented as a chart to exhibit evidence and substantiate decisions.

5 Discussion

Potentialities of Knowledge Graphs in Educational Settings. We assess that general-purpose KGs (such as DBpedia) are a useful source in educational contexts because of the heterogeneity of covered topics, from sport to art, from science to geography, and it can be exploited by different educational subjects. In Table 1, we report for each subject a list of useful entities and few examples of queries that can be replied by querying DBpedia (by ELODIE).

Required Skills for Knowledge Management by Knowlegde Graphs. The *KM* requires the ability of *actively* extracting information out of data and representing the acquired knowledge by a shareable artefact. It implies that future citizens must learn how to ***locate and query data*** of interest, *critically* evaluate retrieved information, learn how to ***synthesise*** it to reply the initial query, and to decide how ***represent and share*** the knowledge. To address *data literacy*, students must learn how to interpret artifacts, use them as evidence in

Table 1. It provides an overview of entities covered for each educational subject and few examples of queries whose reply can be retrieved by querying DBpedia.

Subject	Covered entities	Examples
Geography	Natural places (e.g., river), place (e.g., continent, country), mountain, volcano	The most populated continent The longest river
Science	Celestial body, unit of work, chemical substance, anatomical structure, species (e.g., animal, plant)	The orbiting bodies around the Sun
History	Military conflict, royalty, concentration camp, politician	Concentration camps turned into museums
Sport	Sport facility, athlete, activity, coach, league, event, sport season, team	Athlete taller than 2 m
Art	Fictional character, artist, museum, colour	Museums of modern art
Religion	Cleric (e.g., pope, saint)	Saint canonised in XXI century
Technology	Device, software programming language	How programming language influence each other
Music	Musical artist, musical work	Which hard rock artist play piano?
Literature	Writer, written work	Who wrote autobiographies?

discussions while arguing with data. It implies to learn how to ***interpret charts*** and be able to defend an opinion by using it as evidence. In working with KGs, future citizens have to try how data of interest can be abstracted and, then, verify how data are actually modelled. For instance, if I am interested in museums and their paintings, I have to realise that the concepts I am interested in are *museum* and *painting*. Then, I have to think about how this information can be linked: the museum *exposes* a painting or a painting *is located in* a museum. Thus, the ***modelling ability*** plays a crucial role.

6 Conclusions and Future Directions

Data are vertiginously growing and it is crucial for future citizens learning how to exploit them to the best. To actively engage students in the knowledge management, we propose ELODIE and the datalet mechanism that could have the potential to guide users in acquiring knowledge step by step. Our proposal spur students in *locating and querying data* of interest, *critically* evaluating and *synthesising* retrieved information and, finally, making decisions. We have already tested out proposal with high-school students, and we obtained positive results [7]. The next step is to test its effectiveness with younger participants.

References

1. Lynkova, D.: 39+ big data statistics for 2020 (2019). https://leftronic.com/big-data-statistics/. Accessed 14 Feb 2020
2. Humby, C.: Data is the new oil. https://ana.blogs.com/maestros/2006/11/data_is_the_new.html. Accessed 29 Mar 2020
3. Mandinach, E.B., Gummer, E.S.: A systemic view of implementing data literacy in educator preparation. Educ. Res. **42**(1), 30–37 (2013)
4. Organisation for Economic Co-operation and Development (OECD): Core Foundations for 2030 (2019). http://www.oecd.org/education/2030-project/teaching-and-learning/learning/. Accessed 21 Feb 2020
5. O'Dell, C., Grayson, C.J.: If only we knew what we know: identification and transfer of internal best practices. Calif. Manag. Rev. **40**(3), 154–174 (1998)
6. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. J. Inf. Sci. **33**(2), 163–180 (2007)
7. De Donato, R., Garofalo, M., Malandrino, D., Pellegrino, M.A., Petta, A., Scarano, V.: Linked data queries by a trialogical learning approach. In: 23rd IEEE International Conference on Computer Supported Cooperative Work in Design (2019)
8. Petrides, L.A., Nodine, T.R.: Knowledge management in education: defining the landscape. Institute for the Study of Knowledge Management in Education (2003)
9. Rodrigues, L., Pai, R.: Preparation and validation of km measurement instrument: an empirical study in educational and it sectors. In: International Conference on Knowledge Management, pp. 583–593 (2005)
10. Chu, K., Wang, M., Yuen, A.: Implementing KM in school environment: teachers' perception. Knowl. Manag. E-Learn. **3**, 139–152 (2011)
11. Yan, J., Wang, C., Cheng, W., Gao, M., Zhou, A.: A retrospective of knowledge graphs. Front. Comput. Sci. **12**(1), 55–74 (2018)
12. Mariia, R.: Knowledge graph application in education: a literature review. Acta Universitatis Lodziensis. Folia Oeconomica **3**, 7–19 (2019)
13. Cui, J., Yu, S.: Fostering deeper learning in a flipped classroom: effects of KGs versus concept maps. Br. J. Educ. Technol. **50**(5), 2308–2328 (2019)
14. Chen, P., Lu, Y., Zheng, V.W., Chen, X., Yang, B.: KnowEdu: a system to construct knowledge graph for education. IEEE Access **6**, 31553–31563 (2018)
15. Dias, S.B., Hadjileontiadou, S.J., Diniz, J.A., Hadjileontiadis, L.J.: Computer-based concept mapping combined with learning management system use: an explorative study under the self-and collaborative-mode. Comput. Educ. **107**, 127–146 (2017)
16. McCrae, J.P.: The LOD cloud (2007). http://lod-cloud.net. Accessed 22 Feb 2020
17. Berners-Lee, T.: 5-star OD (2006). http://5stardata.info. Accessed 22 Feb 2020
18. Damljanovic, D., Agatonovic, M., Cunningham, H.: Natural language interfaces to ontologies: combining syntactic analysis and ontology-based lookup through the user interaction. In: Proceedings of the 7th ISWC: Research and Applications (2010)
19. Ferré, S.: SQUALL: the expressiveness of SPARQL 1.1 made available as a controlled natural language. Data Knowl. Eng. **94**, 163–188 (2014)

20. Dadzie, A., Rowe, M.: Approaches to visualising linked data: a survey. Semant. Web **2**(2), 89–124 (2011)
21. Saddiqa, M., Rasmussen, L., Magnussen, R., Larsen, B., Pedersen, J.: Bringing open data into Danish schools and its potential impact on school pupils. In: Proceedings of the 15th International Symposium on Open Collaboration (2019)
22. Paavola, S., Hakkarainen, K.: The knowledge creation metaphor - an emergent epistemological approach to learning. Sci. Educ. **14**(6), 535–557 (2005)