The Visual Subject Analysis of Library and Information Science Journals with Self-Organizing Map†

Lu An*, Jin Zhang**, and Chuanming Yu***

* School of Information Management, Wuhan University, Wuhan, P.R. China 430072, <whuanlu@yahoo.com.cn>

** International Collaborative Academy of Library and Information Science, Wuhan University, Wuhan, P.R. China 430072, School of Information Studies, University of Wisconsin-Milwaukee, WI, USA 53211, <jzhang@uwm.edu>

*** School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, P.R. China 430073, <yuchuanming2003@126.com>

Dr. Lu An is a lecturer at the School of Information Management, Wuhan University, China. She has published many research papers in international journals such as *JASIST*, *The Canadian Journal of Information and Library Science*, and *The Journal of Academic Librarianship*. Her book *Visual Studies on Academic Journal Subjects* is going to be published by Wuhan University Press in 2011. Her research is supported by China Postdoctoral Science Foundation and Hubei Province Social Science Foundation. Her research interests include knowledge organization and knowledge management, artificial neural network, competitive intelligence, Webometrics and informetrics, and digital libraries.

Dr. Jin Zhang is a professor at the School of Information Studies, University of Wisconsin-Milwaukee. He has published many research papers in prestigious international journals such as *JASIST*, *IPM*, *Journal of Intelligent Information Systems*, *Journal of Information Retrieval*, *Journal of Information Science*, and *IEEE Transactions on Visualization and Computer Graphics*. His book *Visualization for Information Retrieval* was published by Springer in 2008. His research interests include visualization for information retrieval, information retrieval theory and algorithm, Internet information organization, search engine evaluation, consumer health informatics, transaction log analysis, digital libraries, and computer human interface design.

Dr. Chuanming Yu graduated from Wuhan University, China. He is currently an associate professor at the School of Information and Safety Engineering, Zhongnan University of Economics and Law. He has published many research papers in journals such as the *Journal of the China Society for Scientific and Technical Information* and *Information Studies: Theory and Application*, and has had international conference papers indexed by EI Compendex. He has won multiple national academic awards. His research is supported by the National Natural Science Foundation of China (NSFC) under the grant No. 70903047. His current research interests include knowledge organization and knowledge engineering, opinion mining, natural language processing, and digital libraries.

An, Lu, Zhang, Jin, and Yu, Chuanming. **The Visual Subject Analysis of Library and Information Science Journals with Self-Organizing Map**. *Knowledge Organization, 38(4)*, 299-320. 49 references.

ABSTRACT: Academic journals play an important role in scientific communication. The effective organization of journals can help reveal the thematic contents of journals and thus make them more user-friendly. In this study, the Self-Organizing Map (*SOM*) technique was employed to visually analyze the 60 library and information science-related journals published from 2006 to 2008. The U-matrix by Ultsch (2003) was applied to categorize the journals into 19 clusters according to their subjects. Four journals were recommended to supplement library collections although they were not indexed by SCI/SSCI. A novel *SOM* display named Attribute Accumulation Matrix (AA-matrix) was proposed, and the results from this method show that they correlate significantly with the total occurrences of the subjects in the investigated journals. The AA-matrix was employed to identify the 86 salient subjects, which could be manually classified into 7 meaningful groups. A method of the Salient Attribute Projection was constructed to label the attribute characteristics of different clusters. Finally, the subject characteristics of the journals with high impact factors (IFs) were also addressed. The findings of this study can lead to a better understanding of the subject structure and characteristics of library/information-related journals.

Received 1 November 2010; revised 16 February 2011; accepted 18 March 2011

† This work was supported by the China Postdoctoral Science Foundation funded project under grant No. 20090460986.

1. Introduction

Academic journals are important means of scientific communication. With the rapid progress of science and the accumulation of human knowledge, the quantity of academic journals has been increasing at a high speed. The British Library claimed that they currently hold over 260,000 journal titles (*Facts and Figures* 2010), which is more than 10 times the number of titles held in 1973 (Carpenter and Narin 1980). The quantity of journal articles has also increased. Archibald and Line (1991) investigated the journals and articles between the 1950s and 1980s in nine subject areas using the Dewey Decimal Classification, i.e., 1) Philosophy and Religion; 2) Social sciences and Languages; 3) General science and Mathematics; 4) Astronomy, Physics, Chemistry, Earth, and related sciences; 5) Paleontology, Life sciences, Botanical, and Zoological sciences; 6) Technology (Applied sciences); 7) Medical sciences; 8) Arts and Literature; and 9) Geography and History. They found a rising quantity of journal articles in most disciplines.

With the rapid growth of academic journals and articles, more and more journal articles share similar or related thematic contents. The organization of the journals according to subject matter is crucial both for the reasonable selection and acquisition of academic journals by libraries, and for user-friendliness. Researchers can easily identify appropriate journals with relevant topics to their studies, and they can be provided with useful information about their prospective research interests.

In some high-quality journal databases such as *Proquest* (*Proquest* 2009), each article is indexed with several controlled subject terms, which reflect the thematic contents of the corresponding articles. This function greatly facilitates journal analysis based on normalized subject terms. As journal subject analysis usually studies the journals in a wide time span, for example in three years, this may result in a large quantity of the investigated subjects. It is difficult for people to intuitively observe the subject characteristics of the journals. Ordinary statistical or bibliometric measures are capable of calculating the frequencies of specific subject terms and identifying those with high frequencies. However, the relations among the salient subject terms are unknown unless a further classification or clustering process is conducted. Journals may be clustered according to their subject terms using some popular clustering algorithms, such as k-means and hierarchical clustering techniques. The k-means algorithm is computationally expensive, particularly for clustering a large dimensional data set into many categories (Filho et al. 2003). It requires defining the number of clusters first (MacQueen 1967) and depends on the initial cluster centroids (Hoon et al. 2004). According to Janssens et al. (2006), the agglomerative hierarchical clustering technique outperformed the k-means technique on their dataset. However, the agglomerative hierarchical clustering technique exhibits the disadvantage that the wrong merges made by the algorithm in an early stage can never be fixed (Kaufman and Rousseeuw 1990). One of the major concerns is that these traditional clustering methods do not provide an intuitive way to observe the relationships among clusters and the relationships among objects in a cluster from multiple perspectives.

The Self-Organizing Map (*SOM*) (Kohonen 1982), an unsupervised artificial neural network technique, was considered in this study based on the following reasons. First, unlike the traditional clustering methods, the *SOM* technique can effectively cluster a data set without knowing the distributional characteristics of the data set. Through competitive learning, the high-dimensional input data are visualized in the low-dimensional space with its topology structure preserved. Like other visualization techniques, the objects with similar attributes are projected onto a cluster. However, the objects projected onto an individual *SOM* node can be considered as a cluster. The objects projected onto the adjacent *SOM* nodes can also be considered as a cluster. The objects clustered in a node are more relevant to each other than the objects in a cluster defined by an adjacent area. As a result, the size of a cluster and the degree to which the objects are relevant to each other can be controlled. Thus it is intuitive, flexible, and efficient for researchers to observe the overall characteristics of input data. Certainly, some other data analysis technique such as Latent Semantic Analysis (LSA) (Deerwester et al. 1990) may also reveal the subject similarity between two items. The LSA technique can represent the synonyms and the related documents by a similarly weighted combination of indexing variables and position the associated terms in the reduced space in a way that reflects the correlations in their use across documents (Hull 1994). However, the LSA technique has its own weaknesses like high storage and computation costs of the matrix decomposition (Hull 1994), linear restriction in the mapping of LSA (Hofmann 1999), and resultant interpretation difficulty in natural language (Gabrilovich and Markovitch 2007).

The purposes of this study are (1) to identify the salient subjects of journals in Library and Information Science (LIS), (2) to discover journals which are not in the pool of JCR by ISI (*Journal Citation Reports* 2009) and have quite different subject coverage from those journals in the pool of JCR and to supplement journal collections of libraries, and 3) to explore the subject characteristics of the investigated journals, especially the journals with high Impact Factors (IFs). The findings of the study can reveal the salient research topics in the LIS field and lead to a better understanding of the subject structure of journals. This information will be useful to libraries and researchers in terms of the selection and utilization of journals.

2. Related research

2.1 Studies on journal subject analysis

Studies on journal subject analysis have various research objectives. Related researchers may summarize the thematic contents of the journal articles from their abstracts or titles, cluster the journals according to their subjects or cluster the subjects of the journals based on their co-occurrences, identify the salient subjects of the journals in a certain period, analyze the distribution of a specific group of subjects among the journals, or study the developing trends of the journals. According to the different data sources, the journal subject analysis can be divided into several fields, the journal analysis based on co-citation data, expert surveys, and keywords, abstracts, or even full text of the articles.

First, in the field of the journal analysis based on author co-citation data, Chen et al. (2010) introduced a multiple-perspective co-citation analysis method and conducted a comparative author co-citation analysis (ACA), a progressive ACA of a time series of co-citation networks, and a progressive document co-citation analysis (DCA) on the co-citation data of 12 journals published between 1996 and 2008. They found that the multiple-perspective method had the advantage of enhancing the interpretability and accountability of ACA and DCA networks. The classic ACA was enhanced by Zhao and Strotmann (2008a) by employing both orthogonal and oblique rotations in the factor analysis (FA). After comparing different FA methods in ACA on 12 main Information Science (IS) journals from 1996 to 2005, they found that each method had its own advantages when applied to ACA. Similar to ACA, Zhao and Strotmann (2008b) introduced the method of author bibliographic-coupling analysis (ABCA) and applied it on 12 core Information Science (IS) journals from 1996 to 2005. After comparing the results from ABCA and those from ACA, they found that the combination of these two citation-based author-mapping techniques revealed a more complete picture of the intellectual structure of the IS field than either of them. Although the journal analysis based on author co-citation data often took journals as data sources, it is noteworthy that the objectives and results of ACA are quite different from those of this study. The ACA usually identifies the prominent researchers and their specialties and does not aim to reveal the thematic characteristics of journals as this study does.

Besides ACA, the journal co-citation analysis was also done by some other researchers. Marion, Wilson, and Davis (2005) collected the co-citation data of 100 core journals in the field of information systems and categorized these journals into seven thematic clusters, such as computer science, computer networking, etc. They found that Information Science journals, for instance, ASIST publications, *JASIST*, *ARIST*, and *PASIS* were situated between technically oriented and application-focused clusters in the Multidimensional Scaling (MDS) map. Åström (2007) performed a set of co-citation analyses upon 21 library and information science journals to study changes in research fronts between 1990 and 2004. The results showed that the research on informetrics and information seeking and retrieval were stable while webometrics had become a dominating research area in both fields.

In the field of expert surveys, Shewchuk et al. (2006) investigated 147 faculty members of North American health management programs about their perceptions of 34 North American health care-oriented journals. A three-dimensional map with seven clusters was provided which illustrates the perceived similarities of journals. Lowry, Roman, and Curtis (2004) investigated 8,741 faculty members of the global 414 departments of information systems, categorized the supporting disciplines for information systems into 11 groups and summarized the top journals in each discipline. To establish a systematic and complete knowledge map of the IS field, Zins (2007) conducted a Critical Delphi study and surveyed 57 renowned scholars from 16 countries. A map with 10 basic categories was generated through panel discussions.

Finally, in the field of the journal analysis based on keywords, abstracts or full text, Chua and Yang (2008) compared the top keywords of articles published in the *Journal of the American Society for Information Science and Technology* (*JASIST*) between two decade periods, i.e. 1988-1997 and 1998-2007. They found that the distribution of top keywords' occurrences had shifted from core information science towards other sub disciplines, e.g., information technology and sociobehavioral science. Besides keywords, abstracts could also serve as cues to article contents. Kivlighan and Miles (2007) conducted latent semantic analysis of the abstracts of 97 articles in *Group Dynamics* and identified six subjects of the investigated journal. Janssens, Leta ,Glänzel, and De Moor (2006) clustered 1000 articles and research notes published by five LIS journals in the period of 2002-2004 with an agglomerative hierarchical clustering method to generate six clusters. They also applied both Latent Semantic Analysis (LSA) and MultiDimensional Scaling (MDS) to the full-texts of the journal articles and notes to reveal the structure and clusters of journals. The research findings of these studies based on keywords, abstracts, or full text of the articles in the investigated journals were more objective and solid compared with those based on co-citation data or expert surveys. However, the limitation of their study is that only a few journals were examined.

2.2 The SOM technique and its application in journal subject studies

The Self-Organizing Map (*SOM*) introduced by Kohonen (1982) is based on an unsupervised artificial neural network algorithm. It can project a set of high-dimensional input data onto a low-dimensional space where the topology of the input data is preserved. The principles and basic algorithms of the *SOM* technique have been described in literature (Mehotra et al. 1997; Kohonen 1993).

Two basic learning algorithms, i.e., sequential learning and batch learning can be used in the *SOM* technique. Fort, Letremy, and Cottrell (2002) compared the two learning algorithms and found that the batch learning algorithm had the advantages of computational simplicity, efficient training time, less distortion, omission of the tuning adaptation parameter, and stability of results compared with the sequential learning algorithm. Currently, there are two conventional *SOM* displays, the U-matrix and the component plane. The U-matrix was first introduced by Ultsch and Siemon (1990) and later improved by Ultsch (1992). The U-matrix can reveal the internal structure of the input data by measuring the distances among the weight vectors associated with the adjacent *SOM* cells. The component plane reflects the contribution of an identified attribute to the overall *SOM* display (Zhang and An 2010).The *SOM* technique was widely used in many fields, such as visualization of machine states, fault identification (Kohonen et al. 1996), feature extraction, computer vision (Kohonen 1996), exploratory data analysis (An et al. 2009), knowledge discovery in databases (Kohonen 1999), and transaction log analysis (Zhang et al. 2009; Zhang and An 2010).

The studies on journal subject analysis usually involve processing a large number of subjects. The multiple subjects can be used to describe the characteristics of the investigated journals in the vector space. As a result, they determine the dimensionality of the investigated journals in the vector space. After the investigated journals are described in the high dimensional vector space, their subject connections/relationships are kept in the high dimensional space. The subject connections/relationships are crucial in the journal subject analysis. Unfortunately, they are invisible due to the high dimensionality of the vector space. The *SOM* algorithm is a suitable method for this task since it can visualize the high dimensional input data in a low dimensional *SOM* space through a competitive learning process. In this low dimensional space, an individual *SOM* node can be considered as a unique cluster, and the adjacent *SOM* nodes also provide some information about clustering. Then, the subject connections /relationships of the journals are effectively revealed. However, the studies on journal subjects with the *SOM* technique are limited, and most literature focuses on the subject characteristics of journal articles, in which the journal articles are considered as analysis objects. In this study, the journals are taken as analysis objects, and the subject characteristics of the journals are revealed.

The examples of employing the *SOM* technique to analyze journal articles are summarized as follows. Guerrero-Bote, Moya-Anegón, and Herrero-Solana (2002) selected 202 documents from the LISA database in 1996 by excluding those documents that contained none of the seven descriptors, i.e., Acquisitions, Artificial Intelligence, Business Management, Computerized Information Storage and Retrieval, Conferences, Periodicals, and World Wide Web. The documents and terms were trained by the *SOM* algorithm, and the authors suggested that it was possible to organize the documents with the *SOM* technique. Users could expand their search queries or results through browsing the *SOM* display. In another study, Börner, Chen, and Boyack (2003) employed the *SOM* technique in combination with other information visualization methods to demonstrate the semantic relationships among the four research areas, Citation Analysis, Semantics, Bibliometrics, and Visualization, based on 2,764 journal articles from the SCI and SSCI databases. The objective of these studies was to reveal the relationships among the subject terms, articles, or specific research fields. Few studies on journal analysis have used individual journals as analysis objects and used the *SOM* technique to reveal the subject relationships among the journals.

Some researchers took journals as analysis units in the journal co-citation analysis. Moya-Anegon, Herrero-Solana, and Jimenez-Contreras (2006) collected the journal (also author) co-citation data of 73 journals in the field of Library and Information Science from 1992 to 1997. They utilized the *SOM* technique, in conjunction with the MDS and Ward's clustering methods, and categorized the investigated journals into four domains: Information Science, Library Science, Science Studies, and Management. The results were compared with those by White and McCain (1998), who analyzed the author co-citation data of 12 journals in the field of information science from 1972 to 1995 with the *SOM* and MDS approaches. Moya-Anegon et al. (2006) claimed that the *SOM* representation was "comfortable" to observe object relationships (Costa 1998), and their findings confirmed what Kaski (1997) called 'ordered display.' In these studies, the *SOM* technique was usually used as a cluster tool. However, the distinctions among the different subject categories were vague, and an academic journal may involve multiple disciplines (Leydesdorff and Bensman 2006). Some researchers even believed that it was impossible to provide a simple classification for the journals according to their subjects (Bensman 2001). Thus, it is necessary to identify the salient subjects of the journals and to explore the characteristics of the journals in terms of the salient subjects. The current *SOM* display, the U-matrix, and the component plane have failed to accomplish the tasks. Thus, in this study, we propose a novel *SOM* display named "Attribute Accumulation Matrix" (AA-matrix) to identify the salient subjects among the journals. To label the journal clusters, we came up with a method called Salient Attribute Projection.

3. Research method description

3.1 The Proquest *journal article record description*

The data source in this study comes from the *Proquest Academic Research Library*. *Proquest* is a comprehensive academic journal database, which incorporates 3,811 academic journals and newspapers in the field of library and information science (LIS), business and economics, education, computers, etc. Each journal article record contains article title, authors, journal title, serial number, article URL, publication year, volume, issue, starting page, abstract, and several controlled subject terms. The number of keywords provided by authors is usually very large, and the keywords are too specific to categorize them. The keywords from authors are not normalized. There is no synonym control, homographic control, or ambiguity control for these keywords. For these reasons, the controlled subject terms were used in this study, although at the cost of losing newly updated lexicons.

Out of the 53 journals listed in the category of information and library sciences in *Journal Citation Reports*, 2008 edition, 32 journals were found to be included in *Proquest* from 2006 to 2008. These journals were considered in this study. Twenty journals in *Proquest* were selected whose journal titles contain library, libraries, librarian, or information. Another eight journals were retrieved using common terms in the LIS field, such as digital library, and knowledge management, in Document Title Search in *Proquest*.

Thus, a total of 60 journals were selected in this study. The journal titles are listed in Appendix I, in which each journal is labeled based on whether it is indexed by *SCI* or *SSCI*. The impact factors of the journals in 2008 (*Journal Citation Reports* 2009) are also given, if applicable. We collected 12,415 article records ranging from January 1, 2006 to December 31, 2008. From these journal article records, we extracted 3,952 subjects.

3.2 Definitions of the SOM input matrices

In this study, two *SOM* input matrices were constructed. The first matrix is a journal-subject matrix, denoted by *M*1. It has *m* rows (journals) and *n* columns (subjects). Rank all journals alphabetically and number them from 1 to *m*. Also rank all subject terms alphabetically according to their spelling and number them from 1 to *n*. c_{ij} $(i=1,2,\ldots,m; j=1,2,\ldots,n)$ stands for an element of the matrix. Element c_{ij} in the matrix is defined as the sum total of the j^{th} subject term from all articles in the i^{th} journal over the period concerned.

The second matrix is a subject-journal matrix, denoted by *M*2. It has *p* rows (subjects) and *q* columns (journals). Rank all subject terms alphabetically and number them from 1 to *p*. Also rank all journals alphabetically and number them from 1 to *q*. d_{kl} $(k=1,2,\ldots,p; l=1,2,\ldots,q)$ stands for an element of the matrix. Element d_{kl} in the matrix is defined as the sum total of the k^{th} subject term from all articles in the l^{th} journal over the period concerned. In fact, *M*2 is the transformation matrix of *M*1, which means that p=n, q=m. The reason why both *M*1 and *M*2 are constructed in this study is that *M*1 will be processed by the *SOM* technique to reveal the journals' distribution in the *SOM* display and *M*2 will be processed to cluster the subjects that the journals involved. The objects to be projected onto the *SOM* nodes are different.

3.3 Description of the U-matrix in the toroid space

The concept of Unified Distance Matrix (*U-matrix*) was introduced by Ultsch and Siemon (1990) to reveal the differences between the weight vectors associated with a *SOM* node and those with its immediate neighbours. According to its definition, the value of an element of *U-matrix* is equal to the sum of the distances of the corresponding *SOM* node to all its immediate neighbouring nodes normalized by the largest occurring value in the *SOM* grid (Ultsch 1992). In the planar *SOM* display, the *SOM* nodes which reside on the border or at the corner of the *SOM* display have fewer immediate neighbours than those which are located in the central area of the *SOM* display, which causes "border effect" (Kohonen 2001). To attack this problem, Ultsch extended the *U-matrix* to the toroid space (Ultsch 2003), where the upper edge of the *SOM* display is connected with the bottom edge, and the left edge is connected with the right edge. Applying the *U-matrix* to the background colours of the *SOM* display will help identify the borders of the clusters. The high values of the *U-matrix* represent borders between clusters, and the low values represent clusters themselves.

3.4 Definition of the proposed Attribute Accumulation matrix

The *U-matrix* is capable of revealing the structure of the input data. However, it is a difficulty to determine and label the characteristics of the clusters. Freeman and Yin (2004) suggested comparing the element values of the weight vector associated with a *SOM* node to label the cluster. The attribute which corresponds to the element with the highest value in the weight vector was selected as the label of the corresponding *SOM* node. This method is suitable for the circumstances where the attributes of the input data are not related. However, it is a different case in this study. When the journal-subject matrix is applied, some of the attributes, namely the subjects which appeared in a certain journal, are highly related. The individual element value to which a subject corresponds may be relatively small while a group of related subjects with low element values may considerably represent the characteristics of the journal cluster. If the method by Freeman and Yin (2004) is adopted, some important subjects which have a certain number of related subjects may be ignored. Thus, we propose a novel *SOM* display named Attribute Accumulation matrix (*AA-matrix*) to solve the problem.

Suppose a *SOM* configuration with *u* rows and *v* columns. s_{ij} $(i=1, 2, \ldots, u, j=1, 2,\ldots, v)$ represents the weight vector associated with the *SOM* node at the i^{th} row and j^{th} column. It has *n* elements, denoted by $(w_{ij1}, w_{ij2}, \ldots, w_{ijn})$, where *n* is the number of the attributes, namely dimensionality.

The *AA-matrix* also has *u* rows and *v* columns. c_{ij} $(i=1,2,\ldots,u, j=1,2,\ldots,v)$ is the element at the i^{th} row and j^{th} column.

First, the value of $c_{ij}^{'}$ is calculated as Equation (1).

$$c'_{ij} = \sum_{k=1}^{n} w_{ijk} \qquad (1)$$

It is seen in Equation (1) that c'_{ij} equals to the sum of all the elements of the weight vector associated with the corresponding *SOM* node.

Then the value of c_{ij} is calculated as Equation (2).

$$c_{ij} = \sqrt{\frac{c'_{ij}}{\max(c'_{\mathrm{lm}})_{l \le u, m \le v}}} \qquad (2)$$

Equation (2) shows that the element value of the *AA-matrix* equals to the square root of the quotient of c'_{ij} normalized by the largest occurring values. The values of *AA-matrix* are converted to colours and applied to colouring the background of the *SOM* display. The reason why we extract the square root of the quotient is to widen the differences among the element values of *AA-matrix* so that users can easily observe the differences. Otherwise, if most of the element values of *AA-matrix* are too close, the *SOM* display may exhibit large areas of similar colours. The effectiveness of the *AA-matrix* is justified in Section 4.3.

In this study, the *SOM* Toolbox (About *SOM* Toolbox 2005) was adopted as the research tool, and the AA-matrix was implemented in this software in the environment of Matlab. The *AA-matrix* is applied to the background colours of the *SOM* display generated by the subject-journal matrix, and the subjects are projected onto the S*OM* display. It reveals the occurring frequencies of different subjects. The subjects with high *AA-matrix* values appeared the most in the investigated journals and are believed to be suitable for labelling the characteristics of the journal clusters. The subjects with low *AA-matrix* values appeared the least in the investigated journals and are believed to be unimportant for labelling the characteristics of the journal clusters.

3.5 Description of Salient Attribute Projection

Based on the results of the AA-matrix, the Salient Attribute Projection is used to visualize the attribute characteristics of journal clusters. The method is described as follows.

Suppose l input samples. Each input sample has n attributes, denoted by $v_1, v_2, \ldots, v_n$. Take the n attributes as rows and the l input samples as columns, and construct a *SOM* input matrix. Train the input matrix with the *SOM* algorithm and the n attributes are projected onto the *SOM* display. Calculate the *AA-matrix* values for each *SOM* node. Each attribute is projected onto a *SOM* node and associated with an *AA-matrix* value. Rank all the attributes in a descending order according to their *AA-matrix* values. The attribute(s) whose *AA-matrix* value(s) are the largest are considered as the First Salient Attribute(s), denoted by *S1*. The attribute(s) whose *AA-matrix* value(s) are the second largest are considered as the Second Salient Attribute(s), denoted by *S2*. Notice that *S1* or *S2* may contain multiple attributes if multiple attributes are projected onto the same *SOM* node. The same is true with the rest.

Suppose a *SOM* display S which is generated upon the input matrix with the l input samples as its rows and the n attributes as its columns. It has m *SOM* nodes. s_i represents the weight vector associated with the i^{th} node. It has n elements, denoted by $w_{i1}, w_{i2}, \ldots, w_{in}$, where n is the number of attributes of the input data, or dimensionality.

Suppose the First Salient Attribute(s) *S1* contains p attributes, denoted by $v_{j1}, v_{j2}\ldots, v_{jp}$, the Second Salient Attribute(s) *S2* contains q attributes, denoted by $v_{k1}, v_{k2}, \ldots, v_{kq}$, and the Third Salient Attribute(s) *S3* contain r attributes, denoted by $v_{l1}, v_{l2}, \ldots v_{lr}$. Construct three vertical vectors X, Y, and Z as Equations (3)-(5), where $i = 1, 2, \ldots, m$.

$$X = (\sum_{t=j1}^{jp} w_{it})' \qquad (3)$$

$$Y = (\sum_{t=k1}^{kq} w_{it})' \qquad (4)$$

$$Z = (\sum_{t=l1}^{lr} w_{it})' \qquad (5)$$

It is seen in Equation (3) that the X coordinate values equal the sum of the elements in the weight vector associated with each *SOM* node, which corresponds to the First Salient attributes *S1*. The Y coordinate values equal the sum of the elements in the weight vector associated with each *SOM* node, which corresponds to the Second Salient attributes *S2*. The same is true with the Z coordinate.

Project the *SOM* display, which is generated by the input matrix (with l input samples as its rows and n attributes as its columns), into the three-dimensional space. This is the Salient Attribute Projection.

In this study, the *SOM* display, which is generated upon the journal-subject matrix (*M*1), is projected

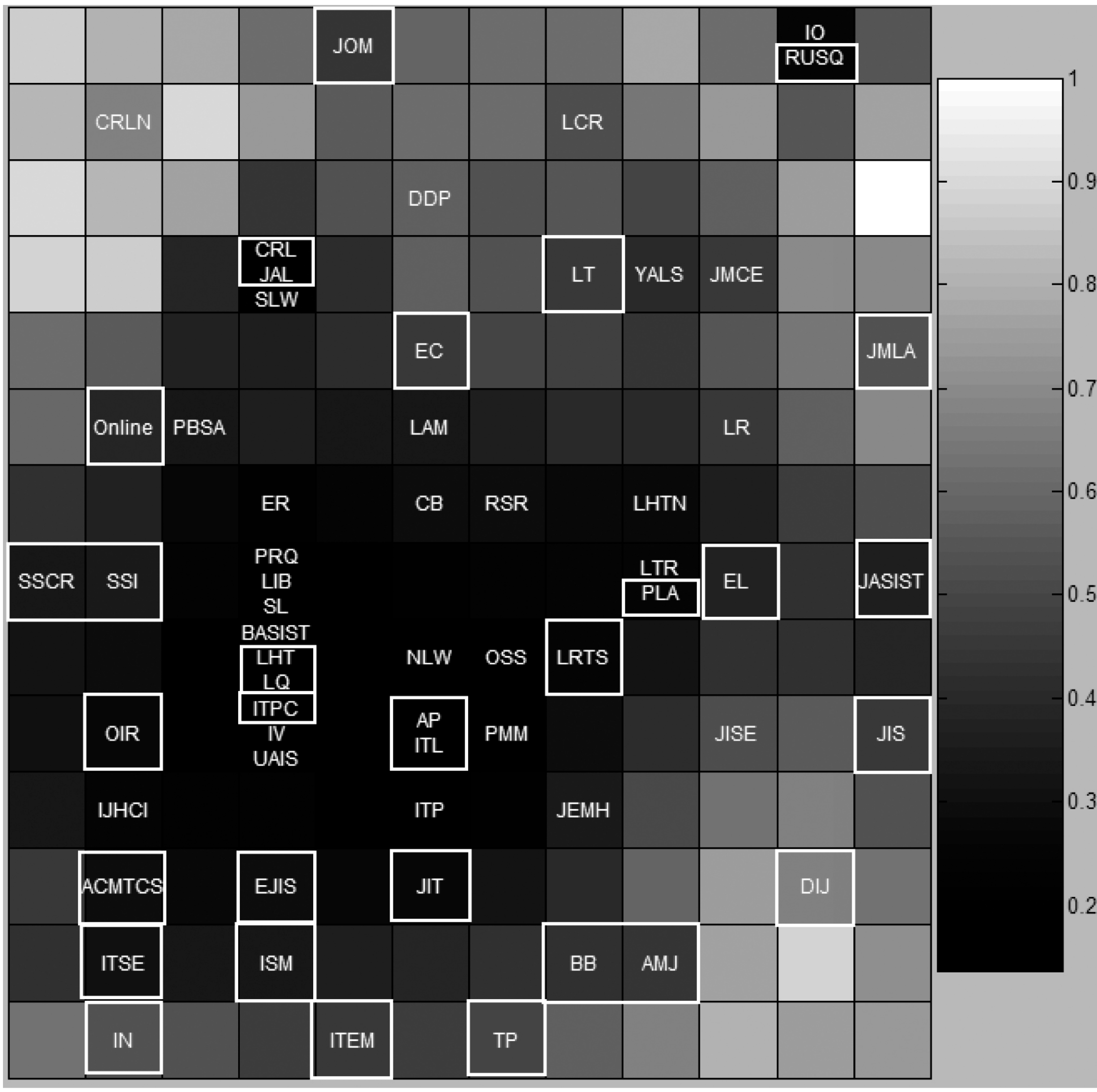

Figure 1. The *SOM* display for the investigated journals

into the three dimensional space constructed by the three Salient Attributes *S1*, *S2*, and *S3*. According to the principle of the salient attribute projection, if a journal cluster is projected onto a high value of *S1*, it means that the subjects that *S1* contains are the important subject labels for this journal cluster. The same is true with *S2* and *S3*. Thus, the characteristics of journal clusters in terms of salient attributes can be intuitively observed.

4. Results Analysis and Discussion

4.1 Journal and subject data description

Since each journal article is indexed with several controlled subjects in the *Proquest Academic Research Library*, the indexed subjects were used in the study. This may yield better processing efficiency as compared to a search for keywords that are listed by the authors. Our source of data was a list of journals from *Proquest*. A total of 60 journals were selected in this study, as described above in section 3.1. From these journal article records, we extracted 3,952 subjects.

4.2 Journal Distribution analysis

The journal-subject matrix *M*1 was constructed, which had 60 rows and 3,952 columns. Because the value ranges of different attributes (columns) were different, and the *SOM* display would be dominated by the attributes with large value ranges, the *SOM* input matrix was first normalized with the 'Var' method (*SOM*_norm_variable 2002), in which the variances of the attributes are linearly normalized to 1. In order to avoid "border effect," the toroid space was applied.

Linear initiation and batch learning were adopted because our pilot study showed that this combination produced the smallest final quantization error among all the combinations of random or linear initiation and sequential or batch learning. The *U-matrix* was employed to generate the background colour for the *SOM* display (See Figure 1 for the results). The colour bar on the right indicated the *U-matrix* values for each colour. The labels (namely the abbreviated form of the journals' names) in Figure 1 represent different journals. See Appendix I for the relationships between the labels and corresponding journals. To observe the distribution of journals indexed by *SCI/SSCI* in the *SOM* display, the journals indexed by *SCI/SSCI* are marked with white squares.

According to the principle of the *U-matrix*, the journals which were projected onto the *SOM* nodes with low values of *U-matrix* tend to have similar subjects while the journals which were projected onto the *SOM* nodes with high values of *U-matrix* tend to have quite different subjects. The similarity of input data is also reflected by the distances between the *SOM* nodes they are projected onto. In this study, journals which were projected onto the *SOM* nodes in close proximity had similar subjects while journals which were projected onto the *SOM* nodes in far distance had dissimilar subjects.

It is seen in Figure 1 that there is a large dark area in the "central" part of the *SOM* display. Several journals, such as *Publishing Research Quarterly* (PRQ), *The Library* (LIB), *The Serials Librarian* (SL), *Bulletin of the American Society for Information Science and Technology* (BASIST), *Library Hi Tech* (LHT), and *The Library Quarterly* (LQ), were projected onto the *SOM* nodes with low values of *U-matrix*. It means that those journals were highly similar in terms of the subjects that they involve.

On the other hand, journals such as *College & Research Libraries News* (CRLN) and *Drug Information Journal* (DIJ) were projected onto the *SOM* nodes with high *U-matrix* values and they are separated from other journals by some empty nodes. It means that these journals had very different subjects from other journals. Although *College & Research Libraries News* (CRLN) is not indexed by *SCI/SSCI*, this journal is recommended to be selected for library collections to supplement the inadequate subjects of other journals.

Therefore, the journal selection policy is to select journals to cover as many different subject clusters as possible. The journals which have very different subject coverage from other journals are recommended to be selected to supplement library collections, even if they are not indexed by *SCI/SSCI*.

To implement this policy, we categorized the 60 journals into different clusters. The method of determining journal clusters is described as follows.

1. If two or more journals were projected onto the same *SOM* nodes (such as [AP] and [ITL]) or neighbouring nodes (such as [SSCR] and [SSI], also including diagonal neighbouring, such as [TP] and [BB]) and all their *U-matrix* values were not higher than 0.5 (the background colour is dark or black), the journals form a cluster.
2. If two or more journals were projected onto *SOM* nodes separated by empty nodes (such as [ITEM] and [TP]) and one of their *U-matrix* values were higher than 0.5 (the background colour is white or grey), the journals form different clusters, respectively.

In this way, the *SCI/SSCI* journals were categorized into different clusters, and some journals which are not in the range of *SCI/SSCI* are recommended to supplement library collections. See Table 1 for the results. The underlined journal labels in Table 1 represent journals indexed by *SCI/SSCI.*

Table 1 shows that the 60 journals in this study were categorized into 19 clusters. The largest cluster, i.e., the 11th cluster contained 14 journals while the smallest cluster only contained one journal, for example, the 2nd cluster. The average cluster size is 3.16.

Four journals, i.e., *College & Research Libraries News* (CRLN), *DTTP, Documents to the People* (DDP), *Libraries & the Cultural Record* (LCR), and *Journal of Information Systems Education* (JISE), involve quite different subjects from other journals and can effectively supplement the subject deficiency of other *SCI/SSCI* journals. Thus they were recommended to be selected as library collections although they were not indexed by *SCI/SSCI*.

4.3 The calculation of the AA-matrix and the justification of its effectiveness

Academic journals usually cover a lot of subjects. Some subjects relevant to hot issues appear frequently in many journals, while some other subjects related to the topics which receive little attention only appear infrequently in a few journals. The former kind of the subjects is salient and can be considered as candidate subjects for labelling the subject characteristics of journal clusters. In this article, we plan to employ the

Cluster No.	Journal cluster	Recommended journal supplements
1	(SSCR)(SSI)(JASIST)*	
2	(CRLN)	(CRLN)
3	(Online) ,(PBSA)	
4	(OIR),(IJHCI), (ACMTCS),(ITSE)	
5	(IN)	
6	(CRL),(JAL),(SLW)	
7	(ER),(PRQ),(LIB),(SL),(BASIST),(LHT),(LQ),(ITPC),(IV),(UAIS)	
8	(EJIS), (ISM),(ITEM),(JOM)*	
9	(DDP)	(DDP)
10	(EC),(LAM),(CB),(RSR)	
11	(NLW),(AP),(ITL),(ITP),(JIT),(OSS),(PMM),(LRTS),(JEMH),(LHTN),(LTR),(PLA),(LR),(EL)	
12	(TP), (BB), (AMJ)	
13	(LCR)	(LCR)
14	(LT),(YALS), (JMCE)	
15	(JISE)	(JISE)
16	(IO),(RUSQ)	
17	(DIJ)	
18	(JMLA)	
19	(JIS)	

Table 1. Journal clusters and recommended journal supplements

Note: * (JASIST) was the neighbouring node of (SSCR) since the *SOM* display is the toroid shape. For the same reason, (JOM) was the neighbouring node of (ITEM).

proposed *AA-matrix* to identify the salient subjects for the investigated journals. First of all, it is necessary to justify the effectiveness of the *AA-matrix*.

The *SOM* input matrix *M*2 was constructed. It consists of 3,952 rows and 60 columns. For the reason explained before, the input matrix was normalized with the 'Var' method and trained by the linear initiation and batch learning algorithm. The *SOM* display was a 22*15 grid. That is, there were 330 nodes. The *AA-matrix* values were calculated according to Equation (1) and (2) and applied to the background colours of the *SOM* display.

In this study, each best matched node was assigned an *AA-matrix* value. According to Equation (1) and (2), the *AA-matrix* value of a best matched node is supposed to be highly correlated with the total occurrences of the subject (in all the investigated journals), which was projected onto the best matched node. If it is true, the subjects with high *AA-matrix* values can be considered to be salient ones and suitable for labelling the characteristics of the journal clusters, while the subjects with low *AA-matrix* values are considered unimportant for labelling the characteristics of the journal clusters.

The values of the total occurrences of some subjects and the corresponding *AA-matrix* values of their best matched nodes are listed in Table 2.

Because the distribution for the total occurrences of the subjects and the *AA-matrix* values is skewed, a Spearman correlation coefficient test was used to examine the relation between the total occurrences of the subjects and the corresponding *AA-matrix* values. See Table 3 for the results.

It is shown in Table 3 that the Spearman coefficient between the total occurrences of the subjects and the *AA-matrix* values equals 0.829, and the Sig value equals 0.000, which is lower than 0.01. It indicates that the correlation between the total occurrences of the subjects and the *AA-matrix* values is statistically significant. Thus, the *AA-matrix* can be effectively to identify the salient subjects.

4.4 Identification of salient subjects

The *AA-matrix* values were applied to the background colours of the *SOM* display. See Figure 2 for the results. The number labels in each *SOM* node represent the number of subjects which were projected onto

Subject No.	the total occurrences of the subjects	The AA-matrix value	Serial No. of the best match node
s_1	5	0 + 0.098908i	78
s_2	1	0 + 0.089169i	188
s_3	1	0 + 0.086815i	292
s_4	5	0.080412	223
s_5	2	0 + 0.086945i	247
…	…	…	…
s_{3948}	7	0.2237	325
s_{3949}	1	0 + 0.044537i	322
s_{3950}	1	0 + 0.08173i	224
s_{3951}	2	0 + 0.08173i	224
s_{3952}	2	0 + 0.083919i	147

Table 2. The total occurrences of some subjects and the corresponding *AA-matrix* values

Correlation Coefficient

			sum_subject_sort	bmus_aa_matrix_sort
Spearman 的 rho	sum_subject_sort	Coefficient	1.000	.829**
		Sig.(2-tailed)		.000
		N	3952	3952
	bmus_aa_matrix_sort	Coefficient	.829**	1.000
		Sig.(2-tailed)	.000	
		N	3952	3952

Table 3. The relationship between the total occurrences of the subjects and the AA-matrix values

the corresponding node. The colour bar on the right indicates the *AA-matrix* values for each colour.

It is seen in Figure 2 that most subjects were projected onto black or dark *SOM* nodes (low *AA-matrix* values), which means that most subjects appeared in the investigated journals with low frequencies. That is to say, a large number of subjects only occupied a small number of occurrences. A certain number of "corner" subjects were projected onto nodes with medium or high *AA-matrix* values. They were salient subjects, which contributed the most to the total occurrences of subjects.

The *SOM* nodes with high or medium *AA-matrix* values and the subjects which were projected onto these nodes were summarized in Table 4, in which $s(i, j)$ ($i=1,2,\ldots,22, j=1,2,\ldots,15$) represents the *SOM* node on the i^{th} row and the j^{th} column.

It is seen in Table 4 that, among the 3,952 subjects, only 86 subjects, approximately 2% of all subjects, were salient subjects which appeared the most in the investigated journals. Considering their geometric vicinity in Figure 2 and semantic meanings, the 86 subjects can be classified manually by the investigators as the following groups:

1. The group of library, including 13_Academic_libraries, 1016_Digital_libraries, 2081_Librarians, 2082_Libraries, 2086_Library_science, 356_Book_reviews, 2491_Nonfiction, 357_Books, 2085_Library_resources, 2084_Library_collections, 3057_Reference_services, 1988_Journals, 2083_Library_cataloging, 301_Bibliographic_literature, 624_Collections, 1103_E-books, 1235_Encyclopedias, 3055_Reference_books;
2. The group of information technology, including 1850_Information_technology, 922_Data_bases, 3594_Technological_change, 3877_Web_sites, 1849_Information_systems, 2360_Models, 3403_Software, 1844_Information_retrieval, 1920_Internet, 3255_Search_engines, 3256_Searches, 3879_Weblogs, 3559_Systems_design, 1922_Internet_access,

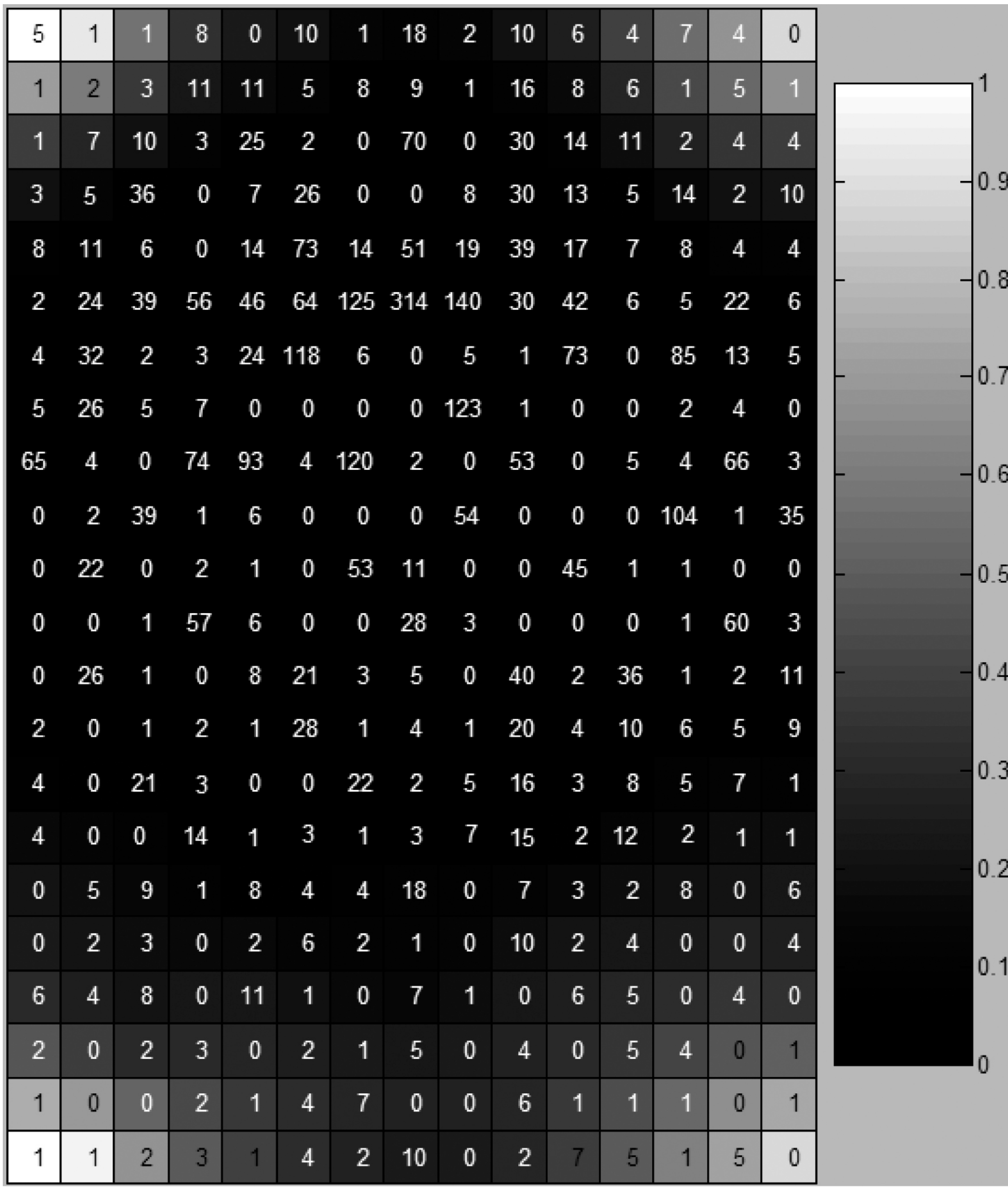

Figure 2. The *SOM* display for subjects (AA-matrix)

2547_Online_data_bases, 2550_Online_information_services, 2557_Open_source_software, 3875_Web_site_design, 3934_World_Wide_Web, 3873_Web_portals, 1836_Information_centers, 1839_Information_industry;

3. The group of information management, including 1841_Information_management, 2009_Knowledge_management, 1843_Information_professionals, 2305_Metadata, 656_Communication, 2584_Organizational_behavior, 287_Behavior, 2662_Performance_evaluation, 1840_Information_literacy, 1924_Internet_resources, 3383_Social_networks, 949_Decision_making, 1846_Information_sharing, 2221_Marketing;
4. The group of education, including 2056_Learning, 636_Colleges_&_universities, 635_College_students, 1140_Education, 1689_Higher_education, 3508_Students, 1055_Distance_learning, 2551_Online_instruction, 3588_Teaching_methods;

Colour	*AA-matrix* value	*SOM* nodes	Subjects
White	>0.8	s(1,1)	13_Academic_libraries, 1016_Digital_libraries, 2081_Librarians, 2082_Libraries, 2086_Library_science
		s(22,1)	356_Book_reviews
		s(22,2)	1850_Information_technology
		s(1,2)	3509_Studies
Very light grey	0.7~0.8	s(21,1)	2491_Nonfiction
		s(21,15)	357_Books
		s(22,14)	738_Conferences, 922_Data_bases, 2056_Learning, 3594_Technological_change, 3877_Web_sites
		S(2,1)	1841_Information_management
		s(1,14)	636_Colleges_&_universities, 1840_Information_literacy, 2084_Library_collections, 3110_Research
		S(22,3)	1849_Information_systems, 2360_Models
		S(2,15)	2085_Library_resources
Light grey	0.6~0.7	S(22,13)	3403_Software
		S(2,2)	1844_Information_retrieval, 1920_Internet
Grey	0.5~0.6	S(1,3)	2009_Knowledge_management
		S(1,13)	635_College_students, 1140_Education, 1689_Higher_education, 3057_Reference_services, 3113_Research_&_development--R&D, 3115_Research_methodology, 3508_Students
		S(21,13)	1843_Information_professionals
		S(2,14)	178_Archives_&_records, 1190_Electronic_publishing, 1988_Journals, 2083_Library_cataloging, 2305_Metadata
		s(20,15)	1705_History
		S(22,12)	480_Case_studies, 1924_Internet_resources, 3255_Search_engines, 3256_Searches, 3383_Social_networks
		S(22,4)	656_Communication, 2584_Organizational_behavior, 2662_Performance_evaluation
		s(20,1)	301_Bibliographic_literature, 2962_Publishing_industry
Dark grey	0.35~0.5	S(21,12)	3879_Weblogs
		S(21,4)	949_Decision_making, 3471_Statistical_analysis
		S(2,13)	624_Collections
		S(20,13)	211_Associations, 2052_Leadership, 2287_Meetings, 2876_Professional_development
		S(1,12)	287_Behavior, 1055_Distance_learning, 2551_Online_instruction, 3588_Teaching_methods
		S(22,5)	3559_Systems_design
		S(3,15)	1103_E-books, 1235_Encyclopedias, 3055_Reference_books, 3221_Scholarly_publishing
		S(22,11)	1629_Guidelines, 1846_Information_sharing, 1922_Internet_access, 2547_Online_data_bases, 2550_Online_information_services, 2557_Open_source_software, 3875_Web_site_design
		S(3,1)	2761_Polls_& _surveys
		S(21,5)	981_Design
		S(2,3)	682_Comparative_analysis, 2659_Perceptions, 3934_World_Wide_Web
		S(21,11)	3873_Web_portals
		S(20,12)	805_Copyright, 1321_Executives, 1836_Information_centers, 1839_Information_industry, 2221_Marketing

Table 4. Salient subjects and the corresponding SOM nodes

5. The group of publishing, including 1190_Electronic_publishing, 2962_Publishing_industry, 3221_Scholarly_publishing;
6. The group of research methods, including 3509_Studies, 3110_Research, 480_Case_studies, 3115_Research_methodology, 3113_Research_&_development--R&D,3471_Statistical_analysis, 682_Comparative_analysis, 2761_Polls_& _surveys, 981_Design;
7. The group of academic activities and organizations, including 738_Conferences, 211_Associations, 2052_Leadership, 2287_Meetings, 2876_Professional_development, 1629_Guidelines; and,

8. The others, including 1705_History, 178_Archives_&_records, 2659_Perceptions, 805_Copyright, 1321_Executives.

The eight salient subject groups revealed that the subject focuses of the 60 library/information-related journals in this study resided in the aspect of library, information technology, information management, education, publishing, research methods, and academic activities and organizations. Specific subject features of these journals need to be investigated in the following study.

4.5 The exploration of the subject features of journal clusters

In Section 4.4, the salient subjects in the investigated 60 journals were identified. In this section, we reveal the specific subject characteristics of each journal. Since only 2 percent of the 3,952 subjects were found to be salient in these journals, a reasonable choice is to label the journal clusters with salient subjects identified in Table 4. Then the characteristics of each journal in terms of salient subjects are revealed.

To this end, six coordinates were constructed using the six groups of salient subjects whose corresponding *AA-matrix* values were mostly higher than the 50^{th} percentile. Two salient attribute projection experiments were respectively conducted upon:

(I) A: Library (LIB),
B: Information technology (IT),
C: Information management (IM).
D: Education (EDU),
E: Publishing (PUB),
F: Research methods (RM).

The combination of salient subject groups will not affect the salient attribute projection results since each salient subject group is independent from the rest.

According to the principle of salient attribute projection as described above, the *SOM* display in Figure 1 was projected into the three-dimensional space constructed by the subject group of library (LIB), information technology (IT) and information management (IM). See Figure 3 for the results, in which the *X* co-ordinate (*S1*) represents the subject group of library, the *Y* co-ordinate (*S2*) represents the subject group of information technology, and the *Z* co-ordinate (*S*3) represents the subject group of information management.

In Figure 3, the *SOM* nodes were dispersed in different positions in the three dimensional space, which enabled users to intuitively observe the differences among journals in terms of salient attributes. Some journal clusters were projected onto high values of a certain co-ordinate (representing a certain group of salient attributes). For example, the journal cluster {*Journal of the American Society for Information Science and Technology* (JASIST)} was projected onto a high value of *S2*, which means that this journal cluster had a prominent subject characteristic of Information Technology. Some journal clusters were projected onto low values of a certain co-ordinate. The corresponding subjects were not the prominent subject characteristics of these journal clusters. For instance, the journal clusters {*Information Systems Management* (ISM)} and {*European Journal of Information Systems* (EJIS)} were projected onto a low value of *S1*, which means that the subject group of library was not the important subject characteristics of these journals.

In the same way, the *SOM* display in Figure 1 was projected into the three-dimensional space constructed by the subject group of education (EDU), publishing (PUB), and research methods (RM). See Figure 4 for the results.

The scales of the six co-ordinates were divided into 3 levels. See the dash lines in Figures 3 and 4. The subject characteristics of each journal cluster seen in Figures 3 and 4 were interpreted in Table 5.

It is seen in Table 5 that Cluster Nos. 2, 6, and 16, including six journals in total, showed strong subject characteristic of library, i.e., *College & Research Libraries News* (CRLN), *College & Research Libraries* (CRL), *Journal of Academic Librarianship* (JAL), *School Librarian's Workshop* (SLW), *Information Outlook* (IO), and *Reference & User Services Quarterly* (RUSQ). Cluster Nos. 3, 13, 14, and 18, including seven journals in total, showed this subject characteristic to a medium extent.

Cluster Nos. 16 and 18, including three journals in total, showed strong subject characteristic of information technology, i.e., *Information Outlook* (IO), *Reference & User Services Quarterly* (RUSQ), and *Journal of the Medical Library Association* (JMLA). Cluster Nos. 1, 2, 3, 4, 15, and 19, including 12 journals, showed this subject characteristic to a medium extent.

Cluster No. 16, namely *Information Outlook* (IO), and *Reference & User Services Quarterly* (RUSQ), showed strong subject characteristics of information management. Cluster Nos. 1, 2, 3, 15, 17, 18, 19, including 10 journals, showed this subject characteristic to a medium extent.

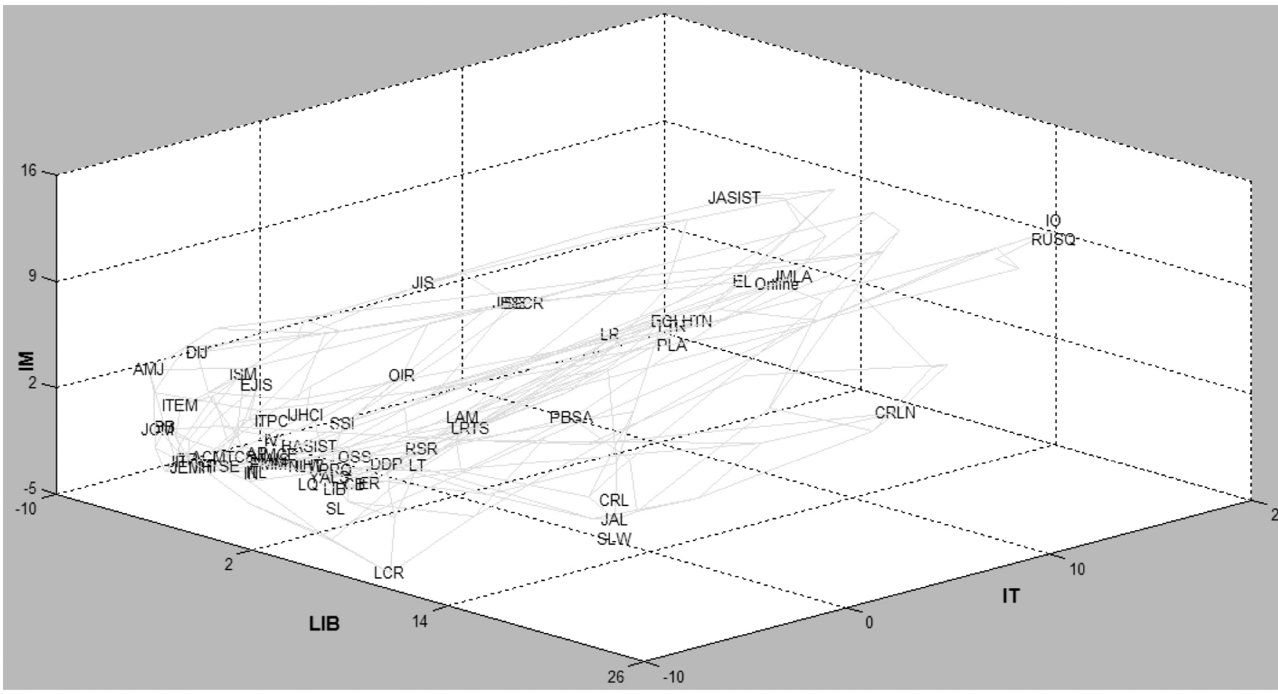

Figure 3. The salient attribute projection for the *SOM* display of journals (I)

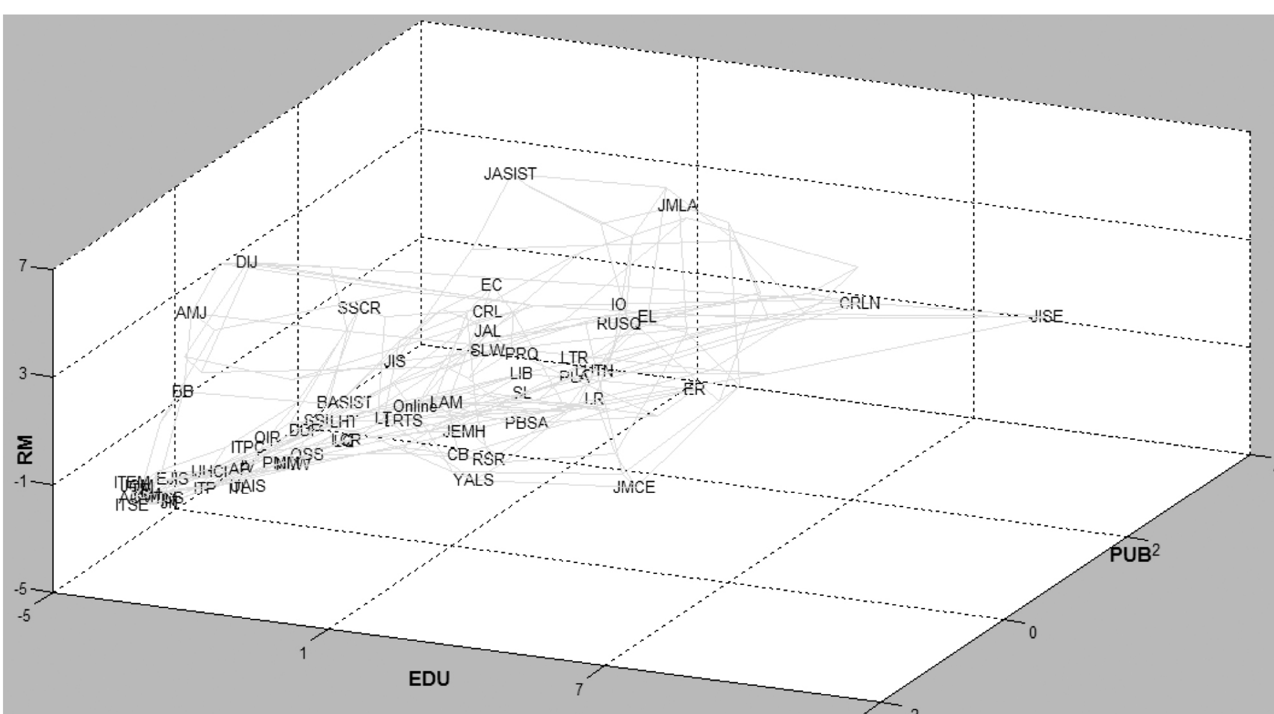

Figure 4. The salient attribute projection for the *SOM* display of journals (II)

Cluster No. 15, namely *Journal of Information Systems Education* (JISE) showed strong subject characteristics of education. Cluster Nos. 2, 14, 16, 18, including 7 journals, showed this kind of subject characteristics to a medium extent.

Cluster Nos. 2 and 7, including five journals in total, showed strong subject characteristics of publishing, i.e., *College & Research Libraries News* (CRLN), *Publishing Research Quarterly* (PRQ), *EDUCAUSE Review* (ER), *The Library* (LIB), and *The Serials Li-*

Cluster No.	Journal labels in each cluster	LIB	IT	IM	EDU	PUB	RM	Subject characteristics
1	(SSCR)(SSI) (JASIST)*	low	2 medium 1 high	2 medium 1 low	low	medium	1 high, 1 medium, 1 low	IT/IM/PUB/RM medium
2	(CRLN)	high	medium	medium	medium	high	low	LIB/PUB high, IT/IM/EDU medium
3	(Online) ,(PBSA)	medium	1 medium 1 high	1 medium, 1 low	Low	medium	low	IT high or medium, LIB/IM/PUB medium
4	(OIR),(IJHCI), (ACMTCS),(ITSE)	low	2 medium 2 low	low	low	low	low	IT medium or low
5	(IN)	low	low	low	low	low	low	entirely low
6	(CRL),(JAL),(SLW)	high	low	low	low	medium	medium	LIB high, PUB/RM medium
7	(ER),(PRQ),(LIB), (SL),(BASIST),(LHT),(LQ), (ITPC),(IV),(UAIS)	low	low	Low	9 low 1 medium	4 high, 3 medium, 3 low	low	PUB medium
8	(EJIS), (ISM),(ITEM), (JOM)*	low	low	3 low, 1 medium	low	low	low	IM medium or low
9	(DDP)	low	low	low	low	low	low	Entirely low
10	(EC),(LAM),(CB), (RSR)	3 low, 1 medium	1 high 1 medium, 2 low	Low	3 low 1 medium	1 high 2 medium, 1 low	1 medium 3 low	IT/PUB high or medium
11	(NLW),(AP),(ITL), (ITP),(JIT),(OSS), (PMM),(LRTS), (EMH),(LHTN), (LTR),(PLA),(LR), (EL)	8 low, 6 medium	9 low, 4 medium, 1 high	8 low, 6 medium	8 low, 6 medium	10 low, 4 medium	7 medium 7 low	RM medium or low
12	(TP), (BB), (AMJ)	low	low	2 low, 1 medium	low	low	2 medium 1 low	RM medium or low
13	(LCR)	medium	low	Low	low	low	low	LIB medium
14	(LT),(YALS), (JMCE)	2 medium 1 low	low	low	2 medium 1 low	low	1 medium 2 low	LIB/EDU/RM medium or low
15	(JISE)	low	medium	medium	high	low	high	EDU/RM high, IT/IM medium
16	(IO),(RUSQ)	high	high	high	medium	low	high	LIB/IT/IM/RM high, EDU medium
17	(DIJ)	low	low	medium	low	low	high	IM/RM medium
18	(JMLA)	medium	high	medium	medium	medium	high	IT/RM high, LIB/IM/EDU/PUB medium
19	(JIS)	low	medium	medium	low	medium	medium	IT/IM/PUB/RM medium

Table 5. The subject characteristics of journal clusters

brarian (SL). Cluster Nos. 1, 3, 6, 18, and 19, including 10 journals, showed this kind of subject characteristics to a medium extent.

Cluster Nos. 15, 16, 17, and 18, including 5 journals in total, showed strong subject characteristics of research methods, i.e., *Journal of Information Systems Education* (JISE), *Information Outlook* (IO), *Reference & User Services Quarterly* (RUSQ), *Drug Information Journal* (DIJ), *and Journal of the Medical Library Association* (JMLA). Cluster Nos. 6 and 19, including four journals, showed this kind of subject characteristics to a medium extent.

Notice that Cluster Nos. 5 and 9 had low values on the six co-ordinates. It does not necessarily mean that these journals were seldom involved in the groups of salient attributes (subjects). The salient attribute projection shows the comparative magnitude of the *SOM* nodes in terms of their corresponding salient attributes. A *SOM* node which was projected onto a low value on the *X* (or *Y, Z*) co-ordinate means that journals projected onto this *SOM* node tended to involve less in subjects in *S1* (or *S2*, *S3*) compared with other journals in the study instead of being compared with other subjects. That is to say, even if a journal cluster was projected onto low values on the six co-ordinates, the journals in this cluster may actually involve quite a lot in some salient attributes (subjects). However, compared with other journals, the journals in this cluster did not excel in these subjects. Thus, these subjects were not suitable for labelling this journal cluster. The subject characteristics of this journal cluster may reside in some subjects other than the six groups of salient attributes.

4.6 The subject characteristics of journals with high Ifs

The impact factors (*IF*s) issued by *ISI* measure the cited ratio of journal articles. Journals with high *IF*s are usually believed to be of high quality and receive more attention from libraries and users than those with low *IF*s. Thus it is interesting to explore the subject characteristics of the journals with high *IF*s. Do they have common subject characteristics or not? The findings will help people understand the subject emphasis of these journals and reveal the cited patterns among the different fields in the library/information-related discipline.

Rank the journals in Appendix I in a descending order of their *IF*s if applicable. Sixteen journals were found to have an *IF* higher than 1.0. See Table 5 for the 16 high *IF*-journals with underlines. Below is the summary of the subject characteristics of high *IF* journals.

The 16 journals with high *IF*s formulate 8 clusters. Several high *IF* journals tend to have similar subject coverage. This includes: *European Journal of Information Systems* (EJIS), *Information Systems Management* (ISM), *IEEE Transactions on Engineering Management* (ITEM), and *Journal of Operations Management* (JOM).

Nearly all journals with high *IF*s in this study showed weak subject characteristics of library except for *Journal of the Medical Library Association* (JMLA) with a medium value.

Two journals, i.e., *Journal of the American Society for Information Science and Technology* (JASIST) and *Journal of the Medical Library Association* (JMLA), showed strong subject characteristics of information technology and three journals, i.e., *Online Information Review* (OIR), *Portal: Libraries and the Academy* (PLA), and *Journal of Information Science* (JIS), showed this kind of subject characteristics to a medium extent.

Six journals, i.e., *Journal of the American Society for Information Science and Technology* (JASIST), *Information Systems Management* (ISM), *Portal: Libraries and the Academy* (PLA), *Academy of Management Journal* (AMJ), *Journal of the Medical Library Association* (JMLA), and *Journal of Information Science* (JIS), showed the subject characteristics of information management to a medium extent. However, only two journals, i.e., *Portal: Libraries and the Academy* (PLA), and *Journal of the Medical Library Association* (JMLA), showed the subject characteristics of education to a medium extent.

Four journals, i.e., *Journal of the American Society for Information Science and Technology* (JASIST), *Portal: Libraries and the Academy* (PLA), *Journal of the Medical Library Association* (JMLA), and *Journal of Information Science* (JIS), showed the subject characteristics of publishing to a medium extent.

Two journals, *Journal of the American Society for Information Science and Technology* (JASIST) and *Journal of the Medical Library Association* (JMLA), showed strong subject characteristics of research methods and four journals, namely *Portal: Libraries and the Academy* (PLA), *Briefings in Bioinformatics* (BB), *Academy of Management Journal* (AMJ), and *Journal of Information Science* (JIS), showed this kind of subject characteristics to a medium extent.

The number of journals which had high, medium or low values on the six salient groups of subjects is illustrated in Figure 5.

In summary, among the six salient groups of subjects, the subject group of library is the least involved in journals with high *IF*s. Subjects related to information technology, research methods, or information management tend to be the subject emphasis of these journals. Only a minority of journals with high *IF*s showed the subject characteristics of education or publishing to a medium extent.

5. Conclusion and Research Limitation

In this study, the *SOM* algorithm was employed to study the clustering structure of journals based on their subjects. The 60 journals were clustered into 19 categories with the help of *U-matrix*. In order to provide suggestions for libraries and users regarding journal selection, the journal(s) in each category were determined whether they were indexed by *SCI* or *SSCI*. Four journals, i.e., *College & Research Libraries News*

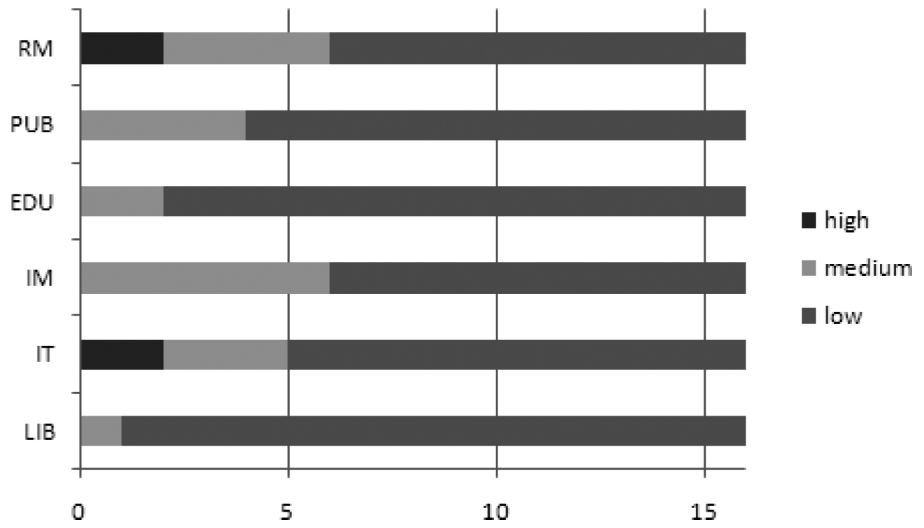

Figure 5. The subject characteristics of journals with high IFs

(CRLN), *DTTP, Documents to the People* (DDP), *Libraries & the Cultural Record* (LCR), and *Journal of Information Systems Education* (JISE), were found to have different subject characteristics from journals indexed by *SCI* or *SSCI*. They are recommended to journal collections of libraries to effectively supplement the subject deficiency of *SCI/SSCI* journals. A journal selection policy is recommended to consider both the quality and the subject coverage of journals.

In order to identify the salient subjects among journals, a novel *SOM* display named *AA-matrix* was proposed and found to correlate significantly with the total occurrences of the subjects in the investigated journals. The *AA-matrix* was applied to the background colours of the *SOM* display generated upon the subject-journal input matrix. Eighty-six subjects, approximately 2% of all subjects were found to be salient subjects which appeared the most in the investigated journals and can be classified into six groups, i.e., library, information technology, information management, education, publishing, and research methods. It is noteworthy that we are not classifying the 60 journals into these six broad subject categories. In other words, the 86 salient subjects did not cover all the subjects that the 60 journals involved. They were actually a sub-set of the latter.

To determine the subject characteristics of these journal clusters, the *SOM* display generated upon journal-subject input matrix was projected into a three-dimensional space twice, each of which was constructed by the two groups of salient attributes (subjects). The subject characteristics of the 19 journal clusters were summarized and discussed based on the observations on the salient attribute projection results.

Finally, an exploration study was conducted to analyze the subject characteristics of journals with high *IF*s. Sixteen journals with *IF*s higher than 1.0 were selected and 8 clusters were formulated. The subject group of library is the least to be involved in journals with high *IF*s. Subjects related to information technology, research methods, or information management tend to be the subject emphasis of these journals. Only a minority of journals with high *IF*s showed the subject characteristics of education or publishing to a medium extent.

The findings in this study can help libraries and users better understand the subject structure of the investigated library/information-related journals and provide suggestions regarding the journal selection and usage. The research methods established in this study can be applied to other academic literature in other disciplines, such as business, education, engineering, etc.

The limitation of this study concerns the investigated journals. As it provides subject indexing for journal articles, *Proquest Research Library* was selected as the data source in this study. However, due to its journal coverage, some important library/information related journals were not investigated in this study, such as *Knowledge Organization*, *Information Processing and Management*, *The Canadian Journal of Information & Library Sciences*, etc. In future research, we hope to find some other academic literature database which also provides subject indexing for journal articles and covers the important library/information related journals omitted in this study so that their subject characteristics can be investigated further. In addition, more experiments need to be conducted upon the data other than those provided by *Proquest* to verify the effectiveness of the *AA-matrix* further.

References

About SOM Toolbox. 2005. http://www.cis.hut.fi/projects/somtoolbox/about.

An, Lu; Zhang, Jin and Li, Gang. 2009. Methodology study of applying Self-Organizing Map to data analysis. *Journal of the China Society for Scientific and Technical Information* 28: 720-26.

Archibald, G. and Line, M.B.1991. The size and growth of serial literature 1950-1987, In terms of the the number of articles per serial. *Scientometrics* 20: 173-96.

Åström, Fredrik. 2007. Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology* 58: 947–57.

Bensman, Stephen. 2001. Bradford's law and fuzzy sets: Statistical implications for library analyses. *IFLA journal* 27: 238-46.

Börner, Katy; Chen, Chaomei and Boyack, Kevin W. 2003. Visualizing knowledge domains. In Cronin, Blaise, ed., *Annual review of information science & technology* 37. Medford, USA, pp. 179-255.

Carpenter, M. P. and Narin, F. 1980. The subject composition of the world's scientific journals. *Scientometrics* 2: 53-63.

Chen, Chaomei; Ibekwe-SanJuan, Fidelia and Hou, Jianhua. 2010. The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology* 61: 1386-1409.

Chua, Alton Y. K. and Yang, Christopher C. 2008. The shift towards multi-disciplinarity in information science. *Journal of the American Society for Information Science and Technology* 59: 2156-70.

Costa, Joan. 1998. *La esquemática: Visualizar la información.* Barcelona, Paidós.

Deerwester, Scott, Dumais, Susan T. Furnas, George W., Landauer, Thomas K., Harshman, Richard. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science* 41: 391-407.

Facts and Figures. 2010. http://www.bl.uk/aboutus/quickinfo/facts/index.html.

Filho, Abel Guilhermino da S., Frery, Alejandro C., Arajo, Cristiano Colho de, Alice, Haglay, Cerqueira, Jorge, Loureiro, Juliana A., Lima, Manoel Eusebio de, Oliveira, Maria das Graas S., Horta, Michelle Matos. 2003. Hyperspectral images clustering on reconfigurable hardware using the k-means algorithm. *Proceedings of the 16th Symposium on Integrated Circuits and Systems Design.* Washington, DC: IEEE Computer Society, pp. 99-104.

Fort, Jean-Claude, Letremy, Patrick, and Cottrell, Marie. 2002. Advantages and drawbacks of the batch Kohonen algorithm. In Verleysen, M. ed., *Proceedings of 10th European Symposium on Artificial Neural Networks (ESANN).* Bruges, Belgium, pp. 223-30.

Freeman, Richard T. and Yin, Hujun. 2004. Adaptive topological tree structure for document organization and visualization. *Neural networks* 17: 1255-71.

Gabrilovich, Evgeniy and Markovitch, Shaul. 2007. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, Hyderabad, India, pp.1606-11.

Guerrero Bote, Vicente P., Moya Anegón, Félix, and Herrero Solana, Victor. 2002. Document organization using Kohonen's algorithm. *Information processing & management* 38: 79-89.

Hofmann, Thomas. 1999. Probabilistic latent semantic analysis. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence* (UAI'99), Stockholm, pp. 289-96.

Hoon, M.J.L. de, Imoto, S., Nolan, J., and Miyano, S. 2004. Open source clustering software. *Bioinformatics* 20:1453-54.

Hull, David. 1994. Improving text retrieval for the routing problem using latent semantic indexing. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* New York: Springer-Verlag Inc, pp. 282-91.

Janssens, Frizo; Leta, Jacqueline; Glänzel, Wolfgang and De Moor, Bart. 2006. Towards mapping library and information science. *Information processing & management* 42:1614-42.

Journal citation reports. 2009. http://www.thomsonreuters.com/products_services/scientific/Journal_Citation_Reports.

Kaski, Sami.1997. *Data exploration using self-organizing maps.* Helsinki: Helsinki University of Technology.

Kaufman, L., and Rousseeuw, P. J. 1990. *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley and Sons Inc.

Kivlighan Jr, Dennis M., and Miles, Joseph R. 2007. Content themes in group dynamics: Theory, research and practice, 1997-2002. *Group dynamics* 11n3: 129-39.

Kohonen, Teuvo.1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics* 43n1: 59-69.

Kohonen, Teuvo. 1993. Things you haven't heard about the Self-Organizing Map. In *Proceedings of International Conference on Neural Networks (ICNN).* Piscataway: IEEE.

Kohonen, Teuvo. 1996. New developments and applications of self-organizing maps. In *Proceedings of International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing.* San Francisco: IEEE.

Kohonen, Teuvo, Oja, Erkki, Simula, Olli, Visa, Ari, and Kangas, Jari. 1996. Engineering applications of the self-organizing map. *Proceedings of the IEEE* 84: 1358-84.

Kohonen, Teuvo. 1999. Analysis of processes and large data sets by a self-organizing method. In *Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials*. Honolulu, USA.

Kohonen, Teuvo. 2001. *Self-organizing maps* (3rd ed.). Berlin: Springer.

Leydesdorff, Loet and Bensman, Stephen. 2006. Classification and powerlaws: The logarithmic transformation. *Journal of the American Society for Information Science* 57: 1470-86.

Lowry, Paul Benjamin; Romans, Denton and Curtis, Aaron. 2004. Global journal prestige and supporting disciplines: A scientometric study of information systems journals. *Journal of the Association for Information Systems* 5n2: 29-77.

MacQueen, James. B. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 281-97.

Marion, Linda S., Wilson, Concepcion S., and Davis, Mari. 2005. Intellectual structure and subject themes in information systems research: A journal co-citation study. In *Proceedings of the 68th annual meeting of the American Society for Information Science and Technology.* Charlotte: American Society for Information Science and Technology.

Mehotra, Kishan, Mohan, Chilukuri K., and Ranka, Sanjay. 1997. *Elements of Artificial neural networks.* Cambridge: MIT Press.

Moya-Anegón, Félix, Herrero-Solana, Victor, and Jimenez-Contreras, Evaristo. 2006. A connectionist and multivariate approach to science maps: The SOM, clustering and MDS applied to library and information science research. *Journal of information science* 32: 63-77.

Proquest. 2009. http://proquest.umi.com/login.

Shewchuk , Richard M., O'Connor, Stephen J., Williams, Eric S., and Savage, Grant T. 2006. Beyond rankings: Using cognitive mapping to understand what health care journals represent. *Social science & medicine* 62: 1192-1204.

SOM_norm_variable. 2002. http://www.cis.hut.fi/som toolbox/package/docs2/som_norm_variable.html.

Ultsch, Alfred and Siemon, H. Peter. 1990. Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of International Neural Network Conference.* Dordrecht: Kluwer.

Ultsch, Alfred.1992. Self-Organizing neural networks for visualization and classification. In *Proceedings of Conference of Society for Information and Classification.* Dortmund, Germany.

Ultsch, Alfred. 2003. Maps for the visualization of high-dimensional data spaces. In *Proceedings of Workshop on Self-Organizing Maps (WSOM'03).* Kyushu, Japan.

White, Howard D. and McCain, Katherine W. 1998. Visualizing a discipline: an author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science* 49: 327–55.

Zhang, Jin, An, Lu, Tang, Tao, and Hong, Yi. 2009. Visual health subject directory analysis based on users' traversal activities. *Journal of the American Society for Information Science* 60: 1977-94.

Zhang, Jin and An, Lu. 2010. Visual component plane analysis for the medical subjects based on a transaction log. *The Canadian journal of information & library sciences* 34: 83-111.

Zhao, Dangzhi, and Strotmann, Andreas. 2008a. Information science during the first decade of the Web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology* 59: 916-37.

Zhao, Dangzhi, and Strotmann, Andreas. 2008b. Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic coupling analysis. *Journal of the American Society for Information Science and Technology* 59: 2070-86.

Zins, Chaim. 2007. Knowledge map of information science. *Journal of the American Society for Information Science and Technology* 58: 526–35.

Appendix I

The list of investigated journals in this study

Label	Journal Title	SSCI	SCI	IF
AMJ	Academy of Management Journal	Y	N	6.079
ACMTCS	ACM Transactions on Computer Systems	N	Y	2.391
AP	Aslib Proceedings	Y	Y	0.493
BB	Briefings in Bioinformatics	N	Y	4.627
BASIST	Bulletin of the American Society for Information Science and Technology	N	N	-
CB	Collection Building	N	N	-
CRLN	College & Research Libraries News	N	N	-
CRL	College & Research Libraries	Y	N	0.781
DIJ	Drug Information Journal	N	Y	0.504
DDP	DTTP, Documents to the People	N	N	-
EC	EContent	Y	N	0.271
ER	EDUCAUSE Review	N	N	-
EJIS	European Journal of Information Systems	N	Y	1.202
IN	IEEE NETWORK	N	Y	3.068
ITEM	IEEE Transactions on Engineering Management	Y	Y	1.156
ITPC	IEEE Transactions on Professional Communication	Y	N	0.609
ITSE	IEEE Transactions on Software Engineering	N	Y	3.569
IO	Information Outlook	N	N	-
ISM	Information Systems Management	N	Y	1.242
ITP	Information Technology & People	N	N	-
ITL	Information Technology and Libraries	Y	Y	0.703
IV	Information Visualization	N	N	-
IJHCI	International Journal of Human - Computer Interaction	N	N	-
JAL	Journal of Academic Librarianship	Y	N	0.667
JEMH	Journal of Educational Multimedia and Hypermedia	N	N	-
JIS	Journal of Information Science	Y	Y	1.648
JISE	Journal of Information Systems Education	N	N	-
JIT	Journal of Information Technology	Y	Y	1.966
JOM	Journal of Operations Management	Y	Y	2.420
JASIST	Journal of the American Society for Information Science and Technology	Y	Y	1.954
JMLA	Journal of the Medical Library Association	Y	N	1.669
JMCE	Journalism & Mass Communication Educator	N	N	-
LCR	Libraries & the Cultural Record	N	N	-
LAM	Library Administration & Management	N	N	-
LHTN	Library Hi Tech News	N	N	-
LHT	Library Hi Tech	Y	N	0.344
LRTS	Library Resources & Technical Services	Y	N	0.698
LR	Library Review	N	N	-
LTR	Library Technology Reports	N	N	-
LT	Library Trends	Y	N	0.239
NLW	New Library World	N	N	-
OSS	OCLC Systems and Services	N	N	-
OIR	Online Information Review	Y	Y	1.103

Label	Journal Title	SSCI	SCI	IF
Online	Online	Y	N	0.352
PMM	Performance Measurement and Metrics	N	N	-
PLA	Portal: Libraries and the Academy	Y	N	1.146
PRQ	Publishing Research Quarterly	N	N	-
RUSQ	Reference & User Services Quarterly	Y	N	0.339
RSR	Reference Services Review	N	N	-
SLW	School Librarian's Workshop	N	N	-
SSCR	Social Science Computer Review	Y	Y	0.714
SSI	Social Science Information	Y	N	0.341
TP	Telecommunications Policy	Y	Y	1.244
EL	The Electronic Library	Y	N	0.393
LQ	The Library Quarterly	Y	N	0.364
LIB	The Library	N	N	-
PBSA	The Papers of the Bibliographical Society of America	N	N	-
SL	The Serials Librarian	N	N	-
UAIS	Universal Access in the Information Society	N	N	-
YALS	Young Adult Library Services	N	N	-