

# Creating a Digital Library to Handle Malay Manuscripts Using Greenstone

**Mohd Hilmi Md Rifin and A.N.Zainab**

Library and Information Science Unit  
Faculty of Computer Science and Information Technology  
University of Malaya 50603 Kuala Lumpur, Malaysia  
E-mail: mhilmi@um.edu.my; zainab@um.edu.my

## Abstract

*The study describes the requirements analysis for the development of a digital library that can handle manuscript collections. This involves finding out how selected Malaysian repositories organized their Malay manuscript collection and observing the main features provided by current digital library of manuscripts collection available on the web. To ensure accessibility and use of manuscript collections an open source software, Greenstone is chosen. The paper describes the design and output of the implementation.*

**Keywords:** Malay manuscripts; Digital libraries; Greenstone digital library software; Open source digital library software

## 1. Introduction

The Malay manuscript is defined as a handwritten document in the Malay language, written in *Jawi* script, which surfaced in the beginning of the 14th century and ended in early 20th century. Manuscripts generally refer to unpublished original sources and have the characteristic of being unique. Unlike books, journals and other resources typically collected by libraries, manuscripts typically do not exist in multiple copies. Malay manuscripts that are currently held by Malaysian institutions are physically distributed among various libraries, museums, government agencies and some are kept by individuals as their private collections. Currently, many of these invaluable sources are physically stored in secure archives and are accessible in a very restricted way to a handful of experts only, due to their fragile nature.

As an important source of information for cultural and genealogical studies, it becomes an obligation for libraries that hold such collection to preserve and provide effective access to Malay manuscripts in anticipation of demand from clients whom the libraries serve. The problem of access to these materials is exacerbated by the fact that the collection of historical Malay manuscripts are scattered in repositories around the world.

Apart from the problem of accessing the Malay manuscripts, another problem arises where most of the manuscripts have not been studied because of the lack of researchers who can read old *Jawi* script. Only trained experts have the skills and patience to comprehend the complicated script though with no less difficulty.

Moreover, the manuscripts themselves are susceptible to natural process of deterioration due to exposure to unsuitable environment (humidity, temperature, chemical reaction) over the years. Base materials onto which the content was written such as paper, parchment, deerskin and others do have their limit and various preservation measures have been undertaken such as the use of chemical compound, framing or temperature control within specialized room area.

These challenges have been addressed in different ways by the repositories. Over the past several years, libraries have moved towards using digitization as a method for



reformatting endangered and fragile paper-based materials to simultaneously preserve and provide access to such library collections. Digital library of manuscripts provide new possibilities and paradigms for access to information, knowledge dissemination, and to stimulate collaborative learning and research.

## 2. Related studies

A prominent study by Ding (1986) discloses that there is hardly any standardization in the catalogue description of Malay manuscripts among libraries. Physical arrangement of the individual items is also different from one institution to the other. Ding is of the opinion that it is generally believed that library materials, including Malay manuscripts once used has a high probability of being sought again from time to time by patrons. Most of them, being research students or scholars, have very vague idea of the relevance of the library's materials to their research before visiting the library. Practically all the institutions with a large collection of Malay manuscripts have placed them separately and in closed access in a Special Collection. This situation presents problems of convenience of access to both staff and readers since these materials are non-circulating.

In another study, Ding (1993) mentioned that, an increasing number of Malay manuscripts in libraries especially those in private collections, are disintegrating. This problem has been with the Malay World since the very beginning. Deterioration may be due to acid embrittlement of the paper, unsuitable storage environment, biological attacks or human negligence and mishandling. A number of these manuscripts are already brittle to a point of being unusable.

As mentioned by Aytac (2002), flooding, terrorism, bomb threats, explosions, transportation accidents, earthquakes, fire and war also prove dangerous threats to the written cultural heritage. It has been well publicized how much effort has been expended in order to restore the losses experienced by Bosnian and Kosovan libraries. The artifacts in Baghdad National Museum and Baghdad National Library are no longer available for the benefit of future generations and humanity as a whole. The world cultural heritage is continually at risk because of natural and unnatural disasters.

Many digital library initiatives focus on preserving various cultural heritages. Often, the initiatives comprise of various materials such as pictures, maps and paintings, and manuscripts are one of the materials included. Shafi (2004) listed some Western and Indian digital library of manuscripts initiatives; among others MASTER (Manuscript Access Standards for Electronic Records), The Digital Scriptorium (<http://sunsite.berkeley.edu/scriptorium/>), Oxford University Manuscripts, (<http://image.ox.ac.uk>), Medieval Manuscripts of Syracuse University (<http://libwww.syr.edu/digital/collections/m/Medieva>), The Bibliotheque National De France (<http://www.bnf.fr/enluminures/>), "Medical Manuscripts in NLM", "European Manuscript Server Initiative" (EMSI), and "Unesco Memory of World"

Those projects have used photographic and digital methods to develop a full or partial archival copy of the manuscripts available on-line using different software with JPEG format for image processing in association with different metadata initiatives. Shafi (2004) argues that these attempts could be very useful in evolving a common strategy for digitization of vast manuscripts resources.

Chepesiuk (2001) in his article mentioned that some of the leading research libraries in United States are taking advantage of digital technology to reach out to the global community after years of limiting the accessibility of their rare books, manuscripts, and documents to on-site visitors. He briefly describes a number of digital library initiatives such as University of Michigan Library "Making of America Project



(moa.umdl.umich.edu/), New York Public Library "Digital Library Collection" (digital.nypl.org/), University of Virginia Library "American Civil War Collections" (etext.lib.Virginia.edu/civilwar/) and several other projects. He concluded his article with a remark that the bottom line is: Digitizing can be costly and time-consuming, but most archives and special collections could do it in some form that is feasible to them.

The development of a digital library for Malay manuscripts would provide significant insurance against the loss of these valuable materials for use to the learning community. The storage of digitized documents in the open Internet environment is an important mechanism for allowing valuable and rare documents to be accessed easily by a large number of geographically distributed users.

### 3. Methodology and Scope

To elicit the preliminary requirements, of the digital library for manuscripts, questionnaires were sent to the five main Malay manuscripts repository located in the Klang Valley which is National Library of Malaysia, University of Malaya Library, Tun Seri Lanang Library (UKM Library), Dewan Bahasa dan Pustaka (The Institute of Language and Literature) and Islamic Arts Museum Malaysia. The aim of the questionnaire was to find out about the state of the art of Malay manuscripts organization amongst repositories in Malaysia.

Since the sample is small and the aim is to get as much information as possible from the responsible institution, a face-to-face meeting was arranged at all five institutions to distribute the questionnaire. The questionnaire were administered to the professional member of staff that directly responsible for the Malay manuscripts collection in their organization. The professional staff responded to the questionnaire during the meeting and the session was followed with an unstructured interview to obtain additional information.

After the required data has been collected, the next step is to implement a prototype of digital library of Malay manuscripts using Greenstone, an open source digital library software. The aim is to overcome the barriers that exist in accessing Malay manuscripts collection through traditional channel that being practiced by the repositories nowadays.

### 4. Results and Discussion

#### (a) Malay Manuscripts Collection

Table 1 indicates the number of Malay manuscripts copies held by each institution. The Malay Manuscripts Centre in the National Library of Malaysia have the biggest collection since it is the official Malay manuscripts repository in Malaysia and the acquisition is a part of a coherent collection development policy. On the other hand, the large manuscript collection in the Islamic Arts Museum Malaysia are mostly made up of manuscripts covering subjects related to Islamic teaching.

Table 1: Malay Manuscripts Collection Held at Five Selected Repositories

| <i>Institutions</i>          | <i>Malay manuscripts collection<br/>(no. of copies)</i> |
|------------------------------|---|
| National Library of Malaysia | 3699  |
| University of Malaya Library | 307   |
| Tun Seri Lanang Library      | 57  |
| Dewan Bahasa dan Pustaka     | 226   |
| Islamic Arts Museum Malaysia | <3500   |
| <b>Total</b>                 | <b>&lt;7789</b>   |



Table 2 shows that all the repositories being surveyed unanimously provide their Malay manuscripts collection separately on close access. This means that access to these resources is limited and requires the client to contact the librarian in charge, while browsing the collection is also restricted.

Table 2: Method of Shelving the Malay Manuscripts Collection by Institutions

| Institutions                 | Separately on open access | Integrated with other collection on open access | Separately on close access | Integrated with other collection on close access |
|------------------------------|---------------------------|---|----------------------------|--|
| National Library of Malaysia |                           |   | √                          |  |
| University of Malaya Library |                           |   | √                          |  |
| Tun Seri Lanang Library      |                           |   | √                          |  |
| Dewan Bahasa dan Pustaka     |                           |   | √                          |  |
| Islamic Arts Museum Malaysia |                           |   | √                          |  |

Table 3 shows that the usage rate of the Malay manuscript collection held at the University of Malaya Library was more than ten times per month. This infers that the Malay manuscripts collection is not fully utilized by patrons.

Table 3: Institutions and Monthly Usage of the Malay Manuscripts Collection

| Institutions                 | <10 | 10-30 | >30 |
|------------------------------|-----|-------|-----|
| National Library of Malaysia | √   |       |     |
| University of Malaya Library |     | √     |     |
| Tun Seri Lanang Library      | √   |       |     |
| Dewan Bahasa dan Pustaka     | √   |       |     |
| Islamic Arts Museum Malaysia | √   |       |     |

Two of the repositories, namely Dewan Bahasa dan Pustaka and National Library of Malaysia have already initiated digitization effort of their Malay manuscripts collection. Both of them outsourced their digitization projects, which result in very high cost. The other repositories are planning to digitize their Malay manuscripts collection in the near future.

Making information available through digital library is more than just a simple website. The Web lacks the essential features of selection and organization. To consider a collection as a "library", the collection must be maintainable, searchable, and browsable. Most well-organized Web sites are created manually through hand-crafted hypertext linkage structures. But just as adding new acquisitions to a physical library does not involve digging into the books and rewriting parts of them, so it should be possible for new material to be added to a digital library without any need for manual updating of the structures used for access and retrieval.

Prior to presentation, each collection should undergo a building process that, once established, is completely automatic. This process creates all the structures that are used at runtime for accessing the collection. Searching is based on various indexes involving full text and metadata. Browsing is based on various metadata and on phrase structures, and other information, abstracted from the full text of the documents. Support structures



for both are created during the building operation. When new material appears, it can be fully incorporated into the collection by rebuilding.

These exceptionally broad requirements of digital libraries are difficult to achieve by building them in-house, especially in situations where the professional staffs do not have the technical expertise required. To outsource them is another alternative but the cost is a major factor to be considered, and these personalized digital libraries are generally not extensible. To address these problems, the Greenstone digital library software is chosen for the development of the manuscript digital library. Greenstone is a digital library software made public and free. It is issued under the GNU General Public License, and, in the spirit of open-source software, users are invited to contribute modifications and enhancements

## 5. Implementation

### (a) Introduction on Greenstone

The Greenstone Digital Library Software (<http://www.greenstone.org>). provides a way of building and distributing digital library collections, opening up new possibilities for organizing information and making it available over the Internet or on CD-ROM (Witten and Bainbridge, 2003). Greenstone is developed by the New Zealand Digital Library group project (<http://nzdl.org>), Greenstone is intended to lower the bar for construction of practical digital libraries, yet at the same time leave a great deal of flexibility in the hands of the user.

In accordance with the maxim "simple things should be easy, complex things should be possible" new users can quickly put together standard-looking collections from a set of source documents that may be HTML, Word, PDF, or many other formats (Witten, Bainbridge and Boddie, 2001). Given an existing collection, it is easy to clone its structure and populate an identical copy with entirely new documents, provided they are in the same formats as those in the existing collection. A more committed user who studies the options that Greenstone offers can personalize the digital library system and create new kinds of collection that take advantage of available metadata to provide different kinds of browsing facilities, which are akin to different perspectives on the collection. Users with programming skills can extend the system by adding modular units called "plugins" that accommodate new document and metadata formats, and new browsing and document access facilities.

Greenstone has been used to develop many digital library collections. Some were created within the New Zealand Digital Library as demonstration collections. However, the use of Greenstone internationally is growing rapidly, and several web sites show collections created by external users. Most contain unusual and interesting material, presented in novel and imaginative ways.

### (b) Building a Prototype Digital Library of Malay Manuscripts Using Greenstone

Greenstone does enable the creation of local content such as Malay manuscripts, strengthen the mechanisms and capacity of the library's information systems and services. They increase the portability, efficiency of access, flexibility, availability and preservation of content. Once the information is made digital, it could be stored, retrieved, shared, copied and transmitted across distances without having to invest any additional expenditure. Greenstone software runs on a wide variety of platforms such as Windows, Unix / Linux, and Apple Mac. Collections could be built and served locally from the user's own web server, or (given appropriate permissions) remotely on a shared digital library host. In this study, the collection is served on an Apache web server. Figure 1 illustrates the system architecture for the prototype digital library of Malay manuscripts that has been built.



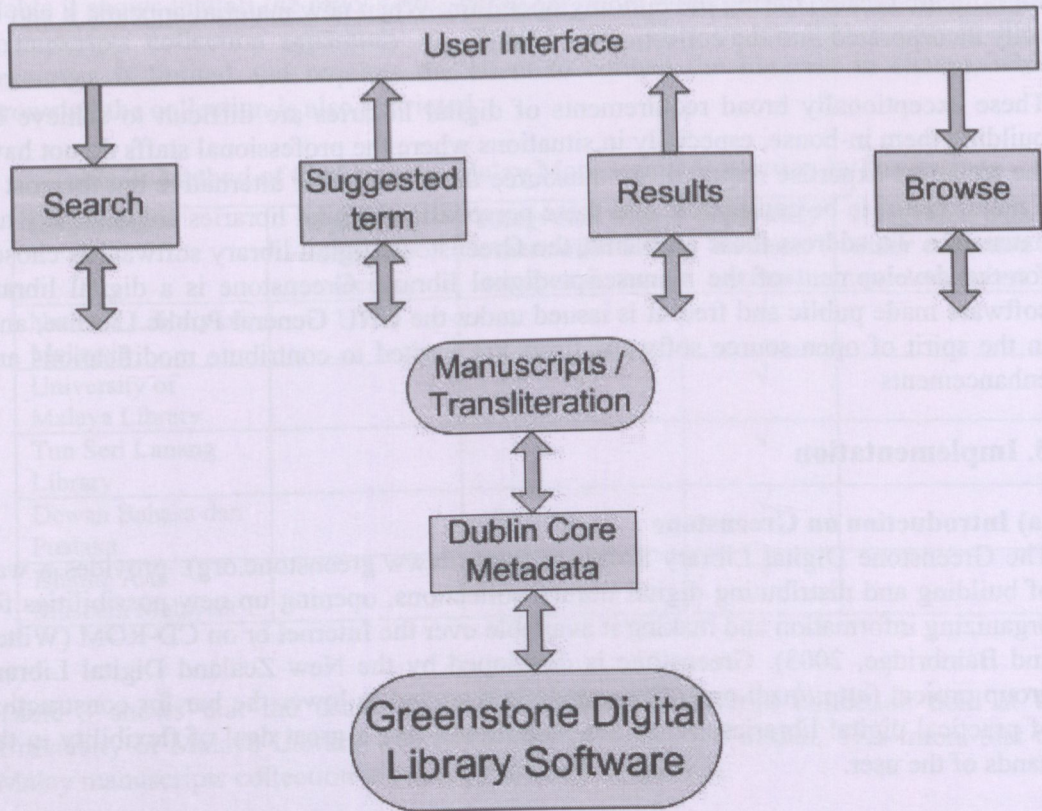


Figure 1: System Architecture of the Prototype Digital Library of Malay Manuscripts

To build the collection, the Greenstone Librarian Interface (GLI) which is a Java based GUI interface is used. It is the easiest way to build a collection. Collection developers can activate the GLI software and use the ‘Gather’, ‘Enrich’, ‘Design’, and ‘Create’ panel for making collection.

The ‘Gather’ panel facilitates extracting the relevant files from the local drive into the collection. The ‘Enrich’ panel explains how metadata is created, edited, assigned and retrieved, and how to use external metadata sources. The ‘Design’ panel facilitates customizing the interface, once the files are marked up with metadata. Using the ‘Design’ panel, user can specify the fields that are searchable, allow browsing through the document, facilitate the languages that are supported, and provide the buttons that are to appear on the page. The ‘Create’ panel facilitates the creation of the collection.

For the purpose of this study, the raw material for building a prototype digital library using Greenstone is provided in machine-readable form where a sample of Malay manuscript entitled “Hikayat Parang Puting (MSS 34)” is acquired with permission through the National Library of Malaysia Digital Resource System (<http://sumberdigital.pnm.my/>). The sample manuscript is in PDF form that is supported by Greenstone. Greenstone also supports many other formats such as HTML, XML, Word, PostScript, RTF, JPG, GIF and MPEG with the use of various extensible plug-ins.

User will get to select desired metadata sets. By default, the system will prompt Dublin Core metadata. Figure 2 illustrates the ‘Enrich’ panel where necessary values were given for the Dublin Core element sets.



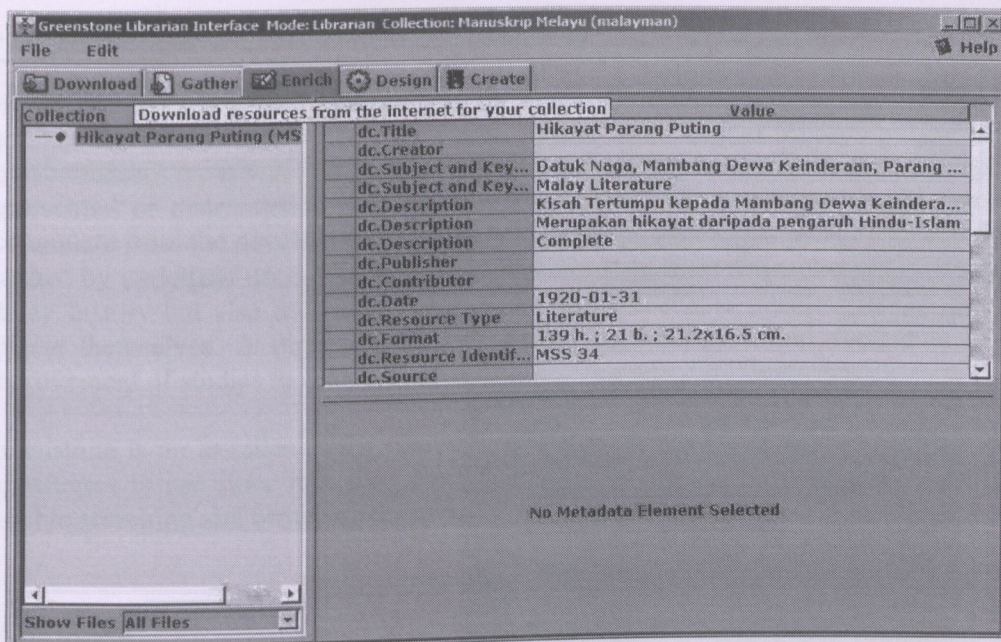


Figure 2: Assigning Metadata in Greenstone

For this study, it is useful to provide a facility for viewing the transliteration of the Malay manuscripts as well. Figure 3 shows the screenshot from the sample "Hikayat Parang Puting", in both image and text form. Both forms are provided in this collection because of the advantage of being able to search for the terms in the full text of the transliteration. The transliteration was obtained from a publication published by Dewan Bahasa dan Pustaka (Jamilah, 1980).

Since this prototype digital library of Malay manuscripts is built using only one sample of manuscript, it is somehow limited in demonstrating the excellent searching, browsing and retrieval capabilities of Greenstone. Information collections built by Greenstone combine full-text search with browsing indexes based on different metadata types. There are several ways for users to find information, although they differ between collections depending on the metadata available and the collection design. The default search interface is a simple one, suitable for casual users. Advanced searching, which allows Boolean expressions, phrase searching, and case and stemming control, can be enabled from a "Preferences" page.

Greenstone supports full-text searching, and, depending on the collection, users may select indexes built from different parts of the full text or the metadata. Some collections have separate indexes of full documents, sections, paragraphs, titles, and section headings, each of which can be searched for particular words or phrases. When browsing, users examine data structures created from metadata: lists of authors, titles, dates; hierarchical classifications; and so on. Structures for both searching and browsing are specified by instructions in the configuration file, and can be rebuilt entirely automatically. No information is inserted by manually.

Each document may be hierarchically organized into logical *sections*, each of which comprises *paragraphs*. Metadata such as author, title, date, keywords, may be associated with documents, or with individual sections. This is the raw material for indexes. It must either be provided explicitly (for example, in an accompanying spreadsheet) or be derived automatically from the source documents. Metadata is stored with the document for internal use.



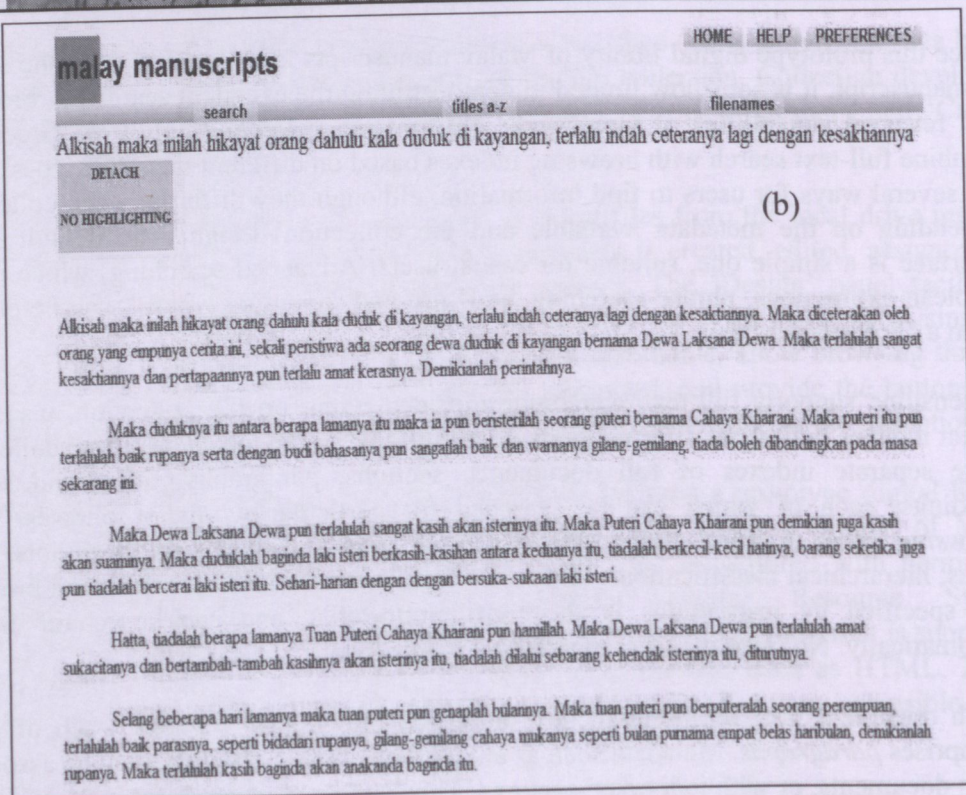
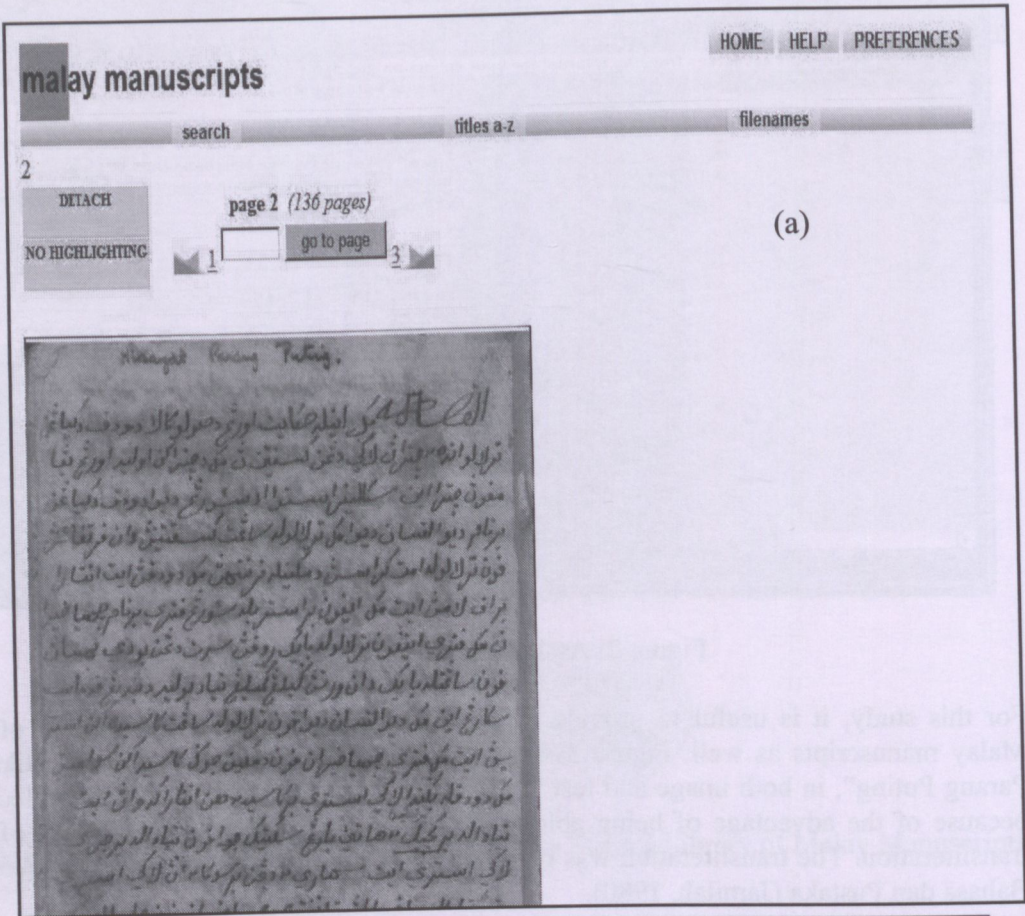


Figure 3: Hikayat Parang Puting: (a) Page Image (b) Text Transliteration



## 6. Conclusion

Digital libraries of Malay manuscripts cater for the critical importance of time and facilitate the satisfaction of the user's need in real time. Digital library collections are not limited to document surrogates, they extend to digital artifacts that cannot be represented or distributed in printed formats. Access is the primary benefit that will accumulate from the development of a digital library of Malay manuscripts that could be reached by variety of users. These users will not only include researchers and experts in Malay history but also individuals with general and aesthetic interests in the physical objects themselves. A digital library for Malay manuscript would contribute to the improvement of studies in Malay history, culture, literature and linguistics.

Greenstone is an excellent open source digital library software for Malay manuscripts repositories to use since it is free, multi-platform, metadata-driven, extensible and has flexible searching and browsing features.

## References

- Aytac, S. 2002. Development of a Web Based Digital Archive System for Ottoman Manuscripts. Paper presented at the *European Summer School Electronic Publishing for Cultural Heritage Studies*, September 9-26, Bulgarian Academy of Sciences, Institute of Mathematics, Sofia, Bulgaria.
- Chepesiuk R. 2001. Digitizing rare materials: Special collections go global. *American Libraries*. 54.
- Ding, C. M. 1986. Access to Malay Manuscripts. Paper presented at the XXXII *International Congress for Asian and North African Studies*, August 25-30, Hamburg.
- Ding, C. M. 1993. Malay manuscripts: materials and problems of conservation *Journal of the Malaysian Branch, Royal Asiatic Society*, 66(pt.2), 81-102.
- Jamilah, H.A. 1980. *Hikayat Parang Putting*. Dewan Bahasa dan Pustaka, Kuala Lumpur.
- Shafi, S.M. 2004. *Digitization Perspective of Medieval Manuscripts*. Paper presented at the 2<sup>nd</sup> Convention PLANNER – November 4-5, 2004, Manipur University, Imphal.
- Witten, I.H., Bainbridge, D. and Boddie, S. 2001. Power to the people: end-user building of digital library collections. *Proc ACM Digital Libraries*, Roanoke, VA.
- Witten, I.H. and Bainbridge, D. 2003. *How to Build a Digital Library*. Morgan Kaufmann, San Francisco, CA.