

Knowledge Pattern Extraction and Their Usage in Exploratory Search^{*}

Andrea Giovanni Nuzzolese^{1,2}

¹ STLab-ISTC Consiglio Nazionale delle Ricerche, Rome, Italy

² Dipartimento di Scienze dell'Informazione, Università di Bologna, Italy

Abstract. Knowledge interaction in Web context is a challenging problem. For instance, it requires to deal with complex structures able to filter knowledge by drawing a meaningful context boundary around data. We assume that these complex structures can be formalized as Knowledge Patterns (KPs), aka frames. This Ph.D. work is aimed at developing methods for extracting KPs from the Web and at applying KPs to exploratory search tasks. We want to extract KPs by analyzing the structure of Web links from rich resources, such as Wikipedia.

1 Problem Statement and Related Work

In the vision of the Semantic Web agents are supposed to interact with Web knowledge in order to help humans in solving knowledge-intensive tasks. Though Linked Data is a breakthrough in Semantic Web it is still hard to build contextualized views over data, which would allow to select relevant knowledge for a specific purpose, i.e., to draw relevant boundaries around data. Let us suppose we are interested in events involving Arnold Schwarzenegger in the artistic context. For example, the movies that Arnold Schwarzenegger starred before starting his political career. We need to recognize “starring” situations over the knowledge about Arnold Schwarzenegger available on the Web. Such situations are represented as complex structures that relate entities and concepts according to a unifying view, e.g., Arnold Schwarzenegger having role of actor in movies during a time period. Such complex structures can be exploited for supporting a variety of knowledge interaction tasks, such as decision support, content recommendation, exploratory search, content summarization, question answering, information visualization, interface design, etc. These complex knowledge structures have been identified and described by Minsky [5], who proposed to conceptualize them as *frames*. Frames, known as *Knowledge Patterns* (KPs), have been repropounded in Semantic Web [3]. A KP can be briefly defined as “a formalized schema representing a structure that is used to organize our knowledge, as well as for interpreting, processing or anticipating information”. This Ph.D. work aims at developing methods for discovering and extracting KPs on the Web and exploit them for supporting exploratory search. Exploratory search is well known in literature. [9] is a survey that presents examples of exploratory

* Advisors: Paolo Ciancarini, Valentina Presutti, and Aldo Gangemi.

search on the Semantic Web. Existing approaches perform exploratory search by organizing knowledge according to scattered elements like classes or relations. Differently from them, we want to experiment with exploratory search based on KPs as a solution for providing summarizations and navigating knowledge.

We focus on analyzing and studying KPs in Wikipedia and some Linked Data data sets. Our aim is: (i) to develop methods for KP discovering and extraction on the Web, and (ii) to apply KP-based strategies to exploratory search and prove that the user experience improves as compared to state of the art tools.

We hypothesize that: (i) KPs can be discovered by analyzing the linking structure of Web resources.¹ In fact, linking things to other things is a typical cognitive metaphor used by humans on the Web for organizing knowledge. Hence, Web links (either hypertextual links or RDF triples) convey rich knowledge that can be used for extracting KPs. (ii) Exploratory search would benefit from using KPs. In our vision KPs can be applied to exploratory search in order to filter knowledge and guide a user to explore contextual relevant knowledge.

2 KP Extraction and Usage in Exploratory Search

We want to develop methods for discovering and extracting KPs by analyzing the linking structure of Web resources and exploit them for supporting exploratory search.

KPs sources. We identify at least two different kind of sources from which to extract KPs:

- **already existing KP or frame sources.** Examples are the FrameNet project [1] or the Ontology Design Pattern project². These resources are modelled with a top-down approach and they are typically designed by domain experts, that firstly formalize the domain semantics and then move to data. We are aimed at transforming these resources in order to represent them homogeneously as they are expressed with different formats and semantics. Possible solutions are reengineering, refactoring based on transformation rules, key concept identification, ontology mapping, etc. Initial results have been obtained and in [6] we present a solution for reengineering FrameNet in order to produce (i) a LOD data set³ and (ii) a collection of reusable KPs available as OWL2 ontologies; (ii)
- **the Web.** The Web provides a lot of heterogeneous sources from which KPs can be extracted. In order to narrow them we need to make some assumption. We want to take into account only text resources including links that (i) are associated with structured data, e.g., Linked Data, and (ii) can be formally interpreted. Wikipedia fits perfectly these assumptions. In fact, it provides rich content deriving from a collaboratively crowd sourcing performing an

¹ Under a number of assumptions (see Section 2).

² <http://www.ontologydesignpatterns.org>

³ <http://stlab.istc.cnr.it/stlab/FrameNetKCAP2011>

encyclopedic task. Furthermore, it has a RDF dump in Linked Data, i.e., DBpedia [4], and ontologies, e.g., YAGO and the DBpedia Ontology, that allow to give formal interpretation to data. For these reasons we want to focus on the extraction of KPs from Wikipedia.

For this purpose some work has been done. In [7] we present a method that we have defined for extracting KPs from Wikipedia by analysing links and their interpretation through the DPpedia Ontology.⁴ The analysis of results shows a bias due to a large number of untyped entities in DBpedia. For that reason we have investigated inductive as well abductive approaches for automatically typing DBpedia entities [8]. Experiments show that NLP is needed in order to address this task. Hence, we have investigated a NLP-based approach showing good results [2]⁵. With this approach we plan to type all DBpedia entities and then re-run the extraction of KP from Wikipedia.

KPs in Exploratory Search. In Exploratory Search tasks users want to discover what they are looking for by exploring knowledge. In some case, they have only a vague idea about what the nature of their search is. In these cases it is important to select relevant data in order to help users to filter and to summarize knowledge during their search. KPs can be used in exploratory search tasks for drawing boundaries around data in order to provide contextual relevant knowledge. This means that, depending on the context, different KPs could be automatically selected on the same data in order to make emerge the knowledge that is relevant to a user's search. We have started to experiment KPs in Exploratory Search with Aemoo⁶. Aemoo selects and organize the core knowledge about a DBpedia entity by applying KPs extracted from Wikipedia as lenses over data.

3 Evaluation

We want to evaluate both the performance of our extraction method and the quality of extracted KPs. On one hand, the evaluation of the method's performance can be easily analyzed by taking into account the execution time of the extraction algorithm. On the other, it is less clear how to evaluate extracted KPs. We think that a solution for evaluating the quality of extracted KPs derives from the combination of the following methods:

Gold standard based evaluation. We want to ask to a group of at least 3 expert users in knowledge representation to build a sample of KPs on some subset of some specific domain. Each expert has to formalize KPs with respect to the tasks proposed, e.g. to provide the core concepts that summarize what an

⁴ We have extracted 231 KPs representing the core knowledge used for describing a specific type of entity.

⁵ Typing precision is around 86%.

⁶ <http://aemoo.org>. Aemoo participated to the last Semantic Web Challenge reaching the final round.

airplane is. The set of KPs proposed will be the same for each expert in order to evaluate the inter-rater agreement among experts. KPs over a certain threshold of agreement will be the gold standard. Hence, by applying our extraction method on the same set will be possible to compare the results to the gold standard for evaluating precision and recall of our method;

User based evaluation. After the extraction of KPs we want to ask to a group of heterogeneous users in terms of education and expertise in knowledge representation to evaluate them. The evaluation will be conducted by asking users to fill a survey about the capacity of assigned KPs to be relevant to some context, exhaustive in summarizing knowledge, appropriate, etc. This will give an idea of soundness and accuracy of extracted KPs;

Task based evaluation. A comparison between an exploratory search application based on KPs and other existing exploratory search applications will be used for evaluating the effectiveness of KPs in exploratory search tasks. In this case will be asked to a group of users to solve tasks of knowledge exploration, learning, summarization, relation finding, definition extraction, etc.

References

1. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet Project. In: Proc. of the 17th International Conference on Computational Linguistics, Morristown, NJ, USA, pp. 86–90 (1998)
2. Gangemi, A., Nuzzolese, A.G., Presutti, V., Draicchio, F., Musetti, A., Ciancarini, P.: Automatic Typing of DBpedia Entities. In: Proc. of the 11th International Semantic Web Conference (ISWC 2012). Springer, Boston (2012)
3. Gangemi, A., Presutti, V.: Towards a Pattern Science for the Semantic Web. *Semantic Web* 1(1-2), 61–68 (2010)
4. Lehmann, J., Bizer, C., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - A Crystallization Point for the Web of Data. *Journal of Web Semantics* 7(3), 154–165 (2009)
5. Minsky, M.: A Framework for Representing Knowledge. In: Winston, P. (ed.) *The Psychology of Computer Vision*, McGraw-Hill (1975)
6. Nuzzolese, A.G., Gangemi, A., Presutti, V.: Gathering Lexical Linked Data and Knowledge Patterns from FrameNet. In: Proc. of the 6th International Conference on Knowledge Capture (K-CAP), Banff, Alberta, Canada, pp. 41–48 (2011)
7. Nuzzolese, A.G., Gangemi, A., Presutti, V., Ciancarini, P.: Encyclopedic Knowledge Patterns from Wikipedia Links. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) *ISWC 2011, Part I. LNCS*, vol. 7031, pp. 520–536. Springer, Heidelberg (2011)
8. Nuzzolese, A.G., Gangemi, A., Presutti, V., Ciancarini, P.: Type inference through the analysis of Wikipedia links. Submitted at the WWW 2012 Workshop on Linked Data on the Web (2012), <http://stlab.istc.cnr.it/documents/papers/ldow2012.pdf>
9. Uren, V., Lei, Y., Lopez, V., Liu, H., Motta, E., Giordanino, M.: The usability of semantic search tools: a review. *The Knowledge Engineering Review* 22(4), 361–377 (2007)