Journal of Network and Computer Applications 36 (2013) 1441–1450

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Intelligent patent recommendation system for innovative design collaboration

Amy J.C. Trappey [a,*], Charles V. Trappey [b], Chun-Yi Wu [a], Chin Yuan Fan [c], Yi-Liang Lin [a]

[a] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Taiwan
[b] Department of Management Science, National Chiao Tung University, Taiwan
[c] Science & Technology Policy Research and Information Center, National Applied Research Laboratories, Taiwan

ARTICLE INFO

Article history:
Received 11 August 2012
Received in revised form
27 December 2012
Accepted 8 February 2013
Available online 15 March 2013

Keywords:
Design collaboration
Patent search
Behavior records
Patent recommendation system
Collaborative filtering algorithm

ABSTRACT

Patents' search is increasingly critical for a company's technological advancement and sustainable marketing strategy. When most innovative designs are created collaboratively by a diverse team of researchers and technologists, patent knowledge management becomes time consuming with repeated efforts creating additional task conflicts. This research develops an intelligent recommendation methodology and system to enable timely and effective patent search prior, during, and after design collaboration to prevent potential infringement of existing intellectual property rights (IPR) and to secure new IPR for market advantage. The research develops an algorithm to dynamically search related patents in global patent databases. The system clusters users with similar patent search behaviors and, subsequently, infers new patent recommendations based on inter-cluster group member behaviors and characteristics. First, the methodology evaluates the filtered information obtained from collaborative patent searches. Second, the system clusters existing users and identifies users' neighbors based on the collaborative filtering algorithm. Using the clusters of users and their behaviors, the system recommends related patents. When collaborative design teams are planning R&D policies or searching patents and prior art claims to create new IP and prevent or settles IP legal disputes, the intelligent recommendation system identifies and recommends patents with greater efficiency and accuracy than previous systems and methods described in the literature.

1. Introduction

Intangible assets such as Intellectual Property Rights (IPR) and trademarks are a significant part of a modern enterprise's net worth. In particular, the intellectual property registered internationally as patents are used to legally protect the proprietary technology and insure the market advantage of the firm, promote further commercialization, royalty, licensing, and sales. When new technology is developed and a patent is issued, official claims insure that the assignee(s) maintain their competitiveness by preventing others from using the patented technology without prior permission. As declared by US Patent Law (US Patent Act), patents are to be used to encourage and promote commercial development by providing legal protection. Whoever without authority makes, uses, imports, or sells any patented invention during the term of the patent stands in violation of the law and infringes upon the patent assignee. Certainly, companies with high quality patents hold a competitive and sustainable market position.

The World Intellectual Property Organization (WIPO) reports that the numbers of applications have reached a record high of a half million patents per year. When companies attempt to search, interpret, compare, and classify patent documents, they are overwhelmed by the difficulty of reviewing, analyzing, and synthesizing the illustrations, information, claims, and technical knowledge. Thus, computer assisted patent information and knowledge management systems are needed to facilitate the manual processing, organization, and knowledge management of relevant patents.

Companies and individual inventors can win or lose substantial profits and market advantage if their innovative designs unknowingly conflict or infringe upon existing technology or if others are misappropriating their pre-existing claims. There is a constant need to search, review, and interpret patents in various patent databases to understand inventions and prior arts (1) prior to creating new patent applications or new product commercialization and (2) to maintain legal authority over existing IP. Therefore, patent search and recommendation methodologies are a critical function of a patent knowledge system and knowledge management program. When users lack specific domain knowledge, their patent search efforts are often impeded and result in useless searches. Patent documents use a plethora of domain concepts to describe the invention for comprehensive IP right protection. Thus, it is difficult for the majority of patent researchers (especially in a

* Corresponding author. Tel.: +886 933035375; fax: +886 35722204.
E-mail addresses: trappey@ie.nthu.edu.tw (A.J.C. Trappey), trappey@faculty.nctu.edu.tw (C.V. Trappey), d9534524@oz.nthu.edu.tw (C.-Y. Wu), cyfan@stpi.narl.org.tw (C.Y. Fan), s9834507@m98.nthu.edu.tw (Y.-L. Lin).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.02.035

multi-lingual global IP environment) to accurately search for and retrieve related patent documents without using a computer based and intelligent recommendation system.

While the numbers of global patent applications continues to grow, the patent recommendation system must be useful for finding related patents rapidly and effectively. Thus, this research provides an intelligent recommendation methodology based on the records of the users' search behavior. The research implements a collaborative filtering algorithm and a clustering method to construct the intelligent patent recommendation system platform. When general users search patents, the platform automatically identifies and recommends related patents. This paper is organized into several sections. In Section 2, the related literature is reviewed with a focus on recommendation systems and web search methods. In Section 3, the proposed recommendation methodology and algorithms are formulated and described. Section 4 describes the system framework, prototype implementation, and the case study based on a solar cell technology patent search with recommendation results. Finally, the Section 5 summarizes the research outcomes, contributions, and provides directions for future research.

2. Literature review

Patents protect the invention and intellectual property rights (IPR) claimed by the inventor. Inventors write their research and development (R&D) results or innovative achievements as patent documents following a required patent document format and specification guideline. The patent documents are then evaluated for qualification by the issuing patent office. During the issuance time, the patent owner (also called the patent assignee) can take legal action to prevent unauthorized manufacturing, selling or usage of the invention. This section describes related literature covering patent search and classification, patent analysis, and patent recommendation systems.

Enterprises often use various patent databases to search patent documents, e.g., the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the World Intellectual Property Organization (WIPO). In addition to these national and international open databases, there are integrated databases which require subscription fees or database purchases, such as Delphion and EPO PATSTAT. These databases generally use three functions in the patent search, i.e., a simple keyword search, an advanced metadata search (e.g., keywords, assignee, inventors, and year), and a patent number search. When users search patents without definite directions or specific keywords, they usually find the search results with many type-1 or type-2 errors where they either miss patents or retrieve the wrong patents. Therefore, many automatic classification methods have been proposed to help facilitate patent search. If patents are classified before searching, time and cost of search can be reduced. The EPO uses an auto-classification method based on the *k*-nearest neighboring and clustering algorithm (Krizer and Zacca, 2002). Researchers have depicted that the patent classification can be more accurate by considering both the patent metadata and the full-text of the patent (Richter and MacFarlane, 2005). Trappey et al. (2010) propose a non-exhaustive clustering method to group patent documents into overlapping clusters. This approach determines whether a given patent can possibly be categorized into multiple clusters, which is consistent with the principle of a patent allowing multiple claims. Further, Chiang et al. (2011) developed an intelligent system for automated binary knowledge document classification and content analysis. The system is constructed using a back-propagation artificial neural network, hierarchical ontology, and normalized term frequency methods to improve patent classification in a binary and hierarchical fashion. Thus, the system iteratively identifies patents until a sufficiently reduced number of highly related patents are collected.

After patent search and classification, the collected patents are analyzed to extract detailed information. Patent analysis includes the identification of technology trends as well as the performance of legal due diligence. There are four general analytical approaches used including time series analysis, patent citation analysis, international patent classifications (IPC) analysis, and the construction of patent maps. Time series analysis is used to analyze the change in the number of patents applied for overtime which in turn can be related to the development of the technology life cycle. Patent citations determine the relation between two different patents by using their citations and other references. For various technologies, IPC analysis helps determine which technology is being or not being developed. If some IPC regions are limited to a small numbers of patents, then there may be a R&D bottleneck and business potential if a solution can be found. The result of IPC analysis provides information to support government or enterprise R&D strategy. Finally, patent maps are used to depict the potential technological relationships between two different patent groups. The patent map describes different patent groups that belong to technology groups or assignees Huang and Li (2010).

User feedback is a type of rating or voting method for users to express their personal satisfaction with a search. This behavior is divided it into explicit or implicit feedback (Nichols, 1997). The explicit feedback represents information provided directly by the user such as personal information, common replies, survey results, and work experience. Implicit feedback is collected indirectly from user and is usually extracted from user's browsing records or query logs. Oard and Kim (1998) report that the user's implicit information can be extracted and classified as three behavioral types which in turn serve as the inference data for the recommendation system.

Researchers have also considered different techniques in defining user behaviors and methods to trace their queries and feedback. For example, a service match maker plays an important role in ensuring the connectivity between the user and the service provider. However, the lack of relevant service domain knowledge and incorrect service queries prevents the semantic service match makers from identifying the service concepts that correctly represent the service requests. Wang et al. (2007) demonstrate that the information coverage and update problems are a common bottleneck for current web search engines. These problems limit the offers of service and make the resolution of complaints difficult to achieve. To solve these problems, a new search algorithm based on DNS is proposed in their research. This system adopts a layered distributed architecture, similar to DNS, which is different from current commercial search engines. Dong et al. (2011) present a novel semantic similarity model for describing the service ontology environment whereas Bouras and Poulopoulos (2012) propose a web personalization mechanism based on dynamic creation and automatic updates of user profiles to better match users preferences. This approach assumes that a user's profile is affected by other user's grouping details which are constructed with similar profiles. As a result, a real-time user-centric document grouping mechanism is implemented to support the web personalization system and provide data for experimental evaluation of the system.

Researchers (de la Torre-Diez et al., 2013) use generic and selective filters set up by the administrator from the module RSS_PROYECT installed in Joomla. The generic filter allows a search of the words included in a series of sources indexed by the user. The filter categorizes all sources that contain the word without exception. Different languages such as PHP, MySQL, HTML, XML and the Application Program Interface (API) of Joomla were used to evaluate the results. The results are favorable for the selective filter and strongly favorable for the generic filter. Better than average processing times were obtained for RSS_PROYECT with respect to other modules using Joomla.

Li et al. (2012) develops a novel framework which is presented and implemented for classifying patents according to the levels of invention as defined by TRIZ theory. Liang et al. (2012) proposed an Issue, Solution and Artifact Layer (ISAL) model for design rationale representation. The research focuses on algorithm design to discover design rationale from design documents according to ISAL modeling. Moreover, Liang et al. (2012) use text mining as their primary method for analyzing issues, solutions and artifact layers. Text mining is useful for identifying keyword terminology, but requires additional analysis to correlate and organize the information. Li et al. (2012) use TRIZ as a method for identifying problem solutions. However, TRIZ must consider many different strategies, such as level of invention, which makes the system less practical for non-domain specialists.

Most recommendation approaches utilize implicit data extracted from user's behavior records. Kelly and Teevan (2003) integrate the research proposed by Oard and Kim (1998) and divide search behaviors into five different types including examine, retain, reference, annotate, and create. For the above approach, it is necessary to record the user's operational history such as searching, browsing, markup, or editing using IT techniques and then transform and save the records in a standard format. From these records, a user's personal profile is created.

Implicit data can also be utilized by a recommendation system. Lee et al. (2008) use the opening time of commodities sold and the time that consumer's buy as a data source, analyze the consumer's preferences, and then make recommendations for further commodities trading. The research considers that when buying times are closer together, the commodity better fits consumer demand. Given the commodity's open selling time, commodities are assigned different weights, and inferences are made that there are other commodities that buyers might be in interested and recommendations are made. This type of data extraction helps avoid personal and subjective influence and is sufficient for model training.

Patent specific search and analysis may be adapted to similar recommendation systems. After a patent search, a user may have difficulty in identifying patents of interest or related patents. Thus, this research proposes an intelligent recommendation methodology and system for patent search. A recommendation system is considered to be an information filtering system that effectively reduces the cost and time of search activities. Resnick and Varian (1997) report that information filtering systems are widely applied in e-business and help users grasp information rapidly. The recommendation system filters and analyzes the feedback of users and helps with the specification and classification of items. The system is constructed as a dynamic model to collect information based on user's requirements. Moreover, most recommendation systems are based on two types of mechanisms (Ansari et al., 2000). One type uses collaborative filtering and the other uses content based filtering. This research focuses on the collaborative filtering since it relies on social or personal network recommendations. This mechanism clusters the target and similar behavior of other users in a closed group. By defining the group's common interests and similar behavior model, the mechanism infers the information or products (in this case, patents) that are of common interest. Bhavnani et al. (2008) propose a qualitative study of experienced patent searchers. The research assumes that the professional searchers will use well-formed search strategies that can rapidly and effectively search for the related patents and identify the novelty of invention. Many researchers have developed mechanisms to build recommendation systems for other applications. There is one for recommending movies to viewers which merges combinations of features and attributes (Nazim-uddin et al., 2009). de Campos et al. (2010) propose a hybrid system which uses a Bayesian network model to determine the weights of the target search and then provides a recommendation result. Barragáns-Martínez et al. (2010) propose a novel Web 2.0 TV program recommendation system. The hybrid approach combines content filtering techniques with collaborative filtering while providing the advantages of a social network. In order to eliminate the most serious limitations of collaborative filtering, an item-based collaborative filtering algorithm was implemented to improve performance. The resulting application simplifies the task of selecting programs to watch on TV.

3. Methodology

This research accumulates users' behavior records and the related patent search information and applies collaborative filtering to recommend patents. The research methodology framework (Fig. 1) is divided into three parts which defines the behavior types and then records users' behavioral operations based on pre-defined types. The methodology summarizes the users' behavior records based on specific search conditions and when the target user searches patents, the methodology filters the results for the patent recommendation system. The following section describes the patent collaborative filtering process as the core of the dynamic patent recommendation methodology.

3.1. User behavior record analysis

In order to define the users' operational complexity under differing search conditions, five behavior types classify and record the user' behavior over time. The users' searching and viewing frequencies are accumulated and the analysis records and historical bookmark records are logged. Finally, these data and records are exported to the users' behavior database. The behavior types,

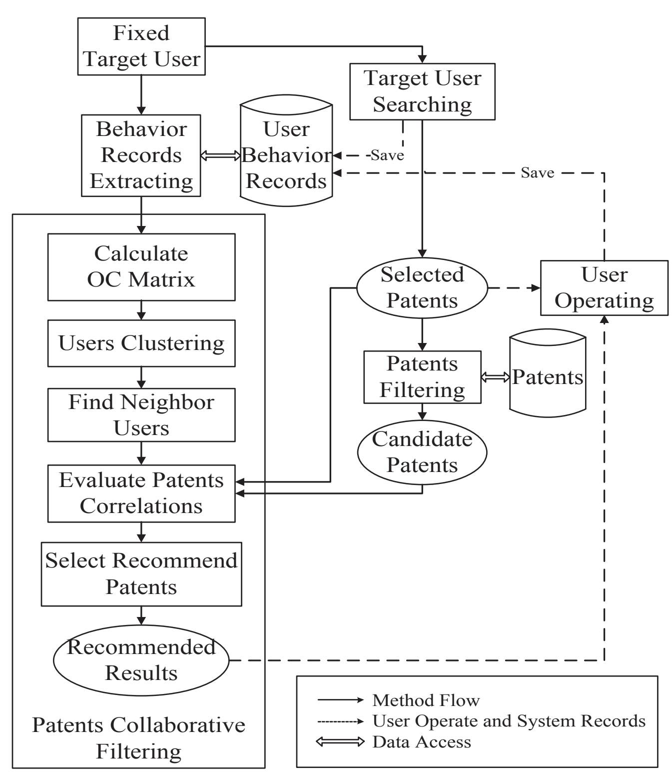


Fig. 1. Research procedure.

Table 1
Behavior type and weight.

Behavior	Sub Behavior	Notation	Weight
Search	Custom search	S1	W_{S1}
	Patent number search	S2	W_{S2}
	Industry patent search	S3	W_{S3}
	Technology patent search	S4	W_{S4}
View	Patent view	V1	W_{V1}
	Patent comparison	V2	W_{V2}
Analysis	Two-dimension chart	A1	W_{A1}
	Three-dimension chart	A2	W_{A2}
	Patent quality analysis	A3	W_{A3}
Bookmark	Bookmark after Custom search	B1	W_{B1}
	Bookmark after Patent number search	B2	W_{B2}
	Bookmark after Industry patent search	B3	W_{B3}
	Bookmark after technology patent search	B4	W_{B4}
Export	Patent list export	E1	W_{E1}
	Single patent export	E2	W_{E2}

their sub-behaviors, notations, and weights are presented in Table 1.

The search is divided into four sub-behavior types by the search type and weights are assigned according to the search complexity. Bookmarks are divided into four sub-behavior types based on the patent search approach. The weights of the bookmark sub-types reflect the operation complexity. The algorithm distinguishes between single patent views and comparative patent views. The comparative patent view is used to select two patents of interest for simultaneous analysis. Thus, a lower weight is assigned to the behavior of single views and higher weights are assigned to the comparison views. The sub-behaviors, such as patent charting and patent quality analysis (Trappey et al., 2012) are an additional behavioral sub-type analysis. Patent charting analyzes the metadata attributes from patents and includes statistical patent trends based on the international patent classification (IPC), assignees, and countries of applicants. The research further creates a two-dimension chart and a three-dimension chart based on the analyses. The more dimensions considered, the greater weight assigned to the behavior. Patent quality analysis is used to evaluate the patent value based on patent indicators. Thus, the patent quality analysis sub-type is given the highest weight among the analysis sub-types. Finally, export behavior is divided into general patent list exports and specific single patent exports. The former exports a series (a list) of all patent information and the latter only extract a single patent's information. Since selecting a single patent requires in-depth understanding and interest, a higher weight is assigned to the single patent export. The user behavior record is used by the system to evaluate each user's patent search process and grade the operational complexity.

3.2. Operation complexity function definition

The Operation Complexity (OC) value defined by Edwards and Barron (1994) is used to assign the weights of the pre-defined behavior types in the previous section (Table 1). The patent's attributes and patent search approach defines several data items such as patent number (PN), international patent classification (IPC), inventors (IN), assignees (AN), key phrase (KP), industry (INDT), and technology type (TECH). The OC function is shown in Eq. (1) and the variables are defined in Table 2.

$$\mathrm{OC} = (W_s \times S + W_V \times V + W_A \times A + W_B \times B + W_E \times E) \quad (1)$$

Table 2
OC function variables' definition and weight.

Variable	Definition	Weight
OC	Operation complexity	–
S	Search behavior Score	W_S
V	View behavior Score	W_V
A	Analysis behavior score	W_A
B	Bookmark behavior score	W_B
E	Export behavior score	W_E

Table 3
Operation complexity matrix.

User	Search Condition (SC)				
	SC(1)	SC(2)	SC(j)	…	SC(s)
$\mathbf{User_1}$	OC_{11}	OC_{12}	OC_{1j}	…	OC_{1s}
$\mathbf{User_2}$	OC_{21}	OC_{22}	OC_{2j}	…	OC_{2s}
$\mathbf{User_i}$	OC_{i1}	OC_{i2}	OC_{ij}		OC_{is}
…		…		…	…
$\mathbf{User_m}$	OC_{m1}	OC_{m2}	OC_{mj}	…	OC_{ms}

where
OC_{Tj} is the target user's operation complexity for the jth search condition;
Index T represents the target use;
Index j represents the jth search condition, and the total number of conditions is s, $j<=s$; and
OC_{ij} represents the ith user's OC value for the jth search condition.

3.3. User cluster analysis

After collecting the users' behavior records, the K-medoids clustering algorithm is used to group the target user and neighbors using the operation complexity (OC) matrix shown in Table 3. K-medoids minimizes outliers from over influencing the clustering result and better selects a real user as a cluster center (not a pseudo center as can occur with the K-means approach).

K-medoids helps decrease the calculation time. Furthermore, using an actual user as the cluster center is more appropriate when analyzing non-numerical data (Basumallick and Wong, 1996). The algorithm selects k objects as the initial centers with the target user (T) among them. Then the distances between the object and centers are calculated and the cluster with the shortest distance between its center and the object is selected. After finishing the first iteration, the algorithm randomly selects the object to replace its original center. If the new center produces better cluster results, then the clustering calculation continues If not, the clustering algorithm stops and provides the final result.

Different clustering results are generated with different cluster numbers (k). To better select k, the minimizing root mean square standard deviation (RMSSTD) and maximizing R-squared (RS) value are used as recommended by Sharma (1996). RMSSTD is used to calculate the data's homogeneity within each cluster. A smaller RMSSTD shows higher homogeneity within clusters. RS is used to measure the average divergence of two different clusters. The larger RS represents larger differences between clusters which is the desired outcome. The formulas for RMSSTD and RS are shown in Eqs. (2) and (3).

$$RMSSTD = \left[\sum_{\substack{i=1\dots nc \\ j=1\dots v}} \sum_{k=1}^{n_{ij}} \frac{(x_k - \overline{x_k})^2}{\sum_{\substack{i=1\dots nc \\ j=1\dots v}} (n_{ij}-1)} \right] \quad (2)$$

n_c: cluster numbers,
v: data's dimension

n_j: dimension j's data numbers
n_{ij}: dimension j's data numbers in cluster i

$$RS = \frac{SS_b}{SS_t} = \frac{SS_t - SS_w}{SS_t} = \frac{\{\sum_{j=1\ldots v}[\sum_{k=1}^{n_j}(x_k - \overline{x_k})^2]\} - \{\sum_{\substack{i=1\ldots c \\ j=1\ldots v}}[\sum_{k=1}^{n_{ij}}(x_k - \overline{x_k})^2]\}}{\sum_{j=1\ldots v}[\sum_{k=1}^{n_j}(x_k - \overline{x_k})^2]} \quad (3)$$

SS_b: sum of square between the clusters
SS_t: total sum of square
SS_w: sum of square within the cluster

3.4. Filtering recommended patents

Since the number of searched patents is quite large, the matrix is difficult to build effectively for system processing. After user clustering, the patents' correlations within a given cluster are calculated as the basis for recommending related patents. Based on the clustering result, the target user and the target user's neighbors are identified as a group. The patents selected by the target user's neighbors are called the selected patents (SPs). Frequently appearing IPCs of selected patents are summarized and the other patents which contain the same IPCs from the database are chosen as the candidate patents (CPs). Thus, patents with little relationship are filtered and eliminated to avoid complex calculations.

After finding neighboring users and CPs, the target user and the neighbors' behavior records are summarized. The collaborative operation index is calculated (Table 4) using notations shown and explained in Table 5. The collaborative operation index *ij* represents the operating count of *i*th patent and *j*th patent under different collaborative operations. The *i*th patent and *j*th patent can be the selected patents (S1, S2… Sm) or the candidate patents (C1, C2… Cn). Thus, the correlation score (called Co Score) is calculated individually by multiplying the pre-defined weights and the count of each collaborative operation index (Formula (4)). After calculating all patent combination correlation scores, the scores are normalized using the maximum value based on collaborative operation indexes. Table 6 shows the patent correlation matrix, where *DCij* represents the degree of correlation between *i*th patent and *j*th patent.

$$\text{Co score} = W_{\text{CoV1}} \times \text{CoV1} + W_{\text{CoV2}} \times \text{CoV2} + \cdots + W_{\text{CoE2}} \times \text{CoE2} \quad (4)$$

Through this process, the candidate patents which are highly related with the selected patents are identified. The recommended patents are sorted by the highest score calculated by the degree of the patents' correlation compared with each selected patent. The final result is a method to automatically recommend patents to the user.

4. System construction and case study

The system integrates patent search and patent recommendation functions. There are four modules within the patent recommendation unit. These four modules are the user's behavior record, related user clustering, patent filtering, and system parameter management. The system framework describes the relation between user, system platform, and database (Fig. 2). The modules and the detailed actions of the patent recommendation function are shown in Fig. 3.

The recommendation system was built for the solar energy alliance (http://www.wheeljet.com.tw/CIGS/) as a case study to validate the methodology. The system collects the related solar technology patents from the USPTO database and includes the key functions of patent searching, technology classification, industry classification, patent analysis, and patent bookmarking. The system also records the users' operation information for patent recommendation calculations. In the case study, the system collects twenty members' behavioral information which is used to cluster the related users and to recommend the potential non-searched patents to a given user.

The patent recommendation system is an innovative platform which recommends specific and relevant patents based on the previous search results of clustered peers. Thus, the system considers different user situations and not only the experimental results of the domain experts. Moreover, in the case study,

Table 4
Collaborative operation indexes.

Patents		**Indexes**				
Patent i	**Patent j**	**CoV1**	**CoV2**	**CoA1**	…	**CoE2**
S1	S2	$\text{CoV1}_{S1,S2}$	$\text{CoV2}_{S1,S2}$	$\text{CoA1}_{S1,S2}$	…	$\text{CoE2}_{S1,S2}$
S2	C1	$\text{CoV1}_{S2,C1}$	$\text{CoV2}_{S2,C1}$	$\text{CoA1}_{S2,C1}$	…	$\text{CoE2}_{S2,C1}$
S1	C1	$\text{CoV1}_{S1,C1}$	$\text{CoV2}_{S1,C1}$	$\text{CoA1}_{S1,C1}$	…	$\text{CoE2}_{S1,C1}$
…	…	…	…	…	…	…
Sm	Cn	$\text{CoV1}_{Sm,Cn}$	$\text{CoV2}_{Sm,Cn}$	$\text{CoA1}_{Sm,Cn}$	…	$\text{CoE2}_{Sm,Cn}$

Table 6
Patent correlation matrix.

	SP(1)	**SP(2)**	…	**SP(m)**	**CP(1)**	**CP(2)**	…	**CP(n)**
SP(1)		$DC_{S1,S2}$	…	$DC_{S1,Sm}$	$DC_{S1,C1}$	$DC_{S1,C2}$	…	$DC_{S1,Cn}$
SP(2)	$DC_{S2,S1}$			$DC_{S2,Sm}$	$DC_{S2,C1}$	$DC_{S2,C2}$	…	$DC_{S2,Cn}$
…	…	…		…	…	…	…	…
SP(m)	$DC_{Sm,S1}$	$DC_{Sm,S2}$	…		$DC_{Sm,C1}$	$DC_{Sm,C2}$	…	$DC_{Sm,Cn}$
CP(1)	$DC_{C1,S2}$	$DC_{C1,S2}$	…	$DC_{C1,Sm}$		$DC_{C1,C2}$	…	$DC_{C1,Cn}$
CP(2)	$DC_{C2,S2}$	$DC_{C2,S2}$	…	$DC_{C2,Sm}$	$DC_{C2,C1}$		…	$DC_{C1,Cn}$
…	…	…	…	…	…	…		…
CP(n)	$DC_{Cn,S1}$	$DC_{Cn,S2}$	…	$DC_{Cn,Sm}$	$DC_{Cn,C1}$	$DC_{Cn,C2}$	…	

Table 5
Collaborative operation description and notation.

Operation	**Description**	**Notation**
Single patent view	Select specific patent to view the detailed content	CoV1
Patent comparison	Choose any two patents to make comparisons	CoV2
Two-dimension patent chart	Choose some patents and two attributes to draw relationship chart	CoA1
Three-dimension patent chart	Choose some patents and three attributes to draw relationship chart	CoA2
Patent quality analysis	Select several patents for patent quality evaluation	CoA3
Patent bookmark	Bookmark the patents after search	CoB1
Non-removed patent bookmark	Do not remove the original bookmark patents after search	CoB2
Patent list export	Choose some patents and export their important information	CoE1
Single patent export	Select specific patent and export the complete patent information	CoE2

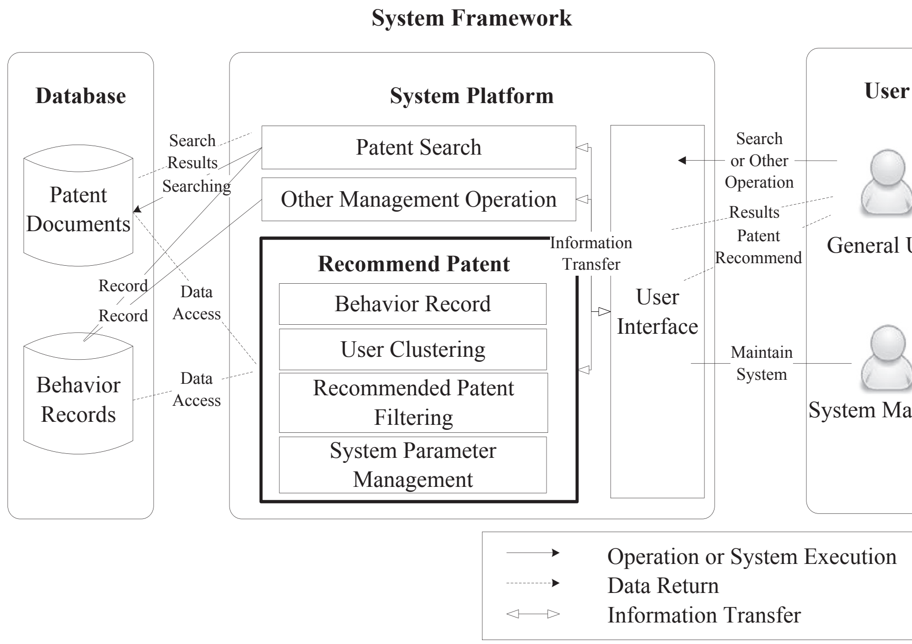


Fig. 2. System framework.

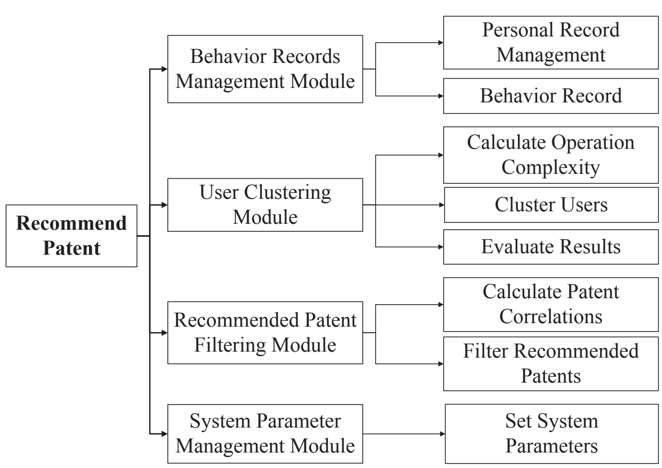


Fig. 3. The modules and actions of patent recommendation function.

sampling is based on the central limit theory. The research considers that users are normally distributed ranging from less experienced patent searchers to very experienced domain experts. Thus, the patent recommendation methodology is proved to work in a common collaborative R&D environment consisting of a wide range of users.

4.1. Calculating operation complexity

The patent search for the semiconductor material Copper Indium Gallium Selenide (CIGS) is used as the case study and methodology test. CIGS are thin-film semiconductors made with copper, indium, gallium, and selenium which are elements used as light absorbing materials for thin-film solar cell production. Nine hundred and eighty-seven related patents were used to build the patent database which includes the behavior records of the twenty researchers that conducted the patent search operations.

The algorithm selected user id “ieAC01” as the target user. The system classifies the users' data according to the key phrases used for searching. After filtering phrases, the system calculates the frequency of key phrases used by the patent search function. The system clusters the users preliminarily into four groups as shown in Table 7 and makes comparisons using the clustering result. Next, the system summarizes the search conditions and counts the number of sub-behavior operations across all behavior types: search, view, analyze, bookmark and export. The system automatically calculates the scores for the five behavior types as shown in Table 8 based on the pre-defined weights of sub-behaviors as listed in Table 1.

The operation complexity (OC) for the different search conditions across different search operations are listed in Table 9. The system normalizes the scores based on the maximum score of each behavior type. Thus, all the scores fall between 0 and 1. Next, the system calculates the OC scores using Eq. (1) with the given parameter values as listed in Tables 10 and 11. Finally, the OC scores for the case are summarized into a matrix as shown in Table 12.

4.2. User clustering

The OC matrix is the input data for the clustering analysis and shows the users' OC scores under different search conditions. The cluster groups are set to sizes of 2, 3 or 4. Table 13 shows their RMSSTD and RS results with size 4 being the best cluster number. The size of 4 clusters best satisfies the statistical requirements for the most suitable number of groups (Table 14). User id “ieAC01” is the center of the cluster group, and the group contains “ieAC03,” “ieAC09,” “ieAC10,” and “ieAC17.” Therefore, these four neighbors “ieAC01” and their records are the references for the recommendation process. The final cluster result can be compared with the user's pre-classification (Table 7). The case study shows that they have the same group numbers, and except for the groups centered on “ieAC14” and “ieAC15” there are only slight differences (ieAC19). The other groups are the same. The clustering analysis performs consistently and the system automatically infers recommendations based on the users' clustering results.

The proposed recommendation approach extracts relevant patents by analyzing users’ behavior records and calculating each patent's collaborative contribution. Thus, the recommendation system clusters the target user with the members of the system according to the related patent search strategies and the

Table 7
User pre-classification.

Commonly Used Key Phrase	Members
coevaporation, evaporation, precursor, etc.	ieAC01, ieAC03, ieAC09, ieAC10, ieAC17
printing, non-vacuum, precursor, etc.	ieAC05, ieAC06, ieAC11, ieAC12, ieAC18
electrodeposition, RTP, annealing, etc.	ieAC07, ieAC08, ieAC14, ieAC16, ieAC19
sputtering, vacuum, RTP, etc.	ieAC02, ieAC04, ieAC13, ieAC15, ieAC20

Table 8
Sub-behavior count summary (partial).

User	**Field**	**Key Term**	**S1**	**S2**	**S3**	**S4**	**V1**	**V2**	**A1**	**A2**	**A3**	**B1**	**B2**	**B3**	**B4**	**E1**	**E2**
ieAC01	Tech	Coevaporation	0	0	1	12	14	0	5	2	0	0	0	0	24	4	0
ieAC01	AN	Sotec Corp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ieAC01	KP	Solar Cell	3	0	0	0	5	0	0	0	0	2	0	0	0	0	0
ieAC01	PN	US4105471	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ieAC01	IPC	H01C	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ieAC01	KP	Coevaporation	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ieAC02	Tech	Sputtering	0	0	0	11	9	5	4	3	0	0	0	0	25	0	0
ieAC02	Tech	Electrodepositon	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ieAC02	PN	US4528082	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Table 9
Operation complexity table (partial).

User	**Search**	**Condition**	**S**	**V**	**A**	**B**	**E**
ieAC01	Tech	Coevaporation	3.6	11.3333	1.5	7.2	1.3333
ieAC01	AN	Sotec Corp.	0.1	0	0	0	0
ieAC01	KP	Solar Cell	0.3	1.6667	0	0.2	0
ieAC01	PN	US4105471	0.4	0	0	0	0
ieAC01	IPC	H01C	0.2	0.3333	0	0	0
ieAC01	KP	Coevaporation	0.1	0	0	0	0
ieAC02	Tech	Sputtering	3.3	6.3333	1.6667	7.5	0
ieAC02	Tech	Coevaporation	0.3	0	0	0	0
ieAC02	Tech	Electrodepositon	0.3	0	0	0	0
ieAC02	PN	US4528082	0.4	0.3333	0	0	0
…	…	…	…	…	…	…	…
Max			4.2	11.3333	3.8333	13.2	2

Table 10
Sub-behavior weights.

Behavior	Sub-behavior	Weight (%)
Search	W_{S1}	10
	W_{S2}	40
	W_{S3}	20
	W_{S4}	30
View	W_{V1}	33
	W_{V2}	67
Analyze	W_{A1}	17
	W_{A2}	33
	W_{A3}	50
Bookmark	W_{B1}	10
	W_{B2}	40
	W_{B3}	20
	W_{B4}	30
Export	W_{E1}	33
	W_{E2}	67

Table 11
The behavior types weights.

Behavior Type	Weight	Behavior Type	Weight (%)
W_S	5.00%	W_B	25.67
W_V	9.00%	W_E	45.66
W_A	15.67%		

Table 12
Operation complexity matrix (partial).

User	Tech: Coevaporation	AN: SunPower Corp.	Tech: Printing	Tech: Sputtering	IPC: H01L
ieAC01	0.63	0	0	0	0
ieAC02	0.0029	0	0.0029	0.2957	0
ieAC03	0.1037	0	0	0	0
ieAC04	0.0029	0	0	0.0376	0
ieAC05	0	0	0.1524	0	0
ieAC06	0	0	0.0956	0	0
ieAC07	0.0213	0.0242	0	0	0
ieAC08	0	0	0	0	0
ieAC09	0.0196	0	0.0055	0	0.0189
ieAC10	0.2833	0	0	0	0

Table 13
Cluster result evaluation.

Cluster Numbers	RMSSTD	RS	RMSSTD+1/RS
2	0.0503	0.2932	3.4609
3	0.0422	0.5298	1.9297
4	0.0389	0.6246	1.6399

evaluation of RMSSTD. Following, the system recommends the related patents by analyzing the information of the patent search collaborative operations. Thus, the recommend patents are extracted by the users who use similar search strategies. Although some previous researches are applying the concepts and principles of collaborative filtering and collaborative operation evaluation (Nichols, 1997; Bhavnani et al., 2008; Nazim-uddin et al., 2009; de Campos et al., 2010; Barragáns-Martínez et al., 2010; Herlocker et al., 2004), this research has developed the unique algorithm specifically for patent recommendation application. This research focuses on recommending patents collected and prioritized based

Table 14
Final cluster result.

Cluster center	Cluster members
ieAC01	**ieAC01, ieAC03, ieAC09, ieAC10, ieAC17**
ieAC12	ieAC05, ieAC06, ieAC11, ieAC12, ieAC18
ieAC14	ieAC07, ieAC08, ieAC14, ieAC16
ieAC15	ieAC02, ieAC04, ieAC13, ieAC15, **ieAC19**, ieAC20

Table 15
Collaborative operation weight.

Collaborative Operation	Weight	Collaborative operation	Weight
W_{CoB1}	2	W_{CoV1}	1.8
W_{CoB2}	0.5	W_{CoV2}	2.5
W_{CoE1}	1.5	W_{CoA1}	1
W_{CoE2}	2	W_{CoA2}	1.2
		W_{CoA3}	1.4

on the patent search characteristics and the search's collaborative operation records of the peers with the common domain interests.

4.3. Patent recommendation and inference

User id “ieAC01” searches patents for low-cost manufacturing solar cells and focuses on the assignee “Georgia Tech Research Corporation.” The search results include two selected patents (SPs), US5510271 and US5766964. The research collects another 985 related patents as candidate patents (CPs). The last step for the methodology is to infer the recommended patents from the CPs which are sufficiently related to SPs. The proposed collaborative filtering is based on neighbor's ratings which calculate the search conditions or user's operating status. The filtering analyzes users' records and selects appropriate patents to recommend. After user clustering, the system confirms the neighbor users' operating status for patent recommendation. The system focuses on the patents used by neighbors' collaborative operations and calculates the collaborative values between any two patents. The system

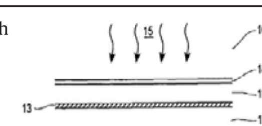
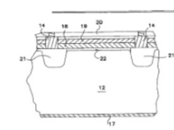
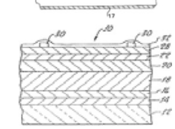
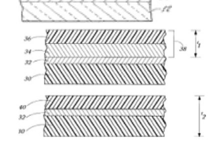
Table 16
Summary of patent collaborative operation counts (partial).

Patent I	Patent II	CoV1	CoV2	CoB1	CoB2	CoE1	CoE2	CoA1	C0A2	CoA3	Co Score
US5510271	US7842882	0	1	1	5	1	0	0	2	0	10.9
US5510271	US5928438	1	0	0	5	0	1	1	1	1	9.9
US5766964	US5871630	0	1	0	1	0	0	4	1	1	9.6
US5766964	US6518086	0	0	1	1	1	0	0	2	2	9.2
US5766964	US6897560	0	1	0	0	0	0	1	3	0	7.1
US5603778	US6897560	1	0	0	1	1	0	1	1	0	6
CO weights		1.8	2.5	2	0.5	1.5	2	1	1.2	1.4	

Table 17
Simplified patent correlation matrix (partial).

Patent	US7842882	US5928438	US5871630	US6518086	US6897560	…
US5510271	1	0.91	–	–	–	…
US5766964	–	–	0.88	0.84	0.65	…

Table 18
Patent recommendation result.

Rank	Recommend result	Patent title	Patent drawing
1	US7842882	Low cost and high throughput deposition methods and apparatus for high density semiconductor film growth	
2	US5928438	Structure and fabrication process for self-aligned locally deep-diffused emitter (SALDE) solar cell	
3	US5871630	Preparation of copper–indium–gallium–diselenide precursor films by electro-deposition for fabricating high efficiency solar cells	
4	US6518086	Processing approach towards the formation of thin-film Cu(In,Ga)Se2	

analyzes the correlation scores and the weights of collaborative operations are defined in Table 15. Table 16 lists the summary for each patent combination and counts for different types of collaborative operations.

The final recommended patent is selected from the candidate patents which are highly related with the selected patents. The research simplifies and normalizes the matrix only for the selected patents versus the candidate patents as shown in Table 17. The normalized score is the final correlation of the patents. After setting a threshold, the system selects the highly related candidate patents for each selected patent. Table 18 shows the recommendation result of the case, and the system further analyzes the recommended patents by providing different ranking values. In summary, these patents are recommended to user “ieAC01” and the recommended patents describe low-cost and highly efficient solar cell manufacturing technology and disclose related technology that decrease manufacturing costs.

4.4. Recommendation and evaluation

Recommendations for patents are automatically generated by the proposed system and the results are judged by field experts. Table 19 shows the overview of four recommended patents. Among four recommended patents, US7842882 and US5928438 are also the choice of experts. After content analysis, the top two selections are both in the area of thin-film improvement processes designed to decrease the manufacturing cost, to increase the production capacity, and to improve power capacity and efficiency. Nonetheless, the recommendation system also identifies US5871630 and US6518086, which are also relevant, yet, receive slightly lower Co Scores than the top two choices. The result indicates the system provides versatile and flexible recommendations comparing to the human experts’ selection.

Many different websites are built for patent analysis and can be divided into two types including patent searching (e.g. Google Patent, PatTools, WIPS) and patent analysis (e.g. SOOPAT, GPSA, IPDSS). The patent recommendation system is different since it does not search directly from the official patent database. The system does not support patent text mining analysis, such as patent clustering or patent claim construction which is to be included in future research and development. However, the advantage of the system is that it automatically recommends patents by analyzing the users' operation behavior and search rules. Moreover, the system continuously collects new domain patents filed in the USPTO database by regular updating the database. The system collects members' patent search behavior information to train the recommendation module. Then, the collected set of users’ behavior is used to analyze and cluster the members of the system and recommend the most relevant patents to users with demonstrated interest in the domain.

5. Conclusion

The intelligent recommendation methodology and system for patent search is based on the analysis of users' behavior records from the patent search platform. The research calculates each user's behavior records according to pre-defined behavior types and analyzes the user clustering results according to decision rules. The recommendation system extracts the most appropriate patents based on the collaborative filtering algorithm and helps users obtain related patents while saving time and costs. The intelligent patent search system recommends patents to users based on the user clustering result and the neighbors' behavior records.

The proposed system requires little time to infer recommendations and works independently without human management. The system proposes the most relevant patents, in a shorter time, which provides users more options. Since the users' behavior records are an input source, personal factors which may influence search results are avoided. Since there are many users on the system, the recommendation system makes inferences with greater consensus and less bias. The proposed system updates the database quickly and recalculates collaborative information effectively. However, there still are restrictions in the analysis since the data only includes patent search result information and users’ behavior records. Users’ behavior records also require sufficiently long periods of time to collect for accurate analysis. Researchers refer to this as the “cold-start” problem which is to say that the recommendation system will not perform well initially and only when there is sufficient data for analysis. However, the recommendation results’ accuracy improves with the long-term data collection (Schein et al., 2001; Herlocker et al., 2004). The recommendation system is trained by evaluating the search rules and analysis strategies of the patent engineers who work in the Green Energy and Environment Research Laboratories at the Industrial Technology Research Institute (ITRI). After collecting the search results of the domain's patent engineers, members in the solar energy alliance use the proposed system to search patents of interests and receive recommended patents. The system clusters the new members and the domain engineers by analyzing

Table 19
Recommended patents content overview.

Patent no.	Recommendation of proposed system Description	**Co Score**	Expert's Choice
US 7842882	Discuss the precursor layer using new material, anticipating decrease the semiconductor thin-film manufacturing cost and increase the battery charge capacity.	10.9	✓
US 5928438	The invention comprises a solar cell with reduced electron–hole recombination performance, relatively high efficiency including relatively low electrode resistance, which can be fabricated at relatively low cost using simplified fabrication techniques resulting in high yield.	9.9	✓
US 5871630	Fabricating the film by electrodepositing copper, indium, gallium, and selenium onto a glass/molybdenum substrate simultaneously, improving the battery energy conversing efficiency.	9.6	
US 6518086	A two-stage production of thin film batteries, including non-heated substrate for the first phase non-crystal deposition layer and the second phase of the pilot short-term treatment. This technique for optical correlation applications.	9.2	

the users' operation behaviors and search strategies. One of the system's advantages is actually preventing following gurus by recommending relevant patents through the collaborative filtering of all clustered users, which consist of new, less experienced, seasoned, and domain-expert engineers.

Acknowledgments

This research was partially supported by the Taiwan National Science Council research Grant (NSC100-2221-E-007-034-MY3 and NSC100-2410-H009-011-MY2), as well as the Ministry of Education Grant (100N2074E1/101N2074E1) for the Advanced Manufacturing and Service Management Research Center at National Tsing Hua University.

References

US Patent Act, Part III. Patents and protection of patent rights. Infringement of patents, Section 271. Infringement of patent. Available from: ⟨http://www.law.cornell.edu/patent/35uscs271.html⟩ [accessed 21.12.12, chapter 28].

World Intellectual Property Organization (WIPO). Available from: ⟨http://www.wipo.int/portal/index.html.en⟩ [accessed 21.12.12].

Ansari A, Essegaier S, Kohli R. Internet recommendation systems. Journal of Marketing Research 2000;37(3):363–75.

Barragáns-Martínez AB, Costa-Montenegro E, Burguillo JC, Rey-López M, Mikic-Fonte FA, Peleteiro A. A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. Information Sciences 2010;180(22):4290–311.

Basumallick S, Wong JSK. Design and implementation of a distributed database system. Journal of System Software 1996;34(4):21–9.

Bhavnani C, Clarkson G, Scholl M. Collaborative search and sensemaking of patents. In: Proceedings of the CHI EA '08 CHI '08 extended abstracts on human factors in computing systems, New York, USA; 2008. p. 2799–804, ISBN: 978-1-60558-012-8.

Bouras C, Poulopoulos V. Enhancing meta-portals using dynamic user context personalization techniques. Journal of Network and Computer Applications 2012;35(5):1446–53.

Chiang TA, Wu CY, Trappey AJC, Trappey CV. An intelligent system for automated binary knowledge document classification and content analysis. Journal of Universal Computer Science 2011;17(14):1991–2008.

Dong H, Hussain F, Chang E. A service concept recommendation system for enhancing the dependability of semantic service matchmakers in the service ecosystem environment. Journal of Network and Computer Applications 2011;34(2):619–31.

Edwards W, Barron FH. SMARTS and SMARTER: improved simple method for multiattribute utility measurement. Organizational Behavior and Human Decision Processes 1994;60:306–25.

Herlocker JL, Konstan JA, Terveen LG, Ridel JT. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 2004;22(1):5–53.

Huang LC, Li Y. Research on technology trend based on patent information. In: Proceedings of IEEE management of innovation and technology (ICMIT) international conference; 2010. p. 209–13.

Kelly D, Teevan J. Implicit feedback for inferring user preference: a bibliography. SIGIR Forum 2003;37(2):18–28.

Krizer M, Zacca F. Automatic categorisation applications at European Patent Office. World Patent Information 2002;24(1):187–96.

Lee TQ, Park Y, Park YT. Time-based approach to effective recommender system using implicit feedback. Expert Systems with Applications 2008;34(4):3055–62.

Li Z, Tate D, Lane C, Adams C. A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge-transfer and patent citation metrics. Computer-Aided Design 2012;44(10):987–1010.

Liang Y, Liu Y, Kwong CK, Lee WB. Learning the "Whys": discovering design rationale using text mining—an algorithm perspective. Computer-Aided Design 2012;44(10):916–30.

Nazim-uddin, M, Shrestha J, Jo GS. Enhanced contented-based filtering using diverse collaborative prediction for movie recommendation. In: Proceeding of 2009 first asian conference on intelligent information and database systems; 2009. p. 132–7.

Nichols DM. Implicit rating and filtering. In: Proceedings of the 5th DELOS workshop on filtering and collaborative filtering; 1997. p. 31–6.

Oard DW, Kim J. Implicit feedback for recommender systems. In: Proceedings of the AAAI workshop on recommender systems; 1998. p. 81–3.

Resnick P, Varian HR. Recommender systems. Communications of ACM 1997;40(3):56–8.

Richter G, MacFarlane A. The impact of metadata on the accuracy of automated patent classification. World Patent Information 2005;27(3):13–26.

Schein AI, Popescul A, Ungar LH, Pennock DM. Methods and metrics for cold-start recommendations. In: Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval; 2001. p. 253–60.

Sharma S. Applied multivariate techniques. New York: John Wiley & Sons, Inc; 1996.

de la Torre-Diez I, Álvaro-Muñoz S, López-Coronado M, Rodrigues J. Development and performance evaluation of a new RSS tool for a web-based system: RSS_PROYECT. Journal of Network and Computer Applications 2013;36(1):255–61.

Trappey AJC, Trappey CV, Wu CY, Lin CW. A patent quality analysis for innovative technology and product development. Advanced Engineering Informatics 2012;26(1):26–34.

Trappey CV, Trappey AJC, Wu CY. Clustering patents using non-exhaustive overlaps. Journal of Systems Science and Systems Engineering 2010;19(2):162–81.

Wang L, Guo YP, Fang M. Web search engine based on DNS. Journal of Network and Computer Applications 2007;30(2):466–78.

de Campos LM, Fernández-Luna JM, Huete JF, Rueda-Morales MA. Combining content-based and collaborative recommendations: a hybrid approach based on bayesian networks. International Journal of Approximate Reasoning 2010;51(7):785–99.