Patent Retrieval

Patent Retrieval

Mihai Lupu

Vienna University of Technology
Vienna 1040
Austria
lupu@ifs.tuwien.ac.at

Allan Hanbury

Vienna University of Technology
Vienna 1040
Austria
hanbury@ifs.tuwien.ac.at

now
the essence of knowledge
Boston – Delft

Foundations and Trends® in
Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is M. Lupu and A. Hanbury, Patent Retrieval, Foundation and Trends® in Information Retrieval, vol 7, no 1, pp 1–97, 2013

ISBN: 978-1-60198-648-1

Foundations and Trends® in Information Retrieval

Volume 7 Issue 1, 2013

Editorial Scope

Foundations and Trends® in Information Retrieval will publish survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians
Foundations and Trends® in Information Retrieval, 2013, Volume 7, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Foundations and Trends® in
Information Retrieval
Vol. 7, No. 1 (2013) 1–97
© 2013 M. Lupu and A. Hanbury
DOI: 10.1561/1500000027

Patent Retrieval

Mihai Lupu[1] and Allan Hanbury[2]

[1] *Vienna University of Technology, Favoritenstraße 9-11/188, Vienna, 1040, Austria, lupu@ifs.tuwien.ac.at*
[2] *Vienna University of Technology, Favoritenstraße 9-11/188, Vienna, 1040, Austria, hanbury@ifs.tuwien.ac.at*

Abstract

Intellectual property and the patent system in particular have been extremely present in research and discussion, even in the public media, in the last few years. Without going into any controversial issues regarding the patent system, we approach a very real and growing problem: searching for innovation. The target collection for this task does not consist of patent documents only, but it is in these documents that the main difference is found compared to web or news information retrieval. In addition, the issue of patent search implies a particular user model and search process model. This review is concerned with how research and technology in the field of Information Retrieval assists or even changes the processes of patent search. It is a survey of work done on patent data in relation to Information Retrieval in the last 20–25 years. It explains the sources of difficulty and the existing document processing and retrieval methods of the domain, and provides a motivation for further research in the area.

Contents

1

Introduction

Innovation is at the core of technological and societal developments. New ideas on how to make things better, faster, cheaper, and more reliable, or simply on how to make totally new things, are the result of many different economical, managerial, and cultural factors. In addition to all these factors, it stands to reason that technology moves forward on the basis of prior technology, and that therefore society as a whole benefits from public availability of detailed descriptions of technical innovation. For this reason, the patent system has been created to encourage inventors to share their know-how, in exchange for a temporary monopoly. Without approaching any controversial topic, this review looks, from the perspective of Information Retrieval (IR) researchers, at methods for benefiting from this amount of information.

The search for innovation, as expressed in patent documents, and for the purposes of obtaining new patents, has two facets: the *search for content* and the *search for legal information*. Most of us, as scientists, are very familiar with the search for content. While performed for different purposes, at its core lies the need to understand a technical process or entity. We achieve this by finding references to similar processes or entities, by analyzing its components and the more general categories of processes or entities of which it is part. Second, there is the search for legal information related to the protection granted to the inventor for a specific invention. The two facets often intermingle in the different search use cases described in Section 1.4, but this review will focus on the former.

We begin the introduction with a brief description of the past and present of the patent system, in order to provide a context for everything that will be discussed further on. This will also give the reader the understanding of the specific terminology used in the following sections. We continue with an overview of the content of patent documents in Section 1.2, illustrated by analyzing an example of a patent in Section 1.3. Section 1.4 gives a description of the most important patent search types. Finally, Section 1.5 gives a brief overview of patent research in the IR community, and the sources of patent documents.

1.1 History and Present

The term *"patent"* stems from the Latin verb *patere* and means *"laying open."* As a noun, it is the short form of *letters patent*, an official document used in the middle ages by an authority to assign specific rights to a person or group. The first patent law in the sense that we would imagine it today, i.e., pertaining to inventions, was issued in Venice in 1474 [138], followed by the British Statute of Monopolies of 1623, the United States in 1790, and France in 1791 [11]. The full history of the patent law is certainly not the focus here, but rather the point that, when approaching this particular field, one has to take into account centuries of practice. The side-effect of this public disclosure of inventions is a library of cultural heritage documenting the development of human technologies from the middle-ages to the present day. All this, and more, is prior-art to any new patent application.

The different laws in different countries result in several possible definitions of a patent. According to the World Intellectual Property Organization (WIPO) [3],

> "a patent is the right granted to an inventor by a State, or by a regional office acting for several States,

> which allows the inventor to exclude anyone else from commercially exploiting his or her invention for a limited period, generally 20 years."

The conditions under which such a right may be granted may also show slight differences between authorities, but generally four conditions have to be met [11]:

novelty: The invention must not have been described or used before the application

inventive step: The invention must also not be a new but obvious combination of existing processes or entities

industrial applicability: It must be possible to build or use it in practice (e.g., no patent for a *perpetuum mobile*)

non-excluded material: It must not refer to areas explicitly excluded by law from patenting (e.g., natural products)

The modern practice of patent law starts with the Paris Convention of 1883 [207]. At the time of writing, there were 174 nations listed as contracting parties, the latest one being Thailand in 2008. The Paris Convention is the first in a series of international agreements that aim to make the patent system a truly global one. For even though most laws take prior art to be any public data anywhere in the world, the practice is essentially a national one, with only one true multinational authority, the European Patent Office (EPO).[1]

The Paris Convention lays down one of the fundamental properties of the current patent system, the *priority*. In essence, the Convention allows the inventor to claim priority on an invention at any patent office of a signatory country, based on a prior application he/she made in any other signatory country, generally within 12 months. This system results in the creation of links between documents issued by different patent offices, in different languages, essentially covering the same invention. It is a fundamental property that, as we will see in the following sections, has found its utility not only in the search methods,

[1] Even in the case of the EPO, the actual patents are issued by national offices, but the procedure is greatly simplified.

but also in machine translation, network analysis and evaluation of IR systems.

Priorities create the possibility of building patent *families* — the set of patents describing the same invention. Depending on how flexible one is in linking the documents based on their priority references, the families can describe a very specific invention, or a general technical field. Families are however not a legal concept, and just to illustrate this flexibility, let us note that the WIPO, in its Handbook of Industrial Property Information and Documentation [206], identifies five types (simple, complex, extended, national, and artificial), while the EPO, on its information Web site,[2] gives three definitions and provides links to how some commercial providers define their understanding of patent families.

This difference in definition notwithstanding, there are at least 250,000 common applications per year among *The Five IP Offices* (IP5)[3] (i.e., the same application filed at more than one IP5 office) [141]. This amounts to a considerable body of comparable multilingual data. And, given the way the patent system currently works (i.e., applications for the same invention made and examined at different patent offices), a set of independent searchers are creating relevance judgments in an ad-hoc pooling-like way.[4]

The size of patent corpora is relatively small when compared to the current web corpora (ClueWeb'09 is 25 terabytes compressed [1], while none of the patent corpora available reach the 1 terabyte mark). However, the research issues are still abundant. This review covers the vast majority of the research already done, as well as points out potential avenues for the future. When talking to a patent expert, it emerges that the work still needing to be done for patent retrieval is a mixture of technology features and legal or administrative issues. While this review certainly focuses on the former, the two are surprisingly difficult to extricate from each other. Often enough, procedures are put in

[2] http://www.epo.org/searching/essentials/patent-families/definitions.html

[3] Five patent offices that have agreed to a tighter collaboration in patent prosecution: European Patent Office (EPO), United States Patent and Trademark Office (USPTO), Japan Patent Office (JPO), Korean Intellectual Property Office (KIPO), and State Intellectual Property Office of the People's Republic of China (SIPO).

[4] There are international efforts underway to eliminate this apparent work duplication in order to speed-up patent prosecution.

place to do a good job of searching with the technology of the 1980s or even earlier. A classical example of this would be the creation of extremely long and complex Boolean queries [24]. These procedures then become part of what the community generally defines as "patent search" and research is done to adapt to this particular scenario. This is a factor to keep in mind when looking beyond the restricted confines of the described use cases.

Adams, in his presentation to the WIPO in 2009 [8] and later in his keynote at the Patent Information Retrieval Workshop in 2011,[5] identified three areas of development for improving search:

(1) Search strategy development — the human factor
(2) Database creation and maintenance
(3) Search engines and information navigation tools

The three areas all interact with and are dependent on each other, but the *Human Factor* and the *Database creation and maintenance* are not the subject of this survey. However, it is important for the IR community to understand that although the core algorithm, its supplementary features and its interfaces are very important, they are just a third of the complete process. Furthermore, studies on information navigation and visualization in the IR community are sparse. We will briefly cover them in Section 4.5.

The need for better search engines is however particularly acute now, as the number of patent applications grows and, together with it, the backlog of patent offices. For instance, as of January 2011, the United States Patent and Trademark Office (USPTO) had a backlog of 1.2 million patent applications [34].

1.2 Domains within a Domain

A common understanding of a *domain-specific search engine* is that it *"limits its index to pages corresponding to a particular subject area, publisher or purpose"* [183]. This definition covers all aspects of what one may define as domain-specific search, provided we are slightly liberal

[5] http://ifs.tuwien.ac.at/pair2011

in its interpretation. The “subject area” component refers to domains such as scientific publications, healthcare, biomedicine, chemistry, etc. The “publisher” could be perceived as a publication-specific medium. Text (hyperlinked or not), images, and combinations thereof such as news feeds, blogs or twitter search are examples that come to mind in this sense. Finally, “purpose” is better understood as *users* or *use case domains* and implies a connection to the user performing the search and his or her motivations and objectives. Let us therefore rephrase this definition:

Definition 1.1. Domain-specific search [engine | process] is a search [engine | process] that fixes one or more of the following three dimensions:

(1) **subject area** (e.g., chemical, biomedical, healthcare)
(2) **publication form or medium** (e.g., blogs, micro-blogs, books)
(3) **users or use case domain** (e.g., patent search, cultural heritage, expert search)

It should be noted that in reality the three axes are not quite orthogonal. Some use case domains require specific subject areas or publication forms. The lack of perfect orthogonality has however never been an obstacle in IR and we should take this definition in this spirit as well.

Figure 1.1 shows a graphical representation of the three axes. As an illustration of a domain, a volume of the space can be used to represent a domain in a qualitative way. The domain of healthcare is shown as an example of a domain that covers a more limited subject area, but targets a large number of users and user scenarios (both medical professionals and non-professionals regularly search for health and medical information in a large number of scenarios). In contrast, the patent domain covers chemistry, mechanical engineering, electrical engineering, and practically all other domains of industry applicable human knowledge, but focuses on a relatively small number of users and use cases. We develop the discussion on these use cases in the patent domain in Section 1.4.

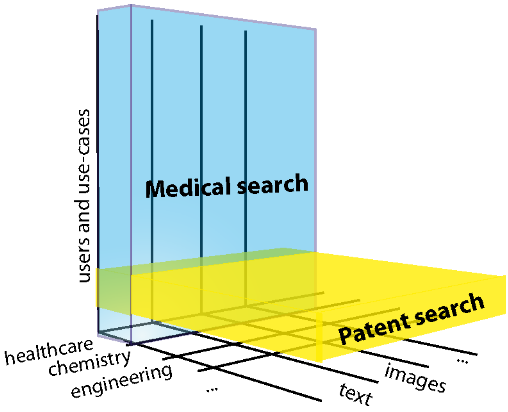


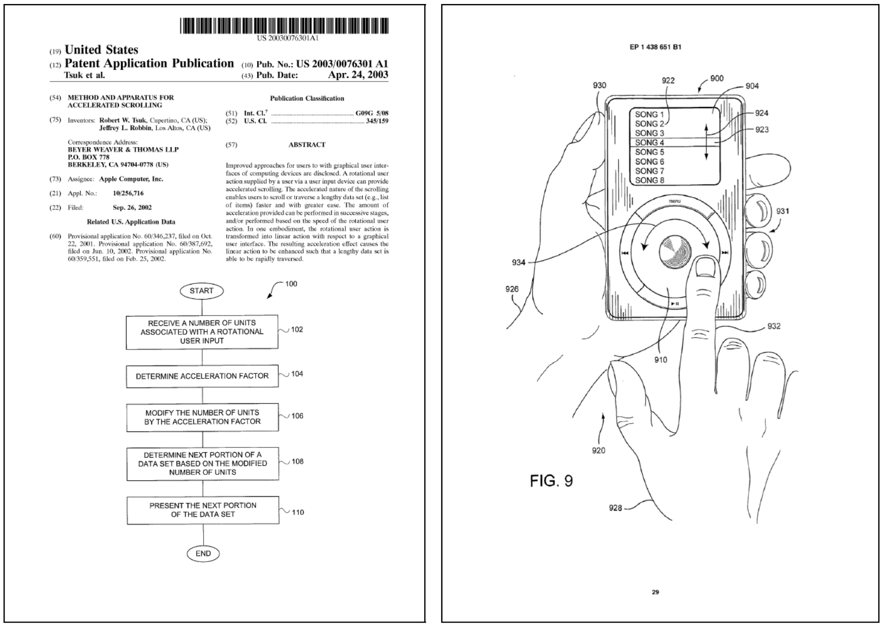
Fig. 1.1 The patent domain cuts across many scientific and technical domains.

1.3 A Patent Example

Before moving on with this survey, it is worth taking a closer look at the patent documents that we will often mention in later sections. As with any example, it does not cover all types of patents and aspects of the patenting process, but it provides the basic understanding necessary for subsequent sections. The reader familiar with the domain may skip this section.

Figure 1.2 shows two pages of the US patent application 10/256716, related to a very well-known consumer product. It is clear from Figure 1.2(b) that this patent application refers to the navigation mode of an iPod. The title (*Method and apparatus for accelerated scrolling*) and generally the first page is much less clear. In fact, this particular application was filed on September 26, 2002, after the first iPod was released, yet there is no mention of the device by its name in the text of the application. Instead, this patent application provides a detailed technical description of how the scrolling mechanism for the iPod works. For obvious reasons, we do not reproduce here the full text of the description, but it is available online.[6] The application requests

[6] http://1.usa.gov/Oq0Wr7



US 20030076301A1

(19) United States
(12) Patent Application Publication
Tsuk et al.

(10) Pub. No.: US 2003/0076301 A1
(43) Pub. Date: Apr. 24, 2003

(54) METHOD AND APPARATUS FOR ACCELERATED SCROLLING

(75) Inventors: Robert W. Tsuk, Cupertino, CA (US); Jeffrey L. Robbin, Los Altos, CA (US)

Correspondence Address:
BEYER WEAVER & THOMAS LLP
P.O. BOX 778
BERKELEY, CA 94704-0778 (US)

(73) Assignee: Apple Computer, Inc.

(21) Appl. No.: 10/256,716

(22) Filed: Sep. 26, 2002

Related U.S. Application Data

(60) Provisional application No. 60/346,237, filed on Oct. 22, 2001. Provisional application No. 60/387,692, filed on Jun. 10, 2002. Provisional application No. 60/359,551, filed on Feb. 25, 2002.

Publication Classification

(51) Int. Cl.^{7} G09G 5/08
(52) U.S. Cl. 345/159

(57) ABSTRACT

Improved approaches for users to with graphical user interfaces of computing devices are disclosed. A rotational user action supplied by a user via a user input device can provide accelerated scrolling. The accelerated nature of the scrolling enables users to scroll or traverse a lengthy data set (e.g., list of items) faster and with greater ease. The amount of acceleration provided can be performed in successive stages, and/or performed based on the speed of the rotational user action. In one embodiment, the rotational user action is transformed into linear action with respect to a graphical user interface. The resulting acceleration effect causes the linear action to be enhanced such that a lengthy data set is able to be rapidly traversed.

START
100
RECEIVE A NUMBER OF UNITS ASSOCIATED WITH A ROTATIONAL USER INPUT
102
DETERMINE ACCELERATION FACTOR
104
MODIFY THE NUMBER OF UNITS BY THE ACCELERATION FACTOR
106
DETERMINE NEXT PORTION OF A DATA SET BASED ON THE MODIFIED NUMBER OF UNITS
108
PRESENT THE NEXT PORTION OF THE DATA SET
110
END
EP 1 438 651 B1
930
922
900
904
924
923
SONG 1
SONG 2
SONG 3
SONG 4
SONG 5
SONG 6
SONG 7
SONG 8
menu
931
934
926
932
910
920
FIG. 9
928
29

(a) The first page

(b) Page 12 of 28

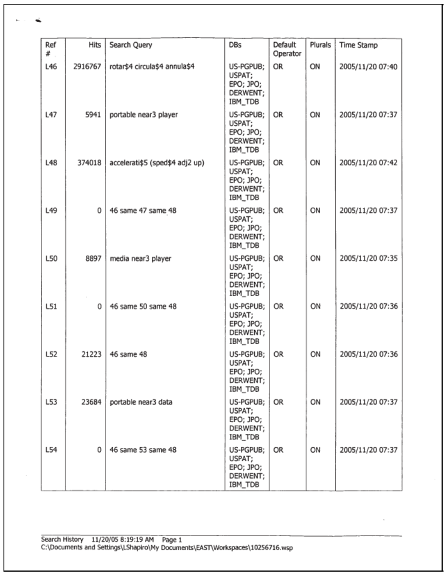
Fig. 1.2 Two pages of patent application 10/256716.

protection for a set of ideas, which it describes in 59 claims. Here are the first five:

(1) A method for scrolling through portions of a data set, said method comprising: receiving a number of units associated with a rotational user input; determining an acceleration factor pertaining to the rotational user input; modifying the number of units by the acceleration factor; determining a next portion of the data set based on the modified number of units; and presenting the next portion of the data set.
(2) A method as recited in claim 1, wherein the data set pertains to a list of items, and the portions of the data set include one or more of the items.
(3) A method as recited in claim 1, wherein the data set pertains to a media file, and the portions of the data set pertain to one or more sections of the media file.
(4) A method as recited in claim 3, wherein the media file is an audio file.
(5) A method as recited in claim 1, wherein the rotational user input is provided via a rotational input device.

As we can see, the claims are written in a particular style, sometimes referred to as *patentese* [19], resembling the language of many legal contracts. In practice this is almost always the case. What one can also see from this example is that there are a number of internal references between claims. In practice, a claim which does not reference any other claim is called an *independent claim* and all others are called *dependent claims*. In the example above, Claim 1 is independent, while the following four are all dependent, forming a tree of references.

Following examination, this patent application was granted a patent in the United States, namely US7,312,785, issued over 5 years after the initial application. In this process, the examining office (i.e., the USPTO, in this case) published over a hundred documents, covering mostly the communication between the office and the applicants, as well as some procedural notes from the office. Among them, the most interesting for IR researchers are probably the *Examiner's search strategy and results*, the list of references cited by the examiner, and the series of decisions made by the office. In the end, the granted patent (US7,312,785) contains 40 claims only. Figure 1.3 shows two examples of documents published by the USPTO in relation to this application.



Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp

L46	2916767	rotar$4 circula$4 annula$4	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:40

L47	5941	portable near3 player	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:37

L48	374018	accelerati$5 (sped$4 adj2 up)	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:42

L49	0	46 same 47 same 48	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:37

L50	8897	media near3 player	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:35

L51	0	46 same 50 same 48	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:36

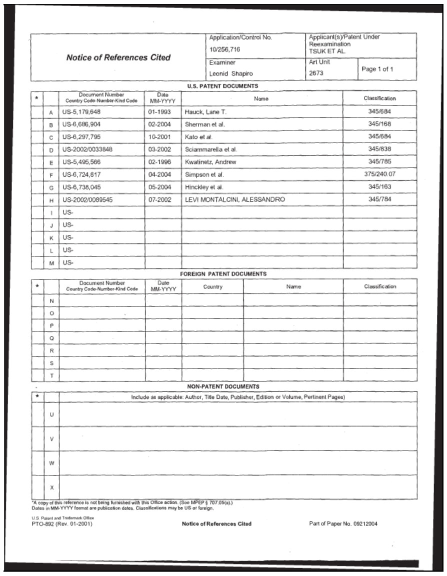
L52	21223	46 same 48	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:36

L53	23684	portable near3 data	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:37

L54	0	46 same 53 same 48	US-PGPUB;
			USPAT;
			EPO; JPO;
			DERWENT;
			IBM_TDB
				OR	ON	2005/11/20 07:37

Search History 11/20/05 8:19:19 AM Page 1
C:\Documents and Settings\LShapiro\My Documents\EAST\Workspaces\10256716.wsp

(a) Examiner's strategy



Notice of References Cited	Application/Control No. 10/256,716	Applicant(s)/Patent Under Reexamination TSUK ET AL.	
	Examiner Leonid Shapiro	Art Unit 2673	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-5,179,648	01-1993	Hauck, Lane T.	345/684
	B	US-6,686,904	02-2004	Sherman et al.	345/168
	C	US-6,297,795	10-2001	Kato et al.	345/684
	D	US-2002/0033848	03-2002	Sciammarella et al.	345/838
	E	US-5,495,566	02-1996	Kwatinetz, Andrew	345/785
	F	US-6,724,817	04-2004	Simpson et al.	375/240.07
	G	US-6,738,045	05-2004	Hinckley et al.	345/163
	H	US-2002/0089545	07-2002	LEVI MONTALCINI, ALESSANDRO	345/784
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

U.S. Patent and Trademark Office
PTO-892 (Rev. 01-2001) **Notice of References Cited** Part of Paper No. 09212004

(b) List of citations

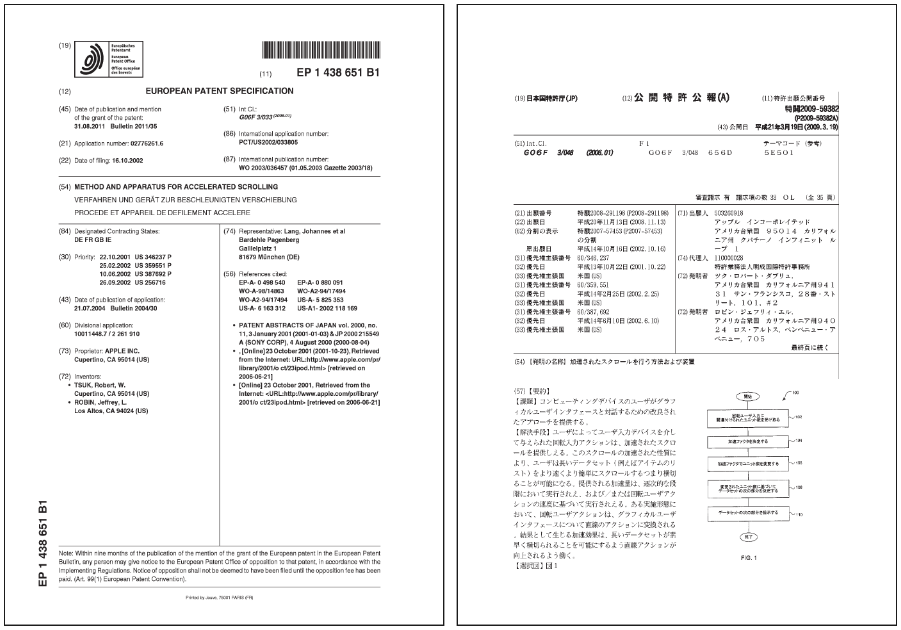
Fig. 1.3 Some documents published by the USPTO in relation to patent application 10/256716.

The US patent offers the owner a monopoly over the manufacture and licensing of the invention only in territories under the US jurisdiction. This is why, 20 days after filing the application with the USPTO, another application was filed with the WIPO: WO03/036457 on October 16, 2002. Even though the WIPO is not a patent office (i.e., it does not grant patents), it is the entry point to the so called *PCT Route*, which is a system designed to facilitate the acquisition of protection in different jurisdictions. The name of the route comes from the Patent Cooperation Treaty (PCT), which establishes the procedures under which an application filed with the WIPO reaches the national patent offices which can grant patents. Without going into the full details of the system,[7] here is the brief description of the route:

(1) an application is filed with the WIPO
(2) an International Search Report (ISR) is created by a Search Authority (an accredited search organization, generally a national patent office), and published. This phase is referred to as *Section 1*
(3) an optional further search is performed by a Search Authority (Section 2)
(4) the applicant decides, based on the search reports, whether to proceed to the national phase. This means that the patent application is passed to the set of offices where the applicant desires protection (they are called *designated states*)
(5) upon receiving the application, the national patent offices start their own examination procedures, optionally taking into account the ISR created in the previous steps.

Whether the application goes through the PCT route or not, in situations where protection is sought in different jurisdictions for the same invention, a so-called family of patents results. Figure 1.4 shows the European and Japanese patents corresponding to the US patent discussed before.

[7] Full details about the PCT and PCT applications are available at http://www.wipo.int/pct/en/. The details of the PCT route are much more complicated than the five bullet points presented here.



(19) Europäisches Patentamt European Patent Office Office européen des brevets

(11) EP 1 438 651 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
31.08.2011 Bulletin 2011/35

(21) Application number: 02776261.6

(22) Date of filing: 16.10.2002

(51) Int Cl.:
G06F 3/033 (2006.01)

(86) International application number:
PCT/US2002/033805

(87) International publication number:
WO 2003/036457 (01.05.2003 Gazette 2003/18)

(54) **METHOD AND APPARATUS FOR ACCELERATED SCROLLING**
VERFAHREN UND GERÄT ZUR BESCHLEUNIGTEN VERSCHIEBUNG
PROCEDE ET APPAREIL DE DEFILEMENT ACCELERE

(84) Designated Contracting States:
DE FR GB IE

(30) Priority: 22.10.2001 US 346237 P
25.02.2002 US 359551 P
10.06.2002 US 387692 P
26.09.2002 US 256716

(43) Date of publication of application:
21.07.2004 Bulletin 2004/30

(60) Divisional application:
10011448.7 / 2 261 910

(73) Proprietor: APPLE INC.
Cupertino, CA 95014 (US)

(72) Inventors:
- TSUK, Robert, W.
Cupertino, CA 95014 (US)
- ROBIN, Jeffrey, L.
Los Altos, CA 94024 (US)

(74) Representative: Lang, Johannes et al
Bardehle Pagenberg
Galileiplatz 1
81679 München (DE)

(56) References cited:

EP-A- 0 498 540	EP-A- 0 880 091
WO-A-98/14863	WO-A2-94/17494
WO-A2-94/17494	US-A- 5 825 353
US-A- 6 163 312	US-A1- 2002 118 169

- PATENT ABSTRACTS OF JAPAN vol. 2000, no. 11, 3 January 2001 (2001-01-03) & JP 2000 215549 A (SONY CORP), 4 August 2000 (2000-08-04)
- , [Online] 23 October 2001 (2001-10-23), Retrieved from the Internet: URL:http://www.apple.com/pr/library/2001/o ct/23ipod.html> [retrieved on 2006-06-21]
- [Online] 23 October 2001, Retrieved from the Internet: <URL:http://www.apple.com/pr/library/2001/o ct/23ipod.html> [retrieved on 2006-06-21]

EP 1 438 651 B1

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Printed by Jouve, 75001 PARIS (FR)

(a) European granted patent

(19) 日本国特許庁(JP)　(12) 公開特許公報(A)　(11) 特許出願公開番号
特開2009-59382
(P2009-59382A)
(43) 公開日　平成21年3月19日(2009.3.19)

(51) Int.Cl.	F I	テーマコード（参考）
G06F 3/048 (2006.01)	ＧＯ６Ｆ　3/048　６５６Ｄ	５Ｅ５０１

審査請求 有　請求項の数 33　ＯＬ　（全 35 頁）

(21) 出願番号	特願2008-291198 (P2008-291198)
(22) 出願日	平成20年11月13日 (2008.11.13)
(62) 分割の表示	特願2007-57453 (P2007-57453)の分割
原出願日	平成14年10月16日 (2002.10.16)
(31) 優先権主張番号	60/346,237
(32) 優先日	平成13年10月22日 (2001.10.22)
(33) 優先権主張国	米国 (US)
(31) 優先権主張番号	60/359,551
(32) 優先日	平成14年2月25日 (2002.2.25)
(33) 優先権主張国	米国 (US)
(31) 優先権主張番号	60/387,692
(32) 優先日	平成14年6月10日 (2002.6.10)
(33) 優先権主張国	米国 (US)

(71) 出願人	503260918 アップル　インコーポレイテッド アメリカ合衆国　９５０１４　カリフォルニア州　クパチーノ　インフィニット　ループ　１
(74) 代理人	110000028 特許業務法人明成国際特許事務所
(72) 発明者	ツク・ロバート・ダブリュ． アメリカ合衆国　カリフォルニア州９４１３１　サン・フランシスコ，２８番・ストリート，１０１，＃２
(72) 発明者	ロビン・ジェフリィ・エル． アメリカ合衆国　カリフォルニア州９４０２４　ロス・アルトス，ベンベニュー・アベニュー，７０５

最終頁に続く

(54)【発明の名称】加速されたスクロールを行う方法および装置

(57)【要約】
【課題】コンピューティングデバイスのユーザがグラフィカルユーザインタフェースと対話するための改良されたアプローチを提供する。
【解決手段】ユーザによってユーザ入力デバイスを介して与えられた回転入力アクションは、加速されたスクロールを提供しえる。このスクロールの加速された性質により、ユーザは長いデータセット（例えばアイテムのリスト）をより速くより簡単にスクロールするつまり横切ることが可能になる。提供される加速量は、逐次的な段階において実行されえ、および／または回転ユーザアクションの速度に基づいて実行されえる。ある実施形態において、回転ユーザアクションは、グラフィカルユーザインタフェースについて直線のアクションに変換される。結果として生じる加速効果は、長いデータセットが素早く横切られることを可能にするよう直線アクションが向上されるよう働く。
【選択図】図１

開始
100
回転ユーザ入力に
関連付けられたユニット数を受け取る
102
加速ファクタを決定する
104
加速ファクタでユニット数を変更する
106
変更されたユニット数に基づいて
データセットの次の部分を決定する
108
データセットの次の部分を提示する
110
終了
FIG. 1

(b) Japanese application

Fig. 1.4 Family members of US patent 7,312,785.

All of the application and granted patent documents will generally have the same structure, consisting of:

- **Bibliographical data**: title, metadata related to the specific publication at hand, the inventors, assignees, agents or applicants, as well as relations to other documents
- **Abstract**: a very brief summary of the invention
- **Description**: a detailed description of the invention, including prior work, examples, related technologies
- **Claims**: the legal description of the invention. Adams [11] defines the claims as a *"Sequence of paragraphs at the end of a patent application defining the scope of monopoly sought. After substantive examination, the same section of the granted patent defines the legal rights of the proprietor."*

We will often make references to these sections in the coming sections.

1.4 Patent Search Processes

While, in principle, the search process is always about finding relevant documents to satisfy a particular information need, patent search has specialized into different processes, differentiated as a function of the input (an idea, a disclosure of innovation, a patent application, a claim, a granted patent) and the needed output (a large set of scientific publications covering a domain, a set of patents, a single patent). In relation to patent search one will therefore often hear names such as *State of the art*, *Pre-filing patentability*, *Novelty*, *Freedom to operate*, *Validity*, or *Due diligence* search. Their precise names and definitions vary between different practitioners.[8] Alberts et al. [16] describe in detail five of these search types, but also demonstrate the variability in the definition, by providing a table with seven search types. Adams [11] adds another type of search (*Alerting*) and slightly regroups the rest.

Generally, these types of search are also related to eDiscovery because of their legal nature, which puts a large emphasis on finding *all* relevant documents. The greatest difference between the practice of patent searchers and legal staff is perhaps the amount of metadata available to patent searchers. As we have seen in the previous example, each published patent document is the result of a specific process and comes associated with a rich set of metadata.

The different types of patent search are summarized in Table 1.1, which shows the type of search and alternative names for it, as well as the search specification (what the search begins from) and the corpora in which the search is conducted. A short description of each search type is given in italic text. Figure 1.5 shows the life-cycle of an innovation, with the searches that are directly related to it. The figure describes the path from having an idea to do something new, i.e., an innovation, to obtaining (and defending) a patent. It follows four of the six types of search described in Table 1.1, in the order in which they occur, and shows the most important documents that are a result of this process. The rectangles denoting the searches also indicate who typically

[8] By practitioners, we understand here all those who deal with patents in their professional life. This generally includes corporate librarians, information specialists, private patent searchers, patent examiners at any patent office, and patent lawyers.

Table 1.1. Types of patent search.

Search type	Other names	Search specification	Corpora
State of the art	Technology survey	An idea	All public documents
To obtain a general understanding of the field surrounding the innovation at hand			
Pre-filing patentability		A fairly well defined innovation disclosure	All public documents
Similar to above, but with a more precise request for information and potentially more focus on patent documents			
Patentability	Novelty, Prior Art	A patent application	All public documents until the date of the application
Identify whether a specific patent application satisfies the conditions for granting			
Freedom to operate	Infringement, Right-to-Use, Clearance	A product and related methods or technologies	The set of patents in force in a particular jurisdiction
Identify any patent in force in a particular jurisdiction which may prevent a product from being commercialized in that jurisdiction			
Validity	Invalidity, Enforcement Readiness, Opposition	A granted patent	All public documents prior to the priority date of the patent in question
Identify whether a granted patent satisfied the granting criteria at the earliest priority date (i.e., the moment when a first application was registered for the invention described therein)			
Patent portfolio search	Due diligence, Patent landscape	A company, a technology area	All public documents
Obtain a general understanding of the patents, both in force and expired, in a specific technology area and/or jurisdiction			

performs them. Note that a search represented by a rectangle could take place over a number of hours or even weeks. The diamonds represent decision points, at which the question "Relevant item(s) found?" is asked, relating to the search just performed. If the response is positive, then a previous step in the process must be repeated.

Figure 1.5 does not show the *Freedom to operate* and *Patent portfolio* because, as can be seen from Table 1.1, these use cases do not

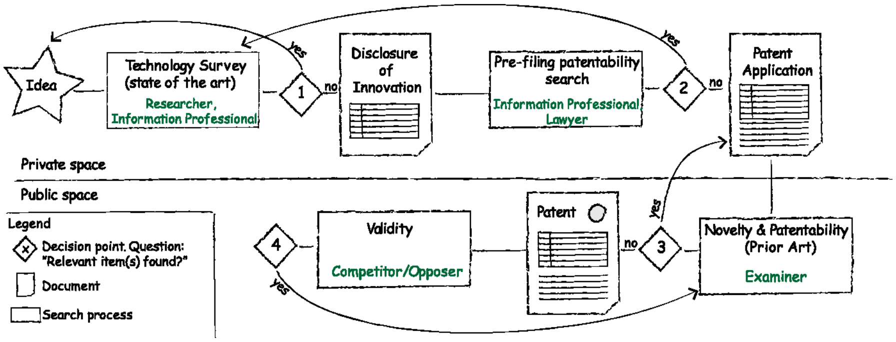

Idea
Technology Survey (state of the art)
Researcher, Information Professional
1
no
yes
Disclosure of Innovation
Pre-filing patentability search
Information Professional, Lawyer
2
no
yes
Patent Application
Private space
Public space
Legend
x Decision point. Question: "Relevant item(s) found?"
Document
Search process
4
yes
Validity
Competitor/Opposer
Patent
no
3
yes
Novelty & Patentability (Prior Art)
Examiner

Fig. 1.5 High-level view of the life-cycle of a patented idea.

have a document at the basis of their request for information. These two types of search can in fact be performed at any time in the course of the development of an idea into a patented product, as well as any time thereafter.

Finally we should note that in Table 1.1, we refer to *all public documents* as a particular corpus, but what should be clear to the reader is that theoretically any publicly disclosed knowledge, not necessarily in written form, can be used to invalidate a patent.

The IR scientist reader may by now realize that for the core IR engine design, these different types of patent search do not appear to make a difference. The principal differences, as we have listed them, have to do with the target data collection or the form in which the request for information is expressed. The differences lie in the attitude with which the search process is conducted and the tools that assist the user in achieving the different objectives of these different patent searches.

Understanding these processes is important for the success of the resulting search system and surveys among professional patent searchers have been conducted in this sense. Hansen [77] have interviewed patent examiners at the Swedish patent office. Tseng and Wu [191] have interviewed 43 patent searchers, 18 of which agreed to a follow-up experimental observation of their search behavior. Most recently, Azzopardi et al. [20] followed up with an online survey which received 81 responses. Characteristically, the set of patent search types identified was also different from the sets of Alberts et al. [16] and Adams [11] mentioned before.

Based on these and similar surveys in the context of the PROMISE project[9] [94] (Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation), the scenario outlined in Listing 1.1 has been crystallized to model the prior art search performed by a patent examiner. This is perhaps the most common type of search, if we do not take into account the technology survey done in the context of research environments. Nevertheless, the patent examiner at a patent office is not the only actor in the field. As Figure 1.5 and Table 1.1 indicate, scientists, librarians, and lawyers are also potentially involved.

Such user studies reveal aspects that the previous taxonomy did not identify. The different types of search identified by patent experts themselves have to do, as we have seen, with the final objectives of the search task. From an IR researcher's point of view, however, the different tasks

Listing 1.1. Prior Art Search use case.

```
1.  User receives a patent application document to evaluate
2.  User enters the search system
3.  User enters a text query , potentially with Boolean operators
    and specific field filters
4.  System presents a result list , sorted by relevance , with
    snippets , metadata information , and links to full documents
5.  User inspects and assesses all the documents
6.  User clicks on one element of the list for further inspection
7.  System presents the full document , with any metadata ,
    attached images and text
8.  User inspects and assesses . Finds the document potentially
    relevant and saves it to a bucket
9.  User clicks on the 'Back' button to return to the list of
    results
10. System presents the list , with the already viewed documents
    visibly identifiable
11. Jump to Step 6 , unless new query query is required or User
    satisfied
12. Based on a potential new understanding , User inputs a new
    query
13. Jump to Step 3 , unless User satisfied
14. User saves the list and creates a search report
15. Use case ends
```

[9] http://www.promise-noe.eu

are distinguished by the formulations of the queries (Boolean, keywords only, full text, metadata, images), by the number of sessions required in the process, and by the level of collaboration in achieving the desired goal. The core difficulty for the IR scientist has to do with the patent documents themselves, even if many search types require the corpus to be the entire, publicly available, human knowledge.

1.5 Patents in the IR Community

This section reviews the beginnings of IR research in the patent domain, and lists a number of sources of information on IR in the patent domain, as well as sources of patent collections for use in IR research. The first recognition of the IR community of the need for special attention to be dedicated to the issues of patent retrieval was in the organization of the first Workshop on Patent Retrieval by Kando and Leong [97]. However the subject had been approached before by Sheremetyeva and Nirenburg [171] and by Larkey [110]. Other studies, for specific domains (particularly chemistry [32, 99]) and for business analysis (for example, [42]) had appeared even before, but outside of the core IR field.

1.5.1 Sources of Knowledge

Work in this domain has been encouraged through evaluation campaigns, first through the NII[10] Test Collection for IR Systems (NTCIR) [57, 60, 93, 145, 146], and later through the Text Retrieval Conference (TREC) (for the chemical domain [129]) and continuing at the moment of writing this review, through the Cross-Language Evaluation Forum (CLEF) [156, 157, 158, 161].

This survey relies heavily on the proceedings of the SIGIR Workshop on Patent IR in 2000 [97]; the ACL workshop on Patent Corpus Processing [4]; the special issue of the Information Processing & Management journal [59]; the PaIR series of workshops [5], generally co-located with the International Conference on Information and Knowledge Management (CIKM); the Advances in Patent Information Retrieval (AsPIRe) Workshop [76] co-located with the European Conference on

[10] National Institute of Informatics (Japan).

Information Retrieval (ECIR) in 2010; as well as the reports from the evaluation campaigns mentioned above. A number of other articles have appeared in various journals, specific to the technology described, and at least another survey has appeared in the *World Patent Information* (WPI) journal [35]. The WPI journal is an important source of domain related information and while its audience is mostly made up of patent information professionals, the IR researcher interested in the domain would find its articles interesting. Most recently, in 2011, an edited book has collected a series of articles focusing on the patent domain [127].

A number of other symposia do not have proceedings, but publish the slides online. The Information Retrieval Facility[11] (IRF) has organized a symposium between 2007 and 2010, bringing together IR researchers and IP professionals.[12] The various patent offices also organize information events and training sessions.[13]

1.5.2 Sources of Data

Working in the patent domain implies having access to collections of patent data. Aside from those made available in the various evaluation campaigns already mentioned previously, patent data can be obtained directly from the patent offices, or from research collections.

While all patent offices make available patent data (it is one of their core responsibilities), it is not always easy to obtain it. The USPTO has its full text database available online[14] for manual search, and for bulk download via Google.[15] The European Patent Office[16] is one of the most active in this area, offering both researchers and commercial organizations free access to their data as well as data they have collected from other offices, via their Open Patent Services (OPS).[17] A fair-use policy applies in this case.

[11] http://www.ir-facility.org

[12] http://www.irfs.at

[13] http://www.wipo.int/meetings/en/index.jsp, http://www.epo.org/learning-events.html, http://www.uspto.gov/products/events/index.jsp

[14] http://patft.uspto.gov/

[15] http://www.google.com/googlebooks/uspto.html

[16] http://www.epo.org

[17] http://ops.epo.org

The Matrixware Research Collection (MAREC), first made available by the IRF, consists of approximately 19 million patent documents in XML format, covering four patent offices (USPTO, EPO, JPO, and WIPO). MAREC is now available for download under a Creative Commons license.[18] Further sources are the datasets, queries, and relevance judgments made available in the NTCIR patent track, and CLEF-IP and TREC-CHEM evaluation campaign tracks.

1.6 Structure of the Survey

The rest of this survey looks in detail at the various aspects of IR in the patent domain. After this introductory section, we continue in Section 2 with a detailed description of evaluation best practices in the field. We do this because, on one hand, evaluation is based on the understanding of search processes just described, and, on the other hand, because in this way we lay the ground for the discussions of results in future sections. The main focus of the survey is in Section 3, on text indexing and retrieval. We cover there both bag-of-words approaches, as well as those supported by Natural Language Processing (NLP) methods. Section 4 follows up on the text retrieval discussion with details on metadata associated with patent documents and how such metadata assists the search process. We cover both existing metadata, as well as experiments on creating new metadata.

Section 5 moves away from textual information and introduces the specific issues related to image and chemical structure retrieval in the patent domain. We discuss the importance of the information contained in the non-textual parts of the patent, as well as algorithms that have been developed to make use of this information in search.

Finally, we summarize the domain in the Conclusions section, and provide a set of research and development trends observed in recent years in relation to patent IR.

[18] http://www.ifs.tuwien.ac.at/imp/marec.shtml

References

[1] "The clueweb dataset," http://lemurproject.org/clueweb09.php/index.php.

[2] "Manual of patent examination procedure, section 608.01(m)," Revision July 2010, http://www.uspto.gov/web/offices/pac/mpep/index.htm.

[3] "Understanding intellectual property," http://www.wipo.int/about-ip/en/.

[4] *PATENT '03: Proceedings of the ACL-2003 Workshop on Patent Corpus Processing*. Association for Computational Linguistics, 2003.

[5] *PaIR '11: Proceedings of the Workshop on Patent Information Retrieval*. ACM, 2011.

[6] S. Adams, "Comparing the IPC and the US classification systems for the patent searcher," *World Patent Information*, vol. 23, pp. 15–23, 2001.

[7] S. Adams, "Electronic non-text material in patent applications-some questions for patent offices, applicants and searchers," *World Patent Information*, vol. 27, pp. 99–103, 2005.

[8] S. Adams, "New methodologies for patent searching; what do we need?," in *Proceedings of the Global Symposium of Intellectual Property Authorities*, 2009.

[9] S. Adams, "The text, the full text and nothing but the text: Part 1 — standards for creating textual information in patent documents and general search implications," *World Patent Information*, vol. 32, pp. 22–29, 2010.

[10] S. Adams, "The text, the full text and nothing but the text: Part 2 — standards for creating textual information in patent documents and general search implications," *World Patent Information*, vol. 32, pp. 120–128, 2010.

[11] S. Adams, *Information Sources in Patents*. G. Saur, 3rd ed., 2011.

[12] M. Agatonovic, N. Aswani, K. Bontcheva, H. Cunningham, T. Heitz, Y. Li, I. Roberts, and V. Tablan, "Large-scale, parallel automatic patent annotation," in *Proceedings of Workshop on Patent Information Retrieval*, pp. 1–8, 2008.

[13] F. Aiolli, "A preference model for structured supervised learning tasks," in *Proceedings of IEEE International Conference on Data Mining*, pp. 557–560, 2005.

[14] F. Aiolli, R. Cardin, F. Sebastiani, and A. Sperduti, "Preferential text classification: Learning algorithms and evaluation measures," *Information Retrieval*, vol. 12, pp. 559–580, 2009.

[15] F. Aiollo and A. Sperduti, "Learning preferences for multiclass problems," in *Advances in Neural Information Processing Systems*, pp. 17–24, 2005.

[16] D. Alberts, C. B. Yang, D. Fobare-DePonio, K. Koubek, S. Robins, M. Rodgers, E. Simmons, and D. DeMarco, "Introduction to patent searching — practical experience and requirements for searching the patent space," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, pp. 3–43, 2011.

[17] L. Andersson, "A vector space analysis of Swedish patent claims with different linguistic indices," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 47–56, 2010.

[18] M. Annies, "Full-text prior art and chemical structure searching in e-journals and on the internet — a patent information professional's perspective," *World Patent Information*, vol. 31, pp. 278–284, 2009.

[19] K. H. Atkinson, "Towards a more rational patent search paradigm," in *Proceedings of Workshop on Patent Information Retrieval*, 2008.

[20] L. Azzopardi, W. Vanderbauwhede, and H. Joho, "Search system requirements of patent analysts," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 775–776, 2010.

[21] L. Azzopardi and V. Vinay, "Accessibility in information retrieval," *Advances in Information Retrieval*, pp. 482–489, 2008.

[22] R. H. Baayen, R. Piepenbrock, and L. Gulikers, "The CELEX Lexical database (release 2)," 1995.

[23] R. Bache and L. Azzopardi, "Improving access to large patent corpora," in *Transactions on Large-Scale Data- and Knowledge-Centered Systems II*, Springer, pp. 103–121, 2010.

[24] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval.* Addison Wesley, 2nd ed., 2010.

[25] J. M. Barnard and G. M. Downs, "Use of Markush structure techniques to avoid enumeration in diversity analysis of large combinatorial libraries," http://www.daylight.com/meetings/mug97/Barnard/970227JB.html, last checked, March 2012.

[26] J. M. Barnard and P. M. Wright, "Towards in-house searching of Markush structures from patents," *World Patent Information*, vol. 31, pp. 97–103, 2009.

[27] S. Bashir and A. Rauber, "On the relationship between query characteristics and IR functions retrieval bias," *Journal of the American Society for Information Science and Technology*, vol. 62, pp. 1515–1532, 2011.

[28] D. Becks, T. Mandl, and C. Womser-Hacker, “Phrases or terms? the impact of different query types,” in *Cross-Language Evaluation Forum (Notebook Papers/Labs/Workshop)*, 2010.

[29] K. Benzineb and J. Guyot, “Automated patent classification,” in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, pp. 239–261, 2011.

[30] K. Beuls, B. Pflugfelder, and A. Hanbury, “Comparative analysis of balanced winnow and svm in large scale patent categorization,” in *Proceedings of Dutch-Belgian Information Retrieval Workshop (DIR)*, pp. 8–15, 2010.

[31] N. Bhatti and A. Hanbury, “Image search in patents: A review,” *International Journal on Document Analysis and Recognition (IJDAR)*, 2012. doi:10.1007/s10032-012-0197-5.

[32] M. Blackman, R. Honeywood, and K. Milne, “Searching organic chemical structures: A comparison of online access to chemical abstracts (CAS) and the corresponding United Kingdom Patent Office search system (c2c),” *World Patent Information*, vol. 8, pp. 20–28, 1986.

[33] A. Blanchard, “Understanding and customizing stopword lists for enhanced patent mapping,” *World Patent Information*, vol. 29, pp. 308–316, 2007.

[34] G. H. Blosser, N. Arshadi, and S. Agrawal, “A critical assessment of the USPTO policies toward small entity patent applications,” *Technology and Information*, vol. 13, 2011.

[35] D. Bonino, A. Ciaramella, and F. Corno, “Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics,” *World Patent Information*, vol. 32, pp. 30–38, 2010.

[36] N. Bouayad-Agha, G. Casamayor, G. Ferraro, S. Mille, V. Vidal, and L. Wanner, “Improving the comprehension of legal documentation: The case of patent claims,” in *Proceedings of International Conference on Artificial Intelligence and Law*, pp. 78–87, 2009.

[37] A. Browne, A. McCray, and S. Srinivasan, “The specialist lexicon,” Technical Report, National Library of Medicine, 2000.

[38] A. Buchanan, N. H. Packard, and M. A. Bedau, “Measuring the evolution of the drivers of technological innovation in the patent record,” *Artificial Life*, vol. 17, pp. 109–122, 2011.

[39] L. Cai and T. Hofmann, “Hierarchical document categorization with support vector machines,” in *Proceedings of International Conference on Information and Knowledge Management*, pp. 78–87, 2004.

[40] L. Cai and T. Hofmann, “Exploiting known taxonomies in learning overlapping concepts,” in *Proceedings of International Joint Conference on Artifical intelligence*, pp. 714–719, 2007.

[41] A. Ceausu, J. Tinsley, A. Way, J. Zhang, , and P. Sheridan, “Experiments on domain adaptation for patent machine translation in the pluto project,” in *Proceedings of Conference of the European Association for Machine Translation*, 2011.

[42] A. Chakrabarti, I. Dror, and N. Eakabuse, “Interorganizational transfer of knowledge: An analysis of patent citations of a defense firm,” *IEEE Transactions on Engineering Management*, vol. 40, pp. 91–94, 1993.

[43] L. Chen, N. Tokuda, and H. Adachi, "A patent document retrieval system addressing both semantic and syntactic properties," in *Proceedings ACL Workshop on Patent Corpus Processing*, 2003.

[44] A. Chu, S. Sakurai, and A. F. Cardenas, "Automatic detection of treatment relationships for patent retrieval," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 9–14, 2008.

[45] G. Csurka, J.-M. Renders, and G. Jacquet, "XRCE's participation at patent image classification and image-based patent retrieval tasks of the CLEF-IP 2011," in *Cross-Language Evaluation Forum (Notebook papers/Labs/Workshop)*, 2011.

[46] H. Cunningham, V. Tablan, I. Roberts, M. Greenwood, and N. Aswani, "Information extraction and semantic annotation for multi-paradigm information management," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, pp. 307–327, 2011.

[47] M. Das, R. Manmatha, and E. M. Riseman, "Indexing flower patent images using domain knowledge," *IEEE Intelligent Systems*, vol. 14, pp. 24–33, 1999.

[48] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, pp. 5:1–5:60, 2008.

[49] M.-C. de Marneffe, B. MacCartney, and C. D. Manning, "Generating typed dependency parses from phrase structure parses," in *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, 2006.

[50] E. D'hondt and S. Verbene, "CLEF-IP 2010: Prior art retrieval using the different sections in patent documents," in *Cross-Language Evaluation Forum (Notebook Papers/LABs/Workshops)*, 2010.

[51] E. D'hondt, S. Verberne, W. Alink, and R. Cornacchia, "Combining document representations for prior-art retrieval," in *Cross-Language Evaluation Forum (Notebook Papers/Labs/Workshop)*, 2011.

[52] C. Emmerich, "Comparing first level patent data with value-added patent information: A case study in the pharmaceutical field," *World Patent Information*, vol. 31, pp. 117–122, 2009.

[53] C. J. Fall, A. Törcsvári, K. Benzineb, and G. Karetka, "Automated categorization in the international patent classification," *SIGIR Forum*, vol. 37, pp. 10–25, 2003.

[54] I. V. Filippov and M. C. Nicklaus, "Optical structure recognition software to recover chemical information: Osra, an open source solution," *Journal of Chemical Information and Modeling*, vol. 49, pp. 740–743, 2009.

[55] A. Fujii, "Enhancing patent retrieval by citation analysis," in *Proceedings International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 793–794, 2007.

[56] A. Fujii and T. Ishikawa, "Patent retrieval experiments at ULIS," in *Proceedings of NII Test Collection for IR Systems-3*, 2002.

[57] A. Fujii, M. Iwayama, and N. Kando, "Overview of patent retrieval task at NTCIR-4," in *Proceedings of NII Test Collection for IR Systems-4*, 2004.

[58] A. Fujii, M. Iwayama, and N. Kando, “Test collections for patent-to-patent retrieval and patent map generation in NTCIR-4 workshop,” in *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, 2004.

[59] A. Fujii, M. Iwayama, and N. Kando, “Introduction to the special issue on patent processing,” *Information Processing and Management*, vol. 43, pp. 1149–1153, 2007.

[60] A. Fujii, M. Iwayama, and N. Kando, “Overview of the patent retrieval task at the NTCIR-6 workshop,” in *Proceedings of NII Test Collection for IR Systems-6*, 2007.

[61] S. Fujita, “Revisiting document length hypotheses: A comparative study of Japanese newspaper and patent retrieval,” *ACM Transactions on Asian Language Information Processing*, vol. 4, pp. 207–235, June 2005.

[62] G. W. Furnas, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum, “Information retrieval using singular value decomposition model of latent semantic structure,” in *Proceedings of SIGIR*, 1988.

[63] D. Ganguly, J. Leveling, and G. Jones, “United we fall, divided we stand: A study of query segmentation and PRF for patent prior art search,” in *Proceedings of PaIR*, 2011.

[64] A. Gibbs, “Boolean patent search: Comparative patent search quality/cost evaluation super Boolean vs. legacy Boolean search engines,” Technical Report, http://patentcafe.com, 2006.

[65] M. Giereth, S. Brügmann, A. Stäbler, M. Rotard, and T. Ertl, “Application of semantic technologies for representing patent metadata,” in *International Workshop on Applications of Semantic Technologies*, 2006.

[66] M. Giereth, S. Koch, Y. Kompatsiaris, S. Papadopoulos, E. Pianta, L. Serafini, and L. Wanner, “A modular framework for ontology-based representation of patent information,” in *Proceedings of the Conference on Legal Knowledge and Information Systems*, 2007.

[67] M. Giereth, S. Koch, M. Rotard, and T. Ertl, “Web based visual exploration of patent information,” in *International Conference on Information Visualization, 2007 (IV ’07)*, pp. 150–155, July 2007.

[68] J. Gobeill, E. Pasche, D. Teodoro, and P. Ruch, “Simple pre and post processing strategies for patent searching in CLEF intellectual property track 2009,” in *Proceedings of the Cross-language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments*, Springer, pp. 444–451, 2009.

[69] E. Graf, I. Frommholz, M. Lalmas, and K. van Rijsbergen, “Knowledge modeling in prior art search,” in *Advances in Multidisciplinary Retrieval*, Springer, pp. 31–46, 2010.

[70] T. Grego, P. Pezik, F. M. Couto, and D. Rebholz-Schuhmann, *Identification of Chemical Entities in Patent Documents*, Vol. 5518 of *LNCS*. pp. 942–949, Springer, 2009.

[71] H. Gurulingappa, B. Mueller, M. Hofmann-Apitius, and J. Fluck, “Information retrieval framework for technology survey in biomedical and chemistry literature,” in *Proceedings of Text Retrieval Conference*, 2011.

[72] H. Gurulingappa, B. Mueller, M. Hofmann-Apitius, R. Klinger, H.-T. Mevissen, C. M. Friedrich, and J. Fluck, "Prior art search in chemistry patents based on semantic concepts and co-citation analysis," in *Proceedings of Text Retrieval Conference*, 2010.

[73] H. Gurulingappa, B. Müller, R. Klinger, H.-T. Mevissen, M. Hofmann-Apitius, J. Fluck, and C. Friedrich, "Patent retrieval in chemistry based on semantically tagged named entities," in *Proceedings of Text Retrieval Conference*, 2009.

[74] J. Guyot, G. Falquet, and K. Benzineb, "UniGE experiments on prior art search in the field of patents," in *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, Springer, pp. 502–507, 2010.

[75] A. Hanbury, N. Bhatti, M. Lupu, and R. Mörzinger, "Patent image retrieval: A survey," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 3–8, 2011.

[76] A. Hanbury, V. Zenz, and H. Berger, "1st international workshop on advances in patent information retrieval (AsPIRe'10)," *SIGIR Forum*, vol. 44, pp. 19–22, 2010.

[77] P. Hansen, "The information seeking and retrieval process at the Swedish patent office: Moving from lab-based to real life work-task environment," in *Proceedings of Workshop on Patent Retrieval*, 2000.

[78] D. Harman, *Information Retrieval Evaluation.* Morgan & Claypool Publishers, 2011.

[79] C. G. Harris, R. Arens, and P. Srinivasan, "Comparison of IPC and USPC classification systems in patent prior art searches," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 27–32, 2010.

[80] C. G. Harris, R. Arens, and P. Srinivasan, "Using classification code hierarchies for patent prior art searches," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, pp. 287–304, 2011.

[81] C. G. Harris, S. Foster, R. Arens, and P. Srinivasan, "On the role of classification in patent invalidity searches," in *Proceedings of Workshop on Patent Information Retrieval*, pp. 29–32, 2009.

[82] Z.-L. He and M. Deng, "The evidence of systematic noise in non-patent references: A study of New Zealand companies' patents," *Scientometrics*, vol. 72, pp. 149–166, 2007.

[83] B. Herbert, G. Szarvas, and I. Gurevych, "Prior art search using international patent classification codes and all-claims-queries," in *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, Springer, pp. 452–459, 2010.

[84] J. D. Holliday and P. Willet, "Representation and searching of chemical-structure information in patents," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, 2011.

[85] V. Hristidis, E. Ruiz, A. Hernández, F. Farfán, and R. Varadarajan, "Patentssearcher: A novel portal to search and explore patents," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 33–38, 2010.

[86] B. Huet, G. Guarascio, N. J. Kern, and B. Mérialdo, "Relational skeletons for retrieval in patent drawings," in *Proceedings of International Conference on Image Processing*, pp. 737–740, 2001.

[87] D. Hull, S. Aït-Mokhtar, M. Chuat, A. Eisele, E. Gaussier, G. Grefenstette, P. Isabelle, C. Samuelsson, and F. Segond, "Language technologies and patent search and classification," *World Patent Information*, vol. 23, pp. 265–268, 2001.

[88] N. Ide and K. Suderman, "Integrating linguistic resources: The american national corpus model," in *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, 2006.

[89] H. Itoh, H. Mano, and Y. Ogawa, "Term distillation in patent retrieval," in *Proceedings of ACL Workshop on Patent Corpus Processing*, 2003.

[90] M. Iwayama, A. Fuji, and N. Kando, "Overview of classification subtask at NTCIR-6 patent retrieval task," in *Proceedings of NII Test Collection for IR Systems-6*, 2007.

[91] M. Iwayama, A. Fujii, and N. Kando, "Overview of classification subtask at NTCIR-5 patent retrieval task," in *Proceedings of NII Test Collection for IR Systems-5*, Tokyo, Japan, 2005.

[92] M. Iwayama, A. Fujii, N. Kando, and Y. Marukawa, "An empirical study on retrieval models for different document genres: patents and newspaper articles," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 251–258, 2003.

[93] M. Iwayama, A. Fujii, N. Kando, and A. Takano, "Overview of patent retrieval task at NTCIR-3," in *Proceedings of ACL Workshop on Patent Corpus Processing*, 2003.

[94] A. Järvelin, G. Eriksson, P. Hansen, T. Tsikrika, A. G. S. de Herrera, M. Lupu, M. Gäde, V. Petras, S. Rietberger, M. Braschler, and R. Berendsen, "Deliverable 2.2 revised specification of the evaluation tasks," Technical Report, PROMISE Network of Excellence, 2012.

[95] C. Jochim, C. Lioma, and H. Schütze, "Expanding queries with term and phrase translations in patent retrieval," in *Multidisciplinary Information Retrieval*, Springer, pp. 16–29, 2011.

[96] H. Joho and M. Sanderson, "Document frequency and term specificity," in *RIAO: Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, 2007.

[97] N. Kando and M.-K. Leong, "Workshop on patent retrieval (SIGIR 2000 workshop report)," *SIGIR Forum*, vol. 34, pp. 28–30, 2000.

[98] I.-S. Kang, S.-H. Na, J. Kim, and J.-H. Lee, "Cluster-based patent retrieval," *Information Processing and Management*, vol. 43, pp. 1173–1182, September 2007.

[99] N. Kemp and M. Lynch, "Extraction of information from text of chemical patents. 1. identification of specific chemical names," *Journal of Chemical Information and Computer Sciences*, vol. 38, 1998.

[100] Y. Kim, J. Seo, and W. B. Croft, "Automatic Boolean query suggestion for professional search," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 825–834, 2011.

[101] K. Kishida, "Pseudo relevance feedback method based on Taylor expansion of retrieval function in NTCIR-3 patent retrieval task," in *Proceedings of the ACL Workshop on Patent Corpus Processing*, 2003.

[102] K. Kishida, K.-H. Chen, S. Lee, H.-H. Chen, N. Kando, K. Kuriyama, S. H. Myaeng, and K. Eguchi, "Cross-lingual information retrieval (CLIR) task at the NTCIR workshop 3," *SIGIR Forum*, vol. 38, pp. 17–20, July 2004.

[103] I. Klampanos, H. Azzam, and T. Roelleke, "A case for probabilistic logic for scalable patent retrieval," in *Proceedings of Workshop on Patent Information Retrieval*, 2009.

[104] R. Klinger, C. Kolářik, J. Fluck, M. Hofmann-Apitius, and C. M. Friedrich, "Detection of iupac and iupac-like chemical names," in *Proceedings of International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 2008.

[105] S. Koch, H. Bosch, M. Giereth, and T. Ertl, "Iterative integration of visual insights during scalable patent search and analysis," *Transactions on Visualization and Computer Graphics*, vol. 17, 2011.

[106] C. H. A. Koster, "Text mining for intellectual property," in *Proceedings of Dutch-Belgian Information Retrieval Workshop (DIR)*, 2010.

[107] C. H. A. Koster, M. Seutter, and J. Beney, "Multi-classification of patent applications with Winnow," in *Perspectives of System Informatics*, Springer, pp. 546–555, 2004.

[108] A. Krishnan, A. F. Cardenas, and D. Springer, "Search for patents using treatment and causal relationships," in *Proceedings of Workshop on Patent Information Retrieval*, New York, NY, USA, pp. pp. 1–10, 2010.

[109] J.-C. Lamirel, S. A. Shehabi, M. Hoffmann, and C. Francois, "Intelligent patent analysis through the use of a neural network: Experiment of multi-viewpoint analysis with the multisom model," in *Proceedings of the ACL Workshop on Patent Corpus Processing*, 2003.

[110] L. S. Larkey, "A patent search and classification system," in *Proceedings of ACM Conference on Digital Libraries*, pp. 179–187, 1999.

[111] A. Leach and V. Gillet, *An Introduction to Chemoinformatics.* Springer, 2007.

[112] G. Leech, "100 million words of english: The british national corpus," *Language Research*, vol. 28, 1992.

[113] J. Leveling, W. Magdy, and G. J. Jones, "An investigation of decompounding for cross-language patent search," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1169–1170, 2011.

[114] L. Li and C. L. Tan, "Associating figures with descriptions for patent documents," in *Proceedings of the IAPR International Workshop on Document Analysis Systems*, pp. 385–392, 2010.

[115] Y. Li and J. Shawe-Taylor, "Advanced learning algorithms for cross-language patent retrieval and classification," *Information Processing and Management*, vol. 43, pp. 1183–1199, 2007.

[116] Y.-R. Li, L.-H. Wang, and C.-F. Hong, "Extracting the significant-rare keywords for patent analysis," *Expert Systems with Applications*, vol. 36, 2009.

[117] W. A. Lise, "An investigation of terminology and syntax in Japanese and US patents and the implications for the patent translator," http://www.lise.jp/patsur.html, 2011. Last visited: September, 4, 2012.

[118] J. List, "How drawings could enhance retrieval in mechanical and device patent searching," *World Patent Information*, vol. 29, pp. 210–218, 2007.

[119] P. Lopez, "Automatic extraction and resolution of bibliographical references in patent documents," in *Advances in Multidisciplinary Retrieval*, Springer Berlin/Heidelberg, pp. 120–135, 2010.

[120] P. Lopez and L. Romary, "Experiments with citation mining and key-term extraction for prior art search," in *Cross-Language Evaluation Forum (Notebook Papers/Labs/Workshop)*, 2010.

[121] P. Lopez and L. Romary, "Patatras: Retrieval model combination and regression models for prior art search," in *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, Springer, 2010.

[122] B. Lu, B. K. Tsou, T. Jiang, O. Y. Kwong, and J. Zhu, "Mining large-scale parallel corpora from multilingual patents: An English-Chinese example and its application to SMT," in *Proceedings of the CIPS-SIGHAN Joint Conference on Chinese Language Processing*, pp. 79–86, 2010.

[123] X. Lu, S. Kataria, W. J. Brouwer, J. Z. Wang, P. Mitra, and C. L. Giles, "Automated analysis of images in documents for intelligent document search," *International Journal on Document Analysis and Recognition*, vol. 12, pp. 65–81, 2009.

[124] M. Lupu, "The status of retrieval evaluation in the patent domain," in *Proceedings of Workshop on Patent Information Retrieval*, 2011.

[125] M. Lupu, "Patolympics — an infrastructure for interactive evaluation of patent retrieval tools," in *Proceedings of CIKM Workshop on Data infrastructures for Supporting Information Retrieval Evaluation (DESIRE)*, 2011.

[126] M. Lupu, Z. Jiashu, J. Huang, H. Gurulingappa, I. Filipov, and J. Tait, "Overview of the trec 2011 chemical ir track," in *Proceedings of Text Retrieval Conference*, 2011.

[127] M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds., *Current Challenges in Patent Information Retrieval. Information Retrieval Series.* Springer, 2011.

[128] M. Lupu, F. Piroi, and A. Hanbury, "Aspects and analysis of patent test collections," in *Proceedings of Workshop on Patent Information Retrieval*, 2010.

[129] M. Lupu, F. Piroi, J. Huang, J. Zhu, and J. Tait, "Overview of the trec chemical ir track," in *Proceedings of Text Retrieval Conference*, 2009.

[130] M. Lupu, R. Schuster, R. Mörzinger, F. Piroi, T. Schleser, and A. Hanbury, "Patent images — a glass-encased tool: Opening the case," in *Proceedings of International Conference on Knowledge Management and Knowledge Technologies*, pp. 16:1–16:8, 2012.

[131] W. Magdy and G. Jones, "A study on query expansion methods for patent retrieval," in *Proceedings of Workshop on Patent Information Retrieval*, 2011.

[132] W. Magdy and G. J. Jones, "Pres: A score metric for evaluating recall-oriented information retrieval applications," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 611–618, 2010.

[133] W. Magdy and G. J. F. Jones, "Examining the robustness of evaluation metrics for patent retrieval with incomplete relevance judgements," in *Proceedings of the 2010 International Conference on Multilingual and Multimodal Information Access Evaluation: Cross-language Evaluation Forum*, Berlin, Heidelberg, pp. 82–93, 2010.

[134] W. Magdy, J. Leveling, and G. J. F. Jones, "Exploring structured documents and query formulation techniques for patent retrieval," in *Proceedings of the Cross-language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments*, Springer-Verlag, pp. 410–417, 2009.

[135] P. Mahdabi, L. Andersson, A. Hanbury, and F. Crestani, "Report on the CLEF-IP 2011 experiments: Exploring patent summarization," in *Cross-Language Evaluation Forum (Notebook Papers/LABs/Workshops)*, 2011.

[136] P. Mahdabi, M. Keikha, S. Gerani, M. Landoni, and F. Crestani, "Building queries for prior-art search," in *Multidisciplinary Information Retrieval*, Springer-Verlag, pp. 3–15, 2011.

[137] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognition*, vol. 36, pp. 1725–1736, 2003.

[138] L. Molà, *The Silk Industry of Renaissance Venice*. JHU Press, 2000.

[139] A. Moldovan, R. I. Bot, and G. Wanka, "Latent semantic indexing for patent documents," *International Journal of Applied Mathematics and Computer Science*, vol. 15, 2005.

[140] G. Moradei and P. C. Contessini, "An evaluation of some semantic tools for simple patent searching," in *Proceedings of Patent Information Conference*, 2011.

[141] N. Morey, "Global business solutions for patent prosecution," in *Proceedings of the Symposium of Intellectual Property Authorities*, WIPO, 2011.

[142] R. Mörzinger, A. Horti, G. Thallinger, N. Bhatti, and A. Hanbury, "Classifying patent images," in *Cross-Language Evaluation Forum (Notebook papers/Labs/Workshop)*, 2011.

[143] S. Mukherjea and B. Bamba, "Biopatentminer: An information retrieval system for biomedical patents," in *VLDB '04: Proceedings of International Conference on Very Large Data Bases*, VLDB Endowment, pp. 1066–1077, 2004.

[144] B. Müller, R. Klinger, H. Gurulingappa, H.-T. Mevissen, M. Hofmann-Apitius, J. Fluck, and C. Friedrich, "Abstracts versus full texts and patents: A quantitative analysis of biomedical entities," in *Advances in Multidisciplinary Retrieval*, Springer, pp. 152–165, 2010.

[145] H. Nanba, A. Fujii, M. Iwayama, and T. Hashimoto, "Overview of the patent mining task at the NTCIR-7 workshop," in *Proceedings of NII Test Collection for IR Systems-7*, 2008.

[146] H. Nanba, A. Fujii, M. Iwayama, and T. Hashimoto, "Overview of the patent retrieval task at the NTCIR-8 workshop," in *Proceedings of NII Test Collection for IR Systems-8*, 2010.

[147] H. Nanba, H. Kamaya, T. Takezawa, M. Okumura, A. Shinmori, and H. Tanigawa, "Automatic translation of scholarly terms into patent terms," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, pp. 373–388, 2011.

[148] H. Nanba, S. Mayumi, and T. Takezawa, "Automatic construction of a bilingual thesaurus using citation analysis," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 25–30, 2011.

[149] J.-Y. Nie, *Cross-Language Information Retrieval.* Morgan & Claypool Publishers, 2010.

[150] N. Oostdijk, E. D'hondt, H. van Halteren, and S. Verberne, "Genre and domain in patent texts," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 39–46, 2010.

[151] N. Oostdijk, S. Verberne, and C. Koster, "Constructing a broad-coverage lexicon for text mining in the patent domain," in *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, 2010.

[152] M. Osborn, T. Strzalkowski, and M. Marinescu, "Evaluating document retrieval in patent database: A preliminary report," in *Proceedings of Conference on Information and Knowledge Management*, pp. 216–221, 1997.

[153] P. Parapatics and M. Dittenbach, "Patent claim decomposition for improved information extraction," in *Proceedings of Workshop on Patent Information Retrieval*, ACM, pp. 33–36, 2009.

[154] J. Perez-Iglesias, A. Rodrigo, and V. Fresno, "Using bm25f and kld for patent retrieval," in *Cross-Language Evaluation Forum (Notebook Papers/LABs/Workshops)*, 2010.

[155] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007 (CVPR '07)*, pp. 1–8, 2007.

[156] F. Piroi, M. Lupu, A. Hanbury, A. Sexton, W. Magdy, and I. Filippov, "Clef-ip 2012: Retrieval experiments in the intellectual property domain," in *Cross-Language Evaluation Forum (Notebook papers/Labs/Workshop)*, 2012.

[157] F. Piroi, M. Lupu, A. Hanbury, and V. Zenz, "Clef-ip 2011: Retrieval in the intellectual property domain," in *Cross-Language Evaluation Forum (Notebook Papers/Labs/Workshop)*, 2011.

[158] F. Piroi and J. Tait, "Clef-ip 2010: Retrieval experiments in the intellectual property domain," in *Cross-Language Evaluation Forum (Notebook papers/Labs/Workshop)*, 2010.

[159] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," *Foundations and Trends in Information Retrieval*, vol. 3, 2009.

[160] S. E. Robertson, "The probability ranking principle in ir," *Journal of Documentation*, vol. 33, 1977.

[161] G. Roda, J. Tait, F. Piroi, and V. Zenz, "Clef-ip 2009: Retrieval experiments in the intellectual property domain," in *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, Springer, pp. 385–409, 2010.

[162] J. Rousu, C. Saunders, S. Szedmák, and J. Shawe-Taylor, “Kernel-based learning of hierarchical multilabel classification models,” *Journal of Machine Learning Research*, vol. 7, pp. 1601–1626, 2006.

[163] J. Ryley, “Latent semantic indexing for patent information,” http://cogprints.org/5710/1/ryley.html, 2007. Last accessed: September, 4, 2012.

[164] M. Sahlgren, P. Hansen, and J. Karlgren, “English-japanese cross-lingual query expansion using random indexing of aligned bilingual text data,” in *Proceedings of NII Test Collection for IR Systems*, 2002.

[165] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing and Management*, vol. 24, pp. 513–523, 1988.

[166] M. Sanderson, “Test collection based evaluation of information retrieval systems,” *Foundations and Trends in Information Retrieval*, vol. 4, 2010.

[167] M. Sanderson, M. L. Paramita, P. Clough, and E. Kanoulas, “Do user preferences and evaluation measures line up?,” in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2010.

[168] I. Schellner, “Japanese File Index classification and f-terms,” *World Patent Information*, vol. 24, pp. 197–201, 2002.

[169] M. W. Seeger, “Cross-validation optimization for large scale hierarchical classification kernel methods,” in *Proceedings of Neural Information Processing Systems (NIPS)*, pp. 1233–1240, 2007.

[170] S. Sheremetyeva, “Natural language analysis of patent claims,” in *Proceedings of ACL Workshop on Patent Corpus Processing*, 2003.

[171] S. Sheremetyeva and S. Nirenburg, “Knowledge elecitation for authoring patent claims,” *IEEE Computer*, vol. 29, pp. 57–63, 1996.

[172] S. Sheremetyeva, S. Nirenburg, and I. Nirenburg, “Generating patent claims from interactive input,” in *Proceedings of International Workshop on Natural Language Generation (INLG’96)*, pp. 61–70, 1996.

[173] A. Shinmori, M. Okumura, Y. Marukawa, and M. Iwayama, “Patent claim processing for readability — structure analysis and term explanation,” in *Proceedings of The ACL Workshop on Patent Corpus Processing*, 2003.

[174] P. Sidiropoulos, S. Vrochidis, and I. Kompatsiaris, “Content-based binary image retrieval using the adaptive hierarchical density histogram,” *Pattern Recognition*, vol. 44, pp. 739–750, 2011.

[175] A. Singh, S. Hallihosur, and L. Rangan, “Changing landscape in biotechnology patenting,” *World Patent Information*, vol. 31, pp. 219–225, 2009.

[176] A. Smeaton, “Nlp and ir,” in *European Summer School on Information Retrieval*, 1995.

[177] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, 2000.

[178] H. Smith, “Automation of patent classification,” *World Patent Information*, vol. 24, pp. 269–271, 2002.

[179] S. Taduri, G. T. Lau, K. H. Law, H. Yu, and J. P. Kesan, "Developing an ontology for the U.S. patent system," in *Proceedings of Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pp. 157–166, 2011.

[180] S. Taduri, G. T. Lau, K. H. Law, H. Yu, and J. P. Kesan, "An ontology-based interactive tool to search documents in the u.s. patent system," in *Proceedings of the Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, New York, NY, USA, pp. 329–330, 2011.

[181] S. Taduri, H. Yu, G. Lau, K. Law, and J. Kesan, "Developing a comprehensive patent related information retrieval tool," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 6, pp. 1–16, August 2011.

[182] T. Takaki, A. Fuji, and T. Ishikawa, "Associative document retrieval by query subtopic analysis and its application to invalidity patent search," in *Proceedings of Conference on Information and Knowledge Management*, 2004.

[183] T. Tang, N. Craswell, D. Hawking, K. Griffiths, and H. Christensen, "Quality and relevance of domain-specific search: A case study in mental health," *Information Retrieval*, vol. 9, pp. 207–225, 2006.

[184] J. Tangelder and R. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools and Applications*, vol. 39, pp. 441–471, 2008.

[185] W. Thielemann, "Ocr errors in patent full-text documents," in *Proceedings of the IRF Symposium*, 2007.

[186] D. Tikk, G. Biró, and A. Törcsvári, "A hierarchical online classifier for patent categorization," in *Emerging Technologies of Text Mining: Techniques and Applications*, (H. A. do Prado and E. Ferneda, eds.), IGI Global, pp. 244–267, 2007.

[187] A. Tiwari and V. Bansal, "Patseek: Content based image retrieval system for patent database," in *Proceedings of International Conference on Electronic Business (ICEB)*, pp. 1167–1171, 2004.

[188] K. Tombre, "Analysis of engineering drawings: State of the art and challenges," in *Graphics Recognition — Algorithms and Systems*, Springer, pp. 257–264, 1998.

[189] A. Trippe and I. Ruthven, "Evaluating real patent retrieval effectiveness," in *Current Challenges in Patent Information Retrieval*, (M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, eds.), Springer, 2011.

[190] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," *Information Processing and Management*, vol. 43, pp. 1216–1247, September 2007.

[191] Y.-H. Tseng and Y.-J. Wu, "A study of search tactics for patentability search: A case study on patent engineers," in *Proceedings of Workshop on Patent Information Retrieval*, pp. 33–36, 2008.

[192] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," in *Proceedings of International Conference on Machine Learning*, pp. 104–111, 2004.

[193] J. Urbain and O. Frieder, "Trec chemical ir track 2009: A distributed dimensional indexing model for chemical patent search," in *Proceedings of Text Retrieval Conference*, 2009.

[194] USPTO, "Acceptance, processing, use and dissemination of chemical and three-dimensional biological structural data in electronic format," *Biotechnology Law Report*, vol. 24, pp. 638–644, 2005.

[195] A. T. Valko and A. P. Johnson, "Clide pro: The latest generation of CLiDE, a tool for optical chemical structure recognition," *Journal of Chemical Information and Modeling*, vol. 49, pp. 780–787, 2009.

[196] S. Verberne and E. D'hondt, "Prior art retrieval using the claims section as a bag of words," in *Proceedings of the Cross-language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments*, Springer, pp. 497–501, 2009.

[197] S. Verberne, E. D'hondt, N. Oostdijk, and C. Koster, "Quantifying the challenges in parsing patent claims," in *Proceedings of Workshop on Advances in Patent Information Retrieval (AsPIRe)*, 2010.

[198] S. Verberne, M. Vogel, and E. D'hondt, "Patent classification experiments with the linguistic classification system lcs," in *Working Notes of the CLEF Labs*, 2010.

[199] M. Verma and V. Varma, "Applying key phrase extraction to aid invalidity search," in *Proceedings of International Conference on Artificial Intelligence and Law*, pp. 249–255, 2011.

[200] S. V. Vishwanathan, N. N. Schraudolph, and A. J. Smola, "Step size adaptation in reproducing kernel Hilbert space," *Journal of Machine Learning Research*, vol. 7, pp. 1107–1133, 2006.

[201] S. Vrochidis, S. Papadopoulos, A. Moumtzidou, P. Sidiropoulos, E. Pianta, and I. Kompatsiaris, "Towards content-based patent image retrieval: A framework perspective," *World Patent Information*, vol. 32, pp. 94–106, 2010.

[202] M. Z. Wanagiri and M. Adriani, "Prior art retrieval using various patent document fields contents," in *Cross-Language Evaluation Forum (Notebook Papers/LABs/Workshops)*, 2010.

[203] X. Wang, X. Zhang, and S. Xu, "Patent co-citation networks of fortune 500 companies," *Scientometrics*, vol. 88, pp. 761–770, 2011.

[204] L. Wanner, R. Baeza-Yates, S. Brügmann, J. Codina, B. Diallo, E. Escorsa, M. Giereth, Y. Kompatsiaris, S. Papadopoulos, E. Pianta, G. Piella, I. Puhlmann, G. Rao, M. Rotard, P. Schoester, L. Serafini, and V. Zervaki, "Towards content-oriented patent document processing," *World Patent Information*, vol. 30, pp. 21–33, 2008.

[205] H. Wongel, "The reform of the IPC — consequences for the users," *World Patent Information*, vol. 27, pp. 227–231, 2005.

[206] World Intellectual Property Organisation, "Glossary of terms concerning industrial property information and documentation, appendix iii to part 10, of the wipo handbook on industrial property information and documentation," WIPO Publication No. CD208, 2003.

[207] World Intellectual Property Organisation, "Wipo handbook on intellectual property, chapter 5," http://www.wipo.int/export/sites/www/about-ip/en/iprm/pdf/ch5.pdf, last visited August, 2012.

[208] X. Xue and W. B. Croft, "Automatic query generation for patent search," in *Proceedings of Conference on Information and Knowledge Management*, pp. 2037–2040, 2009.

[209] X. Xue and W. B. Croft, "Transforming patents into prior-art queries," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 808–809, 2009.

[210] M. Yang, G. Qiu, J. Huang, and D. Elliman, "Near-duplicate image recognition and content-based image retrieval using adaptive hierarchical geometric centroids," in *International Conference on Pattern Recognition, 2006 (ICPR 2006)*, pp. 958–961, 2006.

[211] S.-Y. Yang and V.-W. Soo, "Comparing the conceptual graphs extracted from patent claims," in *IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC)*, pp. 394–399, 2008.

[212] Y. Yang, L. Akers, T. Klose, and C. Barcelonyang, "Text mining and visualization tools — impressions of emerging capabilities," *World Patent Information*, vol. 30, pp. 280–293, December 2008.

[213] F. Zaccá and M. Krier, "Automatic categorisation applications at the European Patent Office," *World Patent Information*, vol. 24, pp. 187–196, 2002.

[214] V. Zenz, S. Wurzer, M. Dittenbach, and E. Ambrosi, "On the effects of indexing and retrieval models in patent search and the potential of result set merging," in *Proceedings of Workshop on Advances in Patent Information Retrieval (AsPIRe)*, 2010.

[215] L. Zhao and J. Callan, "Formulating simple structured queries using temporal and distributional cues in patents," in *Proceedings of Text Retrieval Conference*, 2009.

[216] L. Zhao and J. Callan, "How to make manual conjunctive normal form queries work in patents search," in *Proceedings of Text Retrieval Conference*, 2011.

[217] Z. Zhiyuan, Z. Juan, and X. Bin, "An outward-appearance patent-image retrieval approach based on the contour-description matrix," in *Proceedings of Japan-China Joint Workshop on Frontier of Computer Science and Technology (FCST)*, IEEE Computer Society, pp. 86–89, 2007.

[218] Y. Zhou and C. L. Tan, "Chart analysis and recognition in document images," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1055–1058, 2001.

[219] G. Zhu, X. Yu, Y. Li, and D. Doermann, "Learning visual shape lexicon for document image content recognition," in *Proceedings of European Conference on Computer Vision (ECCV)*, Springer, pp. 745–758, 2008.