World Patent Information 32 (2010) 291–299

Contents lists available at ScienceDirect

World Patent Information

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

Review

Patinformatics as a business process: A guideline through patent research tasks and tools

Martin G. Moehrle *, Lothar Walter, Isumo Bergmann, Sebastian Bobe, Svenja Skrzipale

IPMI – Institute of Project Management and Innovation, University of Bremen, Germany

ARTICLE INFO

Keywords:
Patinformatics
Patent research tasks
Patent research tool
Process modeling
Event-driven process chain
Pre-processing
Patent analysis
Discovered knowledge

ABSTRACT

Since Trippe [1] introduced the term 'patinformatics' a lot of progress has been made in this particular field of information science. However, there is still need for a more comprehensive framework to structure the variety of tasks related to 'patinformatics', to highlight essential functions within the patinformatics process and to identify those process parts, which are supported by currently available software tools, and others, which are not. In this paper we apply business process modeling to describe the patinformatics process for supporting managerial decision making. The process model enables an overview of major tasks within patinformatics and links them to currently available tools. This paper provides a guideline through patent research for most users of patent information. It may also be employed as a fundamental model for the comparison of patinformatics software applications and approaches.

1. Introduction

Trippe [1] introduced the term patinformatics which encompasses all forms of analyzing patent information. This introduction of a new term proved quite valuable in two respects: firstly, patent information may in several ways be distinguished from other forms of information like journal information or webpage information, especially in terms of the (more or less) standardized structure of documents, the relevance for managerial decisions in companies as well as in other institutions like research labs, and the relevance for technical developers as a major source of documented technical knowledge. And secondly, for patent information specific analysis processes and tools may be applied, that make use of these characteristic attributes.

Trippe [1,2] as well as more recently Tseng et al. [3], Yang et al. [4] and Bonino et al. [5] provide different entries to the tasks of patinformatics, thus facilitating the completeness of functions and tasks of the framework developed later on. Trippe [1] generally differentiates between two perspectives: that of the patent searcher and that of the patent analyst. While searchers 'are trained to find a needle in a haystack' on a microscopic level of patent data, analysts work on a macroscopic level and 'want to identify haystacks from space' [1]. Then Trippe [2] defines different tasks related to patinformatics as conducted by patent analysts: list cleanup and grouping of concepts, list generation, co-occurrence matrices and circle graphs, clustering of structured data, clustering of unstructured data, mapping document clusters, adding temporal components to cluster map, citation analysis and subject–action–object structures. He surveys the coverage of these tasks by means of state-of-the-art tools.

On the other hand, Tseng et al. [3] regard this from the perspective of a patent analyst (according to [1]) and define a typical scenario that encompasses the sequence of the following activities: task identification, searching, segmentation, abstracting, clustering, visualization and interpretation. Furthermore, they describe a series of text mining techniques which conform to the patinformatics process while each step in this process has an application of its own.

Also, Yang et al. [4] focus on new technological possibilities, i.e., text mining and visualization. They compare key text mining and visualization tools concerning text mining capabilities, perceived strengths, potential limitations, applicable data sources, and the output of results. This comparison opens up an additional domain-specific perspective by comparing tools with regard to chemical, biological and patent information. As an essential outcome of the comparison they highlight the uniqueness of features provided by each application.

Finally, Bonino et al. [5] propose a user-oriented focus on patinformatics tasks. They pool a variety of common patinformatics tasks to three main classes: patent search, analysis and monitoring. Further, they differentiate between motives of using patent information and specify when and why a patinformatics task is conducted. The paper discusses challenges and opportunities of patinformatics software and especially highlights semantic-based solutions that involve ontologies, thesauri and taxonomies.

* Corresponding author.
E-mail addresses: martin.moehrle@innovation.uni-bremen.de (M.G. Moehrle), lothar.walter@innovation.uni-bremen.de (L. Walter), isumo.bergmann@innovation.uni-bremen.de (I. Bergmann), sbobe@uni-bremen.de (S. Bobe), svenjas@uni-bremen.de (S. Skrzipale).

0172-2190/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wpi.2009.11.003

Beside these approaches, a more comprehensive, process-oriented framework could be beneficial to structure this extension and give insights into it. Such a framework can be used to highlight essential functions within a patinformatics process and their succession. It can also be helpful to identify process parts, which are supported by currently available commercial software tools, and functions, which are not supported.

This paper is divided into four sections. Subsequent to the introduction (Section 1), in which we introduce the field of research and review some collections of relevant tasks, Section 2 will contain the presentation of a method for modeling business processes. In Section 3 we shall introduce the three core-processes of patinformatics, i.e., pre-processing, patent analysis, and established knowledge, and describe these core-processes with their functions in detail. In Section 4 we shall discuss some limitations of our approach and give a perspective concerning the domain of management research.

2. Method of modelling business process

We aim at providing a systematic overview regarding different tasks and currently available tools related to patinformatics. For this purpose we have initially identified and characterized a large number of essential tasks related to patent research. Secondly, we matched state-of-the-art programs and tools with these tasks. This matching was carried out by means of modeling a business process. We allocated the tasks to functions as elements of a process chain and arranged them in a logical order. The outcome of this work is a comprehensive framework, which provides a quick overview by structuring and visualizing the variety of tasks and tools within the patinformatics process, and thus particularly supports managerial decision making. Here, for patinformatics, we have selected the approach of an event-driven process chain and apply a well-known ARIS-approach for business process modeling [6]. Patinformatics are integrated into an *event-driven process chain*.

2.1. Choice of method

For the modeling of business processes a multitude of methods and related tools are available. Hence, the choice of the method and, consequently, the choice of the tool depend on their expediency [7]. The modeling method based on *event-driven process chains* is commonly used because it represents all essential elements of a complex business process, i.e., functions, organizational units, software tools and other resources in a simple graphical flow chart, that makes this method comprehensible even to non-specialists of business process analysis [7].

2.2. Modeling aspects

In general, a business process can be seen as a set of business activities organized in a sequence or a more complex process structure. These business activities need input and generate output by means of different resources. In general an event-driven process chain can be split into two or more parallel flows of activities, which can be connected in the later process by means of logical connectors. The business activities are called functions in *event-driven process chains*. Events within an *event-driven process chain* are changes in status, which trigger a function that follows and generates a new status when finished. Within the patinformatics process events induce functions as follows: e.g., the event 'patent is registered' induces the function 'defining the adequate classification' to enable the storage of patent documents. Organizational units and other objects of information are connected to each of their related functions. This graphical representation enables the documentation of the work flow and the identification of obvious system discontinuities during the organizational routine.

3. Patinformatics as a business process

The main goal of patinformatics is to provide efficient access to patent information and educe business decisions on the basis of diverse patent analyses [2]. The main goal defines the perspective we will choose for developing our framework, and therefore we will look at the topic from the view-point of all users of patent information, including companies, patent attorneys, patent information providers, inventors and patent offices, to name only the most important ones. The main goal also leads to a multitude of tasks. For instance, efficient access presupposes a classification and storage of patent documents. The support of decision making requires a comprehensive and visual representation of patent analysis results. Since the requirements are diverse and complex, a systematically designed business process provides orientation concerning all forms of patent analysis and related activities. Major requirements on the patinformatics process are: preparation of the data for efficient access, analysis of the data by means of different methods and utilization of analysis results for decision making.

According to these requirements, three core-processes are devised for the patinformatic process: (i) pre-processing, (ii) patent analysis and (iii) discovered knowledge. This covers essential tasks conceived to meet requirements defined at the top and encompasses tools which serve to support these tasks. The core-processes represent the first level of the top-down business process modeling:

(i) The core-process of 'pre-processing' is concerned with the preparation of patent documents before the analysis is started. Its object is to provide basic information of high quality, correctness and completeness. This process encompasses tasks and tools for the classification of patents, the digitalization and the storage of patent documents in databases.
(ii) The second core-process, 'patent analysis' aims at a high quality access to patent sources. It focuses on tasks around accessing patent databases and conducting analyses regarding the contents of patents or relationships between patents, e.g., text mining and citation analysis.
(iii) The third core-process is called 'discovered knowledge'. Its goal is to provide analytic results of high quality provisionally. It outlines different modes of visualizing research results, e.g., by means of a patent map.

Apart from the logical order of tasks, additional benefit of this process model can be gained from a systematic arrangement of tasks on the one hand, and patent research tools on the other. Nevertheless, our framework is limited in two ways.

(i) We have modeled patinformatics as a linear process. This is advantageous primarily for didactical reasons, in the course of practical application certain functions and related tasks may be skipped, undertaken simultaneously (for instance visualisation and documentation of research results), or proceeding in loops (for instance, a query may be repeated, if the results of a previous attempt are not sufficient). This model provides the basis for customizing the business process to user-specific requirements. A patent officer may have different requirements than a R&D expert. Both the order of the tasks as well as additional tasks or work flows in this model is open to be adapted.
(ii) Our framework refers to existing methods and tools. But the field of patinformatics is subject to a dynamic development.

Therefore, functions, tasks and tools may change and decline or increase in quantity. Also in this context the model is open for being extended by additional tasks and new tools. New tasks, e.g., symbol recognition, are still undergoing development. Notably the field of non-text analysis possesses enormous potential of further advancement. Analysis techniques such as recognition (and information extraction) of gene-codes as well as physical and chemical formulae are imaginable. New algorithms will be needed for patent data analysis. Hopefully, this will result in a better performance of information retrieval in general. Efforts of further development are also necessary with regard to pre-processing. Because of OCR-failures, inaccurate indexing and other insufficiencies in data pre-processing, research results lack relevance, quality and reliability.

In the following each of these three core-processes will be downscaled to an event-driven-process-chain and described in detail.

3.1. The core-process of pre-processing

The core-process of 'pre-processing' focuses on the procedure of preparing patent documents for further tasks, i.e., conducting a patent analysis. Functions and tasks of this process are supposed to be performed once per patent document. The input of this process is a patent document. The outcome of this process is a well-prepared basis of patent data. Patent documents have to be classified and stored in databases to enable retrieval. They represent surplus value added by supplementary information to enhance the quality of the retrieval results. Finally the database has to be revised to ensure efficient access.

Pre-processing is not only a necessity for patent offices or patent information providers, as one might assume. Many companies possess their own patent database, some even with their own classification and their own surplus value (for instance: internal annotations on certain aspects of a patent). Therefore, we have decided to define pre-processing as a core-process of its own in the context of patinformatics.

Pre-processing comprises the following five functions (see Fig. 1).

3.1.1. Function 1 – definition of patent classification

The definition of patent classification is a fundamental function within the patinformatics process, as further tasks are based on this function. Patent classification enables a systematic archiving of patent documents and assists their retrieval from databases [8,9]. The International Patent Classification (IPC), the European Classification System (ECLA), the United States Patent Classification (USPC) or the FI-Terms of the Japanese Patent Office are established classification systems. Many efforts have been undertaken to harmonize these classification systems [10]. There are also approaches to the computer-aided classification of patent documents [11]. However, automation in patent classification is still a challenging task and it is currently limited to a manually controlled semi-automated classification conducted by patent officers.[1]

3.1.2. Function 2 – digitalization of patent documents

The digitalization of patent documents facilitates the computer-aided access to patent information. Although today many patents are already filed through electronic systems like the EFS[2] (electronic filing system) of the USPTO or eOLF[3] (electronic Online Filing for PCT-procedures) of the EPO and are available in machine-readable formats, i.e., XML, html or pdf, still a lot of patents are filed as hard copies. There is also some backload of elder patents that are still not in the database. In these cases patents have to be converted into digital documents by use of scanning-systems. This function comprises the tasks (i) OCR and (ii) *n*-gram analysis:

(i) Optical character recognition (OCR) enables the automated transformation of a printed document into a digital format. A digitalized patent document enables computer-aided full-text searches in databases.

(ii) Errors in the original document by typing or by OCR-failures are commonly known phenomena. Patent researchers have to bear in mind that a computerized digitalization of patent documents may contain a multitude of mistakes. There are methods of correcting such flaws. One possible method is to carry out an *n*-gram analysis. The *n*-gram analysis enables predicting the next item in a sequence and is applied to segment words and match them with lexical datasets to offer proposals for adjustments [12].

3.1.3. Function 3 – storage of patent documents in databases

Patent information is stored in databases. Various commercial databases are available, each of which meets special research needs, i.e., domain-specific research. Apart from Derwent World Patent Index (DWPI, Thomson Reuters, New York/US) other well-known databases are PatentWeb (Thomson Reuters, New York/US), CAS databases (American Chemical Society, Washington/US), PatFT (Patent Full-Text Database, USPTO), esp@cenet (European Patent Office) and INPADOCDB (International Patent Documentation Database). Many companies also have their own databases to enable company specific surplus value.

3.1.4. Function 4 – generation of added value (metadata)

This function comprises the tasks (i) XML-conversion, (ii) content aggregation and (iii) indexing. The quality of research results could be improved by equipping singular patent documents with extra indexing keywords. A patent researcher may gain several benefits from using metadata. Furthermore, the results will be more structured and clearly arranged, and the researcher can thus acquire a faster overview of relevant patent documents and their contents.

(i) Extensible mark-up language (XML) is a modern and well-supported document interchange format that includes a system for annotating a text by different mark-up elements and allows specification of lexical grammar and parsing requirements. The XML document is divided in two parts, (i) mark-up and (ii) content, which is distinguishable in syntax. To enable interoperability between tools, it is necessary to share not only the content, but also the mark-up (scheme). XML tags are independent of natural language, and it is possible to write XML with different tools.

(ii) The summary of contents may be divided into abstracts and reviews [13]. An abstract is written by the author of a patent, whereas a review is written by a third party. Reviews also contain metadata, such as information about the author of the original patent document, figures, tables, drawings or definitions of employed terms that are not explained in the patent specification.

(iii) Indexing may be accomplished manually or automatically [14–16]. It is a method for furnishing patent documents with significant keywords, to simplify the recovery of documents from databases. Human or manual indexing was established a long time ago, and there are some basic requirements to be taken into account [14]. Firstly, in

[1] http://www.wipo.int/ipccat/ipc.html.
[2] http://www.uspto.gov/ebc/efs_help.html.
[3] http://www.epo.org/about-us/publications/user-guides/online-filing.html.

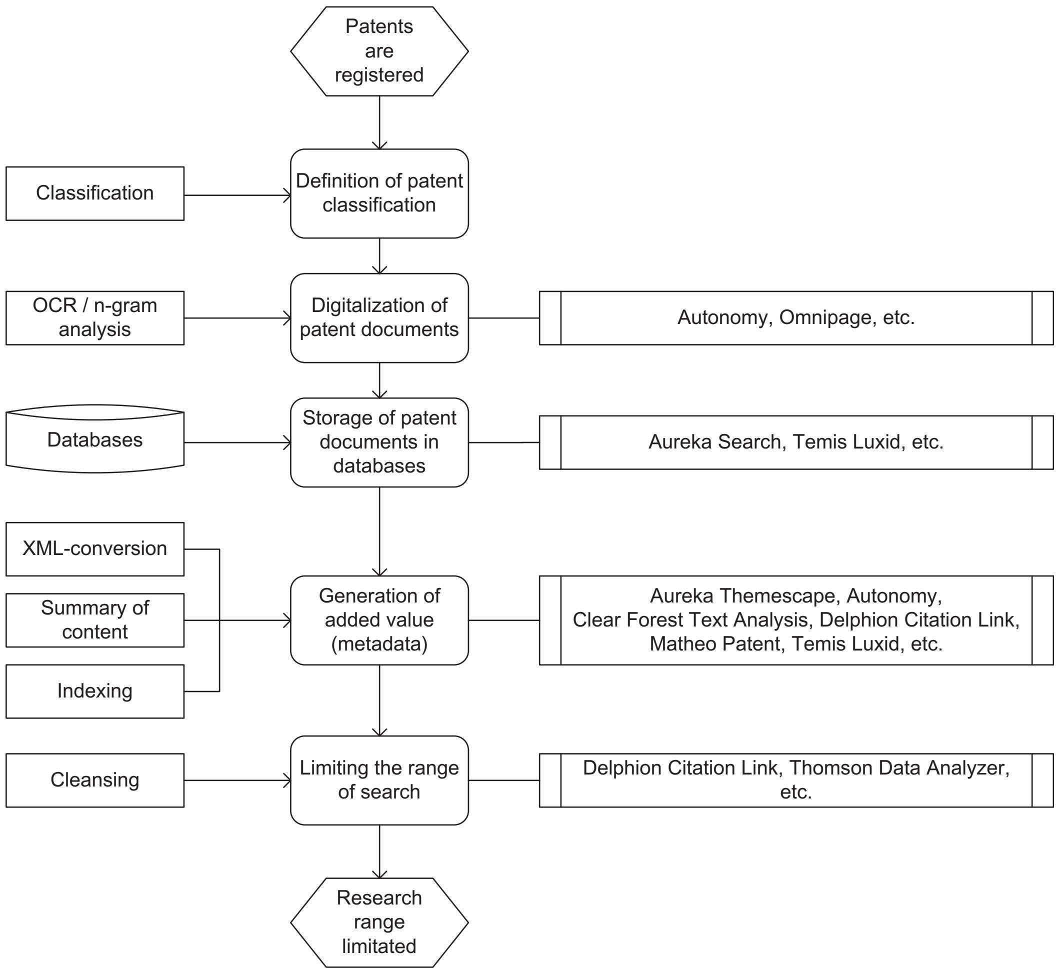

Fig. 1. The core-process of pre-processing.

order to identify the essential contents of a patent, peripheral and trivial contents have to be excluded. Secondly, paraphrases contained in the text have to be lexicalized to gather as many generally accepted keywords as possible. Finally, technical terms should be identified, and if ambiguities occur, they have to be adjusted to the requirements of later users.

3.1.5. Function 5 – limiting the range of research

The object of this function is to eliminate irrelevant documents from the research list. In this context, the task of cleansing is carried out, which encompasses the identification of inaccurate parts of the data and modifying or deleting this data to avoid inconsistencies in databases [17]. Different types of failures affect reliability of research results are lexical errors, domain format errors, irregularities, constraint violation, missing value, missing tuple, duplicates and invalid tuple. Cleansing increases the quality in the sense of completeness, validity, schema conformity, uniformity density and uniqueness [18].

3.2. The core-process of patent analysis

The second core-process of 'patent analysis' uses the previously qualified patent data as input. It leads to the output of finished analyses. The main object of patent analysis is to provide access to relevant patent documents with regard to a research question; in terms of information retrieval the goal can be interpreted as having a question related optimum between high recall and high precision rate.

Differences between media are of particular interest in the core-process of 'patent analysis' (referring to [4], chapter 1). The most notable difference occurs between text analysis (e.g., text mining) and non-text analysis (e.g., picture recognition). While text analysis focuses on full-text documents, non-text analysis explores patents on the basis of information other than text, e.g., images or tables and symbols as well as chemical notations [19]. Non-text information is essential for an additional demarcation of a patent claim, i.e., if the written claim is partly infringing on another claim [8].

Four functions have been defined for patent analysis (see Fig. 2).

3.2.1. Function 1 – document query

Patents have to be retrieved from databases. Firstly, the patent researcher selects a database according to different criteria, e.g., the supported output format of patents or the costs of using a commercial database. Next, relevant documents have to be queried with a view to recall and precision. There are different ways of generating and supporting a query. There are some retrieval concepts and the query may also be extended by using structured knowledge representation. Additionally more sophisticated methods can be applied that combine different techniques. This is realised through (i) Boolean and (ii) Extended Boolean Retrieval are commonly used to formulate the query by combining terms. By means of (iii) synonyms, (iv) thesauri and (v) ontologies the query can be expanded. (vi) Latent Semantic Indexing describes queries as a vector.

(i) Boolean Retrieval [20,21] is based on a dichotomous concept. It combines or excludes search items in the course of the query by means of an operator (AND, OR, NOT). Although

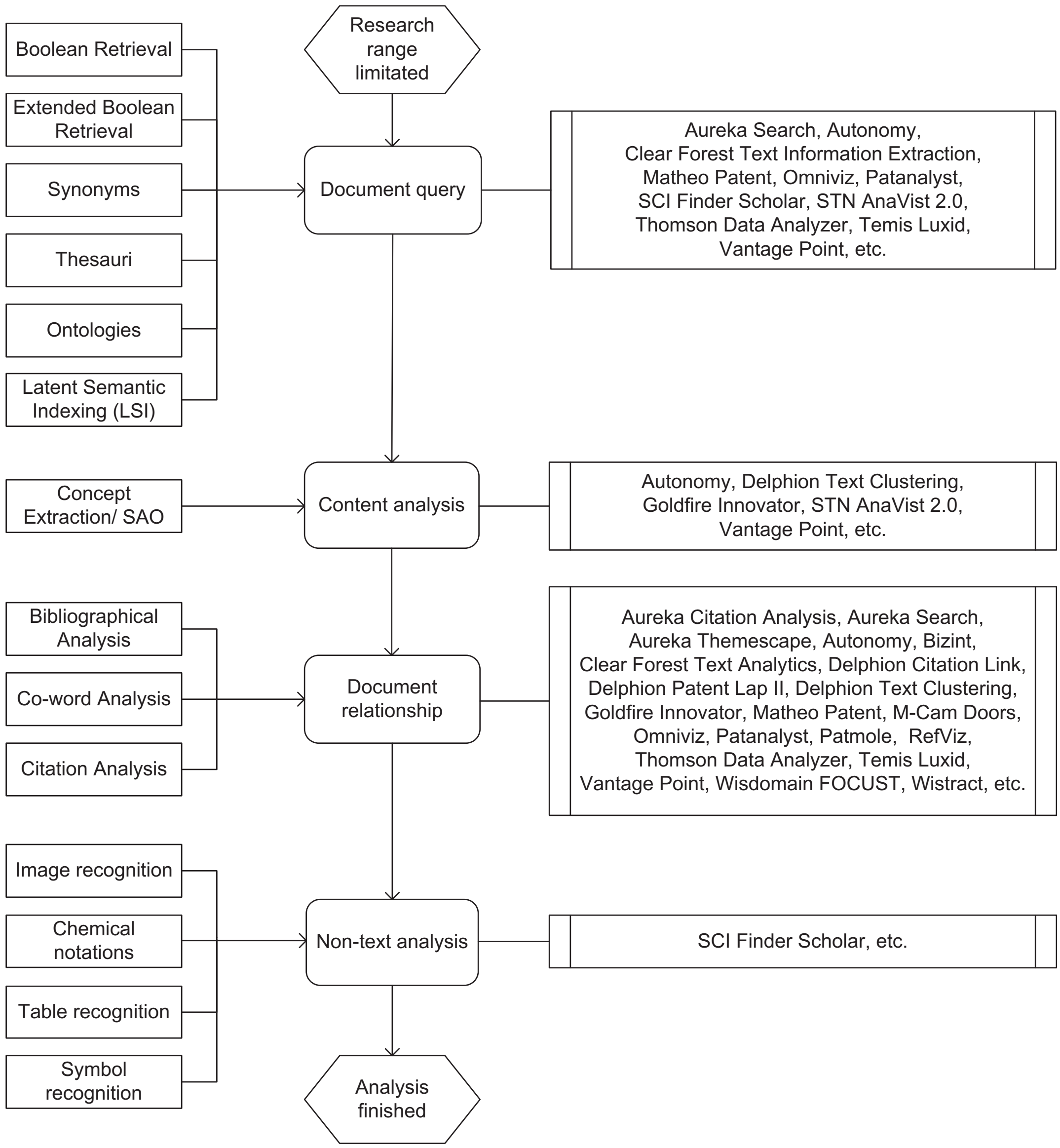

Fig. 2. The core-process of patent analysis.

it is not the most effective method for formulating a search query, the Boolean Retrieval is presumed to be a reliable method and widely used in patent retrieval, sometimes in combination with other retrieval methods.

(ii) Extended Boolean Retrieval [22] additionally allows a weighting of search terms and enables the ranking of search results according to relevance [23].

(iii) Synonyms refer to linguistic expressions (linguistic symbols, lexical symbols and, above all, words) that differ in morphology but convey the same meaning [25]. The use of synonyms for expanding the query helps to increase the recall.

(iv) Thesauri are based on regulated vocabulary including synonym lists and can be used for expanding a query as well. Unlike natural language, this vocabulary consists of predefined and authorised terms that have been preselected. Thesauri have become a capable support for subject cataloguing and for discovering patents within retrieval processes.

(v) Ontologies are hierarchically defined systems of connected terms and can be used for query expansion. Various approaches include relevance feedback, corpus dependent knowledge models and corpus independent knowledge models [24].

(vi) Latent Semantic Indexing (LSI) is a retrieval method that bases on Vector Space Models (VSM). The queries and documents are described as vectors to identify term-document relations based on vector similarity. Latent Semantic Indexing returns a high recall of documents as synonyms are also included, which are generalised to concepts [26]. Latent Semantic Indexing is also seen as an approach that in combination with Boolean Retrieval builds a powerful workflow for helping the user find documents relevant to their needs as quickly and reliably as possible [27].

3.2.2. Function 2 – content analysis

In most cases the content analysis of patent documents is carried out manually. A semi-automated concept extraction, e.g., by extracting SAO (subject–action–object) structures [28,29,2] accelerates the reading process and enables a systematic analysis of contents. The results of this function represent essential input for a further analysis of the relationships between patents, e.g., based on a similarity of contents.

3.2.3. Function 3 – document relationship

The relationship between documents can be analysed by means of different criteria such as co-occurrences of particular keywords or names as well as other similarity indicators. The aim of this task is to identify similar patents in order to group them. The researcher is able to evaluate the patents, e.g., concerning their applicants. Search results may be visualized in a subsequent step. The main tasks are (i) bibliographical analysis, (ii) co-word analysis and (iii) citation analysis. It is possible within all these tasks to apply clustering, when analysing the relationship.

(i) Bibliographical analysis is based on information comprising, e.g., assignee name, filing date, issue date, cited documents and references is additionally useful to highlight document relationships. The choice of clustering criteria depends on specific purpose [33]. A range of statistical multivariate methods support identifying clustering structures: i.e., hierarchical cluster analysis, factor analysis or multi dimensional scaling.

(ii) Co-word analysis is a bibliometric method and explores the co-occurrence of keywords in different patent documents. The use of co-word analysis enables the quantification of identical words between patents, which can be seen as an indicator of similarity [30,31].

(iii) Citation analysis uses references cited in patents and covers both patented literature (PL) and non-patent literature (NPL) to describe the state of the art. Relations between patents may be opened up by both backward citations and forward citations [32]. Generally network-diagrams are used for visualising patent-relationships.

3.2.4. Function 4 – non-text analysis

Non-text analysis explores patents on the basis of information apart from text. Non-text information does not possess a uniform layout like purely textual information, which consists of letters and figures. Still, they represent a source of knowledge in their own way. The following tasks (i) image recognition, (ii) chemical notations, (iii) table recognition and (iv) symbol recognition are specified:

(i) Along with text information images represent essential patent information for the demarcation of claims. Image recognition makes use of relational graph similarity measure and structural histogram approach to search the 2D line drawings [34].

(ii) A query for patents by way of chemical notations (instead of, e.g., keywords) offers to find chemical notations that do not exist within textual information [35].

(iii) Table recognition is used to identify tables within patents and analyse them regarding their composition and the structured data contained in lines and columns [36]. Once the structure of the table has been identified the researcher can evaluate whether the table contains any terms of importance for his research.

(iv) Symbol recognition is a task that offers to identify symbols such as mathematical functions and the like for a subsequent analysis. Symbol recognition enables the researcher to discover particular symbols in the explored patent documents. This retrieval method works by means of special symbols (e.g., Greek or Latin letters).

3.3. The core-process of discovered knowledge

The third core-process is referred to as ‘discovered knowledge’. In the course of this process, analysis results are evaluated further and edited to support operational patent decisions or strategic decision making. An essential aspect of this process is the visualisation of the data, which gives a quick impression of the current situation. Based on this situation analysis, the value of patents may be rated in a subsequent step with reference to further criteria. Finally, the results of the entire patinformatics process have to be stored to pro-

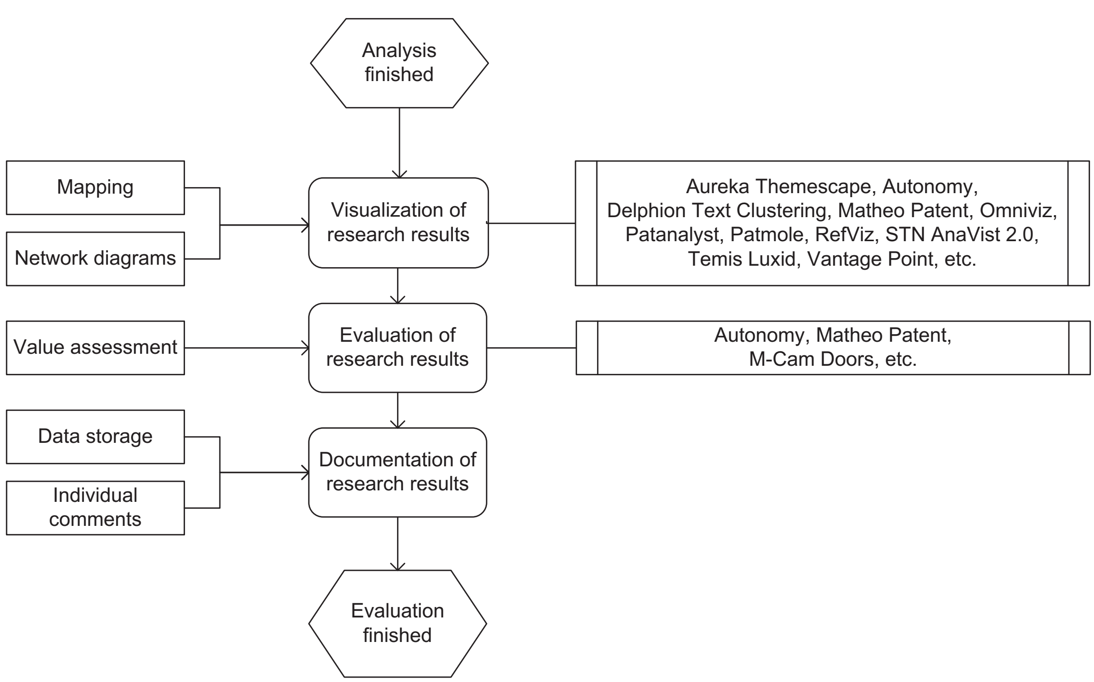

Fig. 3. The core-process of discovered knowledge.

vide a basis for future revision and exploration, especially if they are to be used in lawsuits concerning cases of infringement.

The process of 'discovered knowledge' comprises the following three functions (see Fig. 3).

3.3.1. Function 1 – visualisation of research results

Unlike lists and charts the graphical processing of data arranges a variety of information into a single impression. Thus, the visualisation of research results provides a quick overview and supports strategic decision making. There are two methods of visualisation, namely (i) mapping and (ii) network-diagrams, which are suitable for this purpose:

(i) Mapping permits a graphical visualization of relationships between patent documents, e.g., in the form of patent landscapes or clusters. It enables a quantitative display of fundamental interrelationships or the development of stored data [37].

(ii) Network-diagrams are a concomitant of mapping. These diagrams can highlight direct and indirect relationships between the patent documents [37]. The relationships are given by matching data as provided by co-occurrence or citations and similarity measures.

3.3.2. Function 2 – evaluation of research results

The evaluation of research results basically consists of the task of value assessment. Value assessment of patents includes the benchmarking of patents. The value of patents may be rated with reference to their technological or economical advantages [32].

3.3.3. Function 3 – documentation of research results

Documentation enables a review of the research process. Research strategies and research results become comprehensible to others and provide a source for subsequent research. Accordingly, a repetition of the same research can be avoided, while additional modified repetitions are facilitated. The research results become comparable and the impact of adjusted elements becomes obvious. An additional function of documentation is to annotate the research results with individual comments. These comments may guide following researchers and be of help in judicial proceedings.

4. Conclusions

In this paper we have presented a framework that can be employed for structuring essential tasks related to patinformatics. This framework enables an assignment of important functions and a localization of coverage by supporting computer implemented tools.

Apart from these prospects, two perspectives are to be emphasized. (i) In our framework we have focused on a large number of essential tasks regarding the patent as a source of information. We have not yet discussed in detail all purposes the information may be used for, such as supporting mergers and acquisitions on a strategic level or helping patent attorneys defend a patent claim on the operative level. An exploration of these purposes by way of an event-driven process chain would lead to an ever deeper understanding of patents and their impact on managerial decisions. (ii) Our work shows, that there is not one single tool that fulfills all functions of the patinformatics process. On the other hand, the variety of applications available in the market might overstrain the users. This opens a research space for IPR services, and the role of former information suppliers will change dynamically to that of one stop information system providers [38–42]. There is already a range of services, e.g., the advanced patent processing service with integrated patent data-warehouse (PATExpert, Fraunhofer Gesellschaft, Munich/DE), the content-similarity-based patent analysis service providing patent maps for supporting managerial decisions (PatVisor, IPMI Insitute for Projectmanagement and Innovation, Bremen/DE), value-added search services framed by seminars and lecturing activities (PATON patent information service centre, Illmenau/DE), advanced decision support and patent quality analysis software for strategic patent portfolio management (PatentCafe, Sacramento/US) and as well as application service providers based on user-editable ontologies (IntelliPatent, IntelliSemantic, Torino/IT) to name only a few professional patent information system providers.[4] The questions of what future IPR services might look like and how they would support managerial decisions should be analyzed in future research both empirically and from a conceptual point of view.

Acknowledgments

The authors wish to thank Steffen Lohmann, Axel Meier, Heiko Meinardus and Julia Wente for helpful discussions.

Appendix A. Tools and URL of software producers

Tool	URL
Aureka Citation Analysis	http://scientific.thomsonreuters.com/products/aureka/
Aureka Search	http://scientific.thomsonreuters.com/products/aureka/
Aureka Themescape	http://scientific.thomsonreuters.com/products/aureka/
Autonomy	http://www.autonomy.com/
Bizint	http://www.bizcharts.com/
Clear Forest Text Analytics	http://clearforest.com/Technology/TechnologyOverview.asp
Clear Forest Text Information Extraction	http://clearforest.com/Technology/TechnologyOverview.asp
Delphion Citation Link	http://www.delphion.com/products-research
Delphion Patent Lap II	http://www.delphion.com/products-research
Delphion Text Clustering	http://www.delphion.com/products-research
Goldfire Innovator	http://invention-machine.com/GoldfireInnovator.htm
Matheo Patent	http://www.matheo-patent.com/
M-Cam Doors	http://www.m-cam.com/
Omnipage	http://www.scansoft.de/omnipage/professional/
Omniviz	http://www.biowisdom.com
Patanalyst	http://www.patanalyst.com/
Patmole	http://www.cineca.it
RefViz	http://www.refviz.com/
SCI Finder Scholar	http://www.cas.org/SCIFINDER/SCHOLAR/
STN AnaVist 2.0	http://www.stn-international.de/stninterfaces/stnanavist/stn_anavist.html
Temis Luxid	http://www.temis.com
Thomson Data Analyser	http://scientific.thomsonreuters.com/products/tda/
Vantage Point	http://www.thevantagepoint.com/
Wisdomain FOCUST	http://www.wisdomain.com/index.htm
Wistract	http://www.wistract.com/

[4] All URL of information system providers mentioned in this paper are listed in Appendix B.

Appendix B. Services and URL of information system providers

Service	URL
IntelliPatent	http://www.intellisemantic.com/
PatentCafe	http://www.patentcafe.com/
PatExpert	http://www.patexpert.org/
PATON	http://www.paton.tu-ilmenau.de/
PatVisor	http://www.innovation.uni-bremen.de/

References

[1] Trippe AJ. Patinformatics: identifying haystacks from space. Searcher 2002;10(9):28.
[2] Trippe AJ. Patinformatics: tasks and tools. World Patent Inf 2003;25(3):211–21.
[3] Tseng YH, Lin CJ, Lin YI. Text mining techniques for patent analysis. Inf Process Manage 2007;43:1216–47.
[4] Yang YY, Akers L, Klose T, Barcelon Yang C. Text mining and visualization tools – impressions of emerging capabilities. World Patent Inf 2008;30:280–93.
[5] Bonino D, Ciaramella A, Corno F. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Inf 2010;32(1):30–8.
[6] Scheer AW. ARIS: business process modelling. Berlin: Springer-Verlag; 1998.
[7] Becker J, Kugeler M, Rosemann M. Process management – a guide for the design of business processes. Berlin: Springer-Verlag; 2003.
[8] Foglia P. Patentability search strategies and the reformed IPC: a patent office perspective. World Patent Inf 2007;29(1):33–53.
[9] Vijvers W. The international patent classification as a search tool. World Patent Inf 1990;12(1):26–30.
[10] Rampelmann J. Classification and the future of the IPC – the EPO view. World Patent Inf 1999;21:183–90.
[11] Krier M, Francesco Z. Automatic categorisation applications at the European patent office. World Patent Inf 2002;24(3):187–96.
[12] Kukich K. Techniques for automatically correcting words in text. ACM Comput Surv 1992;24:377–439.
[13] Fugmann R. Content exploration by means of indexing: principles and practical experience (German). Deutsche Gesellschaft fuer Dokumentation, Frankfurt am Main; 1999.
[14] van der Meulen WA, Janssen PJFC. Automatic versus manual indexing. Inf Process Manage 1977;13(1):13–21.
[15] Pulgarín A, Gil-Leiva I. Bibliometric analysis of the automatic indexing literature: 1956–2000. Inf Process Manage 2004;40:365–77.
[16] Anderson J, Pérez-Carbello J. The nature of indexing: how humans and machines analyse messages and texts for retrieval. Part I: research, and the nature of human indexing. Inf Process Manage 2001;37:231–54.
[17] Maletic J, Marcus A. Data cleansing: a prelude to knowledge discovery. In: Maimon O, Rokach L, editors. The data mining and knowledge discovery handbook. Springer-Verlag; 2005.
[18] Müller, Heiko, Freytag, Johann-Christoph. Problems, methods, and challenges in comprehensive data cleansing. Humboldt-Universität zu Berlin, Germany. Online resource: <http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/techreports/2003-hub_ib_164-mueller.pdf> [download from 17.08.2009].
[19] Derendyaev B, Bogdanova T, Piottukh-Peletsky V, Makarov L. Fast taxonomy of chemical structures selected from IR spectral database. Anal Chim Acta 2003:210–6.
[20] Große D, Fey G, Drechsler R. SATRIX: Algorithmen fuer Boolesche Erfuellbarkeit. Shaker Verlag; 2007.
[21] Salton G. A simple blueprint for automatic Boolean query processing. Inf Process Manage 1988;24(3):269–80.
[22] Fox E, Betrabet S, Koushik M, Lee W. Extended Boolean models. In: Frakes W, Baeza-Yares R, editors. Information retrieval: data structures and algorithms. Englewood Cliffs, NJ: Prentice Hall; 1992. p. 393–418.
[23] Lee JH, Kim MH, Lee YJ. Ranking documents in thesaurus-based Boolean retrieval systems. Inf Process Manage 1994;30(1):79–91.
[24] Bhogal J, Macfarlane A, Smith P. A review of ontology based query expansion. Inf Process Manage 2007;43:866–86.
[25] Dixon RMW. A new approach to English grammar, on semantic principles. Oxford University Press; 1992.
[26] Deerwester S, Dumais S, Furnas G, Landauer T, Harshman R. Indexing by latent semantic analysis. J Am Soc Inf Sci 1990;41(6):391–407.
[27] Ryley JF, Saffer J, Gibbs A. Advanced document retrieval techniques for patent research. World Patent Inf 2008;30:238–43.
[28] Bergmann I, Butzke D, Walter L, Fuerste JP, Moehrle MG, Erdmann VA. Evaluating the risk of patent infringement by means of semantic patent analysis: the case of DNA-chips. R&D Manage 2008;38(5):550–62.
[29] Moehrle MG, Geritz A. Developing acquisition strategies based on patent maps. In: Khalil, Tarek, Hosni, Yasser, editors. IAMOT 2004 – new directions in technology management: changing collaboration between government, industry and university. Washington, DC; 2004. p. 1–9.
[30] Rip A, Courtial J. Co-word maps of biotechnology: an example of cognitive scientometrics. Scientometrics 1984;6:381–400.
[31] Callon M, Courtial J, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics 1991;22:155–205.
[32] von Wartburg I, Teichert T, Rost K. Inventive progress measured by multi-stage patent citation analysis. Res Policy 2005;34(10):1591–607.
[33] Giri NC. Multivariate statistical analysis: revised and expanded. 2nd ed. CRC Press; 2004.
[34] List J. How drawings could enhance retrieval in mechanical and device patent searching. World Patent Inf 2007;29(3):210–8.
[35] Dickerson R, Gray H, Haight G. Prinzipien der Chemie. Walter de Gruyter, issue 2; 1978.
[36] Zanibbi R, Blostein D, Cordy J. A survey of table recognition models, observations, transformations, and inferences. IJDAR 2004;1–16.
[37] Morris S, DeYong C, Wu Z, Salman S, Yemenu D. DIVA: a visualization system for exploring document databases for technology forecasting. Comp Ind Eng 2002;43:841–62.
[38] Radauer A, Walter L. Elements of good practice of publicly funded patent database and information services for SMEs – selected results of a benchmarking exercise. World Patent Inf 2010;32(3):237–45.
[39] Sternitzke C, Bartkowski A, Schramm R. Regional PATLIB centres as integrated one-stop service providers for intellectual property services. World Patent Inf 2007;29:241–5.
[40] Wurzer A, Hundertmark S. Route to profitable European PATLIB centres with a process oriented business model in intellectual property services. World Patent Inf 2005;27:302–8.
[41] Krestel H, Goetz B. From information supplier to system provider: the diversification of patent information services at patent information centres. World Patent Inf 2002;24:143–5.
[42] Andrick J. Providing modern information services for small and medium-sized enterprises. World Patent Inf 1998;20:107–10.

Prof. Dr. Martin G. Moehrle has been the director of the Institute for Project Management and Innovation (IPMI) at the University of Bremen (Germany) since 2001, while at the same time holding the chair of Innovation and Competence Transfer. His area of work is technology and innovation management, and his preferred topics are patent management, technology roadmapping, future research, innovation process management, and methodical invention (TRIZ). Prof. Moehrle obtained his doctorate and qualified as a university lecturer on transitional themes between technology management and business informatics.

Dr. Lothar Walter works as a research associate at the Institute for Project Management and Innovation (IPMI), University of Bremen (Germany). Before this he filled a position as project and sales manager for a German medical engineering enterprise. He is a physicist with research interests comprising (i) methodical inventing, (ii) intellectual property management, (iii) semantic patent analysis and (iv) teaching in these areas, and additionally in mathematics for economists.

Dipl.-Oek. Isumo Bergmann is an academic collaborator at the Institute for Project Management and Innovation (IPMI), University of Bremen (Germany). Before this, he filled a position as consultant for Business Process Management for German utility companies. His field of research comprises intellectual property management and semantic patent analysis. He is additionally involved as coach for entrepreneurship at the University of Bremen.

Sebastian Bobe studies Industrial Engineering and Management at the University of Bremen (Germany) since 2004. His major subjects are project and innovation management as well as product development. In 2008 he wrote a student research project about patent research in collaboration with the Institute for Project Management and Innovation (IPMI), University of Bremen.

Svenja Skrzipale wrote a student project about patent research in cooperation with the Institute for Project Management and Innovation (IPMI), University of Bremen (Germany) in 2008. She studies Industrial Engineering and Management at the University of Bremen since 2004. Her major subjects are project and innovation management as well as product finding and development.