INCORPORATING A DOMAIN ANALYSIS APPROACH IN THE INDEXING PROCESS

Oleh: Alfida[*]

Abstract

Peledakan informasi yang terjadi saat ini membuat pengindeksan sebagai proses yang penting dalam penelusuran informasi. Masalah-masalah yang berkenaan dengan analisis isi dan kebutuhan pemakai muncul saat pembuatan indeks. Berbagai penerapan telah dilakukan, namun selalu saja muncul masalah lain seperti tidak terjawabnya kebutuhan pemakai dan kesulitan dalam pemenuhan kriteria dalam pengindeksan. Pendekatan domain analysis merupakan suatu alternatif yang menjawab masalah-masalah tersebut di mana pengindeks dengan mudah mengevaluasi kebutuhan informasi pemakai; menentukan quality judgment; menggali topik dan istilah dalam isi suatu informasi. Tulisan ini akan memaparkan bagaimana pendekatan domain analisis digunakan dan manfaat yang akan diperoleh dari penggunaannya.

Introduction

The proliferation of today's information makes information retrieval play an important role effectively and efficiently. Keeping with the wise saying "right information at the right time", librarians or indexer and people, who work in information world are struggling to provide the best to the users by generating a critical product to facilitate information retrieval. The problem with time consuming to satisfy a request is major reason to what index is really needed. We cannot abandon index to the success of information retrieval.

Indexing has existed along with written records kept by humans. When alphabetizations, keywords and cross-references were used at the beginning of the 13th century, modern indexing appeared (Wellisch, 1992:3). Since then, various efforts have been made to create index to facilitate users finding information they need.

In information world, index has proven to be the one of the facilitating factors for users. This is a way to make documents accessible to users by utilizing methodological approaches, thinking, analysis, and qualitative aspects to identify the document. A reader's advisor in general reference services is like the indexer, and is guided by the scope of the subject and domain, mission and objectives of the information retrieval databases, which if well formulated, will reflect the needs and desires of potential users.

---

[*] *Penulis adalah seorang pustakawan di perpustakaan UIN Jakarta dan sebagai dosen tidak tetap pada Jurusan Ilmu Perpustakaan – FAH UIN Jakarta.*

Index entries indicate the contents of document. This leads to the concept that indexing is actually the process of producing surrogates for documents, summarizing their contents (Fidel, 2003:14). Index is made to help users find what they want by using keywords search or index terms.

As a procedure that produces entries in an index (Cleveland &Cleveland, 1983: 63), indexing is deemed to be costly process and one needs some assurance with reverence to product quality. This task is always close to problems regarding content analysis, as an important step, which deals with representation of the thought-content of the document and terms or identifiers that are capable of representing document content, which can be used as access points to that document, and subsequent corresponding area of user's need. This work is, therefore, assumed to be tough, which makes an indexer need to peruse in every step of indexing process.

There much has been done to deal with indexing approach. Indexers anticipate that the effort and time spent in generating index would be very helpful for user's need of information. The adequacy of current indexing quality, however, is deemed to be unclear to meet the challenge created by today's information revolution. It is too document- oriented or too user-oriented. As users, with index generated, they can identify keywords and phrases needed in information retrieval and filter information that meet their interests. They are trying to find an approach which can address the problems.

Indexing Process

As an art of organizing information, indexing is a process of constructing document surrogates by assigning identifiers to text items. Indexing is the notion of storing data in a particular way, in order to locate and retrieve the data as efficiently as possible. It is the process of analyzing the information content in the language of the indexing system, which is used in information retrieval systems.

It is commonly presumed that the existence of automatic indexing eliminates underlying problems in the indexing process, but this is not always the case. Some problems found regarding indexing approach. Automated index is quite popular because it is easy to undertake. The usage of automated indexing for potential users is considered to be ineffective to satisfy users, because it is too document centered. It is made by pulling terms or phrases out from document in hand.

It is worth to see whether the process of indexing has reflected that steps taken by looking problems surrounding the process. Cleveland suggests steps in indexing which cover: deciding which topics in the item are relevant to the potential user of the document; deciding which topics truly capture the content of the document; determining the terms that come as

close as possible to the terminology used in the document; deciding index terms and the specificity of those terms; grouping references to information that is scattered in the text of the document; combining headings and subheadings into related multilevel headings; and directing the user seeking information under terms not used to those that are being used by means of *see references* and to related terms with *see also references.* In brief, the process of indexing has two components: *first,* content analysis that results in the selection of the concepts to represent the document: and *second,* translation, which is articulating the concepts selected in the index language used by the information system. The steps also reflect at the first time indexer need to dig up topic relevant to users' need. This is very important that indexer finds ways on how he or she decides the topics.

Document centered and Users Need

Over the years, librarian and information specialist or indexer have been trying to yield a reliable and best approach to what really works effectively in preparing index terms, which reflect the criteria of indexing and deal with the potential users. Pros and cons on this issue attempt to consider indexing approach.

Most of indexing process is undertaken by analyzing document in hands. This is what we call document-centered approach. Indexer is directed to the entity and its faithful description (Soergel,1985:227). Lancaster (2003: 37) points out that it is much better for indexer to stick to the text and the authors' claims since it is really the function of the indexer rather than bring potential applications to the potential users as most of information specialist to do so. If indexers tend to assign a given index term to a document to the same extent to which there is a tendency for the term to appear in requests to which the document is relevant, we say the indexer-requester consistency is high; otherwise it is low (Cooper:270). It implies that the criteria to validate underlying indexing process such as depth of indexing, exhaustively, specificity should be considered.

Rolling (1981:23), however, believes that those criteria are often difficult to apply. Assumingly, the approach does not provide sufficient guidance for indexers on how to deal with the complexity, or to identify which factors are the most important (May, 2004:47). Also indexers focus on the object itself without the understanding of the users' information needs and terminology in mind when determining the subject matter of the document as well as when selecting index terms for the document. It is not necessarily that users are excluded in document centered approach. The centrality of the user is not a new notion, and as information seekers, they do not play a central role.

The idea of Information retrieval is how users can find what they want effectively. So users' factor by many of those with expertise in this field agree that it is critical to the user's needs when designing an index for the potential future use of the documents. So, in user-centered approach, indexer tried to find ways on how he or she should make document are visible to potential users and what terms should be used to convey its knowledge to those interested (Albrechtsen, 1993:222). Extra issues should be considered including: who will be using it and for what purpose, what purpose the author intended it to be used for, and for whom or for what the librarian or other manager of messages acquired it. (Fairthorne,1971:38; Wellisch, 2000:621). Hjorland (2001:775) points out that subject analysis should establish priorities based on the specific user groups served. Fidel (1994:573) recognized the need of mentalism and social context rules in indexing. The concept of "every reader his book and every book its reader" articulated by Green (1997:80) takes a user-oriented approach, picturing retrieval as a process of matching users and their attendant needs. This intellectual indexing is time consuming, expensive, and inhomogeneous because of the different background knowledge and expertise of the individual.

Although much has been said about how indexing must be aimed at the user, little is known about the needs of the users in terms of the information conveyed by the documents and the effect of that information on the user once the document is read. Cleveland (2001:34) mentions that neither indexer nor any set of indexing policies assure that every user's request can be anticipated. Strategies to get close to a user-based approach include knowing the user's target to help focus the indexing process. Indexers should have questions when they undertake the indexing process, such as how do users approach information file, and how their perceptions or problems influence the interpretation of the index they use. Bates (1996:27) argues that the document-centered approach is problematic because it blocks out context-dependent factors in the indexing process. Regardless of the method, user-centered indexing cannot be developed before searching behavior is better understood. In the mean times, Anderson (2002:258) reveals that a user-centered approach is very difficult to implement with human, a priori indexing.

To get a better understanding of the users, Cleveland recognizes three types of user's needs including: overt information related to the item such as its author or title; a subject need that is specific and well defined; and a vague and ill-defined need. Studying the user's cognitive processes, which focuses on how people think and use feedback in searching for information, coupled with design rule-based systems meaning that users will follow the rules to find the information they need is deemed to be sufficient methods. Another problem suggested by Park (1994:13,) is the users' specific

knowledge, so that labeling the users need according to their knowledge is important. Sauperl (2004 :60) claims that some indexers do not "consider the different uses someone could make of a particular book. However, a comprehensive and well-structured indexing of the vocabulary will be ineffective if it is not used effectively in information retrieval.

Given those aforementioned problems, a model of the intellectual procedures involves in indexing should be elaborated to bring together the aspect of successful of information retrieval through indexing process. Domain analysis approach is regarded as an alternative way to deal with the problems.

Domain Analysis Approach: How It Works

Domain analysis approach is not new in information science. It is believed to be a traditional approach, which reflect as a innovative way (Hjorland,2001:771) The domain analysis paradigm is a functionalist approach, attempting to comprehend the implicit and explicit functions of information and communication and to outline the instrument underlying informational behavior (Hjorland and Albrechtsen,1995:421)

The problems found in indexing process are suggested to be the fact that indexing process is not really started in a holistic way to get a whole picture of the domain and document to be indexed. Various aspects, ranging from social world, historical, economic, political context, etc in the domain of document where about should be seen as interrelated. Studies on the larger structures of disciplines, coupled with theoretical and philosophical analysis to empirical analysis should be reintroduced and play a greater role in information science (Hjorland and Albrechtsen (1995:422). He further explains that the best way to understand information in information science is to study the knowledge –domain as thought or communities discourse.

Before we discuss further on this subject, the idea of domain must be clearly identified to delineate the domain of discourse. Domain in indexing process means groups of people or professional or an organization or an institution, which have identical goals. In its process subject begins in a state of initial knowledge and move through the problem space to a final state or solution. Negotiation between users and intermediaries as in the public library setting formulates development of the book house retrieval system form a framework for work-centered evaluation and design of information system (Nielsen 2001, p.). In other words, domain approach starts from a real-life problem situation.

The cognitive task of domain analysis approach of indexing process takes place as a top down method and begins with the analysis of the task domain and moves on to the user and involves factors like goals, task situations, division of work, social organization, decision making, strategies,

94

actor, resources and values. Mai (2005:560) suggests that the core of the domain analytic approach is to study the activities and products of domains to gain insights into existing structure and meanings and also includes an understanding and analysis of the users' information needs and information usages. The domain-analysis approach has a continuum of problems from very simple problems that are easy to explore empirically to very complex theoretical problems (Hjorland and Albrechsten,1995:422). Hjorland (2001:779) advocates the use of activity theory and domain analysis as the best framework, model, or paradigm for pursuing information retrieval. He contrasts domain analysis and activity theory with other current models/paradigms of Information retrieval and labels the domain approach as getting: the central action for an indexing process which is the information object paradigm for messages and texts; the behavioral paradigm for human behavior, especially *vis a vis* information; the cognitive paradigm for the functioning of the individual human mind, and the communication paradigm for information retrieval, which is seen as an aspect of human communication. The best way to support users is to provide indexing of documents that is in line with the discourse in their domains. The framework of system communication structures of knowledge domain will help understand users' need of information.

By understanding the domain, indexer will get the knowledge about activities e.g. design of information systems, construction of controlled vocabularies, strategies for evaluation, and indexing of documents. The domain approach is bridging the gap between some who only concern about the documents and those who consider users' needs is the first range. The important on domain approach is also visible as terms used by everyone can be different because of his or her individual educational background, professional experience and organizational membership. It is also obvious, that different languages such as Indonesian, English, and French use different words. Within any given language, different domains use different vocabularies. These differences are often more visible when come to certain space. When we search for "relativity and gravitation" in the INSPEC (the leading English-language bibliographic information service on the world's scientific and technical literature), we may find different place in the Library of Congress Subject Heading. So, the use of special vocabulary or jargon is usually used to facilitate communication between specialized groups.

Indexing is a continuous process whereby each stage is interrelated. Nevertheless, certain step needs more certain degrees of knowledge. Example of domain approach steps can be seen as analyzing these factors, domain-users-indexers-document. Inspired by Ludwig L. Geismar' work about *Family and community functioning: a manual of measurement for social-work*

practice and policy, the writer would like to evaluate information on the topic mentioned using domain analysis approach.

At the first stage, indexer usually scans the document to recognize its knowledge. In domain analysis approach, indexer should know the notion or space of social work. For example, social work is a profession for those with a strong desire to help improve people's lives. This profession functions manage their environment, deal with their relationships, and solve personal and family problems. Social workers often see clients who face a life-threatening disease or a social problem. These problems may include inadequate housing, unemployment, serious illness, disability, or substance abuse. Social workers also assist families that have serious domestic conflicts, including those involving child or spousal abuse. Usually, social workers often provide social services in health-related settings that recently are governed by managed care organizations. To contain costs, these organizations are emphasizing short-term intervention, ambulatory and community-based care, and greater decentralization of services.

The next step of domain analysis approach would be an understanding the users, who will take more advantages of index being made. The users will know much about the jargon they use. Here, indexers become more knowledgeable about the users of this index while they understand the domain of social work.

The following step is that how indexers evaluate their roles on doing this task. This step stimulates the indexer to question about the users' position in the larger domain to spot a special topic. Perhaps, there are that few things, which are of prospective interest to the users; of particular bias that the users interested in; of particular level of specificity that the users need. These questions are important to ensure that the document will be put into the context of both the domain and the users.

Then, indexers will evaluate the document by analyzing the content. The knowledge of the domain will dig up at the topic, which captures the content of the document. At the conception selection step, knowledge of the domain and its macrostructure are useful in selecting the most central concepts. With the title aforementioned, indexer may determine terms regarding the terminology used in the document, such as, Medical and public health; hospitals; nursing and personal care facilities; individual and family services agencies; or local governments, vulnerable populations; psychosocial support; family caregivers, counsel patients, home services, etc. Another thing to consider is that judgment made by indexer to direct the user seeking information under terms used or not used. It is also helpful to add the value of our contribution to information retrieval by expressing qualitative judgment and by highlighting what we believe is best and most appropriate for the user.

Knowing the domain of social work, enable the indexers to effectively evaluate index and its relevance as a function of terms in the thesaurus.

Similarly, in revision stage, knowledge of elements of the context helps indexer to respect current policies in their work environment, to determine whether the selected descriptors meet the needs of social workers of their particular institution. Finally, the indexer to present the index systematically will end the indexing process.

Domain analysis approach: Promises and Challenges

Various advantages can be acquired from this approach. The indexers will have a clear structure and a comprehensive understanding for making decisions with the indexing process. It guarantees that indexing is consistent with user's use of information, without neglecting the ordinary process of indexing. Domain-centered approach offers a guide for indexers that can help them cope with the complexity of indexing. Thus, it hopefully provides effective results. Domain analysis approach suggests the indexer to learn and learn to get more knowledge about aspect or domain being indexed. It is our responsibility to build a foundation for indexing analysis in more satisfactory way. It will improve effective retrieval on the database development where we can organize both the documents and the interface to help users get the better results. So we can say that beside it satisfies users' information needs in supporting users' activities, it also represents subject matter objectively and neutrally.

However, the domain analysis approach may be good at times, but it doesn't mean it is without problems. The creation of multiple indexes for different domains in the similar database is not economically fortunate even though it promises significant improvement for delivering services. The indexer needs a substantial understanding of the nature of the domain. This knowledge can be gained by studying the activities and discourse of the domain. The exact methods for this needs further exploration. The domain analysis approach to indexing offers a framework that manages the complexity of indexing process. Domain is the focal point for analysis to index documents according the users' activities.

In order to be useful in information retrieval, domain analysis should be made more practical and applicable. Only then we will be able to investigate empirically domains in order to identify their structures shaped by the corresponding community discourse. Also, Indexers should stay in direct contact with the community they serve or be part of it to ensure agreement with their viewpoints. There remains much work to do to improve indexing system which the whole picture of the document and also represents the keywords or phrases needed by the users. It is hoped that the success of domain analysis approach will be followed to improve search engine.

Conclusion

Domain analysis approach doesn't mean to neglect common indexing process. It exists at the first stage of the process to direct the indexers to be more cognizant of what beyond the document on their hands. When they have a clear picture, it will be easier to undertake their tasks. Similarly, the users' needs are also taken into consideration. While there remains much more to analyze the domain analysis aproach, it is clear that this approach can explain aspects of intellectual process of indexing which have been received little attention in the indexing process on the subject. Although it needs more energy and time, it would be built more effectively if we want to contribute the important of this mean to enhance information retrieval.

Bibliography

Anderson, J. D and Rowley, F.A. "Building end-user thesauri from full text." Advance in classification research. 1992, 2

Anderson, James D and José Pérez-Carballo. "The nature of indexing: how humans and machines analyze messages and texts for retrieval: part I: research, and the nature of human indexing." Information Processing and Management: an International Journal. 2002, 37

Bates, Marcia J. "Learning about the information seeking of interdisciplinary scholars and students navigating among the disciplines: The library and interdisciplinary inquiry". Library Trends, 1996,11

Cleveland, Donald B and Ana D. Cleveland. "Introduction to indexing and abstracting." Colorado: Library unlimited, 2001

Cooper. W. S. Is interindexer consistency a hobgoblin? American Documentation, 1999, 20

Fairthorne, R. A. "Automatic retrieval of recorded information."Computer journal. 1958,1

Fidel, R. "User-centered indexing." Journal of the American society for information science.1994, 45

Green, Rebecca. "The role of relational structures in indexing for the humanities." Knowledge Organization. 1997, 24

Hjorland, B. and Albrechtsen, H. "Toward a new horizon in information science: domain-analysis."Journal of the American society for information science. 1995, 46

Hjorland, B. "Towards a theory of aboutness, subject, topicality, theme, domain, field, consistency, and relevance." Journal of the American society for information science and technology. 2001, 52

Lancaster, F.W. "Indexing and abstracting in theory and practice." London: Facet Publishing, 2003

Mai, Jens-Erik. "Deconstructing the Indexing Process". Advances in Librarianship. 2000, 23

Mai, Jens-Erik. "Analysis in indexing: document and domain centered approaches". Information Processing and Management 2005, 41

Mai, Jens-Erik. "Semiotics and Indexing: An Analysis of the Subject Indexing Process." journal of Documentation, 2001, 57

Nielsen, Marriane Lykke. "The Role of Domains, Documents, and Decisions in Indexing. in Knowledge Organization and the Global Information Society." Proceedings of the Eight international ISKO Conference. Advances in Knowledge Organization. 2004, 9

Park , Taemin Kim. "Toward a Theory of User-Based Relevance: A Call for a New Paradigm of Inquiry." Journal of the American society for information science. 1994, 4(

Sauperl, Alenka. "Catalogers' common ground and shared knowledge." Journal of the american society for information science and technology. 55.1 (2004): 55-63

Tennis, Joseph T. "Two Axes of Domains for Domain Analysis." Information school university of Washington. 2005. Accessed March 1, 2005 <http://www.informationschooluniversity/washington/Ten/pdf>.

Wellisch, H.H. "Book and periodical indexing". Journal of the American society for information science. 1994, 45