Multilingual Searching in Digital Archives: A case study of Mahatma Gandhi University On-line Theses Archive

Jasimudeen S; Maghesh Rajan M; Suresh Kumar TV

Mahatma Gandhi University Library, Kottayam

Corresponding author email:jasimndd@gmail.com

Abstract

The Paper examines the multilingual search capabilities of Mahatma Gandhi University on-line theses archive. The study finds that the multilingual search features of the Mahatma Gandhi University archive helps the user with more relevant hits.

Introduction

Remote access to full text resources is much easier than ever before in today's academic arena. Open source initiatives and OA Philosophy encouraged Electronic Theses and Dissertations (ETDs) projects. "An ETD is a document that explains the research or scholarship of a researcher/student. It is expressed in a form simultaneously suitable for machine archives and worldwide retrieval. The ETD is similar to its paper predecessor. It has figures, tables, footnotes, and references. It has a title page with the authors' name, the official name of the university, the degree sought, and the names of the committee members. It documents the author's years of academic commitment. It describes why the work was done, how the research relates to previous work as recorded in the literature, the research methods used, the results, and the interpretation and discussion of the results, and a summary with conclusions. The ETD is different, however. It provides a technologically advanced medium for expressing your ideas".[1]

Mahatma Gandhi University (MGU) Online Theses Archive

More than 1150 theses have been digitized and uploaded in the Mahatma Gandhi University Online Theses Digital Library with standard search utilities. A special archiving package named Nitya D' Arch is used for hosting the digital copies of doctoral dissertations emanated out from Mahatma Gandhi University. Deep indexing especially for documents in Indian Languages is a unique feature of Nitya D' Arch Package[2]. The archive supports Multilingual search facility and is OAI-PMH (Open Access initiative - Protocol for Metadata Harvest) compliant.
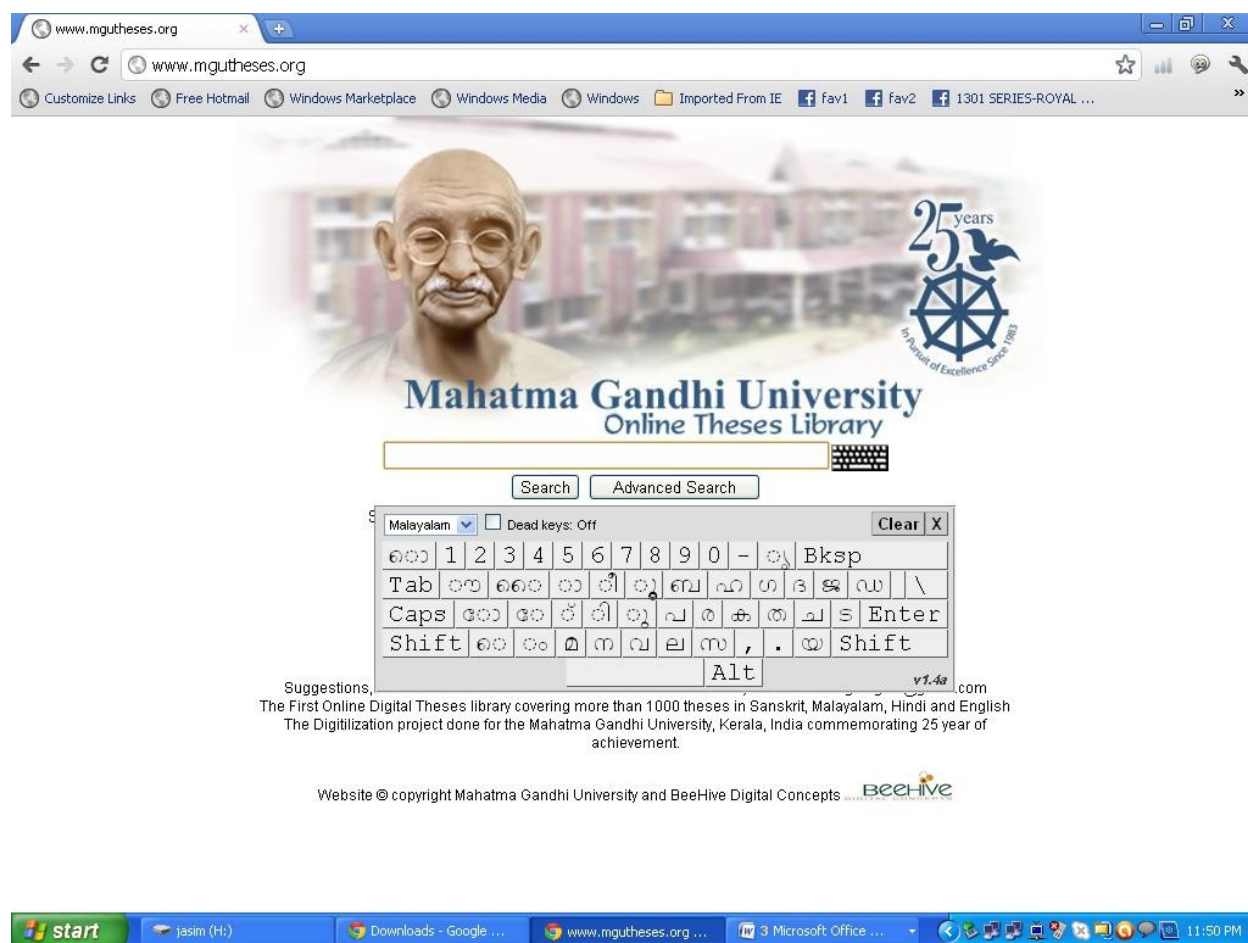
OBJECTIVES OF THE STUDY

   1. Study of the Multilingual Search capability of the MGU theses archive.

Search using Indic Script (Malayalam)

Figure 1

MGU theses archive screenshot (Malayalm multilingual keyboard selected)

   A comparison of **recall/relevance hits** in Nitya D'Arch by using random keywords in Malayalam   is given in the table 1

Table 1

SEARCH RECALL/RELEVANCY USING INDIC SCRIPTS

Using Malayalam scripts with random key words

keyword	hits	Relevant items	keyword	Hits	Relevant items
ബഷീര്	5	5	Basheer	3	3
നാടകം	7	7	Drama	67	3
മലയാള കവിത	15	15	Malayalam poetry	39	25
മാധവികുട്ടി	4	4	Madhavikutty	2	2
പഠനം	42	41	Study	110	17

The table 1 shows that the use of the Malayalam scripts in the Nitya D'arch in which the recall and relevancy is high. If the Malayalam script is used in English the search relevancy will be less as mentioned in the table1. E.g. The word "ബഷീര്" is used as search key word and the number of hits will be 5 and all those 5 documents are relevant. The term " നാടകം" in Malayalam search will provide the relevant items only where as in the term in English will bring all related documents on drama on various subjects of interest. It proves the concept of relevancy in the searching of the regional language.

Search using Indic Script (Hindi)

   A comparison of recall/relevance hits in Nitya D'Arch by using random keywords in Hindi is given in the table 2

Table 2

Using Hindi scripts with random key words

keyword	hits	Relevant items	keyword	hits	Relevant items
हिन्दी कविता	15	15	Hindi poetry	29	12
कबीरदास	1	1	Kabir das	0	0
प्रेम	10	10	love	5	2
नारी	4	4	women	3	3
प्रकृति	1	1	environment	0	0

The table 2 is having the explanation when the search is made in Hindi.

The Hindi term will be searched as it is, and the result of hits will be in Hindi which will be the same of the relevant items. If the search is in English, the relevancy will be less than the hits. It shows that the use of Indic scripts in the Nitya D'Arch is highly accommodating the user's want and demand based on his/her query.

Conclusion

The multilingual search capabilities and retrieval efficiency are found to be excellent in Mahatma Gandhi On-line theses archive. The information searcher is able to find the exact information what they search for and thus the aims and objectives of initiation of archive is almost acheived.

References

UNESCO ETD Guide. Retrieved on 12 December 2011 from:http://etdguide.org/

Raman Nair, R. Mahatma Gandhi University Open Access Digital Library of PhD Theses: User Guide V4.0., 2010