The Information Retrieval Series Volume 37

More information about this series at http://www.springer.com/series/6128

Mihai Lupu • Katja Mayer • Noriko Kando • Anthony J. Trippe
Editors

Current Challenges in Patent Information Retrieval

Second Edition

Springer

Editors
Mihai Lupu
Institute for Software Engineering & Interactive Systems
Vienna University of Technology
Vienna, Austria

Katja Mayer
Research Platform Responsible Research and Innovation in Academic Practice
University of Vienna
Vienna, Austria

Noriko Kando
Information & Society Research Division
National Institute of Informatics
Tokyo, Japan

Anthony J. Trippe
Patinformatics, LLC
Dublin, OH
USA

ISSN 1387-5264
The Information Retrieval Series
ISBN 978-3-662-53816-6 ISBN 978-3-662-53817-3 (eBook)
DOI 10.1007/978-3-662-53817-3

Library of Congress Control Number: 2016963218

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

Patent information retrieval is an economically important activity. Today's economy is becoming increasingly knowledge based and intellectual property in the form of patents plays a vital role in this growth. According to the WIPO IP Statistics Data Center, between 2004 and 2014, the number of patent applications filed worldwide grew by more than 70 %. With the exception of 2009, the year immediately after the economic collapse, every year has shown an increase in the number of filed applications. The number of granted patents worldwide continues to increase, even in 2009, reaching in 2014 1,176,600 grants versus only 625,100 grants in 2004 (an 88 % increase). The substantial increase in patents granted is due, in part, to efforts by patent offices to reduce backlogs as well as the significant growth in the number of patents granted by China and, to a lesser extent in more recent years, by the Republic of Korea. According to these statistics, the total number of patents in force worldwide at the end of 2014 was approximately 10.2 million (WIPO Report 2015). A prior art search might have to cover as many as 100 million patents. By combining data from Ocean Tomo's Intangible Asset Market Value Survey and Standard and Poor's 1200 Index, we can estimate that the global value of patents exceeds US$12 trillion in 2015. In the United States alone, a 2012 study by the Commerce Department found that 'intellectual property intensive industries support at least 40 million jobs and contribute more than US$5 trillion to, or 34.8 percent of, US gross domestic product'.

A patent is a contract between inventors and the state. The inventors must teach the community how to perform the invention and use the techniques they have invented in return for a limited monopoly that gives them a predefined time to exploit the invention and realise its value. Patents are used for many reasons, e.g. to protect inventions, to create value and to monitor competitive activities in a field. Much knowledge is distilled through patents, which is never published elsewhere. Thus patents form an important knowledge resource—e.g. much technical information represented in patents is not represented in scientific literature—and are at the same time important legal documents. In fact, a study done by one of the editors of this volume found that in the chemical domain 95 % of patented substances did not appear in non-patent literature references. In the context of today's drive towards

open innovation, particularly in the European Union and its framework programmes, it seems that patent search should take a more visible role, speeding up knowledge discovery for tackling societal changes.

In the past 15 or 20 years, search technology in general and Web search engines in particular have made tremendous advances. Yet still, we see a considerable gap between the technologies emerging from research labs and in use by major Internet search engines and the systems in day-to-day use by the patent search communities. This gap is unlikely to ever completely disappear, simply for the reasons of corporate practice, whereby only proven systems make their way, through a relatively complicated adoption procedure, to the regular processes of professional searchers. Nevertheless, we have observed in the last 5 years an increasingly active drive towards adoption of the search technology state of the art in several major players in the field.

In 2010, just before the publication of the first edition of this book, a study commissioned by the US Federal National Institute of Standards and Technology (NIST) estimated that since 1991, when the Text Retrieval Conference (TREC) evaluation campaign began, the available information retrieval and search systems have improved 40 % or more in their ability to find relevant documents. And yet the technologies underlying the patent search system were largely unaffected by these changes. Patent searchers generally used the same technology as in the 1980s. Boolean specification of searches and set-based retrieval are still common. Nevertheless, tools have improved over the years, as have the requirements and expectations of the users. Semantic search (under its various interpretations) is now on practically every provider's table. And yet there had not been the kind of revolution in patent search which Google had represented for Web search. Perhaps there will never be a revolution, and we should indeed expect gradual transition to systems that are first well tested in other domains.

This first edition of this book, which appeared in 2011, was part of the development of a joint understanding between IR researchers and IP specialists, understanding that resulted from a series of symposia organised by the Information Retrieval Facility (IRF) in Vienna, Austria, between 2007 and 2011. Its origins lie in the idea of producing post-proceedings for the first IRF symposium. That idea was not fully followed up, in part because of pressure to produce more practical, action-oriented work, and in part because many of the participants felt their approaches were at too early a stage for formal publication. In the course of the following years, it became apparent there really was a demand to produce a volume which was accessible to both the patent search community and the information retrieval research community, to provide a collected and organised introduction to the work and views of the two sides of the emerging patent search research and innovation community and to provide a coherent and organised view of what has been achieved and, perhaps even more significantly, of what remains to be achieved.

A secondary result of the efforts invested by the IRF was an uptake in the academic community of the patent search problem. While the IRF stopped operating in 2011, research continued across the world, with the term 'patent search' being indelibly added to the Call for Papers of major conferences in the field. Since

that time, a number of PhD theses have been written on the topic by outstanding young researchers. We found that a second edition of the book was indeed needed to showcase these as well as the other research advances of the past half-decade.

At the same time, this second edition revisits some of the original chapters from the first edition, as it maintains the original objective to allow the IR researchers to better understand why the patent domain has different needs and what it means in practice. Furthermore, it is our hope that these two books will also be a valuable resource for IP professionals in learning about current approaches of IR in the patent domain. It has often been difficult to reconcile the focus on useful technological innovation from the IP community, with the demands for scientific rigour, and to proceed on the basis of sound empirical evidence, which is such an important feature of IR (in contrast to some other areas of computer science).

Moreover, patent search is an inherently multilingual and multinational topic: the novelty of a patent may be dismissed by finding a document describing the same idea in any language anywhere in the world. Patents are complex legal documents, even less accessible than the scientific literature. These are just some of the characteristics of the patent system which make it an important challenge for the search, information retrieval and information access communities.

Even more than the first edition, the second edition of the book has had a lengthy and difficult gestation: the list of authors has been revised many times as a result of changes in institutional, occupational and private circumstances. Although we, the editors, do feel we have succeeded in producing a volume which will provide important perspectives of the issues affecting patent search research and innovation at the time of writing, as well as a useful, brief introduction to the outlook and literature of the community accessible to its members, regardless of their background, there will always be some areas that are not covered, mostly because there is, at this time, insufficient research on the topic. Most importantly here are the applications of new statistical semantics methods on the patent domain. While these are extremely popular methods in current research, there exist only inconclusive studies on their application in the patent field.

On the other hand, we are very happy to have managed to include in this edition something we had missed in the previous one: a chapter on NTCIR, the first of the evaluation campaigns to focus seriously on patents.

Several of the chapters have been written jointly by intellectual property and information retrieval experts. Members of both communities with a background opposite to the primary author have reviewed the chapters. It has not always been easy to reconcile their differing viewpoints: we must thank them for taking the time to resolve their differences and for taking the opportunity to exchange their knowledge across fields and disciplinary mindsets and to engage in a mutual discourse that will hopefully foster understanding in the future.

Finally, we would like to thank the IRF for making the first edition possible and triggering much research in the area; the publisher, Springer, and in particular Ralf Gerstner, for the patience with which he accepted the numerous delays; as well as the external reviewers who read each chapter and provided the authors with valuable advice.

The editors are very grateful to the following persons, who agreed to review the manuscripts of the two editions of this book:

Stephen Adams, Linda Andersson, Leif Azzopardi, Geetha Basappa, John M. Barnard, Shariq Bashir, Helmut Berger, Katrien Beuls, Ted Briscoe, Ben Carterette, Suleyman Cetintas, Chen Chaomei, Paul Clough, Bruce Croft, Szabolcs Csepregi, Barrou Diallo, Ramona Enache, Nicola Ferro, Árpád Figyelmesi, Karl A. Froeschl, Norbert Fuhr, Eric Gaussier, Julio Gonzalo, Jacques Guyot, Allan Hanbury, Christopher G. Harris, Ilkka Havukkala, Bruce Hedin, Peter Johnson, Cornelis H.A. Koster, Mounia Lalmas, Aldo Lipani, Patrice Lopez, Teresa Loughbrough, Ilya Markov, Marie-Francine Moens, Anastasia Moumtzidou, Roland Mörzinger, Henning Müller, Masaaki Nagata, Iadh Ounis, Doug Oard, Florina Piroi, Keith van Rijsbergen, Patrick Ruch, Georg Thallinger, Philip Tetlow, Henk Thomas, Ingo Thon, Steve Tomlinson, Suzan Verberne, Ellen M. Voorhees, Jianqiang Wang, Peter Willett and Christa Womser-Hacker

Vienna, Austria — Mihai Lupu
Vienna, Austria — Katja Mayer
Tokyo, Japan — Noriko Kando
Dublin, OH, USA — Anthony J. Trippe
August 2016

Contents

Contributors

Doreen Alberts Theravance, Inc., South San Francisco, CA, USA

Linda Andersson TU Wien, Vienna, Austria

Cynthia Barcelon Yang Bristol Myers Squibb, Princeton, NJ, USA

Shariq Bashir Department of Computer Science, Mohammad Ali Jinnah University, Islamabad, Pakistan

Lou Boves Radboud University Nijmegen, Nijmegen, The Netherlands

Gabriela Csurka Xerox Research Centre Europe, Meylan, France

Eva D'hondt LIMSI-CNRS UPR 3251, Orsay, France

Dominic DeMarco DeMarco Intellectual Property, LLC, South Arlington, VA, USA

Barrou Diallo European Patent Office, Rijswijk, The Netherlands

Geoff M. Downs Digital Chemistry Ltd., Sheffield, UK

Igor V. Filippov VIF Innovations, LLC, Rockville, MD, USA

Denise Fobare-DePonio Amgen, Thousand Oaks, CA, USA

Atsushi Fujii Department of Computer Science, Tokyo Institute of Technology, Tokyo, Japan

Allan Hanbury TU Wien, Vienna, Austria

Bruce Hedin San Francisco, CA, USA

John D. Holliday Information School, The University of Sheffield, Sheffield, UK

Makoto Iwayama Hitachi, Ltd., Tokyo, Japan

Noriko Kando National Institute of Informatics, Tokyo, Japan

Luciano Kay Center for Nanotechnology in Society, University of California, Santa Barbara, CA, USA

Ken Koubek Koubek Information Consulting Services, Wilmington, DE, USA

Josep Lladós Dept. Ciències de la Computació, Computer Vision Center, Bellaterra, Spain

Mihai Lupu TU Wien, Vienna, Austria

Nils Newman Intelligent Information Services Corporation, Atlanta, GA, USA

Douglas W. Oard University of Maryland, College Park, MD, USA

Nelleke Oostdijk Radboud University Nijmegen, Nijmegen, The Netherlands

Florina Piroi TU Wien, Vienna, Austria

Alan L. Porter School of Public Policy, Georgia Institute of Technology, Norcross, GA, USA

Ismael Ràfols Ingenio (CSIC-UPV), Universitat Politècnica de València, Valencia, Spain

SPRU, University of Sussex, Brighton, UK

Andreas Rauber TU Wien, Vienna, Austria

Suzanne Robins Patent Information Services, Inc., Westborough, MA, USA

Matthew Rodgers CPA Global, Alexandria, VA, USA

Marçal Rusiñol Dept. Ciències de la Computació, Computer Vision Center, Bellaterra, Spain

Ian Ruthven Department of Computer and Information Sciences, University of Strathclyde, Glasgow, UK

Michail Salampasis Alexander Technology Educational Institute (ATEI), Thessaloniki, Greece

Alan P. Sexton University of Birmingham, Birmingham, UK

Edlyn Simmons Simmons Patent Information Service, Fort Mill, SC, USA

Veronika Stefanov TU Wien, Vienna, Austria

John Tinsley Iconic Translation Machines, Ltd., Dublin, Ireland

Stephen Tomlinson Open Text Corporation, Ottawa, ON, Canada

Anthony Trippe Patinformatics, LLC, Dublin, OH, USA

Suzan Verberne Radboud University Nijmegen, Nijmegen, The Netherlands

Peter Willett Information School, The University of Sheffield, Sheffield, UK

Jan Youtie Enterprise Innovation Institute & School of Public Policy, Atlanta, GA, USA

http://www.springer.com/978-3-662-53816-6

Current Challenges in Patent Information Retrieval

Lupu, M.; Mayer, K.; Kando, N.; Trippe, A.J. (Eds.)

2017, XIII, 455 p. 88 illus., 44 illus. in color., Hardcover
ISBN: 978-3-662-53816-6