

Semantic Views – Interactive Hierarchical Exploration for Patent Landscaping

Tatyana Skripnikova¹

KIT - Karlsruhe Institute of Technology

Hidir Aras²

FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

Anna Weißhaar, Sebastian Blank, Hans-Peter Zorn³

inovex GmbH

Abstract

In this paper, we investigate whether a semantic representation of patent documents provides added value for a multi-dimensional visual exploration of a patent landscape compared to traditional approaches. Word embeddings from a pre-trained model created from patent text are used to calculate pairwise similarities for representing each document in the semantic space. Then, a hierarchical clustering method is applied to create several semantic aggregation levels for a collection of patent documents. For visual exploration, we have seamlessly integrated multiple interaction metaphors that combine semantics and additional metadata for improving hierarchical exploration of large document collections.

Keywords: Patent landscaping, word embeddings, hierarchical clustering, text analysis, semantic exploration, multi-dimensional exploration, visual user interface.

1. Introduction

The number of written works describing scientific progress is steadily increasing, which necessitates the development of supportive tools for their efficient analysis. Developing a visualization approach to facilitate the examination proves to be a challenging task. This is due to the complexity

of these documents, which are not only characterized by their textual content, but also by a number of metadata attributes of various kinds, including information about relationships between them.

Patent landscaping [1] is an example of a task in which complex document explorations take place. With the help of patent landscaping, companies acquire competitive advantages and steer their research and development efforts. However, with hundreds to thousands of patent documents that have to be considered per patent landscaping report, human perception must be aided in the task of finding patterns in data to prevent cognitive overload.

We propose an approach that allows for a multi-dimensional visual exploration [2] based on both semantics and metadata from the patent documents. Semantic embeddings [3] are widely used in natural language processing to capture relationships between text documents. Nonetheless, when trying to visualize those relationships, we face the problem that positions and distances in the embedding

¹<https://www.kit.edu>

²<https://www.fiz-karlsruhe.de>

³<https://www.inovex.de>

Published in "Proceedings of The 1st Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech 2019)", Eds. L. Andersson, H. Aras, F. Piroi, A. Hanbury, PatentSemTech 2019, 12 September 2019, Karlsruhe, Germany. <https://doi.org/10.34726/pst2019.6> — Authors 2019. CC BY 4.0 License.

An extended version has been submitted to the World Patent Information Journal - Virtual Special Issue: Text Mining and Semantic Technologies in the Intellectual Property domain.

space are not easily explainable and can hardly be understood by themselves. As for creating a patent landscape the question what is "in/out" of a focused topical region is crucial, we utilize semantic similarity of documents for creating a patent landscape [4] followed by clustering [5] the documents at 3 aggregation levels employing hierarchical agglomerative clustering.

2. Semantic Exploration of a Patent Landscape

In order to semantically explore patent documents, two challenges exist which must be reflected in our approach:

1. visually presenting high-dimensional *semantic representations* of documents in a way that is intuitively understood, and
2. supporting *semantic interactions*, which means that the display adapts to the intentions of the user with regard to information density and level of detail [6].

The user interface and the interaction metaphors it offers are designed to handle these challenges by utilizing a number of coordinated views which respond to the changes in each other's states. The scatter plot is the main area of the visualization representing the semantic space. At the same time, the histogram and sunburst views display metadata attributes from the dataset in an aggregated form. They enable filtering and highlighting of the data across all views via *brushing and linking*, which means that "the change to the representation in one view affects the representation in the other views as well" [7].

The interactions connecting the views fall into one of three groups: selection, highlighting and re-setting the current selection. The implementation is consistent across all views: 1) clicking means selecting an object/group, 2) hovering with the mouse causes a highlighting of an object/group which is a preview of the selection, 3) clicking on the background of a view resets the selection.

3. Evaluation

In order to evaluate the question of how semantic information in combination with rich metadata can be used to enhance the visual exploration of large document collections, we conducted a summative

study in form of a think-aloud-experiment with several patent experts. The subjects of the study are employees of FIZ Karlsruhe and have acquired significant experience over the years with patent matters in general and with patent research and patent landscaping in particular. This makes them appropriate candidates to study the complex cognitive processes that happen during the exploration.

We evaluated the visualization approach by means of tasks the users had to perform and by means of questionnaires for capturing user feedback.

The usability study indicates that visualization metaphors and interaction techniques were appropriately chosen.

4. Conclusion

We set out to present a novel approach for the hierarchical exploration of large document collections combining semantics and additional metadata. Our research shows that the chosen interaction techniques are consistent and intuitive. The proposed visualization approach provides added value for patent landscaping, and can also be applied to other document exploration tasks.

References

- [1] K. W. Boyack, B. N. Wylie, G. S. Davidson, Domain visualization using vxinsight for science and technology management, *Journal of the American Society for Information Science and Technology* 53 (9) (2002) 764–774. doi:10.1002/asi.10066.
- [2] K. Wittenburg, G. Pekhteryev, Multi-Dimensional Comparative Visualization for Patent Landscaping, in: *IEEE VIS Workshop 2015 BusinessVis15*.
- [3] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations *arXiv:1802.05365*. URL <http://arxiv.org/abs/1802.05365>
- [4] A. Abood, D. Feltenberger, Automated patent landscaping, *Artificial Intelligence and Law* 26 (2) (2018) 103–125. doi:10.1007/s10506-018-9222-4.
- [5] A. Skupin, A Cartographic Approach to Visualizing Conference Abstracts, *IEEE Computer Graphics and Applications* 22 (1) (2002) 50–58. doi:10.1109/38.974518.
- [6] D. Modjeska, *Navigation in Electronic Worlds: Research Review for Depth Oral Exam* (May) (1997) 1–56.
- [7] Marti A. Hearst, *Modern Information Retrieval*, Addison-Wesley-Longman Publishing co., 1999. doi:10.7748/ns.4.42.26.s69.