From industry to scholarly communication: biometric literature over time

Hyoungjoo Park, University of Wisconsin

Abstract
This study investigated the influence between industry and scholarly communication with the comparative analysis of patterns using a large scale dataset on biometric information. To identify whether patterns of cutting edge development in industry affect, are affected by, and/or are studied in parallel with scholarly communication over time, trending topics, word frequency occurrences, and temporal burst detection over time were conducted to assess prominent terms. Patents published in USPTO from 1790 to 2014 were analyzed to represent industry, and published documents such as peer-reviewed journals, conference proceedings and ebooks from both Thomson Reuters Web of Science and IEEE Xplore were analyzed to represent scholarly communication. The results of this study revealed that (1) there are matching trends in the number of publications, (2) transformation points in time are detected using the temporal burst analysis, and (3) patterns of cutting edge developments in industry might not affect, be affected by, and/or develop in parallel with scholarly communication over time in biometric literature.
Keywords: scholarly communication; informetrics; industry; biometric information
Citation: Park, H. (2015). From industry to scholarly communication: biometric literature over time. In *iConference 2015 Proceedings*.

Acknowledgements: I would like to thank Professor Dietmar Wolfram for his important feedback and suggestions of this research.
Contact: hyoungjoo.park@gmail.com

1 Introduction

This study investigated patterns between industry and scholarly communication regarding biometric literature. This research analyzes whether cutting edge developments in industry affects, are affected by, and/or are studied in parallel with scholarly communication over time by analyzing trending topics, word frequency occurrences, and temporal burst detection over time. This research focuses on biometric information because of the increasing needs of biometric information such as fingerprints, facial recognition, and iris recognition to authenticate the identity of people. In border control, biometric information with electronic passports (aka. biometric passports) is required to be captured and to authenticate the identity of travelers using biometric features such as facial recognition and fingerprints for international visitors to enter countries such as the United States, European Union, (EU) and Japan. Other examples might be the use of fingerprints to unlock mobile phones, and a person's iris or weight requirements to enter data centers of corporations.

The primary research question undertaken is whether patterns of cutting edge developments in industry interacts with scholarly communication in biometric literature over time. Based on the patterns found, this research comparatively analyzes whether one affects, is affected by, or matches in parallel with the other.

2 Methodology

Data collection was conducted for both the industry sector and academic sector. As representatives of industry, patent documents from United States Patent and Trademark Office (USPTO) (http://www.uspto.gov) was selected and as representatives of scholarly communication, peer-reviewed journal, proceedings, and books were selected from both Web of Science (WoS) and IEEE Xplore. One thousand and forty-seven patents which have "biometric" in the title from 1790 to 2014 were collected with keyword search from the USPTO website as of September 21, 2014 by manually collecting the title, filed date, and abstract of each patent. Three thousands and nine hundred and thirty-five records which contain "biometric" in the title from all database were collected from both WoS and IEEE Xplore as of October 7, 2014. One thousand seven hundred and thirty-five results from WoS and 2,000 from 2,278 retrieved results without researchers' intervention were analyzed. For the data analysis, the title, publication year and abstract were collected; however, only IEEE Xplore abstracts were collected because WoS does not export abstracts automatically. Word frequency occurrences in the text were revealed for the title and abstract using Voyant (http://voyant-tools.org) tool. Total word occurrence counts excluding stop words were analyzed, as well as temporal burst detection over time as described by Chiappalone et al. (2005) using the Sci2 tool.

3 Findings

3.1 Leading countries

Data analysis of patents shows that 91 percent of the patents on biometrics are published in the United States (48%), European Union (15%), Japan (11%), Republic of Korea (10%), and China (7%) as reported by World Intellectual Property Organization (http://patentscope.wipo.int). Also, data analysis of patents shows that 86 percent of the patents on biometrics are published in the United States (69%), Japan (13%), Germany (2%), and France (2%) as reported by USPTO. This might explain the situations that the United States, European Union and Japan are countries that require strict policy for biometric information in border security control for visitors.

3.2 Matching trends and word frequency occurrences

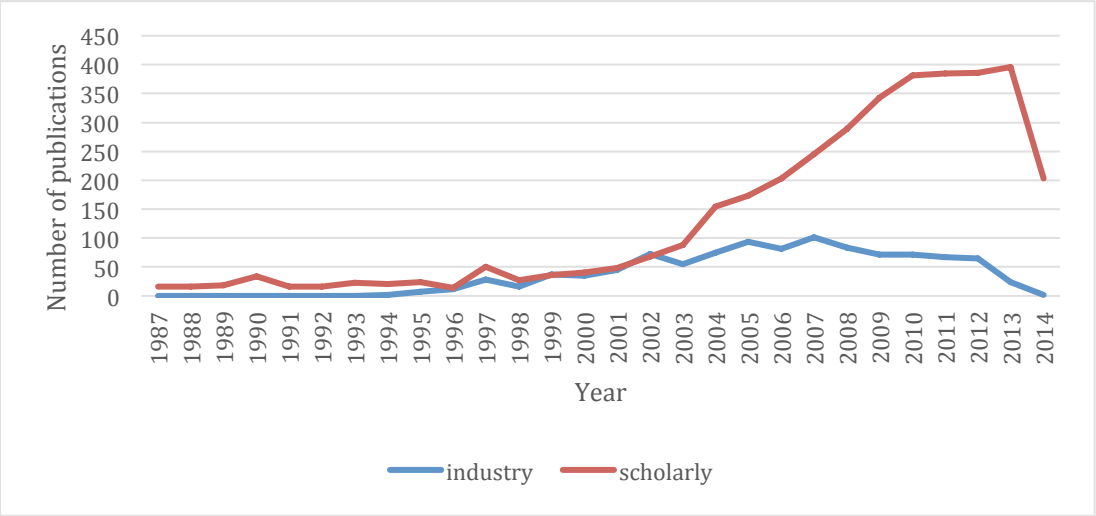

Figure 1. Matching trends over time between industry and scholarly communication

Figure 1 shows the results of the comparison for numbers of publications and patents that contain “biometric” in the title. Biometric information started to be published from 1987 and matches parallel with each other rather than affecting the other.

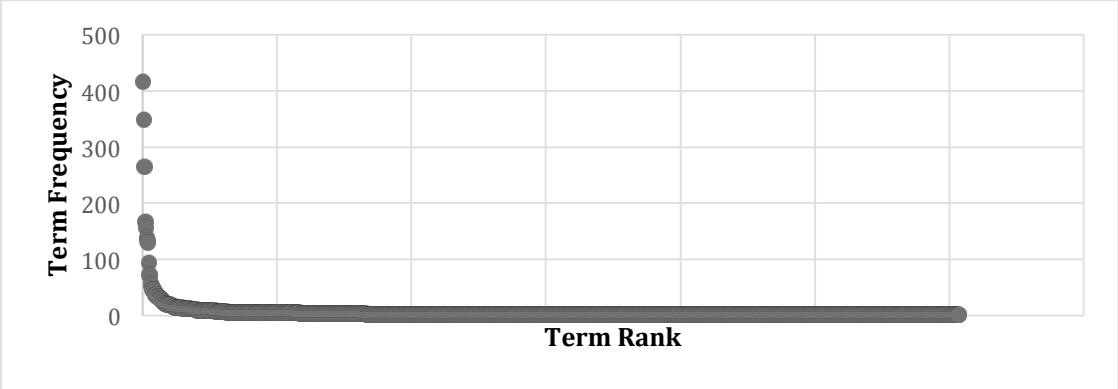

Figure 2. The rank of the terms in USPTO from 1987 to 2014

Figure 2 displays a Zipfian distribution pattern for term frequency distribution of ranked terms based on the text in the patent titles.

3.3 Temporal Burst Analysis

Temporal burst analysis visualizes the increased use of words as a horizontal bar with a start and end date with a text label. The thickness of the bar indicates the weight of each word's burst. This map might show a set of terms that are widely used or a set of terms are never used for the lifespan. Terms may be represented as distinct words or as stems of words that represent variations of a term (e.g., engin representing engine, engines, etc.).

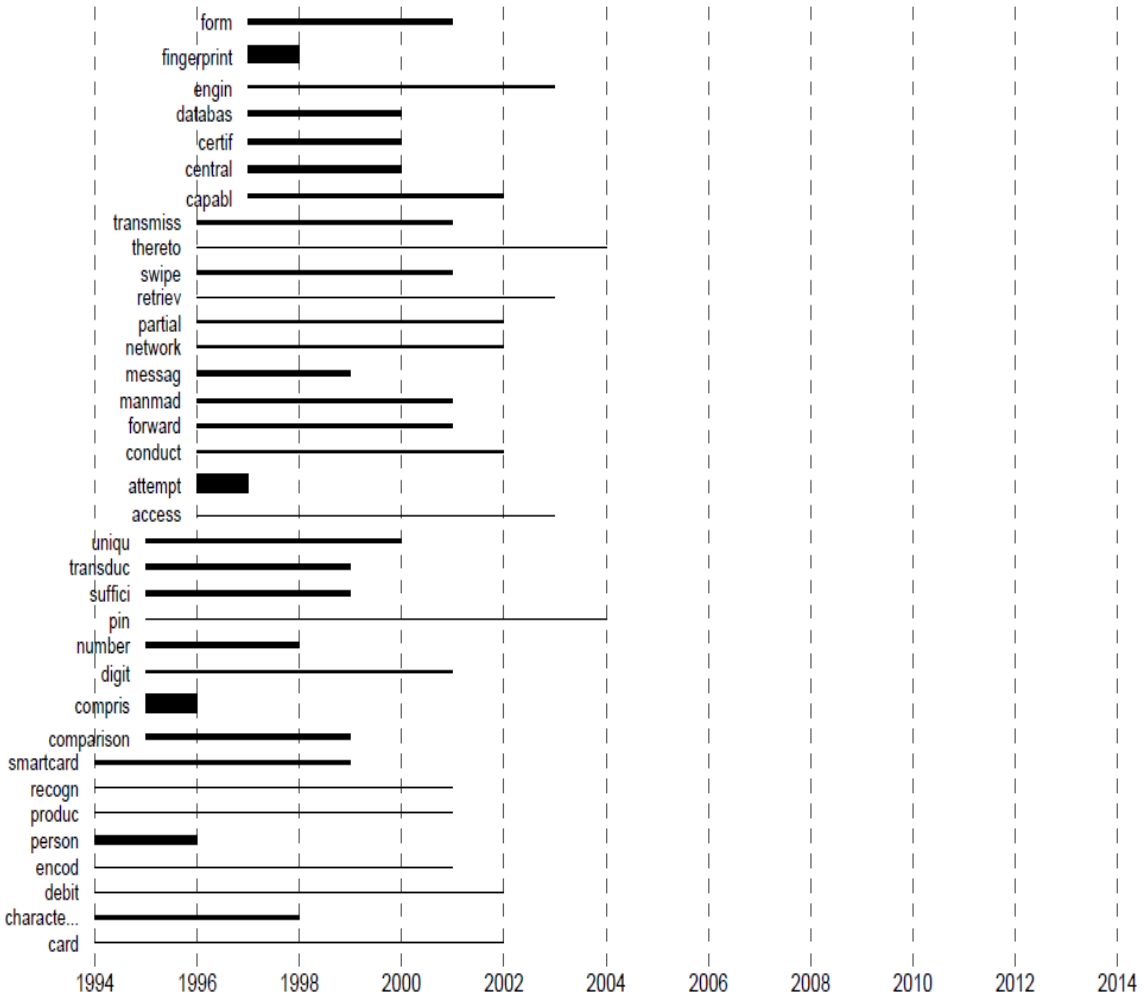

Figure 3. Temporal burst analysis of abstract text in biometric patents (industry)

Figure 3 displays words that peaked and then decreased in usage in industry: "fingerprint*" in 1998, "attempt*" in 1997, and "compris*" in 1996. This might indicate that vocabularies of key terms were used for a long period over time until 2004; however, vocabularies of key terms from 2004 might become much trendier because peaked terms are not found in biometric information as shown in figure 3 from 2004. However, no terms represented after 2004 used in the patents might not be explained in this temporal burst analysis over time.

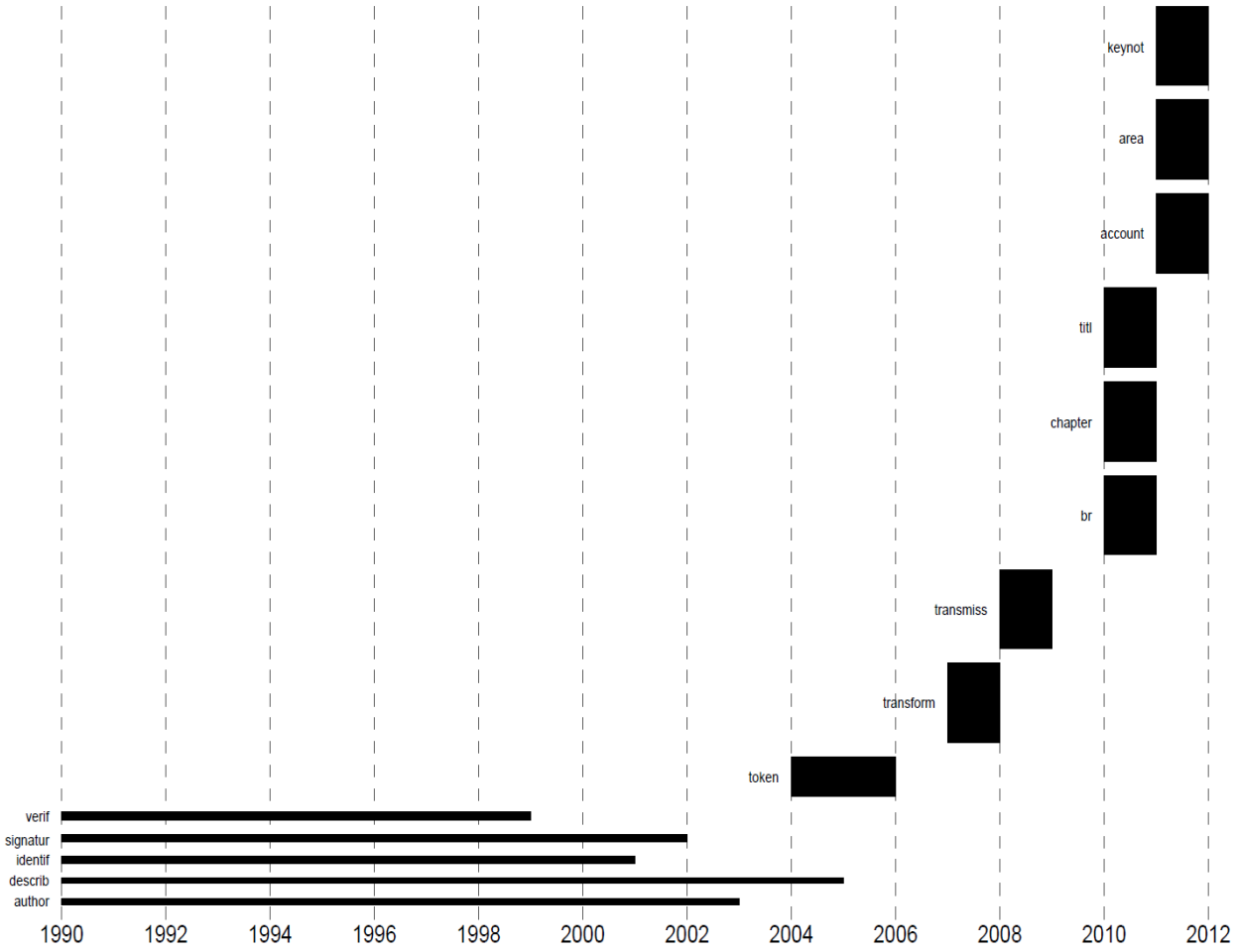

Figure 4. Temporal burst analysis of abstract text in scholarly communication

Figure 4 demonstrates words that peaked and then decreased in usage in scholarly communication sources. These include “token*” in 2006, “transform” in 2008, “transmiss*” in 2009, “titl*, “chapter*” and “br” in 2011, and “keynot*”, “area”, and “keynot” in 2012. Prominent points of difference are shown in 2004: vocabularies of key terms were used for a long period until 2004, however, vocabularies of key terms were used widely in a brief period of time from 2004.

4 Conclusions

The researcher comparatively analyzed both industry and scholarly communication in biometric literature over time. As representatives of industry, 1,047 patents in USPTO from 1790 to 2014 were analyzed. Published documents such as peer-reviewed journals, conference proceedings and ebooks from WoS and IEEE Xplore were analyzed as representatives of scholarly communication. Based on the comparative analysis, this research has found that the transformation point of biometric information might be 2004. However, patterns of cutting edge developments in industry might not affect and/or are affected by scholarly communication because similar terms are not found in each group of publications other than from the comparative analysis of temporal burst detection over time. This study may be useful for industries in which specific technologies need to be invested to penetrate and increase market share. For example, in the medical sector, employee identification using biometric information could be employed to access sensitive patient medical information in electronic medical systems. In scholarly communication, this method will be useful to identify specific topics that are making a significant contribution to or are trending in the field of biometric information. Potential future research may be conducted with a comparative analysis of the full text literature to identify the influences between industry and scholarly communication in biometric literature over time.

References

Chiappalone, M., Nove llino, A., Vajda, I., Vato, A., Martinoia, S. & van Pelt, J. (2005). Burst detection algorithms for the analysis of spatio-temporal patterns in cortical networks of neurons. *Neurocomputing*, 65, 653-662.

European Patent Office (EPO). Retrieved September 21, 2014 from http://www.ipo.gov.uk

IEEE Xplore Digital Library. Retrieved October 7, 2014 from http://ieeexplore.ieee.org/search/searchresult.jsp?action=search&sortType=&rowsPerPage=&searchField=Search_All&matchBoolean=true&queryText=%28%22Document%20Title%22:biometric%29

Science of Science (Sci2). Retrieved October 9, 2014 from https://sci2.cns.iu.edu/user/index.php

Thomson Reuters Web of Science (WoS). Retrieved October 7, 2014 from http://apps.webofknowledge.com/Search.do?product=UA&SID=1BEVc9FbMsxZco5ZmUA&search_mode=GeneralSearch&prID=c12a9053-587f-477c-9c0f-c1323340f783

United States Patent and Trademark Office (USPTO). Retrieved September 21, 2014 from http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=0&p=1&f=S&l=50&Query=TTL%2Fbiometric&d=PTXT

Voyant. Retrieved October 8, 2014 from http://voyant-tools.org/

World Intellectual Property Organization (WIPO). Retrieved October 10, 2014 from http://patentscope.wipo.int

Table of Figures