

Bibliometrics Method for Knowledge Mapping: An Application on the Scientific Research in Indonesia

By Ambar Yoganingrum*



1. Introduction

Knowledge mapping is an activity to illustrate the relationships of certain aspects of knowledge. A map of knowledge depicts a natural and cognitive structure of science and tries to look at certain aspects of sciences from the top. Eugene Garfield used the term 'on the shoulders of giants' for researchers who conducted research in science mapping in a conference in 1998. Knowledge mapping is part of information visualization aimed at disclosing hidden patterns from abstract data (Chaomei Chen, 2003).

A method to build a map of knowledge is called bibliometrics. The term 'bibliometrics' was introduced by Pritchard in 1969. Bibliometrics is a quantitative method to measure the development of science and technology based on numbers of publications and patents. This method is an extension of library sciences. It is now applied on many disciplines of sciences including library and information sciences as well as other disciplines such as biology, public health, medicine, engineering, and economics to analyze the development of those subjects.

* Researcher in Center for Documentation and Information – Indonesian Institute of Sciences (PDII-LIPI). Jl. Gatot Subroto 10. Jakarta 12190. Indonesia.

Email: ambar_anugrah@yahoo.com

Bibliometrics method was first used to analyze the development of science statistically from collection of books, journals, periodicals etc in a library. A scientific publication, whether it is a journal article, a book, a research report or a patent document, contains information, which could explain the concepts and cognitive aspects of the publication. If the entities of the publications such as words, authors, institution of authors, names of journals, or subjects are plotted on a map, they will show ideas, concepts, the relationship of concepts, scholarly communications, science development, research directions, the intellectuals' links of knowledge, and may indicate the abstract and invisible aspects of the specific science.

Bibliometrics uses scientific publications as data sources. Analysis used in bibliometrics included co-word and co-citation. Both are the most fundamental techniques for science mapping. The methods have become invaluable tools for researchers across a variety of disciplines. The combination of the two methods could show the way for a clearer picture of the cognitive content of publication (Chaomei Chen, 2003).

A technique to represent the concept and cognitive aspects of knowledge is multidimensional scaling (MDS) and clustering. They are standard statistical methods used on multivariate data. Visualization with MDS is based on the distance or the similarity and the dissimilarity between the objects. The objective of using MDS is to give a clearer picture of the relationships between the objects.

This paper describes how the bibliometrics method works in order to map the science. In this case the paper uses the example of research on knowledge mapping in Indonesia. The paper also explains the use of the map as a tool for any purposes such as to manage the technology, to evaluate research program, and to look at the development process of science and etc.

The structure of this paper is as follows, section one is introduction. Section two explains how the bibliometrics methods work to map knowledge using scientific publications as data sources. Section three explains the use of the map. Section four provides the conclusions.

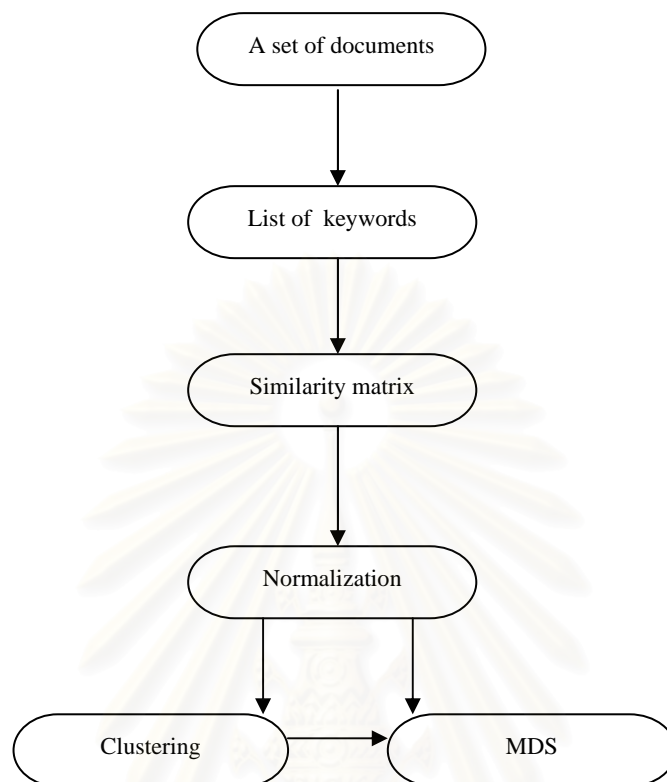
2. Bibliometrics methods for Knowledge Mapping

There are several ways of analysis in using bibliometrics for knowledge mapping. The prominent methods are co-word and co-citation.

2.1. Co-word

This method is developed to avoid the subjectivity of experts in identifying the development of science and technology. Co-word analysis reveals the patterns and trends of a specific discipline by measuring the association strength of the concept among relevant publications. Words on a scientific literature carry and represent the concept existing in the publication. Two keywords co-occurring within the same paper are an indication of a link between the concept to which they refer. Depicting the word on a graph of network would provide a structure of knowledge. If the map is built chronologically, it will show the shift of particular topic or the dynamics of knowledge. There are several ways to build bibliometrics map, but co-word map provides an immediate picture of the actual content of research topics dealt with in the literature (Ding et al, 2000). The steps of co-word analysis are described in Figure 1. The outcome of this method, in general, is a visualization of concept networking. A bibliometrics map, which is built based on words, is called co-word map.

Figure 1. Steps on co-word analysis



Extracting keywords from publications could be conducted by an indexer or using a software. Extracting keywords with software would avoid inconsistencies and judgment of indexer in keyword selection. Using software is also to avoid the time problem that emerges in selecting keyword by indexer, because keyword selection by indexer has to be standard to keep it from synonym, ambiguity, broad and narrow term of the words. One kind of software to extract words directly from full-text documents is Nptools (Ding et al, 2000). Meanwhile Leximappe, a pioneer on software for information visualization based on co-word analysis, was developed in 1980s. Nowadays several countries have developed the similar software such as Content Analysis and Information Retrieval (CAIR), Bibliometric Technology Monitoring (BibTechMon) and Data View* (Ding et al, 2000; Widhalm, 2001).

* <http://crrm.u-3mrs.fr/commercial/software/dataview/dataview.html> tanggal akses: 22 Juli

In the case of Indonesia, the knowledge mapping is still a new area of research so it needs time to socialize it that this area is important. It means that we have to work hard and need enough money to provide tools for science mapping, to develop manpower, and to build or upgrade existing databases and documentation systems in libraries in order to make them compatible with the mapping tools. Therefore a cheaper alternative such as using indexer instead of software is still a choice. In this matter, University of Indonesia collaborated with Indonesian Institute of Sciences has developed a prototype for indexing and co-word technique since 2003 (Yoganingrum, 2003).

As mention above that visualization with MDS is based on the distance or the similarity or the dissimilarity between the objects. The similarity or distance between keywords is measured based on how many times the two keywords appear together in the same publication. The higher the similarity value of a pair of keywords the closer the relation between them. This procedure will produce similarity matrix. Several researchers make the matrix values to be normal by using several coefficients such as Jaccard Index, Inclusion and Proximity Index. The formulas of those coefficients are as follow:

Jaccard Index: $J_{ij} = C_{ij} / (C_i + C_j - C_{ij})$

Inclusion Index: $i_j = C_{ij} / C_i$

Assuming $c_i < c_j$,

Proximity Index: $P(ij) = N C_{ij} / C_i C_j$

C_{ij} = Co-occurrence of Keywords i and j

C_i = The total number of occurrence of keyword i

C_j = The total number of occurrence of keyword j

N = The number of co-occurrences of word i and j

Subsequently clustering of data is taken to overcome the limitation of visualizing the number of data by MDS in addition to have the clearer picture of the field of specific discipline. Then keywords with high frequencies in each cluster were chosen to represent the cluster. Figure 2 shows the example of cluster map. The map depicts the Anthropology researches conducted by foreign researchers in Indonesia using data from 1997 to 2003. Afterwards, Salton Index is applied to count the centrality and density value. Centrality is defined as the mean of strength linked among clusters or and density is the mean of strength linked within a cluster (Ding et al, 2000). The association between clusters is mentioned strong if it has Salton Index of more than 0.2. The map on Figure 2 uses line to indicate the strong linkages according to the Salton Index. For instance cluster 4 has strong relation with clusters 2 and 5. Keywords Healers (cluster 4) and Healing (cluster 5) are separated in difference clusters, however those clusters have a strong linkage. The formula of Salton Index is as follow:

$$\text{Salton Index: } S = C_{ij} / \sqrt{(C_i C_j)}$$

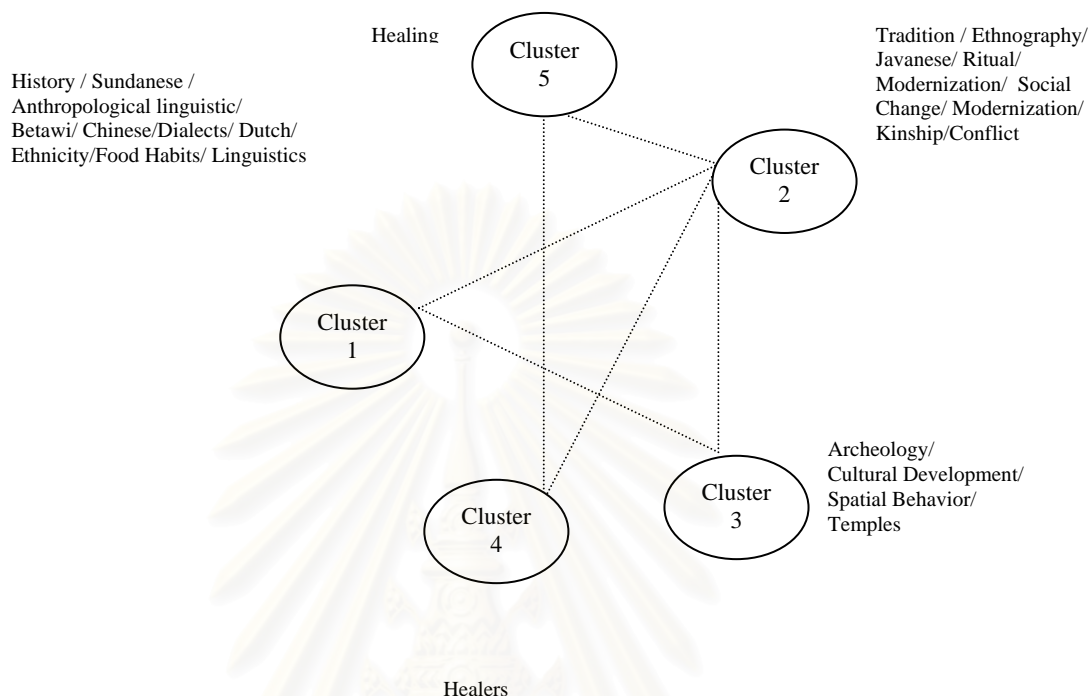
C_{ij} = Co-occurrence of Keywords i and j

C_i = The total number of occurrence of keyword i

C_j = The total number of occurrence of keyword j

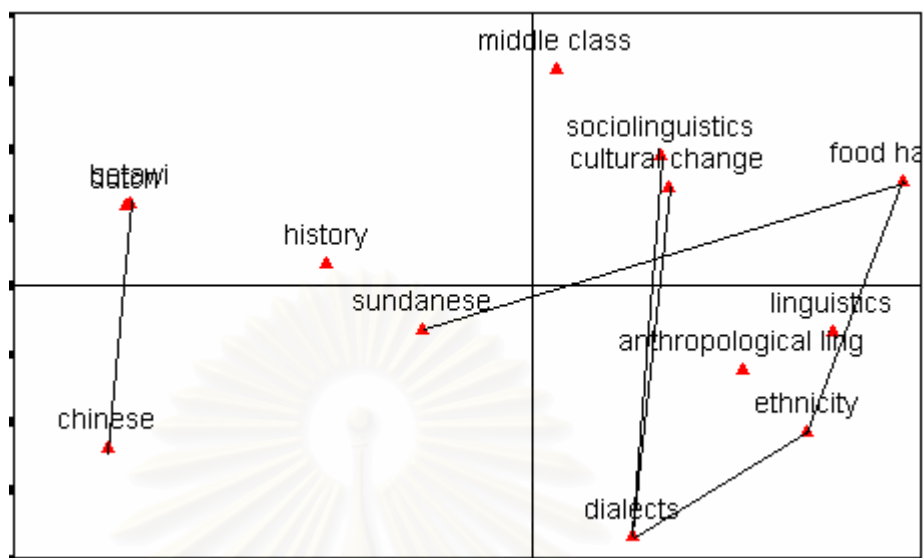
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Figure 2. Co-word based map of Anthropology researches conducted by foreign researcher in Indonesia, 1997-2003 (Yoganingrum et al, 2004)



After that each cluster is mapped by MDS to have a detailed map. This procedure is called multi level mapping (Ding et al, 2000). Figure 3 shows the detailed map of cluster 1 in Figure 2. Salton Index is used also to calculate the correlation among keywords on detailed map. The map shows that the foreign researchers in Indonesia focused on topics of ethnic groups (Chinese, Sundanese and Betawi) in Indonesia related to their food habits and linguistics. The map gave information that there was a close relationship, pointed by a line of salton index, between 'Chinese' and 'Betawi', 'Sundanese' and 'food habits', and 'dialects' and 'cultural changes'.

Figure 3. The detailed map of Cluster 1 (Yoganingrum et al, 2004)



2.2. Co-citation

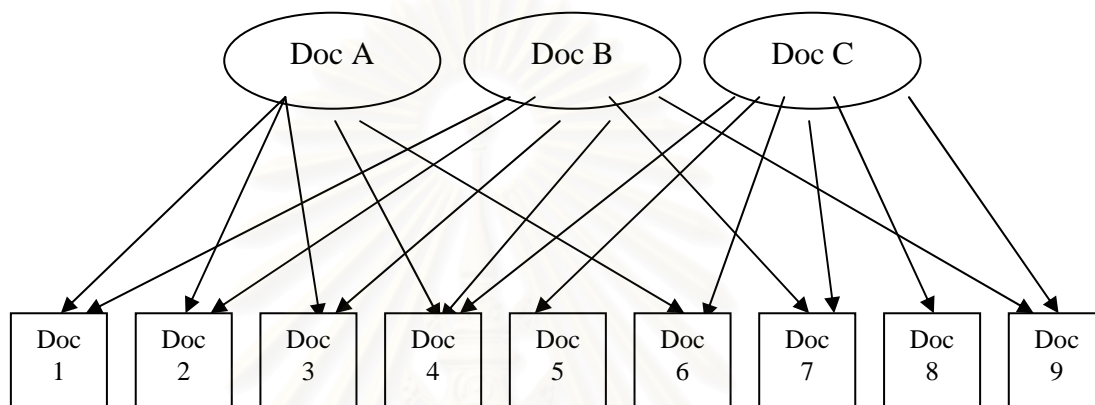
Co-citation is a technique to measure the structure and the dynamics of science and technology from citation pattern of scientific document. The citation pattern can provide insight knowledge of an invisible college (Chaomei Chen, 2003). Co-citation analysis consists of document co-citation analysis (DCA), author co-citation analysis (ACA) and journal co-citation analysis (JCA). One kind of softwares to map science based on co-citation analysis is Essential Science Indicator (ESI)*.

The unit analysis in DCA is document. The similarity matrix value is counted based on the frequency of a pair of documents cited together in the article. Meanwhile the unit analysis in ACA is author. The values of matrix are measured on the frequency of selected author pairs cited together in articles. The unit analysis in JCA is journal. The similarity matrix value is counted based on the frequency of a pair of journal cited together in the article.

* <http://www.thomsonisi.com/demos/esi/>

To make it clearer, the process of co-citation analysis is presented in Figure 4 which is a set of documents. A, B and C have a list of cited document 1 to 10. It shows that document 1 and 2 cited together by document A and document B. It means that document 1 and 2 have distance or similarity value 2. The steps of co-citation analysis are figured out on figure 5.

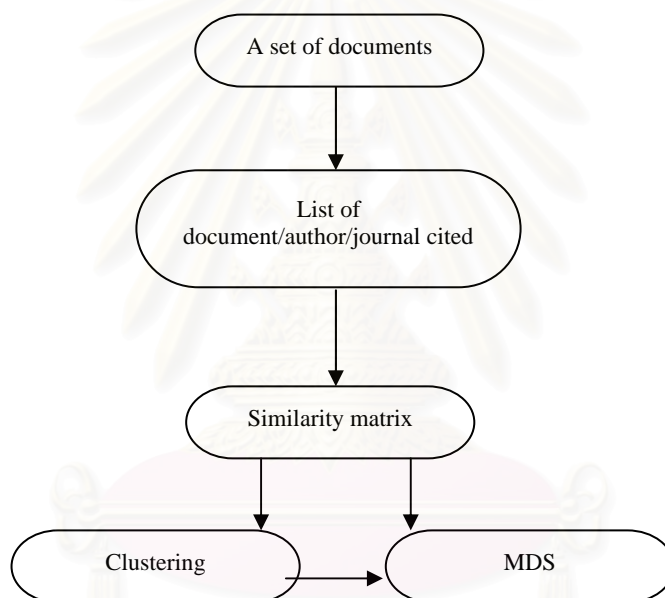
Figure 4. The process of co-citation analysis



There are several objections against co-citation analysis such as loss of relevant papers, inclusion of non-relevant papers, overrepresentation of theoretical papers, time lag between emergence of new specialties and capturing of them in a co-citation map, and subjectivity in threshold setting (Braam, 1991; Chaomei Chen, 2003). Other objection was that co-citation cluster did not represent the entire body of publication that comprised a specialty (Chaomei Chen, 2003). Nevertheless the creation of ACA offers a useful alternative starting point for co-citation analysis. ACA gives a new approach in understanding the structure and the dynamics of science. ACA could be used to fit out and be as a comparison of the result of DCA (Chaomei Chen, 2003).

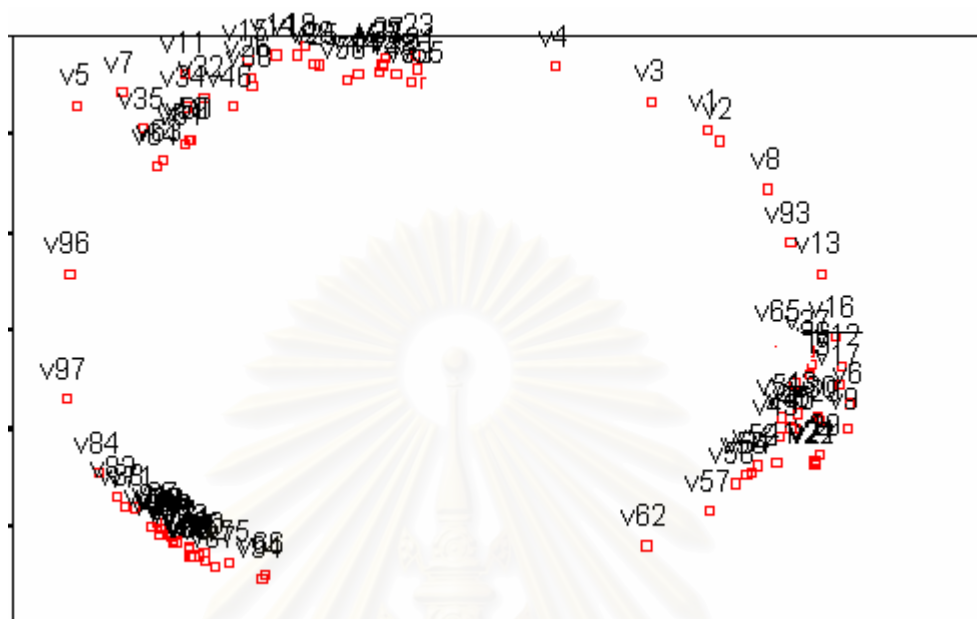
Figure 6 shows the co-citation map of research on biological sciences conducted by LIPI in the period of 1995-2000. Symbol “v” on the map is the document cited. The map was built based on dissimilarity value (not similarity value) between the objects. So the nearer the distance of two objects, the weaker the relation is. It looks that the majority of subjects pile up on same location. It means that the association strength of the majority of objects are weak. It is concluded that a set of researches on biological sciences conducted by LIPI in the period of 1995-2000 is in various domains. The researches did not focus on specific domains.

Figure 5. General steps on co-citation analysis



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Figure 6. The JCA map of the Indonesian Institute of Sciences (LIPI)'s researches in biological sciences, 1995-2000 (Royani, 2005)



3. The Application of Knowledge Mapping

The knowledge mapping of bibliometrics is not only to show the domains of knowledge but also it is able to identify the overlapping, interconnection and interaction between the domains. Some examples of the applications of knowledge mapping are as follows:

3.1. Tools to manage technology

Technology management means to make several aspects of technology in harmony such as the Research and Development (R&D), product engineering, process engineering, manufacturing, system engineering, and services, so that it could make decisions concerning technology. Each technology may have different characters, strategies, or properties so it needs different decisions. To be able to make appropriate decisions it is important to have a clear view of a set of knowledge. Map of technology represents the knowledge domain of a specific technology. It is necessary for practical purposes in technology management such as to define the research program, to design knowledge base structure and to decide technological activities (Pelc, Karol I, 1986).

The definition of research program is needed to fill the gaps between the current knowledge and the technology required. Bibliometrics map which give a picture of domains development of current knowledge and technology will make it easier to identify the gaps and to decide on the activities to reach the goal.

3.2. Tools to evaluate research program

If the map is built based on time series, it could demonstrate the development of domains of knowledge periodically. It could also show the focus of research in each period and the shifts in research interests from one period to another. Therefore, the knowledge mapping is a tool to evaluate the synergy between program and reality. As an example (please refer to figure 2), the researches conducted by foreign anthropologist in Indonesia in the period of 1997-2002 focused on the ethnic groups (Chinese, Sundanese and Betawi) in Indonesia related to food habit, language and cultural changes. Perhaps it would give a different picture for another period of map.

3.3. Tools to track science history and to predict the development of specific science

The map built based on time series will show the changes of domains of knowledge from one period to another. Those changes are perhaps the interaction between two domains that may result in new domains or interconnections of knowledge domains that may give information about the path of the development of the domains. Therefore the map could be used to identify the origin of science and the process of the development of the science.

3.4. Tools to identify the contribution or the productivity of a research institution or an author on a specific discipline

By depicting the domains of researches of an institution or a country on a map, it may demonstrate the contribution of the institution on a specific discipline. It also shows in which domains of knowledge at which an institution focuses on. So the map could reveal the research direction of the institution. Through the map it also could compare the focus of area of research of each institution.

4. Conclusions

Knowledge mapping is an important tool to analyze the development of science, the direction of research, and to track the science history so it may be used to evaluate the research program, to manage the technology, to predict the development of a science and to identify the contribution or the productivity of a researcher or a research institution.

Bibliometrics is a method for knowledge mapping based on quantitative analysis of scientific publication. The eminent techniques of bibliometrics are co-word and co-citation. The method is flexible enough to accommodate changes because of interconnection and interaction between the knowledge domains. The method also is able to represent the level of knowledge with multi level mapping technique. Bibliometrics map can show the phenomenon occurred in the development of science.

5. References

- B M Gupta & Sujit Bhattacharya. 2004. "A Bibliometric Approach towards Mapping the Dynamics of Science and Technology." **DESIDOC Bulletin of Information Technology** 24: 1, 3-8.
- Braam, Robert R; Moed, Henk F and Van Raan, AFJ. 1991. "Mapping of Science by Combined Co-Citation and Word Analysis I: Structural Aspects." **Journal of the American Society for Information Science** 42, 4: 233- 251.
- Chaomei Chen. 2003. **Mapping Scientific Frontiers: The Quest for Knowledge Visualization**. Springer Verlag, London.
- <http://crrm.u-3mrs.fr/commercial/software/dataview/dataview.html>. Tanggal akses: 22 Juli 2002.
- <http://www.thomsonisi.com/demos/esi/>. Tanggal akses: 22 Juli 2002.
- <http://www.cwts.nl/ed/method/method01.html>. Tanggal akses: 1 March 2005
- Pelc, Karol I. 1986. "Knowledge Mapping: a Tool for Management of Technology: Chapter 13." In **Handbook of Management Technology**. Mc Graw Hill. New York.

- Royani, Yupi. 2005. "Identification of Focus on the Research Program in the Biology Field Based on the Document Structure Analysis of the Result of the Research Published in the LIPI Magazine of Biology, 1995-2000." **Thesis**. Department of Library Science. Faculty of Humanities. University of Indonesia (Indonesian version).
- White, Howard D and Mc Cain, 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, 4: 327-355.
- Widhalm,C.; Gigler, U.; Kopsca, A and Schiebel, E. 2001. "Co-Occurrence and Knowledge Mapping to Identify Hot Topics and Key Players in the Fields of Mobility and Transport." In **Proceeding 8th International Conference on Scientometrics and Informetrics 16/7-20/7, 2001**. Faculty of Commerce and Economics The University of New South Wales. Sydney. Australia.
- Ying Ding; Chowdhury, Gobinda G; Foo, Schubert. 2000. Bibliography of Information Retrieval Research by Using Co-word. **Information Processing and Management** 37: 817-842.
- Yoganingrum, Ambar. "Development of System for Mapping on Public Health in Indonesian Institute of Sciences (LIPI)." **Thesis**. Program of Health informatics. Faculty of Public Health. University of Indonesia (Indonesian version).
- Yoganingrum, Ambar; Rr. Endang SRR and Sri Hartinah. 2004. "Anthropology mapping of foreign researchers in Indonesia." In **Proceeding 4th International Workshop of Webometrics, Informetrics, Scientometrics**. Roorkee, 28 Feb-6 March 2004. Central of Library. Indian Institute of Technology, Roorkee. India.