ORGANIZATIONAL KNOWLEDGE MAPPING BASED ON LIBRARY INFORMATION SYSTEM

IRANDOC CASE STUDY

Ammar Jalalimanesh[a,*], Elaheh Homayounvala[a]

[a] *Information engineering department, Iranian Research Institute for Information Science and Technology, 1090, Enqelab St., PO Box: 13185-137,1 Tehran, Iran,*

ABSTRACT

One of the most popular techniques for identifying knowledge in organizations is knowledge mapping. It can help decision makers to better understand the knowledge flow within the organizations. Mapping organizations knowledge, especially in research institutes, has attracted much attention from senior managements in recent years. Libraries, among the most important parts of research institutes, have a significant role in scientific advances. Due to this important role, many knowledge operations take place in collaboration with libraries. All of library transactions including users borrowing and returning logs and also books metadata are recorded in library information systems. Users' transaction logs are rich resources to extract information about knowledge operations in an organization.

In this paper we propose a new methodology for drawing knowledge map, based on library information system logs. Our proposed methodology contains five steps including data collection and making data warehouse, data preprocessing and refinement, applying knowledge mapping algorithm for extracting input data for mapping, drawing knowledge map and finally analyzing the results. According to this methodology, we have drawn the IRANDOC knowledge map emphasizing interdisciplinary domains based on library information system users' logs. IRANDOC knowledge map shows most studied subjects and also interrelation between them which are invaluable source of knowledge for IRANDOC decision makers in order to initiate research projects.

KEYWORDS:

Knowledge mapping, library information system, log analysis, decision making

[*] Corresponding author. Tel.: 66951430 (Ext. 343), E-mail: Jalalimanesh@irandoc.ac.ir

1. INTRODUCTION

Knowledge is one of the most important strategic resources of organizations (McLure Wasko and Faraj, 2000, Hult and Ketchen, 2006). Due to the importance of knowledge, knowledge management becomes a controversial issue for organizations these days. Knowledge management requires identification, generation, acquisition, diffusion and capturing the benefits of knowledge which provides strategic advantages to the organization (Dalkir, 2007). Knowledge identification is an important phase in knowledge management and is a prerequisite for other steps. One of the most popular techniques for identifying knowledge in organization is knowledge mapping.

Knowledge mapping can help decision makers to better understand the knowledge flow within the organizations. These maps can be built in order to illustrate the knowledge sources, sinks, and constraints (Chan and Liebowitz, 2006, Grey, 1999). There have been many researches focusing on the importance of knowledge mapping in organizations and society (Eppler, 2001, Huijsen et al., 2004, Jafari et al., 2009, Wexler, 2001). Some of them try to make a framework for method selection (Jafari et al., 2009, Wexler, 2001) and others focus on defining conceptual and practical methods for drawing knowledge maps (Eppler, 2001, Huijsen et al., 2004). Eppler has classified organizational knowledge maps in five categories including Knowledge source, asset, structure, application and development maps (Eppler, 2001). According to his classes, knowledge source maps answer questions about organization capability for handling projects. Knowledge asset maps visually exhibit the existing stock of knowledge of an individual, a team, a unit, or the whole organization. Knowledge structure maps depict the global architecture of a knowledge domain as well as the way in which its parts are inter connected. The application maps demonstrate which type of knowledge should be applied in a specific business situation and finally knowledge development maps can be used to illustrate the necessary stages to develop a certain competence.

Mapping the knowledge of an organization especially in research institutes has attracted much attention from senior managements in recent years. Decision makers in these institutions, are interested in figuring out the areas that mostly fit to their backgrounds. However, the organizational knowledge has several dimensions and it is hard to draw a comprehensive knowledge map. Data analysis of organizational databases can assist decision makers in depicting such maps.

Libraries are among the most important parts of research institutes and have a significant role in scientific advances. Due to this important role, many knowledge operations take place in collaboration with libraries. We believe that users' transaction logs are rich sources of data from which organization's knowledge flows can be extracted. According to this hypothesis, we tried to draw the Iranian Research Institute for Information Science and Technology (IRANDOC) knowledge map emphasizing interdisciplinary domains based on library information system logs.

In this paper we propose a novel methodology for drawing knowledge map based on library information system logs. According to Eppler categories, our knowledge map is a combination of knowledge source, asset and structure maps. The aim of this research is to find interdisciplinary domains that IRANDOC can establish future research projects on them. We analyzed logs of borrowing from LIS[1] and extract the relation between knowledge fields according to user's subjects of study and based on library of congress (LC) classification. We assume that when users borrow from any couple of subjects, then they have knowledge to connect those subjects. More borrowing from both subjects shows that IRANDOC has a strong chance of success, defining interdisciplinary research projects connecting these subjects.

The organization of this paper is as follows. In the second section, we describe our proposed methodology for drawing knowledge map based on LIS logs, including data collection procedure and pre-processing of raw data, as well as the logic behind the mapping algorithm. In section three, we explain our case study for drawing IRANDOC knowledge map. In this section we demonstrate implementation steps and results of the project. In section four, we make discussion and conclusion and finally we present a number of interesting topics for future research.

[1] Library Information System

2. LIS-BASED ORGANIZATIONAL KNOWLEDGE MAPPING

Since knowledge in organizations is hidden, we need to establish a methodology for knowledge discovery from facts. There are some key resources in an organization that can supply these facts and show the knowledge flows. Organization's library is one of these key resources especially in case of research or academic organizations. Since the most important knowledge assets of organizations are their researchers, tracking their information seeking behavior is a good way for understanding the state of organizational knowledge in every field. All of library transactions including users borrowing and returning logs and also books metadata are recorded in library information system. Therefore, by analyzing these logs we can extract valuable information about information seeking behavior of library users. Our methodology for drawing knowledge map from LIS logs which consists of five phases is depicted in figure (1). These phases are as follows: 1-Data collection and making data warehouse, 2-Data preprocessing and refinement, 3-Applying knowledge mapping algorithm in order to extract input data for the map, 4-Drawing knowledge map 5-Analyzing the results..

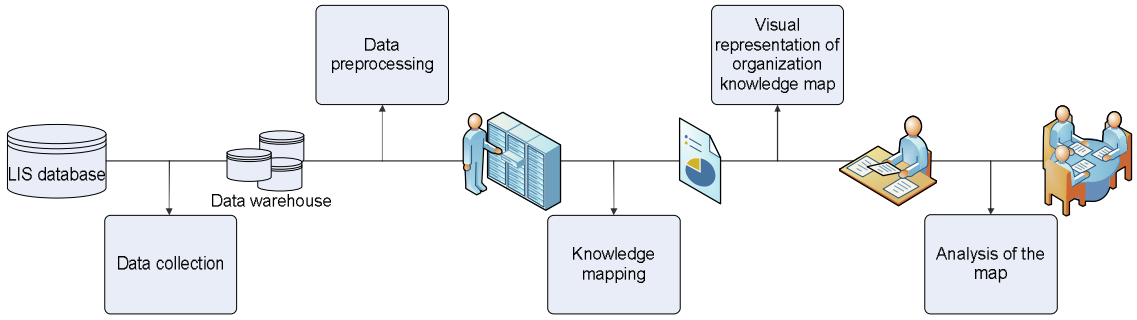

Figure 1: LIS knowledge mapping phases.

2.1 DATA COLLECTION AND PREPROCESSING

The first step is to collect data and establish a data warehouse from LIS. Data warehouse is an integrated data repository including historical data of an organization for supporting decision-making processes(Song and LeVan-Shultz, 2010). In this phase, library transaction data, users' data and books metadata including subjects are extracted from LIS and a new data model based on our needs for decision making process is built in form of a relational database. In the second phase, data has to be preprocessed and prepared for mapping. In data refinement phase incorrect, incomplete and outlier data are cut from repository. Then, in order to prepare exact input data for mapping process, some queries are formed to be answered by our data warehouse. At the next stage, mapping algorithm is applied on collected data to draw organizational knowledge map.

2.2 KNOWLEDGE MAPPING ALGORITHM AND VISUALIZATION

We assume that if a researcher gets a book, he would have knowledge about its topic or is interested in it. Then, if the researcher studies two books from different subject, he would have knowledge in both of them and may have the ability to do research in interdisciplinary fields connecting the two subjects. We need a template for partitioning knowledge to distinct areas for mapping organizational knowledge,. This template also needs to support knowledge flows in the organization. Based on this requirement and since books and other library documents are classified by LC code, we decided to apply Library of Congress (LC) categories as knowledge partitioning template. From this point of view, we calculated the relation between binary combinations of all subjects by counting the amount of

books that every user studied from these subjects. Furthermore, we calculated size of every subject as a measure of organizational knowledge on that subject, by aggregating the number of borrowed books by all library users in the field.

Formulas (1) and (2) show the core of our algorithm for drawing the organizational knowledge map based on the borrowing logs. O_i denotes the size of subject i in the map, where N_{im} is the number of studied books by user m from subject i. Therefore, all the studied books by all users in that subject are aggregated to calculate the size of it. In order to reduce the difference between the largest and smallest subject size, the result were transformed by Neperian logarithm (ln). In formula (2) W_{ij} depicts the weight of relation between subjects i and j where N_{im} and N_{jm} are the number of books that users m borrowed from subjects i and j. We assume that this researcher has interdisciplinary knowledge connecting those subjects equal to minimum number of books that he studied from each of them.

(1) $$O_i = LN\left(\sum_{m=1}^{k} N_{im}\right)$$

(2) $$W_{ij} = \sum_{m=1}^{k} Min(N_{im}, N_{jm})$$

Now we have a cross matrix of relation between subjects W and matrix of subjects size O for drawing a knowledge map. The final output of this algorithm is an undirected graph composed of K nodes as subjects and connections between these nodes, if we have $W_{ij} \neq 0$ per Node i and j.

3. IRANDOC CASE STUDY

Iranian Research Institute for Information Science and Technology (IRANDOC) is an institute affiliated with the Ministry of Science, Research, and Technology (MSRT) which was established to work in the field of science and technology of Information and Librarianship. IRANDOC library books are organized based on the Congress Classification System and is run on the basis of the Open-shelf System. The users of the library are comprised of university professors, students, researchers, and the IRANDOC staff. Based on the Library Collection Policy, the IRANDOC Library, at present, provides the following subjects: Information Science and subjects related to the Library Science, Information Systems Management, Information Technology, Information Analysis, Knowledge & Information Management, Linguistics, Computerized Terminology and Technology.

Since research institutes are among knowledge intensive organizations, it is crucial for them to track and explore knowledge flows in their organization. They also need to know about their abilities to plan vision of future.

According to IRANDOC background, most of its previous research projects were interdisciplinary and related to information science and technology. In addition to that, information science has an interdisciplinary soul (Saracevic, 1995). Hence it is obvious that making decision about future research needs knowledge about interdisciplinary domains that IRANDOC has had more experiences and knowledge assets in them. We decided to draw IRANDOC knowledge map based on our algorithm for finding interdisciplinary knowledge capital of our organization. As noted before we assume that if more researchers study more books from two subjects, then these subjects are connected more than others. Therefore IRANDOC can establish its future research projects based on them.

3.1 DATA COLLECTION AND PREPROCESSING

IRANDOC library has about 14000 Latin books. Librarians manage books collection by use of LIS software. The software has a database based on Microsoft SQL-server and has a report generating module. We started with getting query from LIS and exporting data to Microsoft Excel files. Two main queries were formed on Latin books meta-data and borrowing transaction data since last 5 years.

Based on the gathered data, new data-model was designed. Figure (2) shows entity relationship diagram of our data-model. It is consists of 4 tables. Book, Transaction and patron tables are general tables. We built LC Categories table based on Library of Congress Classification (LCC). This table will help us to make better queries in relation with subjects. All the tables are connected directly or indirectly via other tables, so we can make complex queries to find relations between all dependent fields.

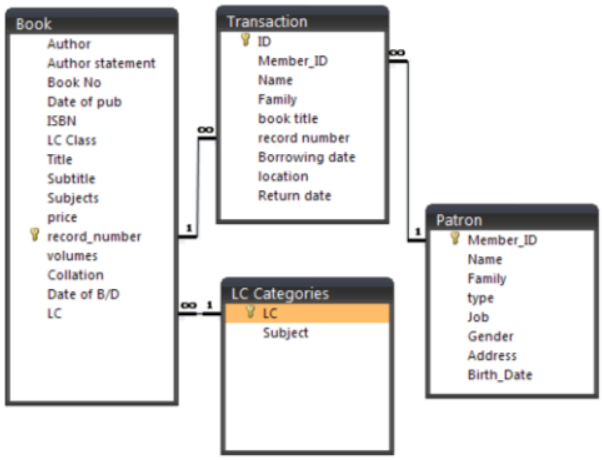

Figure 2: Entity relationship diagram of books borrowing logs.

After building the data-model skeleton in Ms-access, we tried to fulfill it with gathered data in Microsoft Excel format. But there were some problems in data types and also some problems with the data which have been entered in an incorrect way. Some data doesn't have consistency with others, and we had to omit them and also there were some critical incomplete fields which should be filled with correct data. At the end we had data of 12000 Latin books, about 140 users and 4655 borrowing transactions from last 5 years.

3.2 ALGORITHM IMPLEMENTATION AND VISUALIZATION

Our proposed knowledge mapping algorithm is implemented in our case study by utilizing MS-Access and SQL2 scripting language. For making a matrix of subjects size (O), a query was created to aggregates all the books that borrowed in that field by all the users since last five years. Building weight matrix of relation between binary combinations of subjects was more complicated in comparison with previous queries. A complex script was developed to produce the cross matrix between all combinations. Table (1) shows schematic of this matrix for some of high studied subjects.

Table 1: matrix of relation between subjects.

LC	HD	HF	T,	Z,
HD		71	74	124
HF	71		43	48
T,	74	43		54
Z,	124	48	54	

The result of SQL script for producing Subjects relation was 71×71 cross matrix. For drawing the final knowledge map of IRANDOC based on extracted data, an open source add-in for Ms-Excel named NodeXL[2] was used. Figure (3) shows the initial map that we draw by the aid of NodeXL. The map shows the complex graph that many vertices are connected to each other directly or indirectly via few edges.

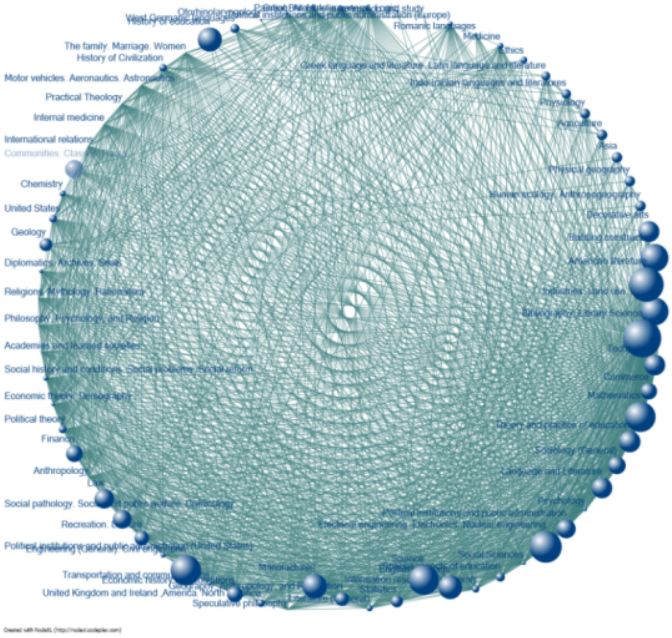

Figure 3: Initial IRANDOC knowledge map based on library users borrowing logs.

Figure (4) shows the same graph filtered to show more related subjects. The size of sphere in each node shows the amount of books that studied by IRANDOC researchers totally. The width of bars that connect subjects together shows the amount of studies as explained in formula (2).

[2] http://www.codeplex.com/NodeXL

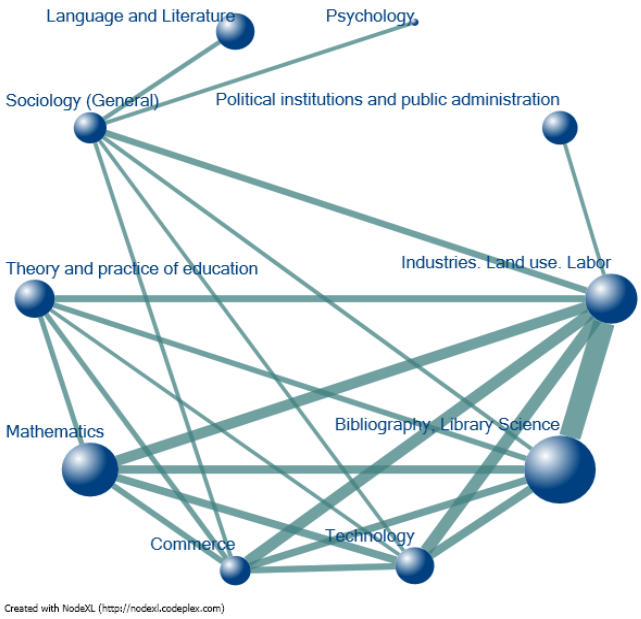

Figure 4: Filtered graph of IRANDOC most related subjects.

4. CONCLUSION AND FUTURE WORKS

Examining the output graphs reveals useful information about knowledge status in IRANDOC especially for interdisciplinary fields. For example, figure (4) shows that HD category (Industries, land use and labor) that contains books about management science, policy making, decision making, organization productivity and industrial management has an important role in connecting other topics together. It means that HD is a hub for relating other topics and IRANDOC can invest on its sub categories in order to handle future projects. It also means that researchers, who have knowledge in this area, would be good assets for managing the research projects. The figure also shows that class Z (Bibliography, Library Science and General Information Resources) is the most studied subject according to historical data and its relation with HD class is more than other subjects. This evidence validate the IRANDOC background in library science and shows the decision makers that their organization have enough knowledge to succeed in research projects in combination of these two subjects.

The map could also be used for other purposes like finding the way to handle interdisciplinary research projects. As another example, if IRANDOC wants to do a research project that needs knowledge of psychology and industrial management, analyzing the visual knowledge map can provide good guidance for such project. According to figure (4), these two subjects are not connected strongly but sociology is connecting them. So it can be inferred that researchers who have knowledge in sociology might play critical role in such projects.

It can be concluded that knowledge mapping in this methodology gives valuable information to organization decision makers. It also clarifies the information seeking behavior of knowledge workers and reveals existing deviations between core strategies and knowledge operations.

Knowledge mapping can be a helpful way for managers and stakeholders to recognize their organization knowledge assets. In this research, we used a straight forward way for mapping knowledge based on real data. Adding more metadata or connecting external databases like research projects will make the map richer. In this study we assumed subjects as knowledge nodes. However, the research groups or even books might be considered as graph vertices. Therefore the network shows the way that groups are connected.

One of the most useful techniques that recently has been widely used in knowledge management is data mining. Techniques such as clustering based on users, books or projects may give us more knowledge about our organization knowledge capitals.

REFERENCES

CHAN, K. & LIEBOWITZ, J. 2006. The synergy of social network analysis and knowledge mapping: a case study. International journal of management and decision making, 7, 19-35.

DALKIR, K. 2007. Knowledge management in theory and practice, Elsevier, Butterworth Heinemann.

EPPLER, M. J. Year. Making knowledge visible through intranet knowledge maps: concepts, elements, cases. In, 2001. IEEE, 9 pp.

GREY, D. 1999. Knowledge mapping: a practical overview. SWS Journal.

HUIJSEN, W. O., DRIESSEN, S. & JACOBS, J. Year. Explicit conceptualizations for knowledge mapping. In, 2004.

HULT, G. T. M. & KETCHEN, D. J. 2006. Knowledge as a strategic resource in supply chains. Journal of Operations Management, 24, 458-475.

JAFARI, M., AKHAVAN, P., BOUROUNI, A. & AMIRI, R. H. 2009. A Framework for the selection of knowledge mapping techniques. Journal of Knowledge Management Practice, 10.

MCLURE WASKO, M. & FARAJ, S. 2000. "It is what one does": why people participate and help others in electronic communities of practice. Journal of strategic information systems, 9, 155-173.

SARACEVIC, T. 1995. Interdisciplinary nature of information science. Ciência da Informação, 24, 36-41.

SONG, I. Y. & LEVAN-SHULTZ, K. 2010. Data Warehouse Design for E-Commerce Environments. Advances in Conceptual Modeling, 374-387.

WEXLER, M. N. 2001. The who, what and why of knowledge mapping. Journal of Knowledge Management, 5, 249-264.