World Patent Information 49 (2017) 34–42

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Visual patent trend analysis for informed decision making in technology management

Qi Han [a, *], Florian Heimerl [b], Joan Codina-Filba [c], Steffen Lohmann [e], Leo Wanner [c, d], Thomas Ertl [a]

[a] Visualization and Interactive Systems Institute, University of Stuttgart, Germany
[b] University of Wisconsin Madison, USA
[c] Pompeu Fabra University, Spain
[d] Catalan Institute for Research and Advanced Studies, Spain
[e] Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany

ARTICLE INFO

Article history:
Received 7 April 2016
Received in revised form
10 April 2017
Accepted 17 April 2017
Available online 10 May 2017

Keywords:
Patent analysis
Trend analysis
Visualization
Visual analysis
Patent information
Natural language processing

ABSTRACT

Patent information is increasingly important for decision makers. Their demand for exploratory trend and competitor analysis poses new challenges with respect to the processing and visualization of patent data. We present PatStream: a highly interactive approach for decision support through patent exploration, which offers a streamgraph-based visualization for trends at different levels of abstraction and facilitates the combined analysis of their various aspects, including patent applicants, IPC distributions and innovativeness. PatStream integrates powerful natural language processing techniques for concept extraction and patent similarity assessment to allow for content-oriented visualization and analysis.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In the past few years, the market saw a new group of users of patent information emerging. These users are, for instance, managers working in the strategic departments of companies as well as policy makers who are interested in accessing information for their work directly from patents. Their needs that target patent information from a strategic perspective are different from the needs of patent experts. This calls for novel approaches to process and present patent information [1]. From the primary questions to be answered by patent analysis according to Ernst [2] and Joho et al. [3], we derived the following desired functionality to support trend analysis and technology management based on patents:

- obtain an overview of different technology topics in a given field;
- identify relevant trends according to individual information needs;
- evaluate the importance of technological trends;
- observe the behavior and productivity of different players relevant to specific trends;
- spot new technologies related to a trend.

To provide those functionalities, we propose a new interactive visual approach for temporal patent analysis: *PatStream*. It provides overview of different technological trends over time, using word clouds to quickly convey their central concepts. Through brushing and linking, it allows users to access various trend aspects comprising their distribution over categories of the International Patent Classification (IPC) [4], ranking of patent applicants relevant to a trend, and the display of technological innovativeness associated with a trend at a certain period of time. In addition, overlays on the streams show the number of patents published by different applicants over multiple years – a feature that enables users to

* Corresponding author.
E-mail addresses: Qi.Han@vis.uni-stuttgart.de (Q. Han), heimerl@cs.wisc.edu (F. Heimerl), joan.codina@upf.edu (J. Codina-Filba), steffen.lohmann@iais.fraunhofer.de (S. Lohmann), leo.wanner@upf.edu (L. Wanner), Thomas.Ertl@vis.uni-stuttgart.de (T. Ertl).

http://dx.doi.org/10.1016/j.wpi.2017.04.003
0172-2190/© 2017 Elsevier Ltd. All rights reserved.

easily compare the filing activities of applicants in different topic areas.

The visual design of PatStream is based on an approach presented by Heimerl et al. [5]. We adapted this approach for patent analysis, such that it enables analysts and managers to reach their goals more effectively. To facilitate the access to the content of patent material, we integrated advanced Natural Language Processing (NLP) techniques. We also included a novel way to depict the usage of IPC classes in a trend, and an interactive approach to compare and analyze patent applicants and the topics in which they file patents. Overall, these adaptations and extensions transform Heimerl et al.'s original proposal into an effective solution to carry out trend analysis on patent datasets.

The rest of this article is organized as follows. Section 2 introduces related work. Section 3 describes data preprocessing for our content-oriented visualization techniques. Section 4 outlines the different views of PatStream and describes how they help users to achieve their analysis goals. Section 5 presents the application of the approach in a concrete use case. Finally, Section 6 concludes with a summary of the techniques used in PatStream and sketches directions for future work.

2. Related work

Visualization has long been part of patent analysis. It helps users improve their understanding of analysis results [6–8]. Thomson Reuters's *Thomson Innovation* [9] features several visualizations to find core patents or interesting topics in a patent collection. Patent iNSIGHT Pro [10], Ambercite [11] and PatSeer [12] provide multiple visual tools, including citation graphs and 2D spatializations to discover hidden relationships between patent documents, without going beyond the use of words for the representation of patent content. Koch et al. iteratively build sophisticated patent queries, visually represented as diagrams [13]. CiteSpace [14] originally supports visual trend analysis in scientific literature, but has been proven to be useful for the analysis of patent collections as well. In contrast to all these works, PatStream focuses on the interactive analysis of patents to support decision making and technology management. For this, it includes novel visualizations to depict technological trends and their evolution over time, using NLP techniques to extract key concepts.

To support users with trend analysis, various ways to create patent-based topic overviews have been proposed. Some of them draw upon metadata to find relationships among patent documents. For instance, Kay et al. [15] suggest to construct graphs of IPC classes based on citation relationships of patents. PatentSight [16] uses the Patent Asset Index [17] to gauge the value of single patents and entire portfolios. In addition, several works use NLP to create visual patent maps [18,19]. Wanner et al. [20] combine visualization and automatic patent claim summarization. Kim et al. [21] propose a method to create an overview of the evolution of a technology. PatStream extends the state-of-the-art with new interaction techniques and the use of content elements (concepts) instead of keywords. Among other features, it allows users to interactively change the number of topics shown in an overview and thus see how topics develop over time in order to select interesting ones for further investigation, e.g., in the context of decision making.

Multiple authors [22–24] find that text mining and visualization can aid patent analysis. Although some initial effort may be necessary in order to use such techniques effectively, the information they contribute to decision making can justify the effort. PatStream therefore seeks to seamlessly integrate concept-based topic analysis and visualization to support decision making in knowledge management.

Streamgraphs [25], which are used in PatStream, have become a popular type of visualization for text datasets. Using a river metaphor, they show the evolution of topic-specific corpus structure. Wei et al. [26] place word clouds onto streams to convey the topic content. Heimerl et al. [5] extend stream graphs to visually link topics in a scientific literature corpus with the cited communities for joint analysis. PatStream can be considered a continuation of this research. It focuses on the visualization of conceptual topics and patenting patterns over time.

In what follows, we first introduce the data preprocessing required for PatStream and then present its visualization methods.

3. Data processing for visualization in PatStream

To enable content-oriented interactive analysis of patent material, patent data are aggregated and processed in a data processing pipeline,[1] which consists of the following major processing stages: 1. aggregation of patents with respect to the IPC hierarchy; 2. concept extraction from patents using advanced NLP techniques; 3. assessment of the similarity between patents and determination of the key concepts of each of them. Let us discuss each of these stages in turn.

3.1. Aggregation of patents with respect to IPC

For aggregation, the patents in a given collection are first situated in the IPC tree. The tree is then recursively reduced starting from the leaves in that all nodes that do not contain IPC codes occurring in the collection are removed. The reduced IPC tree is converted into a binary tree by splitting each n-ary branch of the tree into $n - 1$ binary branches.

3.2. Concept extraction from patents

To be able to go beyond word frequencies, NLP techniques need to be applied to identify the relevant content elements (i.e., concepts) of a patent. Concepts in patents are denoted by multiword terms, as, e.g., *cooking device* and *magnetic record carrier*, which are, from the syntactic point of view, nominal phrases (NPs). The NLP pipeline of PatStream takes this into account; it consists of:

1. *Sentence analysis*, which implies tokenization, part-of-speech (POS) tagging and syntactic parsing. Tokenization is performed by an off-the-shelf module that has been adapted to patent material. For POS tagging and syntactic parsing, we use the dependency parser of Bohnet et al. [27], which has been retrained on patent material and further tuned using domain adaptation techniques [28].
2. *Extraction of NPs* from the parsed dependency tree, drawing upon nominal heads and the topology of the subtrees they govern.
3. *Classification of the individual lexemes that compose an NP* in terms of a predefined category typology, which includes, e.g., 'quantifier', 'part', 'physical measurement', 'figure reference', 'literature citation', 'chemical substance'. Furthermore, each lexeme is marked when it belongs to the "common terms" list, which contains, among other terms, e.g., *apparatus*, *device*, and *mechanism*.
4. *Determination whether a given NP or any of its subtrees constitute a term* and thus a denotation of a concept in the

[1] The patents in the analyzed patent collection are assumed to contain at least the application id, name(s) of the applicant(s), title, IPC classification code, application date, claims, abstract, and description.

Download English Version:

https://daneshyari.com/en/article/4755487

Download Persian Version:

https://daneshyari.com/article/4755487

Daneshyari.com