Decision Support Systems 50 (2011) 415–427

Contents lists available at ScienceDirect

Decision Support Systems

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

Knowledge mapping for rapidly evolving domains: A design science approach

Yan Dang [a], Yulei Zhang [a,*], Paul Jen-Hwa Hu [c], Susan A. Brown [b], Hsinchun Chen [b]

[a] Information Systems Department, Sam M. Walton College of Business, University of Arkansas, Fayetteville, AR 72701, United States
[b] Department of Management Information Systems, Eller College of Management, University of Arizona, Tucson, AZ 85721, United States
[c] Operations and Information Systems Department, David Eccles School of Business, University of Utah, Salt Lake City, UT 84112, United States

ARTICLE INFO

Article history:
Received 9 September 2009
Received in revised form 14 September 2010
Accepted 24 October 2010
Available online 3 November 2010

Keywords:
Knowledge mapping
Design science
Information systems

ABSTRACT

Knowledge mapping can provide comprehensive depictions of rapidly evolving scientific domains. Taking the design science approach, we developed a Web-based knowledge mapping system (i.e., Nano Mapper) that provides interactive search and analysis on various scientific document sources in nanotechnology. We conducted multiple studies to evaluate Nano Mapper's search and analysis functionality respectively. The search functionality appears more effective than that of the benchmark systems. Subjects exhibit favorable satisfaction with the analysis functionality. Our study addresses several gaps in knowledge mapping for nanotechnology and illustrates desirability of using the design science approach to design, implement, and evaluate an advanced information system.

1. Introduction

Knowledge mapping has become increasingly valuable in subject domains that advance rapidly. In general, knowledge mapping offers holistic portrayals of the collective domain knowledge and leverages visualization to present the search and analysis results [23]. Knowledge mapping embraces methods, models, algorithms, and techniques that can be implemented to support individuals in their search and analysis of accumulated knowledge, thereby revealing focal topics, essential subdomains, principal knowledge creators, or emerging trends [4,7]. Visual representations of geographically or temporally oriented knowledge maps can help depict the evolution of a domain over time and can enable researchers and practitioners to access and scrutinize the fast-expanding knowledge [9]. Knowledge mapping is particularly important in fast-growing science or engineering domains where new knowledge is generated at an accelerating pace. In such domains, it is crucial to develop and utilize advanced Web-based systems to support desirable exploitation, analysis, and learning by sifting through vast amounts of information and providing holistic, comprehensive depictions of the overall knowledge [66].

An example of a field in which knowledge has been advancing rapidly is nanotechnology. The nanotechnology domain emerged in the last decades of the 20th century with the development of new enabling technologies for imaging, manipulating, and simulating matter at the atomic scale; it encompasses a broad range of science and engineering activities to develop materials, devices, and systems that exploit the properties of matter at the nanoscale [1]. The fast development of nanotechnology can benefit various fields, including biomedicine, energy, electronics, manufacturing, environmental remediation [1]. As a fast-growth, rapidly evolving scientific domain, nanotechnology research is taking place in academic, government, and corporate institution across the country and around the world. In the past 30 years, a large number of scientific documents on nanotechnology research and development have been generated [42]. For example, the number of nanotechnology-related patent documents published by leading patent offices in the past five years is three times that published since the inception of nanotechnology in the early 1970s [15]. In addition, the overall Federal funding for nanotechnology-related research has tripled since 2001 ($0.46 billion) to 2009 ($1.53 billion) [1]. For scientists, researchers, engineers, and business investors, it is essential to understand the overall knowledge landscape and stay abreast of cutting-edge research, novel developments, frontier-expanding experimentation, killer application development, and essential emerging trends. Knowledge mapping allows detailed analyses and comprehensive understanding of the overall nanotechnology development and its current status. Central to knowledge mapping is thorough analyses of textual documents that report major developments, advancements, or breakthroughs. According to Li et al. [42], knowledge mapping requires the analysis of important documents about patents, funded projects, technical reports, and academic research. By analyzing these documents and exploring their relationships (e.g., co-authors, co-investigators, focal topics, methods, and techniques), we can gain a comprehensive understanding of the overall knowledge development in a field [29].

In spite of its criticality and relevance to information systems (IS) research, knowledge mapping has received limited attention. In

* Corresponding author.
E-mail addresses: ydang@walton.uark.edu (Y. Dang), yzhang@walton.uark.edu (Y. Zhang), paul.hu@business.utah.edu (P.J.-H. Hu), suebrown@eller.arizona.edu (S.A. Brown), hchen@eller.arizona.edu (H. Chen).

0167-9236/$ – see front matter
doi:10.1016/j.dss.2010.10.003

particular, little effort has been devoted to the development or evaluation of advanced knowledge mapping systems that offer comprehensive search support, sophisticated analysis capability, and efficient visualizations. A few systems provide basic knowledge mapping and often have a rather narrow focus as they target a specific document corpus or user queries. Guided by the design science approach by Hevner et al. [27], we develop Nano Mapper, a Web-based knowledge mapping system capable of providing researchers and practitioners with comprehensive depictions of knowledge in the nanotechnology domain. Our choice of the design science approach is appropriate because it provides a robust framework for analyzing and specifying key knowledge mapping requirements in nanotechnology, justifies the system design, connects theory and practice in system development and evaluation, and reinforces important guidelines throughout the build-and-evaluate cycle with the necessary rigor and validity [28]. Nano Mapper includes documents regarding important patents and major funded projects; it supports a wide array of search, has sophisticated built-in analysis functionality, and can present search or analysis results via easily comprehensible displays. We conduct multiple empirical studies to evaluate Nano Mapper that focus on its search functionality and analysis functionality, respectively. Our evaluation studies use measurements from previously validated scales and, when possible, include prevalent systems for benchmark purposes. Overall, our results are encouraging and suggest that subjects can complete knowledge mapping search tasks more effectively and efficiently when using Nano Mapper as compared to using the benchmark systems. Further, subjects exhibit favorable satisfaction with Nano Mapper's analysis functionality.

The remainder of this paper is organized as follows. We first provide an overview of knowledge mapping, review several streams of research fundamental to knowledge mapping, and identify several key challenges that motivate our research. We then describe the design science approach and discuss its advantages and desirability for guiding our system development and evaluation. Next, we detail our system framework, document sources, and system functionality. We then describe our evaluation and highlight key results, followed by a discussion of our contributions to research and practice, study limitations, and recommendations for future work.

2. Knowledge mapping, supporting research, and challenges

2.1. Overview of knowledge mapping

Knowledge mapping is an emerging subfield of information science that has attracted increasing attention from researchers and practitioners alike. It reveals the intellectual structure of a specific domain by synthesizing its accumulated, disparate knowledge into holistic and coherent models and portraits [60,64]. Knowledge mapping depicts the overall development and collective knowledge of a domain, allowing individuals to understand key topics, themes, trends, researchers and their collaboration networks [4,56]. Visualizations are essential to knowledge mapping. They provide users with easily understandable pictures of the knowledge in the target domain as they learn, explore, search, or analyze the entire intellectual space [20]. Knowledge mapping is particularly important in science or engineering fields that advance rapidly. In these domains, it is essential for researchers, practitioners, and business investors to identify leading researchers or institutions, understand their collaboration networks, and become familiar with key topics or emerging trends. Knowledge mapping requires thorough analyses of important explicit tangibles that include patent documents and research project reports. It is critical that knowledge mapping reveals leading researchers, institutions, and their connectedness (e.g., joint patent ownership, collaborated projects, and co-authorships), together with their respective focal topics and major projects [7].

Documents represent a critical repository of knowledge. There are three important types of documents that serve as valuable sources of knowledge: patent documents, funded research documents, and academic research articles. Patent documents are important because they record important research and development (R&D) results that yield substantial commercialization opportunities [29,30,44], and thereby can be used to assess the overall development of a domain [45]. Major patent offices include the United States Patent and Trademark Office (USPTO, http://www.uspto.gov/), the European Patent Office (EPO, http://www.epo.org/index.html), and the Japan Patent Office (JPO, http://www.jpo.go.jp/). Funded research documents are also important as they record major scientific endeavors and findings [31,52]. The National Science Foundation (NSF, http://www.nsf.gov/) is particularly crucial because it is arguably the most influential funding agency for science and technology not directly related to health care. Approximately 6% of the NSF awards granted between 2005 and 2007 were dedicated to nanotechnology research and development [52]. Academic research articles document scientific investigations involving novel approaches, methods, instrumentation, or experimentation, together with key findings [39,55]. The degree to which the number of research articles published in a domain increases over time signifies overall development [4].

2.2. A review of research streams enabling knowledge mapping

Several streams of research are central to knowledge mapping: text mining, network analysis, and information visualization [4]. Text mining extracts important relationships or patterns from a collection of textual documents, and evaluates and interprets those patterns [8]. Text mining is fundamental to knowledge mapping; it reveals important subjects or topics embedded in the title, abstract, or main body of documents. Natural language processing (NLP) and content analysis represent common approaches to text mining. For example, automatic indexing [54] and information extraction [56] follow the NLP approach. Automatic indexing is a noun-phrasing NLP technique that represents document contents using a vector of keywords or terms. Salient noun-phrasing tools include MIT's Chopper, Nptool [62] and Arizona Noun Phraser [61]. Information extraction is capable of extracting important entities of interest from structured texts effectively and efficiently, e.g., names of individuals or locations [56]. Content analysis groups documents on the basis of author, institution, topic area, country, or region, and analyzes them to identify important themes, patterns, or trends [4]. Prevalent techniques include clustering analysis, self-organizing map (SOM), multi-dimensional scaling (MDS), principal component analysis (PCA), co-word analysis, and PathFinder Network (PFNET). Clustering-based techniques group similar documents or topics in a hierarchical structure. SOM [36,37] consists of an unsupervised, two-layered neural network and can be used for clustering or dimension reduction. Chen et al. [11] develop a multi-layered SOM to categorize over 110,000 Web pages on the basis of their respective contents. Kohonen et al. [38] map 6.8 million patent abstracts onto a SOM. MDS and PCA, two classical techniques for dimension reduction, use a low-dimensional Cartesian coordinate space to approximate the corresponding high-dimension vectors. Co-word analysis can depict a network of concepts by generating a matrix of term co-occurrence probabilities for any two terms. PathFinder takes as input estimates of the proximity between pairs of items and selects a network representation of these items by preserving important links only.

Network analysis is also central to knowledge mapping as it can be used to segment subgroups of scientists and researchers, identify key people in a network, reveal their interaction (e.g., collaboration) patterns, and depict the overall network organization or structure [4]. Several essential measures have been developed to characterize each individual node's role in a network; e.g., degree, betweenness, and closeness [63]. The degree of a node describes the number of direct links it has. The betweenness of a node depicts the number of geodesics, i.e., the shortest path between any two nodes passing through the node. The closeness of a node denotes the number of all the geodesics between that node and every other node in the network. Previous research has examined the topological structure and evolution of large real-world networks [46]. For example, Newman [46] reports that the average shortest path length between co-authors in the MEDLINE collection (with 1.5 million nodes) is approximately 4.6, suggesting large networks to have small path lengths between their nodes. Analysis of the MEDLINE co-authorship network shows a coefficient of 0.066 that is several orders of magnitude higher than random associations, suggesting real-world large networks tend to have relatively higher clustering coefficients than do small random graphs.

Knowledge mapping requires effective information visualization to display the mapping results in an intuitive and easily understandable manner [4]. Shneiderman [57] performs a comprehensive review of the existing information representation methods and categorizes them as one-dimensional (1D), two-dimensional (2D), three-dimensional (3D), multi-dimensional, tree-based, network-based, or temporal. Most 1D methods employ one-dimensional visual objects to represent abstract information and display these objects on the screen in a linear or a circular manner [21,25]. With a 2D representation, information is shown as two-dimensional visual objects. Many SOM-based systems adopt a 2D representation to display analysis results [e.g., 11,37,38]. A 3D representation displays information as three-dimensional visual objects, with common metaphors such as rooms [6], bookshelves [6], or buildings [2]. Multi-dimensional representations that use a three-dimensional or a two-dimensional space often project document clusters or themes into that space through dimensionality reduction, e.g., VxInsight system [5]. A tree-based representation also has been employed to show the hierarchical relationships among objects, e.g., tree-map [35], cone tree [51], and hyperbolic tree [41]. The use of a network-based representation is appealing in situations where a simple tree-based structure cannot sufficiently depict the complex relationships. A network representation allows users to visualize the citations among published articles [10] or to understand the linkages among interconnected Web pages on the Internet [2]. A temporal visualization can organize information according to the temporal sequence. Location and animation can be incorporated as visual variables to augment the presentation effectiveness, showing the temporal aspect of information vividly.

2.3. Important challenges of knowledge mapping in nanotechnology

As an example of a rapidly evolving scientific domain, nanotechnology has experienced fast growth in recent years and produced many research streams [32]. It has been shown to have wide implications and significant impacts in various areas related to knowledge generation, industry and biomedical applications, and sustainable environment, and is estimated to be a critical indicator of a country's technological competence [52,53]. More than 60 countries have adopted national projects or programs, such as the United States' National Nanotechnology Initiative (NNI, http://www.nano.gov), to prompt nanotechnology research [52,53]. Nanotechnology has expanded significantly in both scope and specialization granularity, thus making it increasingly difficult for researchers and practitioners to keep up with the explosive knowledge, who usually depend on literature search for monitoring the development of a specific subarea or understanding the current status of new methods or potential applications [32]. Documents reporting major breakthroughs, advancements, or research projects grow in both number and topic diversity overwhelmingly [4,7]. These documents constitute a core knowledge repository as they detail important scientific investigations or development efforts, and report key experiments or analysis results.

Providing interactive search and analysis support has been identified as a major challenge facing knowledge mapping in nanotechnology [12]. Although advanced analytical techniques have been developed to assess the R&D status, few are incorporated to develop Web-based systems capable of supporting interactive searches and analyses. For example, Meyer [44] uses patents issued by the USPTO and scientific literature from the Science Citation Index to examine the interrelationships between academic and industry research. Hullmann [32] employs bibliometric measures for patents and literature to examine the overall status of nanotechnology research and development in the 1980s and 1990s. Huang et al. [29,30] extend previous research by developing a patent analysis framework consisting of bibliometric analysis, content analysis, and citation analysis for evaluating nanotechnology development at country, institution, and specific technology field levels. Narin [45] and Payne and Siow [48] investigate the effects of public funding toward the research and innovation in different domains, reporting that such effects vary across technology fields. By analyzing the relationship between NSF funding and patent publications, Huang et al. [31] show the patents granted to researchers funded by the NSF to have a significantly higher impact on the nanotechnology domain as compared to other reference groups. Although some previous studies have developed or adopted advanced analysis techniques and generated interesting results, few, if any, implement advanced analysis methods in a Web-based system that renders interactive search and analysis support in knowledge mapping.

Both search and analysis are essential to knowledge mapping. Although several Web-based systems have been developed for improved access to nanotechnology-related information, few existing systems embrace a comprehensive corpus of key documents or offer advanced analysis functionality or visualization. Among the existing systems, some target news articles, interviews, and research reports (e.g., Nanotechnology Now at http://www.nanotech-now.com, Nano Tsunami at http://www.nano-tsunami.com); some emphasize hubs of URLs to various Web sites, forums, books, journals, and databases (e.g., ENS Nanotechnology Portal at http://www.ensbio.com/nanotechnologyPortal.html, Nano Scout at http://www.nanoscout.de); yet others focus on convenient access to equipment, education and software (e.g., National Nanotechnology Infrastructure Network at http://www.nnin.org, NanoHUB at http://www.nanohub.org). In general, advanced search methods and analysis techniques for mapping scientific documents on nanotechnology are lacking; few are implemented in Web-based systems. This calls for the development of advanced modeling, analytical methods, and computational techniques and their implementation in knowledge mapping systems that include major source documents. Supported by such systems, researchers and practitioners can identify leading researchers and their collaboration networks, discover focal topics and their interconnectedness, and detect important or emerging trends in nanotechnology. Search support is also important as it provides users with easy access to the sophisticated analysis functionality. Desirably, search support should be comprehensive but without the use of proprietary, platform-specific query language. In addition, visualization matters. Eggers et al. [20] emphasize that effective knowledge mapping must use visual tools to depict big pictures in a knowledge domain. Effective visual displays should be intuitively comprehensible and cognitively efficient, allowing users to explore the collective, intellectual space by tracking the dynamics of the field or identifying new topics or important subareas for further scientific or commercial exploitation. It is therefore crucial to build and evaluate advanced knowledge mapping systems offering comprehensive search support, sophisticated analysis functionality, and effective visualization to researchers, practitioners, business investors, and policy makers in nanotechnology.

Another key challenge is the inclusion of essential document sources to analyze prominent researchers, important topics, and emerging trends [20,59]. As Reid and Chen [50] caution, as the influxes of new, diverse, and disorganized studies continue at an increasing pace, it is challenging to provide knowledge mapping support effectively and efficiently. Analysis shows multiple, diverse document sources essential to the overall development of nanotechnology; they differ in the document collection, document format or structure, system platform, and/or search (query) support. Prior research has identified patent documents, funded project reports, and academic articles as critical to nanotechnology [29,30,44]. For example, patents manifest, with considerable details, significant technological breakthroughs that have promising commercialization prospects; funded projects, particularly those by major funding agencies, are important because they are often led by renowned researchers and pursue high-risk, boundary-expanding scientific explorations or cutting-edge technology developments [31,52]. Academic articles report important investigations by researchers in various institutions that pass the stringent scrutiny of peer reviews [39,55]. According to our literature review, most previous knowledge mapping research targets a specific document source [e.g., 16,29,30,44] rather than including documents from multiple sources.

A review of extant literature suggests that the advances in text mining, network analysis, and information visualization propel the viability and scalability of Web-based knowledge mapping systems capable of portraying the collective knowledge available in a fast-evolving domain. From the perspective of document sources, Web-based knowledge mapping becomes increasingly feasible and appealing as more and more document sources are accessible online. For effective knowledge mapping in nanotechnology, a systematic, integrated, and methodologically rigorous approach is crucial but has not been undertaken by previous research. To address this gap, we design and implement a Web-based system, Nano Mapper, which provides an overall understanding of the knowledge landscape through comprehensive search and advanced analysis functionality supported by appropriate text mining, network analysis, and visualization techniques. Nano Mapper includes patent documents published by major patent offices and the NSF-funded research project documents. In particular, we take the design science approach to guide our system development and evaluation, detailed in the following section.

3. A design science approach for knowledge mapping in nanotechnology

Design science approaches IS research from the lens of building and evaluating IT artifacts targeting specific, real-world problems or challenges. Design science can complement behavioral science research that often emphasizes theory development, justification, or testing [27]. Design science is rooted in a desirable synthesis of the sciences of the artificial [58], engineering design, information systems development [33], system development as a research methodology [47], and executive information system design theory [22]. It offers a conceptual framework for guiding the build-and-evaluate cycle of the IT artifact and elevates visibility of technical IS research with an emphasis on rigor and relevance [27]. This framework explicitly delineates the boundary of design science, elaborates its relationship to behavioral research from both theoretical and empirical perspectives, and advocates a synergistic, complementary cycle between design science and behavioral science to better meet the challenges of IS research at the interface of people, organization, and technology [27,43].

Design science fosters creative exploration and experimentation of novel ideas, modeling methods, analytical techniques, computational algorithms, or visualization designs that can be instantiated (e.g., implemented in a Web-based system) for significant gains in effectiveness or efficiency [19,27]. According to the three-cycle view by Hevner [26], design science embraces relevance, rigor, and design, that together allow researchers and practitioners to generate IT artifacts capable of generating substantially improved utility for targeted problems, in light of theoretically justifiable system development, evaluation, and efficacy. The design science framework enables us to better understand, design and implement an effective knowledge mapping system (i.e., IT artifact) to address the most pressing needs of individuals who will use the system. Existing systems offer little search or analysis support in knowledge mapping in the nanotechnology domain. Typically, they emphasize a specific document source in isolation and support a narrow range of user query. Thus, these systems show the viability of knowledge mapping (in the form of demonstration prototypes) but are not developed in a systematic, integrated manner to provide the utility needed by the end users (e.g., [20,31]). As we summarized in Section 2.3, some prior studies develop advanced search or analysis methods for knowledge mapping; however, few have implemented them in Web-based systems. We follow the design science approach by performing thorough system design and methodological evaluations in the development of Nano Mapper. By taking the design science approach, we can better address the key challenges of comprehensive interactive search support, advanced analysis capabilities, greater document coverage, and effective visualization designs, using theory-based rationales for system requirements and design choices to generate empirical evidence regarding gains in effectiveness and efficiency to demonstrate the value of Nano Mapper to system users. In this study, we use the nanotechnology domain as a demonstration example. However, the framework of the Nano Mapper system is generic and can be applied to other domains.

We implement and evaluate Nano Mapper according to the guidelines set forth by Hevner et al. [27]. Our Web-based system represents an IT artifact instantiated using advanced search support, analysis techniques, and visualization designs, i.e., the "design as an artifact" guideline. To examine and demonstrate desirable system efficacy and utility, we conduct evaluation studies using representative knowledge mapping tasks, prevalent benchmark systems, established empirical methods, appropriate study designs, and previously validated measurement instruments, i.e., the "design evaluation" guideline. By doing so, we contribute to extant knowledge mapping research by providing a real-world knowledge mapping system that includes key documents from different sources and offers comprehensive search support, analysis functionality, and effective visualizations, i.e., the "research contribution" guideline. We construct an advanced knowledge mapping system relevant to nanotechnology researchers, practitioners, business investors, and policy makers, i.e., the "problem relevance" guideline. The development of Nano Mapper is guided by the design science approach and its implementation encompasses different search mechanisms, analytical methods, computational algorithms, and visualization models extended from existing text mining methods and network analysis techniques, i.e., the "research rigor" guideline. Our system design and implementation proactively solicit from nanotechnology experts, knowledge mapping researchers, and general users essential evaluative feedback for improved system design, functionality, and utilities, i.e., the "design as a search process" guideline. Furthermore, we conduct evaluation studies and communicate our empirical results to demonstrate and convey the utility of Nano Mapper, i.e., the "communications of research" guideline.

4. System framework, document sources, and functionality

4.1. System framework

The system is developed based on three layers: database layer, logical control layer, and presentation layer. The three-layer architecture makes the system framework flexible and adaptable to the needs of different data sources and functions for different application domains. Independent modules are developed for different layers.

This structure enables any change in one layer to be independent of and not influence other layers. The database layer stores data extracted from the parsed documents. These documents can be research articles, patents, technical reports, grant documents, or any other type of scientific document from a rapidly involving domain. Each document is parsed and then stored in a relational database.

The logic control layer houses multiple modules for processing user queries or accessing different functions. This layer serves as a middleware connecting the presentation (user front-end) layer and the back-end database layer. On this layer, modules for different search and analysis functions are created. To improve the response time, intermediate tables are generated for storing pre-computed results. Once a user query is captured, the corresponding module will first check the pre-defined results in the intermediate tables and then lower-level tables to conduct detailed processing.

The presentation layer provides user interfaces and thus allows Web access to various functions. Web development languages, such as HyperText Markup Language (HTML), JavaServer Pages (JSP), Active Server Pages (ASP), and Hypertext Preprocessor (PHP), can be used to implement user interfaces. Existing visualization packages and libraries can be customized and embedded in the system to provide dynamic user-interactive visualizations.

4.2. Document sources

Scientific documents represent important intellectual knowledge for a rapidly evolving domain. As to the nanotechnology domain, patents are also important scientific documents since they contain essential research and development (R&D) results about the creation and adoption of nano-related materials and methods, and yield substantial commercialization opportunities. USPTO, EPO, and JPO are acknowledged as the world's leading patent offices; together, they account for approximately 90% of the patents globally [12,40]. The USPTO database provides online full-text access to patents issued since 1976. The EPO database provides access to European patents issued since 1978, and the JPO database contains patents issued since 1976.

Grant documents are also valuable for understanding developments in science and technology (S&T). Specifically, we use the documents summarizing nanotechnology-related projects funded by the NSF since NSF is arguably the most influential funding agency for S&T research that is not directly related to health care. Currently, we do not include academic articles mainly because few publishers offer such articles online, free of charge. As we expand the document corpuses, we intend to include major academic sources as they become freely accessible. Nevertheless, a separate study was conducted that targets academic articles about nanotechnology, using the Thomson Science Citation Index Expanded database [42].

Different data sources typically differ in structure and design, each requiring specific document gathering and parsing techniques. Keyword-based data collection can be conducted to obtain relevant documents for a given domain. Keywords can be used to search against either the entire document (i.e., “full-text” search) or the titles and abstracts (i.e., “title-abstract” search) since sometimes the data sources do not provide access to the entire documents. In general, “title-abstract” search seems to generate more accurate results while “full-text” search offers greater coverage of documents. For Nano Mapper, a set of keywords is obtained from several nanotechnology experts. “Title-abstract” search is conducted in all four source databases. For the USPTO database, we also conduct “full-text” search and “title-claims” search (by matching the keywords against patent title, abstract, and claims). The other data sources do not support either “full-text” search or “title-claims” search.

To the best of our knowledge, few existing systems incorporate both patents and grant documents as data sources. In general, the major patent search systems are the ones maintained by particular patent offices themselves (e.g., USPTO maintains the search system for US patents; EPO maintains the search system for European patents; JPO maintains the search system for Japanese patents). They seldom cover other data sources (e.g., grant documents). Other company-developed patent search systems, such as Google Patents, which supports search across USPTO patents, also use data from a single source. However, nanotechnology experts have explained that in order for them to gain a comprehensive understanding of the current R&D status of an entire area, they must track various sources: patents and grants. Therefore, in this study, we use both patents and grant documents as data sources of Nano Mapper.

4.3. System functionality

In this section, we detail the functions of Nano Mapper. Because of the three-layer architecture, these functions are independent of the data sources. To apply the system framework to another domain, we only need to change the data sources in the database layer, without the need to modify the modules of system functions in the presentation layer. Separate user interfaces for search vs. analysis functionality are developed for each document sources. For improved consistency and a more positive user experience, we adopt the same interface design for each function, regardless of the document source [13].

The search functionality follows a keyword-based search design and the analysis functionality encompasses advanced visualization techniques to provide comprehensive views of nanotechnology research and development in an intuitive, cognitively efficient manner. All the patents (grants) containing the user specified keyword or a set of keywords will be treated as correct answers and returned to the users. The search results are ordered by the patent (grant) publication date — newly published patents (grants) are listed first, on the top of the returned list.

Nano Mapper supports three search functions: patent (grant) number search, quick search, and advanced search.[1] Using the patent (grant) number search, an individual can search a particular patent (grant) using its patent (grant) identifier. The quick search identifies a set of patents (grants) of interest to the user, on the basis of the provided keywords in title, abstract, or (patent) claims. With the advanced search, people use multiple criteria to search, i.e., all available patent (grant) data fields that include patent title, examiner, inventor, assignee, assignee country, classification code, abstract, and claims. Both patent (grant) number search and quick search can support low-complexity search tasks appropriately, whereas the advanced search targets high-complexity tasks. For several data fields that include assignee country, Nano Mapper provides convenient lookup functions that allow users to choose and submit their search criteria. Prior HCI studies have used look-up functions as application activities to better assist a user's information acquisition. In Fig. 1, we show, as an example, the unified interface design for the advanced search function for different document sources.

Nano Mapper offers three distinct analysis functions: statistical analysis, citation network analysis, and content map analysis

[1] The three search functions (i.e., “patent (grant) number search,” “quick search,” and “advanced search”) are different from the three search techniques (i.e., “full-text” search technique, “title-claims” search technique, and “title-abstract” search technique) mentioned in Section 4.2. The three search functions are for users to conduct their patent (grant) search tasks. However, the three search techniques were used to identify nanotechnology related patents and filter out the unrelated ones from the entire USPTO, EPO, JPO, and NSP databases, since these databases contain patents (grants) for many different disciplines, not only the nanotechnology domain. This data filtering step is not visible to end users. Only when users use the more advanced analysis functions will they need to choose a particular dataset to analyze, either the collection obtained using the “full-text” search technique, “title-claims” search technique, or “title-abstract” search technique.

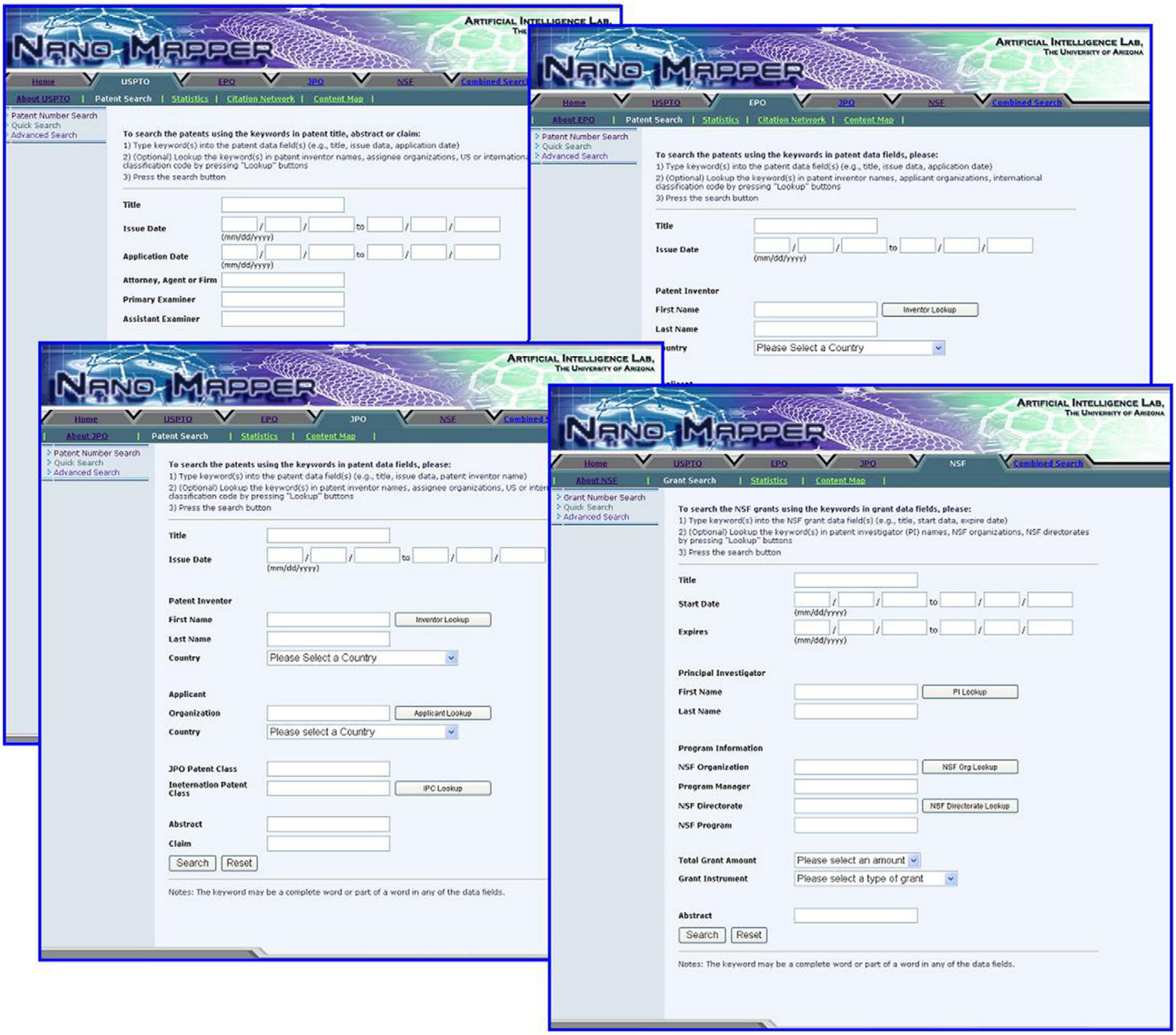

Fig. 1. Advanced search interfaces for nanotechnology documents from different sources.

functions. In Fig. 2, we provide examples of these different analysis functions.

The statistical analysis function generates important statistics and citation information according to the user specified time period (i.e., in month or year) and granularity (i.e., country, institution, inventor, or technology field). Dynamic tables and charts are provided based on Java Applet packages and Chart 2D (http://chart2d.sourceforge.net), a salient open source java library. The dynamic tables allow users to sort analysis results by the number of patents, the number of citations, the average number of citations, the number of grants awarded, or total funding. Dynamic charts present the analysis of annual publication trends in the patents (grants). Users can choose a particular analysis granularity or specify the number of the most productive researchers, shown in a dynamic chart.

The citation network analysis allows users to visualize patent citation networks at different levels of granularity or over various time horizons. It is implemented by customizing Graphviz (http://www.research.att.com/sw/tools/graphviz), an open source graph generation software [24]. To make important citation relations more prominent, the top N (e.g., top 100) relations between or among researchers that have the N largest numbers of citation are displayed. In a citation network, the directionality of a link denotes the exact direction of a citation between two nodes. For example, a link from the "United States" pointing to "Germany" shows that a U.S. patent cites a German patent. Each link has an associated number representing the total number of citation.

The content map analysis function is created using the Topic Map package developed by the Artificial Intelligence Lab, University of Arizona (http://ai.arizona.edu). A map interface contains two components: a folder tree and a hierarchical content map. The folder tree displays the topics identified from patent (grant) documents and the hierarchical content map shows corresponding topic regions. Each topic region is associated with a topic keyword as well as a number denoting the exact number of source documents. The size of a topic region is proportional to the number of documents pertinent to that topic. Related or similar topics are placed adjacent on the map. When the user clicks on a topic region, its subtopics are then shown on the map. This function supports the calculation of the growth of each topic

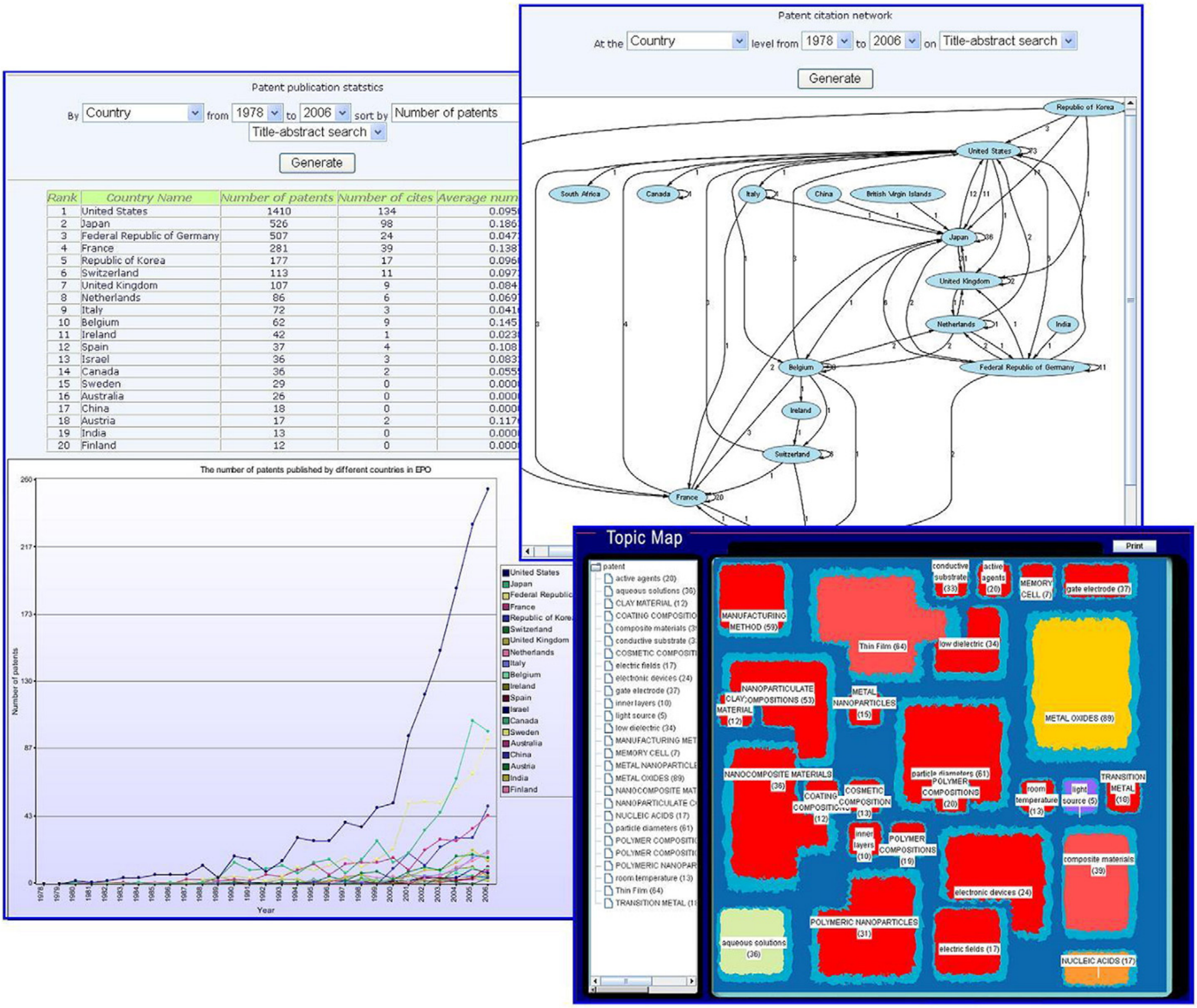

Fig. 2. Examples of Nano Mapper's analysis functionality.

area over two continuous time periods. Specifically, a baseline growth rate is computed at the entire content map level. Topic regions exhibiting a growth rate similar to the baseline rate are displayed in green. Topic regions that have a higher or a lower growth rate are shown in a warmer or a colder color according to the spectrum, respectively. A new topic is shown in a red color. Previous HCI research has shown that color representation can help to attract users' attentions. Warm colors (e.g., red, yellow, and orange) can signify importance and thereby draw an individual's attention, and cold colors (e.g., green, blue, violet, and purple) can be used to denote ordinary objects or instances [49].

5. System evaluations

Following the design science build-and-evaluate approach, we examined the utility and efficacy of Nano Mapper. We performed multiple studies to evaluate Nano Mapper: two examining its search functionality and three assessing its analysis functionality. For search functionality evaluations, each experiment targeted different patent documents (USPTO or EPO) and included a salient system for benchmark purposes. We performed three separate studies to examine the analysis functionality of Nano Mapper, each focusing on a particular patent collection (i.e., USPTO, EPO, or JPO). For each evaluation study, a set of tasks were designed with the assistance of an expert panel, consisting of three nanotechnology experts and two experienced knowledge mapping researchers. These experts used both Nano Mapper (for both search tasks and analysis tasks) and the benchmark system (for search tasks) to develop a “gold-standard” answer for each task. They checked the face validity and content validity of the tasks and answers. For face validity, all five experts helped to make sure that the search tasks and the analysis tasks were indeed measuring the search functions and various analysis functions of the system. For content validity, the nanotechnology experts focused on making sure that the tasks and answers were useful and important for understanding the overall development of the nanotechnology domain, and the knowledge mapping experts mainly focused on making sure that the tasks were clearly stated and that there was no ambiguity in the tasks or “gold-standard” answers. Before conducting the user studies, we performed a pilot study involving five voluntary subjects to fine-tune our study flow and data collection. The subjects were asked to perform all tasks (including both search and analysis tasks) twice in two different days. Consistent task accuracy results were observed, indicating the test–retest reliability of the tasks. In the following sections, we describe the two types of evaluation studies and the analysis results in detail.

5.1. Search functionality evaluations

We performed two experimental studies to examine Nano Mapper's search functionality. We targeted USPTO and EPO patent documents and included the USPTO patent search system (http://patft.uspto.gov/) and the EPO patent search system (http://ep.espacenet.com/) for benchmark purposes. Our benchmark system choices offer desirable symmetry in both source documents and search functions. First, each benchmark system has patent documents also accessible by Nano Mapper. For the USPTO patent collection, we used the USPTO patent search system (http://patft.uspto.gov/) hosted by the United States Patent and Trademark Office as the benchmark system. Specifically, we used the three search functions (i.e., patent number search, quick search, and advanced search) of the USPTO patent search system for the finally granted patent collection (instead of the patent application collection) to conduct the comparison, because the USPTO patent data in Nano Mapper were collected from the same patent database containing the finally granted patents and hosted by USPTO. Similarly, we chose to use the EPO patent search system (http://ep.espacenet.com/) hosted by the European Patent Office as the benchmark system in the search function evaluation study for the EPO patent collection. The EPO patent data in Nano Mapper were collected from the same patent database (containing only the finally grated patents instead of patent applications) hosted by EPO. Second, both benchmark systems, similar to Nano Mapper, support patent number search, quick search, and advanced search, and provide separate search interfaces accordingly. The JPO patent search system was not included as a benchmark because it supports basic patent number search but not quick search or advanced search.

We used both between-groups and repeated-measures designs in the current study. Our between-groups factor is search system that is defined at two levels: Nano Mapper and the benchmark system. Search task complexity is the repeated-measures factor, which is also defined at two levels: low and high. Subjects were randomly assigned to use either Nano Mapper or the benchmark system to perform each task. Compared with alternative experimental designs (such as a "two repeated-measures factors" design), our design allows us to evaluate each system's performance without overwhelming the subjects. We randomized the order of the tasks to be performed by each subject, thus mitigating the potential influences of task sequencing.

We used task performance accuracy to measure each system's effectiveness, which refers to how well a system can support the user to complete a search task correctly. Our accuracy measurement is commonly used in knowledge mapping and information retrieval [14,65]. Specifically, task performance accuracy is calculated as: $Accuracy = \frac{\textit{Number of correctly answered parts}}{\textit{Total number of parts}}$. The expert panel assessed subjects' search results and assigned accuracy scores according to the respective gold-standard answers. We use the amount of time a subject takes to complete a search task to measure task performance efficiency [14,65]. To keep the experiment within a reasonable time span, we introduced a time limit of 15 min for each high-complexity search task and 10 min for each low-complexity task.

Our subjects were undergraduate students enrolled in a sophomore- or junior-level information systems class at a major public university located in the southwest United States. The instructors assisted our recruiting by providing course credit as an incentive for students' voluntary participations in our study. Each subject was asked to use patent number search and quick search to complete three low-complexity tasks, and use advanced search to complete three high-complexity tasks. A subject used either Nano Mapper or the benchmark system to perform all the search tasks, but not both. The order by which the search tasks were assigned to a subject was random, thereby reducing the potential bias introduced by task sequencing. We list in Appendix A the low- and high-complexity search tasks used in the experiment comparing Nano Mapper and the EPO patent search system as an example. After completing all the low- or high-complexity tasks using the assigned system, a subject was asked to report his or her overall satisfaction with that system. We measured user satisfaction using question items adapted from previously validated scales [3], with minor wording changes for tailoring to our subjects and study context. The four items are "very dissatisfied/very satisfied," "very displeased/very pleased," "very frustrated/very contented", and "very terrible/very delighted" with the search function. These question items adopted a seven-point Likert scale, with 7 being "most positive" and 1 being "most negative."

A total of 48 subjects took part in the study comparing Nano Mapper and the USPTO search system. Twenty-four subjects used Nano Mapper, among them, 12 performed the low-complexity tasks first and the remaining received the high-complexity tasks first. The other 24 subjects used the USPTO search system, with 12 performing the low-complexity tasks first and the remaining receiving the high-complexity tasks first. Another 56 subjects participated in the study comparing Nano Mapper and the EPO search system. None of the 56 subjects participated in the study comparing Nano Mapper and the USPTO search system. The same approach was used: half of the subjects evaluated each system and within them half received the low-complexity task first and the other half received the high-complexity tasks first. In Table 1, we provide the subjects' demographic data.

We performed one-tailed t-tests to examine Nano Mapper's search functionality in relation to that of the benchmark system. As summarized in Table 2, the likelihood of a subject's successfully completing low-complexity tasks is significantly higher when using Nano Mapper than using the USPTO system (p-value$<$.01). However, it is not statistically significant for high-complexity tasks (p-value$=$.07). Subjects need significantly less amount of time when using Nano Mapper than using the USPTO system to complete both low-complexity tasks (p-value$<$.05) and high-complexity tasks (p-value$<$.0001). Subjects report a significantly higher satisfaction with Nano Mapper than with the USPTO system (p-value$<$.01 for low-complexity tasks, and p-value$<$.0001 for high-complexity tasks). The results indicate that Nano Mapper outperforms the USPTO search system in task performance accuracy and efficiency as well as user satisfaction. The desirable performance might be partly explained by the USPTO search system's advanced search function that follows a query language-based design. Users must learn the query language syntax before they can use this advanced search function effectively. In our experiment, although we provided appropriate training to help subjects familiarize themselves with the USPTO query language before they performed the search tasks, we anticipated that some subjects might have difficulties using this query language to compose search queries. Several subjects mentioned this during the study. They expressed the concern that the syntax of the query language was strict and sometimes difficult to follow. In addition, some data fields (e.g., patent number, patent title) in the USPTO search system are not labeled clearly, which might also affect its search performance

Table 1
Aggregated subject information of Nano Mapper's search functionality evaluation studies.

Attribute	Subjects' characteristics	
	Nano Mapper vs. USPTO system	Nano Mapper vs. EPO system
Gender	Male: 31; female: 17	Male: 35; female: 21
Age	Mean: 22.31, Std dev: 6.75	Mean: 21.45, Std dev: 1.59
Years of using computer	Mean: 12.23, Std dev: 3.36	Mean: 12.89, Std dev: 3.02
Years of using Internet	Mean: 10.42, Std dev: 3.02	Mean: 10.27, Std dev: 2.20

Table 2
Results of Nano Mapper's search functionality evaluation studies.

Measure	Task complexity	Nano Mapper	USPTO system	Nano Mapper vs. USPTO system *p*-value
		Mean/SD	Mean/SD	
Effectiveness	Low	99.17%/4.08%	85.28%/25.54%	0.0084**
	High	67.64%/26.44%	53.75%/33.68%	0.0727
Efficiency	Low	2.33/0.63	2.71/0.80	0.0462*
	High	3.63/1.23	7.54/2.26	<0.0001**
Satisfaction	Low	6.18/0.80	5.59/0.85	0.0020**
	High	5.02/1.23	2.92/1.21	<0.0001**
Measure	Task complexity	Nano Mapper	EPO system	Nano Mapper vs. EPO system *p*-value
		Mean/SD	Mean/SD	
Effectiveness	Low	95.24%/11.88%	83.21%/24.49%	0.0099**
	High	92.98%/13.71%	85.95%/21.24%	0.0510
Efficiency	Low	2.12/0.46	2.62/0.68	0.0001**
	High	2.88/0.73	3.26/0.91	0.0257*
Satisfaction	Low	6.08/0.89	5.41/1.05	0.0104*
	High	5.82/0.87	5.59/1.16	0.2290

Note. Significance levels $^{*}\alpha=0.05$ and $^{**}\alpha=0.01$. Task performance was measured by averaging performance across tasks. Effectiveness was measured by accuracy. Efficiency was measured in minutes. Satisfaction rating scale ranges from 1 to 7, with 7 being the best.

adversely. During the study, some subjects had difficulty in locating the patent number in the returned patent search page. A few of them also mentioned that it would be helpful if there was a label to clearly indicate the patent number data field in the returned result page. The learning requirements are much lower for using Nano Mapper. Subjects are required to provide keywords to the associated fields only, a more intuitive approach than using a query language. Furthermore, all the data fields in Nano Mapper are clearly labeled, thus increasing its usability and ease of use. During the study, subjects encountered no problems with locating different data fields in patent search result pages.

Compared with the EPO search system, subjects are more likely to successfully complete low-complexity tasks using Nano Mapper (p-value<.01). We note a similar performance differential in high-complexity tasks (p-value=.051). The amount of time subjects needed to complete a search task is significantly less when supported by Nano Mapper than by the EPO search system (p-value<.0001 for low-complexity tasks, and p-value<.05 for high-complexity tasks). User satisfaction is higher with Nano Mapper than with the EPO search system, statistically significant for low-complexity tasks (p-value<0.05) but not for high-complexity tasks (p-value>.05). The performance differences between the two systems could be attributed to the fact that the search results retuned by the EPO search system are ambiguous in some aspects (e.g., publication date). Imaginably, some subjects using the EPO search system might not pay attention to the patent publication date, whereas their counterparts using Nano Mapper received all essential information about each search result, including patent ID, title, and publication date. During the experiment, several subjects commented that in the returned EPO patent summary pages, the patent number and publication date for each patent were displayed together and labeled as "publication info." They noted that this was somewhat confusing. In Nano Mapper, however, each data field in the returned patent summary pages is clearly labeled, and subjects had no problems in understanding these data fields.

5.2. Analysis functionality evaluations

We also conducted three studies to evaluate Nano Mapper's analysis functionality (i.e., statistical analysis, citation network analysis,[2] and content map analysis), each focusing a particular collection of documents (i.e., USPTO, EPO, or JPO).

[2] The JPO patent collection does not have the citation network analysis function since the citation data is not available for the JPO patents.

Our evaluations targeted user satisfaction because user satisfaction has been identified by previous information systems research as a critical measure of system success and thus a relevant factor when examining a particular information system [17,18]. The overall system evaluation was broken down according to the features of functionality and data sources. A total of five evaluation studies were designed and conducted on the combination of different functionalities and different data sources. Specifically, two studies were conducted for search functionality evaluations on the USPTO and EPO patent collections respectively; three studies were conducted for analysis functionality evaluations on the USPTO, EPO, and JPO patent collections respectively. For each study, we modified the wording of the user satisfaction items to indicate the particular functionality and data sources being evaluated.

No benchmark systems are included in the analysis functionality evaluation studies because none of the existing knowledge mapping systems in the nanotechnology domain, including the USPTO, EPO, and JPO patent search systems, offers such analytical functionality. Specifically, we use subjects' evaluative assessments of their satisfaction about Nano Mapper's analysis functions instead of directly comparing these functions with those of a benchmark system.

During the user studies, each subject was asked to perform three different analysis tasks using Nano Mapper's statistics analysis function, citation network analysis function, and content map analysis function, respectively. In Appendix B, we list all the analysis tasks associated with the EPO patent documents. We used the same measurement items [3] from the search functionality evaluation studies. All the items employed a seven-point Likert scale, with 7 indicating the most positive assessment, and 1 indicating the most negative assessment. We use the midpoint (i.e., score of 4) as a cutoff distinguishing positive and negative assessments by subjects. In the evaluation study involving the USPTO patent collection, we recruited subjects from those assigned to use Nano Mapper in the experiment comparing the search functionality of Nano Mapper and that of the USPTO search system. We target these subjects because they are familiar with Nano Mapper and have some experience in using the system for task purposes. Similarly, in the evaluation study involving the EPO patent collection, we recruited subjects from those assigned to use Nano Mapper in the experiment comparing the search functionality of Nano Mapper and that of the EPO system. For the JPO patent collection, we recruited new subjects who had not participated in any search or analysis functionality evaluations. As we mentioned in Section 5.1, since no benchmark system with the same three search functions is available, we did not conduct the search function evaluation study on this set of patent documents. Therefore, different from the subject pools of USPTO and EPO patent collections, we do not have subjects that have already attended the search function evaluation study for the JPO patent collection. We recruited new subjects for the JPO analysis function evaluation. To familiarize these (new) subjects with Nano Mapper, before conducting the evaluation, we first gave them extra time to perform several exercise search tasks using Nano Mapper's search functions for the JPO patent collection. In addition, since we did not directly compare the evaluation results across different patent document collections (i.e., USPTO vs. JPO vs. EPO), we believe that the relative system familiarity difference in different subject pools will not influence the evaluation results. A total of 24, 28, and 13 subjects took part in the studies that involved the USPTO, EPO, and JPO documents respectively. In Table 3, we present the demographics of our subjects.

We performed one-tailed t-tests to assess whether subjects' self-reported satisfaction is significantly higher than the threshold of 4, the midpoint of the seven-point Likert scale. As summarized in Table 4, subjects exhibit favorable satisfaction, significantly higher than 4 (p-value<.0001), across all the patent corpuses under examination. Most of the subjects expressed interest in using these analysis functions. Several commented that they liked the dynamic displays of the tables and charts of the statistical analysis and citation network analysis results. Others said that they liked the user-interactive interfaces of the analysis results of the content map analysis function the most.

6. Contributions and future directions

6.1. Contributions to design science

Hevner et al. [27] emphasized that the novelty of design science research (guideline 4: research contributions) is "solving a heretofore unsolved problem or solve a known problem in a more effective or efficient manner (p82)." Aiming at providing better knowledge mapping support, the framework of the Nano Mapper system integrates various search functions as well as advanced analysis functions (implemented using content map and citation network algorithms) and incorporates various scientific document sources. Although the focus of the current study is not to create specific, new computational algorithms, the overall system framework proposed in this study is new. We haven't seen any existing Web-based, interactive system that provides as comprehensive search and analysis functions as Nano Mapper does for knowledge mapping support in scientific domains. No existing scientific knowledge mapping system has incorporated the patent documents from major patent offices in the world and the NSF grant documents to provide users such a comprehensive scientific document source. In addition, the results of our evaluation studies have shown that the proposed framework can provide knowledge mapping support for a rapidly evolving domain in a more effective and efficient manner.

Table 3
Aggregated subject information of Nano Mapper's analysis functionality evaluation studies.

Attribute	Subjects' characteristics		
	USPTO patent documents	EPO patent documents	JPO patent documents
Gender	Male: 12; female: 12	Male: 15; female: 13	Male: 9; female: 4
Age	Mean: 23.57, Std dev: 9.48	Mean: 21.41, Std dev: 1.87	Mean: 22.25, Std dev: 1.54
Years of using computer	Mean: 11.83, Std dev: 3.31	Mean: 12.63, Std dev: 3.48	Mean: 13.25, S td de: 3.36
Years of using Internet	Mean: 10.17, Std dev: 3.33	Mean: 10.26, Std dev: 2.33	Mean: 10.83, Std dev: 2.82

Table 4
Results of Nano Mapper's analysis functionality evaluation studies.

Measure	Patent documents	Mean/SD	p-value
Satisfaction	USPTO patent documents	5.75/1.02	<.0001**
	EPO patent documents	5.76/0.78	<.0001**
	JPO patent documents	6.25/0.81	<.0001**

Note. Significance levels $^{*}\alpha = 0.05$ and $^{**}\alpha = 0.01$.

Hevner et al. [27] also mentioned that "effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundation, and/or design methodologies (p83, Table 1)." We believe that our study has contribution to the design foundation. Most existing Web-based knowledge mapping systems only focus on providing search functionality. Although it is crucial to have a good search support, we see evidence in our evaluations indicating that providing both search and advanced analysis support via a consistent user interface across different data sources is helpful in knowledge mapping.

Therefore, we believe that our study satisfies the "research contribution" guideline (# 4) of the design science research [27]; in this way, our study is "differentiated from the practice of design (p82)" or routine design.

6.2. Contributions to knowledge mapping

The development of Nano Mapper has made several important contributions by addressing the challenges facing knowledge mapping. First, although previous research has developed some advanced algorithms for knowledge mapping support, few studies, if any, implement Web-based systems for interactive search and analysis support. To address this, we design and implement Nano Mapper, an advanced Web-based knowledge mapping system that incorporates different search or analysis functions by utilizing appropriate text mining, social network, and information visualization techniques. Following the design science approach, we engage in the build-and-evaluate process toward creating an advanced system (i.e., an IT artifact) and then demonstrate its utility and efficacy in mapping nanotechnology documents [19,27]. Our system supports essential knowledge mapping tasks for researchers, technology managers, general users, grant administrators, business investors, and policy analysts and makers.

Second, we contribute to knowledge mapping research and practice by providing advanced analysis functionality with effective visualization displays. A review of the handful existing systems, stand-alone or Web-based, suggests a predominant focus on search support and lacking effective analysis capability and effective visualization design. Knowledge mapping needs to depict the overall development of the rapidly expanding domain, which demands advanced analysis support and easily comprehensible displays. Nano Mapper integrates appropriate text mining and social network analysis techniques to support various analysis tasks by researchers or practitioners, and present the results with easily understandable visualizations. The interactive interface design allows users to select parameter values, execute the analysis, and view the results in an interactive fashion.

Third, contrary to most prior research that focuses on a specific document corpus, we expand the document sources by including patent documents from major patent offices in the world as well as documents regarding the NSF-funded projects. Expanding a single document corpus to multiple document sources is an essential precursor to effective knowledge mapping in nanotechnology as it interfaces multiple disciplines and receives substantial research and development attention globally. Nano Mapper includes multiple, crucial document sources and adopts a uniformed database design and consistent user-end interface designs that allow users to search and analyze nanotechnology knowledge in dispersed, heterogeneous documents available from different sources. For increased document coverage, we will include additional document sources (e.g., academic articles) as they become available online, free of charge.

By addressing several important challenges of knowledge mapping in nanotechnology, we design and implement a system that provides better knowledge mapping support. With its advanced search and analysis functionality, Nano Mapper allows users to gain a good understanding of the overall knowledge landscape in a holistic, comprehensive, and convenient manner. With the unified access to multiple document resources by Nano Mapper, users likely will incur much less efforts and cognitive requirements when searching through vast collections of documents, compared searching the documents in each source separately. Our use of consistent interface designs across difficult document sources makes the system easy to use and facilitates desirable user familiarity. The visualization displays make the important search or analysis more prominent. In addition, functions such as name look-up, analysis result download and print, also help users to accomplish their search or analysis objectives.

6.3. Future research directions

Several directions are worth pursuing in future research. First, investigations of multilingual knowledge mapping are essential as the number of non-English documents keeps increases at a fast pace. Future research can extend the techniques and design of Nano Mapper to support knowledge mapping cross different languages. One promising approach is to include a machine translation module in Nano Mapper for supporting cross-lingual knowledge mapping search and analysis. Second, our evaluation results are obtained from subjects representative of general users. Future evaluations should include non-novice users such as researchers, experienced practitioners, and industry experts. In addition to the controlled experimentation methodology, future research may extend Nano Mapper evaluation by including other empirical methods. For example, conducting quasi-experiments or field studies in an organization or community setting is desirable and can generate more insights into how Nano Mapper can help various users to accomplish their tasks, together with valuable feedbacks for system enhancements. Third, for measuring users' subjective assessment of the performance of Nano Mapper, we used user satisfaction. We also leveraged two objective measures, task performance accuracy and task completion time for evaluating the search functions. While these measures are central metrics for system performance evaluations, we should include additional measurements in our future research, such as feature-specific measures that allow us to obtain in-depth and comprehensive understanding of users' perceptions towards Nano Mapper from various perspectives. Towards that end, the IS success model offers a logical starting point for exploration [17,18]. In addition, future research can leverage previous Knowledge Management literature to examine and identify the major factors related to the success of Web-based knowledge mapping systems (such as Nano Mapper) in the context of evolving scientific domains. Jennex and Olfman's Knowledge Management Success model [34] is one possible starting point. Using the IS success model as a theoretical basis, Jennex and Olfman [34] derived a Knowledge Management Success model from observations generated through a longitudinal study of knowledge management in an engineering organization. Future research can examine the model in the context of knowledge management in evolving scientific domains.

7. Conclusion

In this study, we follow the design science approach to build and evaluate an advanced, Web-based knowledge mapping system, Nano Mapper, which supports a comprehensive range of search and sophisticated analysis functions built upon existing text mining, network analysis, and information visualization techniques in knowledge mapping. Nano Mapper includes patents from major patent offices and grant documents from the NSF, against which comprehensive depictions of the overall development and collective knowledge can be generated. We conduct evaluation studies to examine the overall performance of Nano Mapper. The results show that Nano Mapper's search functionality is more effective than that of the benchmark systems. Furthermore, subjects report favorable satisfaction with the analysis functionality of Nano Mapper. Although we use the nanotechnology domain as a demonstration example, the framework of the Nano Mapper system is generic and can be applied to other domains. In sum, this study illustrates the desirability and viability of the design science approach to design, implement, and evaluate advanced systems that better meet users' requirements and needs.

Acknowledgments

This research is supported by the following awards: NSF, "SGER: Inter-Repository Patent Analysis to Understand Worldwide Nanotechnology Research and Development" CMMI-0738803, and "Mapping Nanotechnology Development," DMI-0533749. We would like to thank USPTO, EPO, JPO, and NSF for making their databases available for research purposes.

Appendix A. Patent search tasks used in the study comparing Nano Mapper and the EPO search system

T1. Please find the patent, of which the patent number is EP1679752. Please write down the title, the publication date, and the inventor(s) of the patent you identified. Please use the Patent Number Search function to finish this task.
(low-complexity)
T2. Please find one patent that has keyword "nanotechnology" in its title. Please write down the patent number, the title, and the publication date of the patent you identified. Please use the Quick Search function to finish this task. (The search function may return more than one patent, you just choose any one of them as your answer.)
(low-complexity)
T3. Please find one patent that has both keyword "nano" and keyword "conductor" in its abstract. Please write down the patent number, the title, and the publication date of the patent you identified. Please use the Quick Search function to finish this task. (The search function may return more than one patent, you just choose any one of them as your answer.)
(low-complexity)
T4. One of the nanotechnology keywords identified by domain experts is "nanotube". Please find one patent with a focus on nanotube and published by the applicant (the organization to which a patent is assigned to) organization "IBM" from United States. Please write down the patent number, the title, and the publication date of the patent you identified. Please use the Advanced Search function to finish this task. (The search function may return more than one patent, you just choose any one of them as your answer.)
(high-complexity)
T5. Please find one patent that has keyword "nanostructure" in its title, and was published/issued on 12/20/2006 (i.e., December 20, 2006). Please write down the patent number, the title, and the inventor(s) of the patent you identified. Please use the Advanced Search function to finish this task. (The search function may return

more than one patent, you just choose any one of them as your answer.)
(high-complexity)

T6. There are two patent classification systems: EPO Patent Classification and International Patent Classification. The International Patent class B32B5/16 represents "Characterized by features of a layer formed of particles, e.g. chips, chopped fibers, powder". Please identify one patent under this International Patent class with keyword "nanofiber" in its title, and has Robert Dubrow being one of its inventors. Please write down the patent number, the title, and the publication date of the patent you identified. Please use the Advanced Search function to finish this task. (The search function may return more than one patent, you just choose any one of them as your answer.)
(high-complexity)

Appendix B. Tasks used in the analysis functionality evaluation study involving EPO patents

T1. Please identify the top five Institutions from 2001 to 2006 based on the numbers of patents published on "title-abstract search". Please write down the name and the number of patents published by each of these top five institutions you identified. Please use the Statistics function to finish this task.

T2. Please identify the most active country (with the most number of links connected from/to it) by reading the country level citation network from 2001 to 2006 on "title-abstract search". Please write down the name of this most active country you identified. Please use the Citation Network function to finish this task.

T3. Content map shows the major and new research topics automatically learned from the patent documents. In a content map, there are some islands. The name on each island is one specific research topic, and the number in the parenthesis shows how many patents are under this topic. The warmer the color (with red being the warmest) of an island, the newer the topic is. Please identify three new research topics during 2000–2004 by reading the related content map. You may identify more than 3 new research topics, you can just pick any three from them as your answer. Please write down the names of the three new research topics you identified. Please use the Content Map function to finish this task.

References

[1] President's Council of Advisors on Science and Technology, The National Nanotechnology Initiative: Second Assessment and Recommendations of the National Nanotechnology Advisory Panel (April 2008).
[2] K. Andrews, Visualizing cyberspace: information visualization in the Harmony Internet browser, Proceedings of IEEE Symposium on Information Visualization (InfoVis'95), IEEE Press, Atlanta, Georgia, 1995, pp. 97–104.
[3] A. Bhattacherjee, Understanding information systems continuance: an expectation-confirmation model, MIS Quarterly 25 (3) (2001) 351–370.
[4] K. Börner, C. Chen, K.W. Boyack, Visualizing knowledge domains, Annual Review of Information Science and Technology Article 37 (2003).
[5] K.W. Boyack, B.N. Wylie, G.S. Davidson, Domain visualization using VxInsight for science and technology management, Journal of the American Society for Information Science and Technology 53 (9) (2002) 764–774.
[6] S.K. Card, G.G. Robertson, W. York, The WebBook and the WebForager: an information workspace for the World Wide Web, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'96), ACM Press, New York, 1996, pp. 111–117.
[7] C. Chen, Mapping Scientific Frontiers: The Quest for Knowledge Visualization, Springer Verlag, New York, 2003.
[8] H. Chen, M. Chau, Web mining: machine learning for Web applications, Annual Review of Information Science and Technology (ARIST) 38 (2004) 289–329.
[9] Y. Chen, Z.-Y. Liu, The rise of mapping knowledge domain, Studies in Science of Science 23 (2) (2005) 149–154.
[10] C. Chen, R.J. Paul, Visualizing a knowledge domain's intellectual structure, Computer 34 (3) (2001) 65–71.
[11] H. Chen, C. Schuffels, R. Orwig, Internet categorization and search: a self-organizing approach, Journal of Visual Communication and Image Representation 7 (1) (1996) 88–102.
[12] H. Chen, M.C. Roco, X. Li, Y. Lin, Trends in nanotechnology patents, Nature Nanotechnology 3 (2008) 123–125.
[13] A. Chevalier, M. Kicka, Web designers and web users: influence of the ergonomic quality of the web site on the information search, International Journal of Human-Computer Studies 64 (2006) 1031–1048.
[14] W. Chung, G. Lai, A. Bonillas, W. Xi, H. Chen, Organizing domain-specific information on the Web: an experiment on the Spanish business Web directory, International Journal of Human-Computer Studies 66 (2008) 51–66.
[15] Y. Dang, H. Chen, Y. Zhang, M.C. Roco, Knowledge sharing and diffusion patterns: international patent and patent family analysis, IEEE Nanotechnology Magazine 3 (3) (2009).
[16] Y. Dang, Y. Zhang, H. Chen, P.J.-H. Hu, S.A. Brown, C. Larson, Arizona Literature Mapper: an integrated approach to monitor and analyze global bioterrorism research literature, Journal of the American Society for Information Science and Technology 60 (7) (2009) 1301–1319.
[17] W.H. DeLone, E.R. McLean, Information systems success: the quest for the dependent variable, Information Systems Research 3 (1) (1992) 60–95.
[18] W.H. DeLone, E.R. McLean, The DeLone and McLean model of information systems success: a ten-year update, Journal of Management Information Systems 19 (4) (2003) 9–30.
[19] P.J. Denning, A new social contract for research, Communications of the ACM 40 (2) (1997) 132–134.
[20] S. Eggers, Z. Huang, H. Chen, L. Yan, C. Larson, A. Rashid, M. Chau, C. Lin, Mapping Medical Informatics Research, Medical Informatics: Knowledge Management and Data Mining in Biomedicine, Springer Science, 2005.
[21] S.G. Eick, J.L. Steffen, E.E. Sumner, Seesoft: a tool for visualizing line-oriented software, IEEE Transactions on Software Engineering 18 (11) (1992) 11–18.
[22] O.A. El-Sawy, Implementation by cultural diffusion: an approach to managing the introduction of information technologies, MIS Quarterly 9 (2) (1985) 131–140.
[23] M.J. Eppler, Making knowledge visible through intranet knowledge maps: concepts, elements, cases, Proceedings of the 34th Hawaii International Conference on System Sciences, 2001.
[24] E. Gansner, S. North, An open graph visualization system and its applications to software engineering, Software Practice and Experience 30 (11) (2000) 1203–1233.
[25] M. Hearst, TileBars: visualization of term distribution information in full text information access, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Press, Denver, Colorado, 1995, pp. 59–66.
[26] A.R. Hevner, A three cycle view of design science research, Scandinavian Journal of Information Systems 19 (2) (2007) 87–92.
[27] A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research, MIS Quarterly 28 (1) (2004) 75–105.
[28] J.B. Holmström, M. Ketokivi, A.-P. Hameri, Bridging practice and theory: a design science approach, Decision Sciences 40 (1) (2009) 65–87.
[29] Z. Huang, H. Chen, A. Yip, G. Ng, F. Guo, Z.-K. Chen, M.C. Roco, Longitudinal patent analysis for nanoscale science and engineering: country, institution and technology field, Journal of Nanoparticle Research 5 (2003) 333–363.
[30] Z. Huang, H. Chen, Z.-K. Chen, M.C. Roco, International nanotechnology development in 2003: country, institution, and technology field analysis based on USPTO patent database, Journal of Nanoparticle Research 6 (4) (2004) 325–354.
[31] Z. Huang, H. Chen, L.J. Yan, M.C. Roco, Longitudinal nanotechnology development (1991–2002): national science foundation funding and its impact on patents, Journal of Nanoparticle Research 7 (4–5) (2005) 343–376.
[32] A. Hullmann, Who is winning the global nanorace? Nature Nanotechnology 1 (2) (2006) 81–83.
[33] J. Iivari, A paradigmatic analysis of contemporary schools of IS development, European Journal of Information Systems Research 1 (4) (1991) 249–272.
[34] M.E. Jennex, L. Olfman, Assessing knowledge management success, International Journal of Knowledge Management 1 (2) (2005) 33–49.
[35] B. Johnson, B. Shneiderman, Tree-maps: a space-filling approach to the visualization of hierarchical information structures, Proceedings of IEEE Visualization'91 Conference, IEEE Press, New York, 1991, pp. 284–291.
[36] T. Kohonen, Self-organization and Association Memory, Springer, New York, 1989.
[37] T. Kohonen, Self-organizing Maps, Springer-Verlag, Berlin, 1995.
[38] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, A. Saarela, Self organization of a massive document collection, IEEE Transactions On Neural Networks 11 (3) (2000) 574–585.
[39] R. Kostoff, J. Murday, C. Lau, W. Tolles, The seminal literature of nanotechnology research, Journal of Nanoparticle Research 8 (2) (2006) 193–213.
[40] T.J. Kowalski, A. Maschio, S.H. Megerditchian, Dominating global intellectual property: overview of patentability in the USA, Europe and Japan, Journal of Commercial Biotechnology 9 (2003) 305–331.
[41] J. Lamping, R. Rao, P. Pirolli, A focus+context technique based on hyperbolic geometry for visualizing large hierarchies, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Press, New York, 1995, pp. 401–408.
[42] X. Li, H. Chen, Y. Dang, Y. Lin, C. Larson, M.C. Roco, A longitudinal analysis of nanotechnology literature: 1976–2004, Journal of Nanoparticle Research 10 (S1) (2008) 3–22.
[43] S.T. March, G. Smith, Design and natural science research on information technology, Decision Support Systems 15 (4) (1995) 251–266.
[44] M.S. Meyer, Patent citation analysis in a novel field of technology: an exploration of nano-science and nano-technology, Scientometrics 51 (1) (2001) 163–183.

[45] F. Narin, Patent bibliometrics, Scientometrics 30 (1) (1994) 147–155.
[46] M.E.J. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Physical Review E 64 (2001) 016132.
[47] J. Nunamaker, M. Chen, T.D.M. Purdin, Systems development in Information Systems research, Journal of Management Information Systems 7 (3) (1991) 89–106.
[48] A.A. Payne, A. Siow, Does federal research funding increase university research output? Advances in Economic Analysis and Policy 3 (1) (2003), (Article 1).
[49] R. Pearson, P.v. Schaik, The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task, International Journal of Human-Computer Studies 59 (2003) 327–353.
[50] E.F. Reid, H. Chen, Mapping the contemporary terrorism research domain, International Journal of Human-Computer Studies 65 (2007) 42–56.
[51] G.G. Robertson, J.D. Mackinlay, S.K. Card, Cone Trees: animated 3D visualizations of hierarchical information, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Press, New York, 1991, pp. 189–194.
[52] M.C. Roco, International perspective on government nanotechnology funding in 2005, Journal of Nanoparticle Research 7 (6) (2005) 707–712.
[53] M.C. Roco, R.S. Williams, P. Alivisatos, Nanotechnology Research Directions, Springer, Dordrecht, 2000.
[54] G. Salton, Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley Longman Publishing Co., Inc., 1989.
[55] J. Schummer, Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology, Scientometrics 59 (3) (2004) 425–465.
[56] R.M. Shiffrin, K. Börner, Mapping knowledge domains, Proceedings of the National Academy of Science 101 (2004) 5183–5185.
[57] B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, Proceedings of the IEEE Symposium on Visual Languages, IEEE Press, Washington, 1996, pp. 336–343.
[58] H.A. Simon, The Sciences of the Artificial, 3rd ed., MIT Press, Cambridge, MA, 1996.
[59] D.F. Sittig, Identifying a core set of medical informatics serials: an analysis using the MEDLINE database, Bulletin of the Medical Library Association 84 (2) (1996) 200–204.
[60] H. Small, Visualizing science by citation mapping, Journal of the American Society for Information Science 50 (9) (1999) 799–812.
[61] K.M. Tolle, H. Chen, Comparing noun phrasing techniques for use with medical digital library tools, Journal of the American Society for Information Science 51 (4) (2000) 352–370.
[62] A. Voutilainen, A Short Introduction to NPtool, available at, http://www2.lingsoft.fi/doc/nptool/intro/1997.
[63] S. Wasserman, K. Faust, Social Networks Analysis: Methods and Applications, Cambridge University Press, Cambridge, 1994.
[64] H.D. White, K. McCain, Visualizing a discipline: an author co-citation analysis of information science, 1972–1995, Journal of the American Society for Information Science 49 (4) (1998) 327–355.
[65] Y. Zhou, J. Qin, H. Chen, CMedPort: an integrated approach to facilitating Chinese medical information seeking, Decision Support Systems 42 (3) (2006) 1431–1448.
[66] D. Zhu, A.L. Porter, Automated extraction and visualization of information for technological intelligence and forecasting, Technological Forecasting & Social Change 69 (2002) 495–506.

Yan Dang is currently a visiting assistant professor in Information Systems Department at the University of Arkansas. She is also a doctoral candidate in Management Information Systems (MIS) at the University of Arizona. She received the BS and MS degrees in Computer Science from Shanghai Jiao Tong University. Her research interests include knowledge management, adoption of information technology, and human–computer interaction. She has published papers in Decision Support Systems, IEEE Intelligent Systems and Journal of the American Society for Information Science and Technology, etc.

Yulei Zhang is currently a visiting assistant professor in Information Systems Department at the University of Arkansas. He is also a doctoral candidate in Management Information Systems (MIS) at the University of Arizona. He received the BS degree in Computer Science and MS degree in Bioinformatics from Shanghai Jiao Tong University. His research interests include social media analytics, Web mining and human–computer interaction. He has published papers in Decision Support Systems, IEEE Intelligent Systems and Journal of the American Society for Information Science and Technology, etc.

Dr. Paul Jen-Hwa Hu is a professor and David Eccles Faculty Fellow at the David Eccles School of Business, the University of Utah. He has a Ph.D. in Management Information Systems from the University of Arizona. His current research interests include healthcare information systems and management, technology implementation management, human–computer interaction, and knowledge management. He has published papers in Decision Support Systems, Journal of Management Information Systems, Decision Sciences, Communications of the ACM, Journal of the American Society for Information Science and Technology, etc.

Dr. Susan A. Brown is an associate professor of Management Information Systems in the University of Arizona's Eller College of Management. She received her Ph.D. from the University of Minnesota and an MBA from Syracuse University. Her research interests include technology implementation, individual adoption, computer-mediated communication, technology-mediated learning, and related topics. Her research has been published in MIS Quarterly, Organizational Behavior and Human Decision Processes, IEEE Transactions on Engineering Management, Communications of the ACM, Journal of the AIS, and others.

Dr. Hsinchun Chen is a McClelland professor of information systems and the director of the Artificial Intelligence Lab, University of Arizona. He received the BS degree from the National Chiao-Tung University, Hsinchu, Taiwan, the MBA degree from SUNY Buffalo, and the PhD degree in information systems from New York University. He has authored/edited 18 books, 17 book chapters, and more than 180 SCI journal articles covering digital library, intelligence analysis, biomedical informatics, data/text/web mining, knowledge management, and web computing. He received the IEEE Computer Society 2006 Technical Achievement Award. He is a fellow of the IEEE and the AAAS.