INDEXING IN MALAY WORLD STUDIES DATA-BASE: SOME CRITICAL ROLES OF RESOURCE CENTERS IN THE CONTEXT OF KNOWLEDGE WORK AND KNOWLEDGE WORKER

Ding Choo Ming, Ph D*

ABSTRAK

Construction of databases, designing search engines and interfaces, offering access and links, are no longer the monopoly of major online services or database publishers. Advances in technology are now allowing greater subject access along with improvements in scanning hardware and software and imaging technology has made visual access to digital materials more common. Adopting the power and capabilities of such technology, ATMA (acronym for the Institute of Alam dan tamadun Melayu) has constructed in-house databases on Malay world studies since April 1999, mimicking the major online databases to offer scholars and researchers a collection of materials that can meet their needs. There are now three databases, namely (i) Single Articles on Malay World Studies, (ii) Malay Proverbs and (iii) Pantun, Shair and Dondang Sayang of Malaysian Baba accessible in http://www.atma.ukm.my. Another two databases are under construction: Adat Pepatih and Malay Architecture. All these and others will be integrated to form a portal called Malaycivilization.com, built with a grant from DAGS (Demonstrator Application Grant Scheme). Now scholars and researchers all over the world are able to enjoy customized information and document retrieval related to the Malay world studies for the first time. Due to copyright problem, we cannot offer full-text services to all the materials as yet. All research centers, including special libraries and resource centers are under constant pressure to provide a quality service in order to prove their worth with the rising expectations of researchers who have come to expect high quality products and recognized standard of control, after using large databases, both on-line and CD-ROM in other fields. The cut back in public funding is forcing us to provide the best customer care in order to stay. The speed and economy achieved in creating databases are paid for by users in time and effort saved. With the increase in end-user searching, search results may be questioned in terms of their relevance, value for money and precision. In short, quality service is our life and blood. But, online searches can falter on the strangest of obstacles: the connection breaks, the login procedure is changed, the password is not valid, the search statement is too long, the manual does not give the help one needs, the users cannot search on certain fields not provided for. These problems accumulate when several online systems are used, the search language and commands are different, the structure of the databases varies. Constant evaluation is thus necessary in terms of limitations in content, lack of reliability and accuracy of data. As mistake and problem can occur at each stage in database building, thus dedication, care, institutional support, among many other critical factors are vital to information workers to perform and deliver their services.

* Pensyarah, Fakulti Teknologi & Sains Maklumat. Universiti Kebangsaan Malaysia.
  e-mail: chooming@pkrisc.cc.ukm.my

Hak Cipta Terpelihara © 2002 – Perpustakaan Negara Malaysia
BACKGROUND

There are many good databases and their quality is improving because of improvements in computer hardware, software and imaging technology. At the moment, the significant contribution of computer to documentation has been the production of databases, with keyword indexes (keyword-in-context or KWIC and keyword-out-of-context or KWOC). Since April 1999, we at ATMA have developed computer-based databases using the above-mentioned state-of-the-art technology. It would have been inconceivable to do it without the application of computer, because of enormous and ever increasing wealth of literature on the Malay world published all over the world since colonial days. It is hoped that the experience and knowledge which we have accumulated in the development in the pioneering project will be applicable to you in designing other systems. So far, lack of commitment, money, personnel and facilities are some of the factors accounting for lack of Malay world studies databases. It is the remoteness of Malay World studies from the mainstream of world information flow that has forced libraries and resource centers to be very active in the use of databases which are only marginally relevant to the Malay world studies in Malaysia. Driven by the need to keep abreast with the latest literature, usage of online databases as well as CD-ROM services in Malaysia have been very strong following automation of libraries in the 1980s. By and large, idea in constructing ATMA’s databases arose from our great and intense frustration in retrieving the relevant information and materials (Ding 2001 & 2002; Ding & Supyan 2000 & 2002; Shamsul et al 2002). All the databases built in ATMA are subject and geographically oriented towards Malay world studies. With this purpose in mind, we have selected materials in all formats, languages and disciplines to help researchers and scholars in retrieving general and specific information on a wide array of topics. To do it, all the materials gathered are integrated, using the relevant fields for storing and searching (Appendix). Ultimately, query on any subject would likely produce information on the subject accessible instantly at ATMA.

Among the challenging tasks encountered in constructing this database is information management, that is indexing, which is to find out what are documents about. It is a process of analyzing the contents of records and expressing it in the language users are familiar and use. It is an intellectual operation representing the texts by making up a list of possible descriptors to represent the documents for information retrieval, thus promising better information retrieval. To do it, the information workers must form a mental picture of what the authors are saying. The problem is concept of a subject may be elusive as all writers have a unique understanding of what the subject is, and also that they have a unique way of expressing themselves and choosing words when they write. Some words can be ambiguous and metaphorical. There would be no indexing problem if we were all robots, programmed to write the same way. Anyway, index is an orderly guide to the intellectual contents and physical location of documents or records. It is a pointer, a guide, employing a set of tags or descriptors which earmark the source of information for which the users are searching. It systematically leads us to previously undetected information, whenever it is needed.

Indexing can be simply described as a process of creating surrogates for documents by summarizing their contents. It is representing documents for information and document retrieval. Promising better retrieval, all indexes has one common objective: to guide users to the intellectual content and physical location of documents. There are many ways to do it. We at ATMA do it by indexing the author, subject, title, keyword and source. By doing so, we provide various options to the end-users to overcoming their difficulty in retrieving the right information at the right time in the right format. The other job of information workers - including librarians - is to help the users retrieve the documents. This can only be done with the necessary finding tool and knowing exactly where the documents are located. The problem may multiply as many users at times are not sure what they are asking for. They always approach database or resource center with the
problem of understanding the topics and question formulation, which can range from the trivial to the profound, thus can be simple or complex.

DATABASES AND INDEXING PROJECTS IN MALAYSIA

Among the successful computerized indexing projects in Malaysia and Singapore is the indexing of medical periodicals, called BILIOMED. It has a history of more than 40 years now. The relevant articles related to medicine and health science published locally and overseas are indexed by the national University of Singapore Medical Library in accordance with Medical Library Subject Headings of National Library of Medicine (MESH).

BILIOMED is a comprehensive bibliography of medicine and related sciences in Singapore and Malaysia published by SEAMTC in 1973. The other equally successful indexing project is PERIND in NUS (National University of Singapore). Between 1960 and 1981, periodical articles on humanities and social sciences relating to Singapore, Malaysia, Brunei and ASEAN were indexed manually in card form. Starting May 1982, the PERIND database was set up, with record input by the staff of the Humanities/Social Sciences/Management Reference Department of the NUS Central Library (for humanities and social sciences), the Medical Library (for medicine and related subjects) and Law Library (for law). Starting 1982, all the relevant materials have been input into it using MINISIS computer system. The NUS Library also participates in Agricultural Information Bank for Asia (AIBA) network and contributed to the compilation of AGRINDEX. A total of 39 local periodicals are scanned regularly for inclusion in AGRINDEX and AGRIASIA databases. The scope covers agriculture and its related fields, including fisheries, food and veterinary science.

Individual articles on Malay World studies database developed at ATMA, accessible online at http://www.atma.ukm, is the first of its kind in the world. It may be evaluated in terms of the amount of time saved compared with the time spent doing a search for the related documents in a library collection. We glean materials from journals, books, seminars, proceedings, theses in English, Malay/Indonesian, Chinese, Dutch, German, Japanese and many more, both past and present, and regardless of whether printed or digital forms. Compilation of articles from such a diverse source is to ensure that we amass as much articles on any subject as possible in order to provide adequate coverage. This multidisciplinary coverage is important for this database to be claimed as not only comprehensive, but also to be compared favorably with other information retrieval systems developed at KITLV (Leiden) or PERIND in Singapore. Our initial target is an arbitrary number of 50,000 articles within a five years period. We assume this figure manageable as well as large enough for certain trends or subjects to be obvious on any topic. The principle is all the materials need to be processed promptly between their receipt and their appearance in the database to avoid duplication. The five years period requested as we can only process about 10,000 articles a year with two full time staff doing data entry, editing, cleaning and updating. In our rough estimation, there are about 2 - 3,000 new research articles in the Malay world studies published every year.

The ultimate aim of ATMA's database base, like any other resource center or indexing service, is retrieval. It involves information management which is a complex task. It requires analyzing the ways documents will be requested and defining indexing parameters according to database criteria. How well it works would again depend on how well the following functions are taken care of:

i. acquisitions of documents,
ii. content analysis,
iii. content representation,
iv. coding of content indicators,
v. creation of a document file,
vi. creation of operational search strategies,
vii. physical dissemination of retrieved results.

The emphasis in this paper is on indexing, another critical role undertaken intellectually by knowledge-workers. Any document not indexed (by a given
parameter field) will not be retrieved (by that parameter). Thus, determination of relevant indexing fields need to be done from the first day the database is planned and built subsequently. This step requires conceptualizing and selecting fields by which the documents can be stored and requested. Doing that does not mean there will be no other problem in retrieving the information or documents. Though indexing is necessary to inform us what has been published on a particular subject, there are complaints that information workers miss important (and new) points in the documents, and add irrelevant terms. This may stem from their insufficient knowledge about new developments, or they have not been active researchers. Similarly, usage of biased index terms can lead to biased search results. This problem is difficult to be solved as there are discrepancies between author and information worker’s preferences, due to conflicting opinions they have on the topics. No matter what it is, all the parties must be responsible. The data must be accurate and reliable as computer-based searching is based on exact matching of words or phrases. If only accurate data allows the data to be re-used, then any error may lead to failure to locate a key reference. Here, accuracy, completeness, consistency and timeliness are vital. They are explained as below:

a. accuracy - may be defined as being true to the original form of the data,
b. completeness - may be loosely defined as covering all materials claimed by the providers of the service,
c. consistency - may be defined as uniform application of a standard set of rules,
d. timeliness - is the time lag between publication of the primary versus the secondary materials.

INDEXING PROCESS AND RULES

Indexing is information management for subsequent retrieval. Any information retrieval system is worthy only if information and documents therein can be satisfactorily traced, located and accessible for use. As mentioned earlier too, indexing is an intellectual work creating a description of documents, usually in some recognized and accepted language. It is a intricate and elaborate work involving the analysis of information. It strives for succinctness. Because of this, B. C. Vickery (1968): defines it "as deriving from a document a set of words that serves as a condensed representation of it. This representation may be used to identify the document, to provide access points in literature search, to indicate its content, or as a substitute for the documents". A quality database, or resource center, or indexing service, is the cumulative result of a series of good information management decisions. The relevant knowledge workers, or librarians or subject specialists, must be certain that all the right terms must be selected, and the unnecessary terms excluded. Including the wrong terms will lead the users to information not wanted or needed, and worse, leaving out the right terms will keep the needed information hidden from discovery. To help the users, they have to try hard also to guess what their users need and how they will react to the index entries. If users have a particular information need, what terms will they use to identify the documents the information workers have in their minds. Closely related to it, if the users decide to use that term, will they be satisfied? As indexing is partly science and partly art, there are rules to follow to ensure quality control. The cardinal indexing rule is 'index all important concepts'. But, what is important? Is it how often an idea is repeated? This comes from the intuitive feeling that important ideas are repeated, while the minor ones are only mentioned! This is especially true in indexing using a computer. The first step in indexing is to have professionally trained and experienced subject experts, namely someone having subject authority. They will decide on the level of indexing and whether the documents are worth indexing in depth and in what way. Such a decision is both an expertise and value judgment, based also on the objectives and policies of the indexing agency, which is dictated by the information need of the clientele. There are at least three things involved here:

i. expertise judgment of the document contents,
ii. the appropriateness of descriptive terms,
iii. the general goals of the indexing policy of the database producers.
The above are expectations from an experienced information worker doing indexing which comprises of a number of basic rules, which are:

a. content analysis,
b. assigning of content indicators,
c. adding location indicators,
d. assembling the resulting entries,
e. choosing the physical form in which the final index will be displayed.

After knowing the rules, the next question is how to do it? Most scholarly papers follow a certain pattern in their publication, thus give information workers an important short-cut to understanding the subject contents or the overall intent of the papers. They are the introduction, the summary and the conclusion. The introduction explains what is going to be read, and the summary and conclusion explain what has been said. It is generally conceded that references do reflect the subject matter of the paper. Considerations should also be given to the references cited too. This is because authors cite other writers who have written on the same topic, both to gain support and to give the readers an opportunity to go to other works that are closely related. No matter what there are, extreme care must be taken to see that the data are recorded correctly, for the obvious reason that incorrect entries cause the documents to be inaccessible. Care is necessary to reduce mistakes which can occur at each stage in database building: data collection, data entry and even data searching. The information workers may misspell a vital entry term or overlook an important concept. The database providers may fail to update it regularly. There are mistakes elsewhere. The original authors may have some facts wrong. The publishers may fail to spot errors. The database producers may miss a file. The telecommunication network may break down. The end users may not search thoroughly. Even if an error is not critical, it can be irritating to the users. We cannot be sure whose fault it is or who is liable.

In content analysis, spot reading is sometimes sufficient for information workers to understand what concepts are dealt with in the documents. On the other hand, other documents may have to be read thoroughly before the information workers may feel confident in identifying the subject contents. The amount of time involved in content analysis will depend on the nature and experience of the information workers. A piece of advice from A. C. Foskett (1982) is: "Scanning a text to decide what it is about is the key operation in indexing, yet it is the least discussed and the least reducible to rule." Scanning the texts is necessary not only in identifying subject concepts, but also the amount of information to be presented to potential readers to help them decide which documents merit their attention. The descriptor choice used is always influenced by depth of indexing or exhaustively, or simply the number of topics that will be covered in the indexing of a document. In a document that covers five topics, and if all the topics are represented, or indexed, the indexing of this document is said to be complete. This means we have indexed in depth. Clearly the more indexing in depth, the more indexing terms will be employed. But some documents may not have main topics, thus deep indexing will not produce many terms, no matter how exhaustive the indexing is. This means that no one is quite sure how to determine the optimal level exhaustively. Depth of indexing will determine how well a retrieval system pulls out all documents that are possibly related to a subject. Will extremely deep indexing retrieve a high proportion of the relevant documents in a collection? As more and more documents are retrieved, the risk of getting extraneous material arises. Thus, when we are aiming at exhaustively, we must bear in mind that at some points they may be negatively affecting the efficiency of the system. Why bother, then, trying to achieve exhaustively if we may run the risk of not uncovering information for the users who are more worried about missing something than in being inconvenienced by having to examine too many irrelevant materials? The cause of failure is not all the documents are indexed in the same depth - those with more words are indexed deeper than the rest. Anyway, depth indexing is dependent on the understanding of information workers of the topics. Though the depth and specificity of indexing is also dependent on the numbers of terms allowed per document, selectivity of terms is also partly based on the policies of the indexing agency which decides what kinds of documents and subjects to be selected.
This means indexing is not based solely on the words that appear in the documents. As the emphasis in ATMA is given to Malay studies, thus the information workers employed by ATMA have been instructed to give priority to subjects related to Malay studies and ignore concepts unrelated or ephemeral to Malay studies.

Information management in ATMA's database is moving away from the traditional pre-coordinate indexing to post coordinating indexing. When Uniterm post-coordinate indexing was first proposed, the idea was to index terms taken from the titles, abstracts and the documents and also allow users to form term combination to fit their individual needs, thereby avoid elaborate cross referencing the complex authority lists, including the Library of Congress Subject Headings. To do the latter, we have provided advanced searches where users are allowed to combine author and title to yield exact search result. Though post-coordinate indexing is not perfect, it has obvious advantages and conveniences both from our, database producer, and end-user's point of view. Among the reasons we go for title indexing, for example, is that titles are important clue to subject contents, or at least they indicate the subject contents of the documents. A good title is, by its very nature, succinct. There are increasing evidences suggesting that titles are becoming more informative, especially in scholarly works. Good titles can be fundamental indicators of subject content, created for the documents as a whole, indicating the level of depth indexing. It is due to the above reasons that some indexes are surrogate of documents. Among the drawbacks are they are always limited in the amount of information they can convey. Thus, usually only the main themes can normally be summarized in the titles, leaving out the specific aspects of that topics. Again, there are some subjects simply cannot be adequately and precisely specified by a short title. Some titles can be badly written and misleading, while the others may be vague, either because they are too generalized or the author trying to be catchy. Worse still there are also titles which are simply not related at all to the subjects dealt with, especially in humanities. Despite their manifold weaknesses, title indexing has two important points which make it attractive in computer databases and the production of printed indexes, based on KWIC indexes: a large number of titles can be processed quickly and cheaply, as it is relatively straightforward, using natural language. Knowing The above weaknesses is counteracted in ATMA by adding keywords extracted from abstracts and texts, all of which accounting for the preference expressed by their authors for them. These additional terms have contributed to making our indexing more exhaustive and specific, and eliminate the problem of uninformative or misleading titles. All the keywords are displayed as headings: one entry for each significant keyword.

Is there no problem with author indexing? Certainly there are. Among them is maintaining the consistency as listed below:

a. the number of names to be allowed per entry when a document has multiple authors,

b. the method of alphabetizing,

c. use of full name or initials,

d. delineation of authors with common names.

To solve them, carefully constructed guidelines are necessary.

What about subject indexing? To manage it well, the information workers must be experts in Malay culture and not merely having a passing familiarity. As indexing requires intimacy and immediacy, it must be done by subject specialists who have cultural authority. Among the difficulties in interpreting the subjects are:

i. it is never easy to understand exactly what the authors mean,

ii. message of the authors may vary with time and environment,

iii. authors may at times misinterpret his or her own intentions,

iv. authors can misinterpret and not fully understand what have been written some time back.

v. selection of subjects is closely related to the policies of the indexing agency.
Having outlined some of the problems, it is up to the training, experience, expertise, commitment and dedication of information workers to perform what are expected of them to ensure adequate retrieval. The kind and level of indexing required would vary according to the background as well as the needs of the end-users. On average, the research assistants at ATMA spend fifteen minutes on each document, looking through the titles, abstracts, documents and picking the relevant terms. Linguistically, the problems they encounter are the multiplicity of languages used in the materials, in addition to the variations in spelling and terminology used in Bahasa Melayu and Bahasa Indonesia. Among the linguistic problems with natural language indexing encountered are:

- a. semantics - variant word forms, antonyms, the possible use of truncation,
- b. homographs - terms with more than one meanings,
- c. hierarchical and other relationships - with no cross-reference and using the terms for a broader or narrower concept.

Our attraction to use natural language, instead of Library of Congress Subject Headings, for instance, is mainly the economic consideration and speed. Thus, retrieval of documents on a search topic relies heavily on the searcher's ability and ingenuity. The main drawback of keyword searching is it complicates searching in several ways. Searching is more difficult and uncertain with synonymous terms, misspellings and general-specific term relationships. Without cross references, general-specific relationship, users have to search from one to another in order to get all the possible approaches to the subject or guess what terms the authors are using. Since one of the qualities of a good index is that it quickly focuses on those entry terms that express the needs of the users and connects the index language to their way of thinking. It means uncontrolled keyword indexing places a burden on the searcher. Other problems beyond the grasp of information workers are that data can be corrupted by file transfer and nobody knows how much data may be lost in file transfer!

Keyword indexing is basically a revolt against the cumbersome pre-coordinating indexing, commonly used in the past. Its simplicity fails to overcome problems arising from the vagaries of natural language: words drop out of vogue, the concepts they represent will, with time, gradually be replaced by new terms. As there is no consistency in terminology used by authors over a number of years, keyword indexing never work well too. Nevertheless, in as far as authors are experts in their fields, the words they use in the titles, abstracts and texts should be accurate, current and correct from time to time. The use of keyword indexing in natural language have the following outcomes, list randomly, both positively and negatively:

- i. the number of potential indexing words can be very much greater in longer documents,
- ii. due to lack of vocabulary control, the greater variety of index words used are more likely to appear that have not necessarily been selected as representative of the document content,
- iii. indexing can be achieved at a detailed level, with many terms per document, with almost no indexing effort,
- iv. we cannot search on words with variant spelling, despite the fact that there is an ability to search on word stems,
- v. in some circumstance, natural language indexing may reflect more closely the terms used by the searchers,
- vi. humando make inappropriate judgments, misinterpret ideas, have lapses in memory or concentration, and generate omissions and inconsistencies in the indexing.

Many of the above problems are beyond the control of information workers. It is common for authors to express concepts in different words and terms, in different ways and at different levels of specificity, which can lead to varying terms chosen for similar documents. Such inconsistency will usually result in poor quality indexing, which in turn hinders the quality of information retrieval. The resultant index can be difficult to use as there is no control over trivial words, inconsistent spellings, abbreviations, word variants and multiple word stems. Thus, there is a great need for editorial control when keywords are
used as search terms. One spelling mistake or typographical error can make the information or document invisible between 'form' and 'from', 'nuclear' and 'unclear'.

EVALUATION + USERS AND SEARCHING TECHNIQUES

As mentioned earlier, using post coordinating indexing system, ATMA's database allow users to combine two or more single index terms to create a new class in advanced search. For example, combing the individual terms 'Seni bina', 'Rumah', 'Melayu', 'Minangkabau' will give rise to a new class 'Seni bina rumah Melayu di Minangkabau'. Users are free to combine the terms with Boolean operators to express the information need as closely as possible. The problem in post-coordinate searching is 'false coordinating', that is we obtain the conceptual opposite of what was expected when the search request was formed. For example, a coordination of 'sepak bola', 'pisang goreng' will also retrieve documents on 'bola sepak' and 'goreng pisang'. This problem can be minimized by stating a query as specifically as possible using the function of 'exact match' in advance searching. Users should choose the correct terms and use all of them, taking advantage of any generic searching capabilities provided by the system. Frankly, there are too many people who are poor users of databases. They do not understand how to fully benefit their potential.

No matter what it is, evaluation of databases and quality assurance must be on-going. There is no such a thing as a good database that has a poor index or no index, like readers have been cheated in purchasing a book with numerous spelling or typographical errors, or where index contains non-existent word or missing pages. The need to address these problems has become more acute with the increase in the use of large databases, both on-line and CD-ROM. Evaluation of database is to judge its effectiveness, efficiency and value, ie how well is it a retrieval tool! We just cannot say this database is good or bad, because it is difficult to determine what is good or bad! Attempts are then made to define good or bad in terms of objectives. Does it fulfill its stated purposes? Are its scope and coverage adequate? Since we are talking about how sufficient the database is in meeting the information requirements of the users, so, we can begin with user reaction and then examine the index for accuracy and consistency. A database can be evaluated either as an individual unit or in comparison with other similar ones. The objective of the former is to rate the database in terms of the needs of the clientele: the subject areas covered, its stated purposes and its cost, among many others. When the latter approach is used, we compare its relative quality and cost. To do that, we must have made a previous judgment about the other databases.

In reality, evaluating database amounts to evaluating the performance of information workers working as a team. This is because a database is the final product of information workers, together with technical and managerial staff.

Users usually have many objectives when they approach a database. Some may want a grand sweep of a topic that pulls together every bit of information written on it, the second fellow may only want a general survey of a topic, while the third person may want to verify a single fact, and even someone who approach the database as a last resort when there is absolutely nowhere else to turn to! No matter what it is, the database must be complete. It must alert the users to everything on the topics. The database must have enough specificity for anyone to get a fact quickly. The database must allow them to retrieve only relevant documents. All of them want to avoid having to examine non-relevant materials. The goal of a good database is to aid the users in finding the relevant information or documents within a reasonable boundary of time and personal effort.

Relevance is the crux to evaluating an database on its performance, even though it is clearly a relative notion as the same documents may be relevant to one, but not the other, or at different time! It is a tricky business, a matter of degree and not simple yes and no decision. The reasons are obvious as below:

i. relevance of a given document may change as a result of other documents, as stock of knowledge at hand changes,
ii. two people with similar background, could approach a system with exactly the same question, get the same answers, and yet be poles apart in relevance judgments, 

iii. even the same person may vary his or her judgment on different times and days, 

iv. not every document retrieved for a given query would be relevant to that query.

Although there are instructions, users have to discover how to use ATMA's database by trial and error. As the onus of successful searching is dependent on the users, 'editing' on the part of users is necessary to bring together scattered similar entries and to remove errors. To do it, there are a number of check points need serious consideration:

a. check spelling, 
b. check pronunciation, 
c. check typography, 
d. check for missing entries, 
e. check for unnecessary entries, 
f. check for headings.

As mentioned earlier, ATMA's database uses basic and advanced search techniques. In the former, users are allowed to browse all the citations retrieved and displayed, while in the latter, they can refine their searches by using Boolean operators in some sequences. Generally, browsing produces fuller retrievals, while selection retrievals are more narrowly relevant. The effectiveness of either search technique depends on user's information searching skills, patience and the relative importance of relevance versus comprehensiveness in the retrieval's satisfaction of the query. Users can search terms of interest, linked by AND, and OR logic. Users can also eliminate certain items by using terms with NOT logic. To enable complicated searching, we are developing searching by truncation and nestling.

We hope we are able to integrate all the databases developed in ATMA under a portal called Malaycivilization.com, a gateway to information and materials on Malay world studies to enable ATMA to be seen as slowly emerging to undertake national and international responsibility for the dissemination and retrieval of materials related to Malay world studies. Making use of them, scholars and researchers can expect to find the maximum number of relevant documents in some justifiable time or in the minimum possible of time. Without them, they would have taken longer time to search the articles scattered all over the places in the libraries. It is less rewarding as looking for them is sometimes like playing the game of seek and hide.

CONCLUSION

As it is, ATMA's database is not as sophisticated as the major commercial retrieval systems available commercially. Thus, we have to improve it continuously in response to the wide national and international interest aroused by the excitement using on-line databases. Among the steps taken to improve it is cleaning up errors and expanding it to be as comprehensive as possible. Evolved originally as a byproduct of a IRPA research, it has been expanded to accommodate the enormous and ever increasing wealth of information on the Malay world published all over the world, regardless of time period and formats. We intend to make this database one-stopping shopping and gateway for the relevant information and materials originated mainly from within the Malay world, to complement that developed in KITLV in Leiden, for instance. A good database (just like a good book) is not to be used once and forgotten forever, but is to be re-used and re-used from time to time whenever there is a need. Thus, it must be maintained. Maintaining it adequately is a continuing expense. Because of the cost, many databases are left to become obsolete. Any database that ever serves as a reference in any sense of the word should have a quality index to make its information available quickly and easily and completely.
REFERENCES

Ding Choo Ming 'Access to digital information: some breakthroughs and obstacles.' Journal of Librarianship and Information Science 32(1) March 2000: 26-32


——— & Supyan Hussin. 'Preservation of culture and knowledge through technology: the experience of UKM.' UKM-UC Seminar on Technology in Development, 25th January 2000, at UKM, Bangi.


