Expert Systems With Applications 42 (2015) 9389–9401

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A framework to explore innovation at SAP through bibliometric analysis of patent applications

Tabitha L. James [a,*], Deborah F. Cook [a], Sumali Conlon [b], Kellie B. Keeling [c], Stephane Collignon [a], Trevor White [a]

[a] Virginia Tech, 1007 Pamplin Hall, Blacksburg, VA 24061, United States
[b] University of Mississippi, 247 Holman Hall, University, MS 38677, United States
[c] University of Denver, 2101 S University Blvd., Denver, CO 80208-8931, United States

ARTICLE INFO

Keywords:
Software
Innovation
Enterprise data management
Business intelligence and social media/data analytics
Text analysis
Cluster analysis

ABSTRACT

Easily accessible patent databases and advances in technology have enabled the exploration of organizational innovation through the analysis of patent records. However, the textual content of patents presents obstacles to gleaning useful information. In this study, we develop an expert system framework that utilizes text and data mining procedures for analyzing innovation through textual patent data. Specifically, we use patent titles representing the innovation activity at one company (SAP) and perform a bibliometric analysis using our proposed framework. Enterprise software, of which SAP is a pioneering developer, must serve a wide assortment of functions for companies in many different industries. In addition, SAP's sole focus is on enterprise software and it is a market leader in the category with substantial patent activity over the last decade. Using our framework to analyze SAP's patent activity provides a demonstration of how our bibliometric analysis can summarize and identify trends in innovation in a large software company. Our results illustrate that SAP has a breadth of innovative activity spread over the three-tier software engineering architecture and a lack of topical repetition indicative of limited depth. SAP's innovation is also seen to emphasize data management and quickly integrate emerging technologies. Results of an analysis on any company following our framework could be used for a variety of purposes, including: to examine the scope and scale of innovation of an organization, to examine the influence of technological trends on businesses, or to gain insight into corporate strategy that could be used to aid planning, investment, and purchasing decisions.

1. Introduction

Innovation is of interest to academics and practitioners alike. It has been examined for decades using several approaches and for a variety of means. Innovation's link to company performance has been extensively studied (Leidner, Lo, & Preston, 2011; Noruzy, Dalfard, Azhdari, Nazari-Shirkouhi, & Rezazadeh, 2013; Srivardhana & Pawlowski, 2007). Research has approached innovation at multiple levels. At the most coarse-grained levels, the influence of innovation on national or global economies is of interest (Hicks, Breitzman, Olivastro, & Hamilton, 2001; Slater, 2012). In many cases, innovation across an industry has been of primary interest (Pullen, de Weerd-Nederhof, Groen, & Fisscher, 2012; Tajeddini & Trueman, 2012). These studies explore innovation as a competitive advantage (Adegoke, Walumbwa, & Myers, 2012), to aid in evaluation of merger activity within an industry (Breitzman & Mogee, 2002; Breitzman & Thomas, 2002), as an assessment of service quality (Song, Song, & Di Benedetto, 2009), to explore emerging technologies within an area (Daim, Rueda, Martin, & Gerdsri, 2006; Kim, Suh, & Park, 2008), as a tool for sustainability/green initiatives (Bos-Brouwers, 2010; Chen & Chang, 2013), or to examine technological strategies within an industry (Ashurst, Freer, Ekdahl, & Gibbons, 2012; Han, Kim, & Srivastava, 1998; Schmoch & Schnoring, 1994). At the organizational or firm level, management of innovation (Sundbo, 1997) and enhancement of business planning (Lee, Yoon, Lee, & Park, 2009; Zippel-Schultz & Schultz, 2011) may be the goal.

Various researchers have approached the study of innovation using patent data. Patent data has been examined using a variety of approaches and procedures for several purposes. Hicks et al. (2001) used patent bibliometrics to investigate the changing composition of innovative activity in the US. They noted patent indicators pointing to significant innovation in information and health technologies. Daim et al. (2006) point to the difficulties of forecasting when there is no historical data available and used a combination of bibliometrics and

* Corresponding author. Tel.: +1 540 231 3163; fax: +1 540 231 3752.
E-mail addresses: tajames@vt.edu (T.L. James), dcook@vt.edu (D.F. Cook), sconlon@bus.olemiss.edu (S. Conlon), kellie.keeling@du.edu (K.B. Keeling), stephane@vt.edu (S. Collignon), tsw90@vt.edu (T. White).

http://dx.doi.org/10.1016/j.eswa.2015.08.007
0957-4174/

patent analysis to forecast emerging technologies in the areas of fuel cells, food safety, and optical storage.

Siwczyk, Warschat, and Spath (2012) note that patents are used to protect innovative ideas and describe the use of patents as a source of new ideas in the early phases of technology management processes. They also discuss the need for software solutions to facilitate patent analysis as patent databases continue to grow. To advance software to explore patents, Kim et al. (2008) used a k-Means clustering algorithm to create patent maps for ubiquitous computing technology. Shih, Liu, and Hsu (2010) propose a patent trend change mining approach to examining patent activity and use it to examine trends in industry and company activities in Taiwan's semiconductor industry. Lee, Yoon, and Park (2009) used a combination of text mining and principal component analysis to identify "patent vacancies" in patent maps to help identify new technology creation opportunities. Yoon and Kim (2011) noted that patents are good sources of information related to technological change while Yoon, Park, and Kim (2013) used natural language processing of patent text to identify technological competition trends finding "patent vacuums" and "technological hot spots". Chen (2009) used experts to assist in the formation of a patent map for car industry patents. Chen (2009) suggests that patent maps can be created using the process detailed in the study by any company that wishes to use the resulting patent map as a strategic tool.

Previous research indicates that analyzing patent data can lead to many beneficial outcomes, such as detecting trends in innovation and forecasting emerging technologies. It has also been suggested in previous research that methodologies for conducting patent analysis, especially textual content, may be useful to companies in identifying strategic opportunities in a company's own patent activity, comparing innovation or R&D approaches between units, companies or industries, or identifying industry opportunities. Patent analysis as a means to assess innovation could also be leveraged in merger and acquisition decisions or large-scale capital purchases where the health and innovation status of the company is important in assessing access to future product development and support.

In the current study, we propose an expert system framework for analyzing innovation through patent data that combines techniques from text and data mining to conduct temporal and textual analyses of patent activity. We utilize a combination of text and data mining procedures similar in nature to many of the procedures seen in the previous expert system approaches to patent analysis described above. However, our approach combines different procedures and stages of analysis than previous studies, that depending on the desired outcome, could provide a more robust and flexible analysis. Our framework combines traditional text and data mining techniques and a combination of open source, freeware, commercial, and research tools is used to implement the procedures. Our framework is flexible enough that any similar tool or relevant procedure could be substituted or added respectively.

To illustrate our framework, we perform a bibliometric analysis of SAP's patent data. SAP is the market leader for enterprise resource planning software (more generically called enterprise software). Enterprise software is by definition complex and wide-ranging. To adequately handle the system infrastructure needs of large companies or organizations, the software must contain logic for all business functions. Enterprise software serves as the information technology backbone for a company, and as such, needs to be customizable and flexible to meet the needs of organizations that vary in size, scope, purpose, etc. Competitors of SAP range from smaller software companies focused on a particular business function (e.g., HighJump Software and SalesForce) to large, diversified software companies that offer enterprise software as part of a larger portfolio of other software products (e.g., Oracle and Microsoft).

Software is a dynamic and innovative subset of the technology industry and SAP provides one of the most complex examples of software that is heavily utilized by business. In addition, SAP's sole focus is to produce and support enterprise system software. SAP is reported as the market leader in the business-management software market with a 24% market share according to Gartner, Inc. (Norton, 2014). Oracle is reported as the next closest competitor with a 12% market share (Norton, 2014). Sage, Infor, and Microsoft are reported to make up a combined 17% of the market with a large number of companies holding market shares of 3% or less (http://www.statista.com/statistics/249637/erp-software-market-share-by-company/) making up the remaining market share. Nearly 80% of the Fortune 500 companies use SAP software and 63% of the financial transactions are at least partially processed with SAP software (http://fortune.com/2012/03/29/inside-saps-radical-makeover/).

SAP has a history of innovation (Leimbach, 2008), currently demonstrated by SAP Predictive Analytics in addressing the ongoing challenges of big data and SAP HANA (High-Performance Analytics Appliance) in the area of the in-memory computing needed to facilitate predictive analytics. It has been noted that SAP's growth has been assisted by its focus on in-memory computing driven by HANA that improves performance by attempting to minimize time spent transferring data from permanent storage, in part because SAP was one of the first to move into this area (Ashford, 2011). Of all other enterprise software companies SAP has the most patent activity and is the only company with substantial patent activity that is focused solely on enterprise software, making SAP the ideal option to study innovation within a single software developer in this industry. Thus, SAP provides a good case study for our framework that may shed insight into the innovation strategies of one of the world's most important corporate software providers.

The rest of the paper is organized as follows. In the next section, we review the history and development of enterprise systems to show the historical evolution of this software category. This provides a historical presentation of a quickly evolving industry and illustrates the pace of technological and business function integration. The review motivates our choice of enterprise software as a case study and assists with the interpretation of the findings from the application of the framework. In Section 3, we will describe the data collection and methodology, following with the analysis and discussion in Section 4. We offer conclusions and directions for future work in Section 5.

2. Background and motivation for case study

2.1. Enterprise systems and SAP

Since the 1960s, computer systems for enterprise management have grown, now encompassing all aspects of an organization, and have branched out into activities related to supply chain management. The development of software programs to help companies manage specific processes (e.g. inventory and order management) presented attractive possibilities as computer use in business proliferated throughout the 1960s. The process of integrating all activities of a firm into a single application eventually evolved into "enterprise systems" (ES) that attempt to allow firms to electronically manage most (or all) of their business activity (Collignon, James, & Cook, 2010). Collignon et al. (2010) developed a time line to illustrate the development of Enterprise Systems from their beginnings as inventory management systems, through their evolution into Material Requirements Planning and then to Enterprise Resource Planning (ERP) software and beyond (Fig. 1).

The term ERP was introduced early in the 1990s to differentiate the fully integrated business applications software of SAP and others from the business software of older companies (Campbell-Kelly, 2003). ERP software systems provide the information backbone to support the infrastructure of many companies. Standard functions such as ordering materials, scheduling production, sales and operations planning, and general accounting functions are managed by