

2016 Satellite meeting - *Subject Access: Unlimited Opportunities*

11 – 12 August 2016

State Library of Ohio, Columbus, Ohio, USA

Improving Resource Discoverability for Non-Roman Language Collections

Magda El-Sherbini

The Ohio State University Library, Columbus, Ohio, USA

E-mail address: el-sherbini.1@osu.edu



Copyright © 2016 by Magda El-Sherbini. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

Abstract:

Linking non-Roman subject terms with LCSH terms through the use of FAST terms opens new possibilities for improving resource discoverability and access to non-Roman script collections. My discussion will be focused on how the MARC21 record can be utilized to display non-Roman script subject terms in the ILS. To maximize discoverability of materials, non-Roman terms are linked to FAST subject terms. I will describe the experience of subject searching of non-Roman script materials using either the FAST English language terms, or their equivalents in non-Roman script languages. The Arabic language subject terms will be used as a prototype that can be applied as best practice for other non-Roman languages.

Keywords: Linking fields, non-Roman languages, subject headings, FAST

Introduction

Providing simultaneous access to multi-script subject headings has been recognized as an issue to be solved by the Association for Library Collections and Technical Services (ALCTS) Task Force on non-English Access as far back as 2007 (ALCTS Task Force on Non-English Access Report, 2007).

It was generally assumed that in the United States the language of the user was English. For this reason the language of the catalogue record and the language of the catalogue interface have generally been English. Although many libraries developed substantial foreign language collections, the language of the library catalogue and its subject access points has been English. The author began to study this topic in 2010 after being appointed to serve as Chair of the ALCTS Steering Committee to Oversee the Implementation of the Recommendations Contained in the Report of the ALCTS Task Force on non-English Access 2007-2009 (ALCTS Steering Committee, 2009). The report included eleven recommendations. The Steering Committee was successful in implementing ten of the recommendations, but recommendation number eleven; “Assign the ALCTS Subject Analysis Committee (SAC)

working with appropriate library organizations to study the needs of library users for multilingual subject access in the appropriate script(s), and to propose steps to address those needs,” was not implemented (ALCTS Task Force on Non-English Access Final Report, 2009).

The ALCTS Steering Committee members explained the reasons for not implementing recommendation eleven in these terms: “this recommendation was extremely vague and broad; a research project that might never be completed; is outside SAC’s scope; SAC doesn’t do much research on end users or their needs” and finally more research is needed.

It became evident that research in this area is needed in order to collect data and evaluate the need for providing multilingual subject access to non-Roman materials. This paper will report on the findings of this research and will focus on the methods of using MARC 21 records to display non-Roman script subject terms in the currently available ILS. To maximize discoverability of materials, the Non-Roman terms are linked to the FAST subject terms. I will describe the experience of subject searching of non-Roman script materials using either the FAST English language terms, or their equivalents in non-Roman script languages. The Arabic language subject terms will be used as a prototype in this study.

Before discussing the research methods, the term “multilingual” needs to be defined. The Cambridge Dictionaries Online provides a simple and precise definition of “Multilingual” by stating “of people or group able to use more than two languages for communication, or (of a thing) written or spoken in more than two different languages” (Cambridge Dictionaries). In general terms “Multilingual” means more than just one language.

In an automated library environment with multilingual collections, the task is to successfully provide access to all materials, including the non-Roman script collections. Currently this is accomplished by applying the Library of Congress MARC 21 Model A (*Vernacular and transliteration*) in the bibliographic record, where the regular MARC 21 fields may contain data in different scripts and in the vernacular or transliterated form of the data. The original-script fields are coded as 880 fields at the end of the bibliographic record, but in public display, and sometimes in staff display, they can be displayed next to the corresponding Romanized field (ALCTS Working Group on Romanization, 2009). Model A is used for the descriptive part of the bibliographic record. However, the subject headings continue to be assigned in English, using the Library of Congress Subject Headings (LCSH) (except some names or corporate bodies where they are assigned as subject and they have the vernacular form established in the Authority File). (See figure 1.)



Figure 1: Subjects are in English

In a survey that was conducted in 2011, the author attempted to assess the need to provide Non-Roman subject access to the library online catalogue. The results of the survey indicated that:

- ⊙ End users were not completely dissatisfied with the current library catalog;
- ⊙ End users and librarians wanted a system that is more open to multilingual subject headings;
- ⊙ Highlighted areas of opportunity for libraries to make significant improvements to the catalog (El-Sherbini and Chen, 2011).

Following the above research, the author received a grant to study the Bibliotheca Alexandrina (BA) multilingual catalogue where the BA provides access to their library collections in three languages (Arabic, English, and French). This research revealed that the BA is using the Library of Congress MARC 21 Model B, where the transcribed text in the bibliographic record is entered only in the script in which it appears. Subject access is recorded based on the authorized thesaurus in each language. For example, the French thesaurus “RAMEAU” is used for assigning subject headings for a book in the French language (El-Sherbini, 2013)

Reasons for providing multilingual subject access to bibliographic records

Information gathered from the work on the ALCTS Steering Committee and my research gave birth to the idea of internationalizing the Ohio State University (OSU) library catalogue. Three factors made this development possible:

⊙ RDA and Multilingual access

RDA was written with “Internationalization” in mind. “RDA promotes internationalization because it accommodates languages, scripts, numerals, and dates other than those in English. So, it facilitates the exchange of data between different language communities. The Joint Steering Committee for the Development of RDA (JSC) realizes that there are some areas (e.g., instructions for musical works and for legal works) that still demonstrate a lack of internationalization; work will continue in those areas” (RDA FAQ).

RDA Chapter 0

Purpose 1: RDA provides a set of guidelines and instructions on recording data to **support resource discovery.**

Purpose 2: The data created using RDA to describe a resource are designed to assist users in performing the following tasks:

- ❖ Find
- ❖ Identify
- ❖ Select
- ❖ Obtain

Purpose 3: Data created using RDA to describe an entity associated with a resource (person, family, corporate body, concept, etc.) are designed to assist users in performing the following tasks:

- ❖ Find
- ❖ Identify
- ❖ Clarify
- ❖ Understand

RDA Person Name, Family Name and Corporate body

- ❖ If the name of a person is found in a script that differs from a preferred script of the agency creating the data, transliterate the name according to the scheme chosen by the agency. 9.2.2.5.3
- ❖ If there is a well-established form of name in reference sources in a language preferred by the agency creating the data, choose that form of name as the preferred name.
- ❖ If different forms are found in reference sources in a language preferred by the agency creating the data, choose the form that occurs most frequently.

Alternative Linguistic Forms of Name for Place -- RDA 16.2.3.7

- ❖ If the name recorded as the preferred name for a place has one or more alternative linguistic forms, record them as variant names.
(Morocco) تطوان
French language form recorded as preferred name: Tétouan (Morocco)

◎ **FAST and Multilingual Access**

Although LCSH is the most commonly used subject vocabulary in most libraries, its complex structure makes searching by subject difficult. This “obstacle” led to the creation of a new subject schema, the Faceted Application of Subject Terminology (FAST), which is derived from the LCSH, but applies a simpler syntax while retaining the richness of the LCSH vocabulary. FAST was developed by a team of researchers at OCLC in conjunction with a team from the Library of Congress. O’Neill and others provided lengthy justifications for developing a different syntax that is based on a controlled vocabulary and at the same time is flexible and

simple to use (O'Neill and others). In 1998, the OCLC Office of Research built a software to automatically process and convert the LCSH authority file and LCSH present in the WorldCat bibliographic record to FAST subject terms (OCLC Research). The main idea was to have OCLC deconstruct the LCSH into discrete headings each containing one facet. As a result of this initiative, most of the records in OCLC include FAST subject terms parallel to LCSH. The FAST terms are found in MARC 6xx second indicator 7 subfield 2 fast. The possibility of linking FAST terms to terms in different languages is greater than linking to LCSH.

◎ **Users' Needs:**

Users can search the library catalogue by subject terms in their preferred languages, e.g., Arabic, Chinese, Japanese, etc. A times, the terms used in the LCSH are not equivalent in meaning to those in other languages. For example, there are several ways of reciting the Koran in Arabic (Weheba, 2012):

- ❖ Koran -- Seven Successive Readings
- ❖ Koran -- Three Completing Readings
- ❖ Koran -- Ten Readings
- ❖ Koran -- Fourteen Readings
- ❖ Koran -- Irregular Readings
- ❖ Koran -- Forged Readings
- ❖ Koran -- Interpolated Readings

However, the LCSH provides a broader term to include all types of readings. (See figure 2.)

010	sh 85072983
040	DLC #b eng #c DLC #d DLC
053	0 BP131.5
130	0 Qur'an #x Readings
430	0 Qur'an #x Qirā'āt
450	Qirā'āt
530	0 Qur'an #x Criticism, Textual #w g
680	#i Here are entered works on the Qur'anic science known as the "science of the readings" (Arabic: qirā'āt) which deals with various renditions of the text of the Qur'an according to different oral traditions.

Broader Term of the Quran types of Readings

Figure 2: Record from the LC Authority File

Ohio State University Library and the Multilingual Subject Access Initiative

All these developments provided an opening for exploring the idea of making the subject search more effective. In addition, the diverse user population of OSU prefers searching the catalogue in their own language. The project to internationalize the OSU catalogue started in February 2015 when Arabic subject terms were added to bibliographic records. Minor configurations to the local Online Catalogue System were made to enable display of the non-Roman subject terms. One example involved indexing the MARC 21 field 6xx with the second indicator 7. Cataloguers were given specific instructions on how to establish terms in the respective languages. The Arabic language cataloguer is using a controlled vocabulary from the printed Arabic subject list. If the subject term is equivalent to the FAST term, the cataloguer links them (figure 3). When the subject term is not found in the thesaurus, the cataloguer assigns a local subject term and indicates in the subject field in the bibliographic record that the term was created locally (figure 4). For personal name, family name, and corporate body, the cataloguers are using the LC Name Authority File as a source of names as subjects (figure 5).

.245	0	0	موسوعة جمال عبد الناصر فى الفقه الاسلامى.
.245	0	0	Mawsū'at Jamāl 'Abd al-Nāṣir fi al-fiqh al-Islāmī.
.246	1		#i Volumes 11-30 have title: #a موسوعة الفقه الاسلامى
.246	1		#i Volumes 11-30 have title: #a Mawsū'at al-fiqh al-Islāmī
264	1		القاهرة : #b المجلس الاعلى للشؤون الاسلامية، #c 1386- [1966-
264	1		al-Qāhirah : #b al-Majlis al-A'lá lil-Shu'ūn
650	0		Islamic law #v Encyclopedias .
650	0		Islamic law #v Terminology .
650	0		Islamic law #x Dictionaries #x
.650	7		الفقه الإسلامى .#2 qrmak
.650	7		Islamic law. #2 fast #0 (OCoLC)fst00979949
.655	7		معاجم .#2 qrmak
.655	7		Dictionaries. #2 fast #0 (OCoLC)fst0142
.655	7		دوائر معارف .#2 qrmak
.655	7		Encyclopedias. #2 fast #0 (OCoLC)fst01423798
.655	7		مصطلحات .#2 qrmak
.655	7		Terminology. #2 fast #0 (OCoLC)fst01423880
.710	1		مجلس الاعلى للشؤون الإسلامية #b مصر.
.710	1		Egypt #b Majlis al-A'lá lil-Shu'ūn al-Islāmīyah.

Figure 3: Arabic Terms are linked to FAST Terms

.245	0	2	المقدم الى الدورة العادية التاسعة والستين للمجلس #b : التقرير المصنف لسوق للاسواق العام
.245	0	3	al-Taqrīr al-niṣf sanawī lil-amn al-āmm : #b al-muqaddam lá al-dawrah al-'ādīyah al-'āsi'ah wa-al-sittin lil-Majlis.
.264	1		القاهرة : #b المجلس الوحدة الاقتصادية العربية الامانة العامة، #c 1999.
.264	1		al-Qahirah : #b Majlis al-Wahdah al-Iqtisādīyah al-'Arabīyah al-'Amānah al-'Āmmah, #c 1999.
610	2	0	Council of Arab Economic Unity .
651	0		Arab countries #x Economic integration .
.610	2	7	مجلس الوحدة الاقتصادية العربية .#2 local/OSU
.610	2	7	Council of Arab Economic Unity. #2 fast #0 (OCoLC)fst00567116
.650	7		التكامل الاقتصادى الدولى .#2 qrmak
.650	7		International economic integration. #2 fast #0 (OCoLC)fst00976879
.651	7		الدول العربية .#2 qrmak
.651	7		Arab countries. #2 fast #0 (OCoLC)fst01240128
710	2		Council of Arab Economic Unity .

Figure 4: Local Subject Term

100	1		ناصر، شحاتة محمد.
100	1		Nāṣir, Shāḥātah Muhammad, #e author.
.245	1	2	شحاتة محمد ناصر / #c حدود التأثر وأهم الملامح #b : السياسة الخارجية الإيرانية فى عهد الرئيس حسن روحانى
.245	1	3	al-Siyāsah al-khārijīyah al-īrānīyah fi 'ahd al-ra'īs Ḥasan Rūḥānī : #b ḥudūd al-ta'thīr wa-ahamm al-malāmīh / #c Shāḥātah Muhammad Nāṣir.
.250			الطبعة الاولى.
.250			al-Ṭab'ah al-ūlá.
.264	1		مركز الامارات للدراسات والبحوث الاستراتيجية، #b : أبو ظبى #c 2014.
.264	1		Abū Ḍaby : #b Markaz al-Imārāt lil-Dirāsāt wa-al-Buḥūth al-Istīrāṭīyah, #c 2014.
.490	1		العدد 191 #v دراسات استراتيجيه :
.490	1		Dirāsāt istīrāṭīyah ; #v al-'adad 191
651	0		Iran #x Foreign relations #y 1997-
651	0		Iran #x Politics and government #y 1997-
.600	1	4	روحانى، حسن.
.600	1	0	Rūḥānī, Hasan .
.650	7		العلاقات الخارجيه .#2 qrmak
.650	7		Diplomatic relations. #2 fast #0 (OCoLC)fst01907412
.650	7		أحوال سياسيه .#2 qrmak
.650	7		Politics and government. #2 fast #0 (OCoLC)fst01919741

Figure 5: The LC Name Authority File is used to assign Personal Names as Subject

How to Search OCLC by Subject in non-Roman scripts?

Users can search OCLC Connexion by the non-Roman subject terms. Figure 6 illustrates a search by a personal name as a subject term. The search was qualified to show only that part of the OSU collection that is about this particular person. The search resulted in four hits about this person (figure 7).

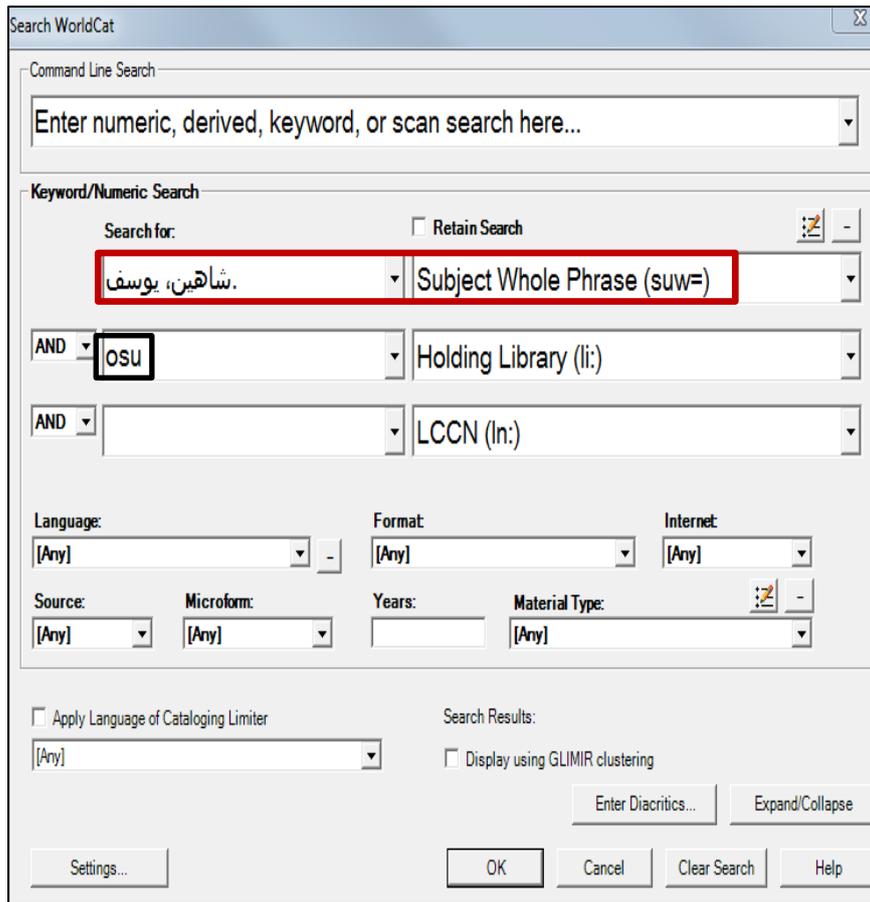


Figure 6: Searching by a Personal Name as Subject

Searching the OSU Catalogue by Non-Roman Scripts

Currently, OSU users can search the online catalogue by subjects in Non-Roman scripts for Arabic. The other Non-Roman scripts will be added in the future. Searching the online catalogue is fairly simple. First, users can download the keyboard from the Internet in the language they prefer. Second, they can change the catalogue menu to search by “subject” as shown in figure 8, and type in the query.

Rec...	Description
1	الكتب [et al.] ... ابن رشد وفيلم العصر ويوسف شاهين وفيلم الآخر وعولمة الجرب ضد الإسلاميين / محمد إبراهيم مبروك [Giza] : [Giza] - [1999] : مركز الحضارة العربية. 179 pages : 24 cm. CatL:eng DLC OCLC: 43293083, Held, Other IRs: 8, Holdings: 20
2	حسن، محمد. بن صلاح أبو يوسف ويوسف شاهين : الرؤية الفكرية السبعينية / محمد حسن. الطبعة العربية 1. القاهرة : مركز الحضارة العربية، 2006
3	Hubaysh, Muhammad. Bayna Salāh Abū Sayf wa-Yūsuf Shāhīn : al-ru'yah al-fikriyyah al-sinimā'iyyah / Muhammad Hubaysh. al-Tab'ah al-'Arabīyah 1. al-Qāhirah : Markaz al-Hadārah al-'Arabīyah, [1999] 179 pages : 24 cm. CatL:eng DLC OCLC: 70158034, Held, Other IRs: 2, Holdings: 12
4	جمال، أمال. يوسف شاهين وجزيرة الإنجاز المشترك / أمال الجمل. الطبعة الأولى. دمشق : ورارة الثقافة المؤسسة العامة للسماع، 2011
5	Jamal, Amal, author. Yūsuf Shāhīn wa-tajrubat al-intāj al-mushtarak / Amal al-Jamal. al-Tab'ah al-'Ulā. Dimashq : Wizārat al-Thaqāfah, al-Mu'assasah al-'Ammah lil-Sinimā, 2011. 248 pages : illustrations (chiefly color) ; 24 cm. CatL:eng DLC OCLC: 824868338, Held, Other IRs: 3, Holdings: 7
6	Kammoun, Mirvet Médini. Youssef Chahine, caméra de tous les combats / Mirvet Médini Kammoun. La Manouba, Tunisie : Centre de publication universitaire, 2010. 391 pages : illustrations ; 24 cm CatL:eng DLC OCLC: 767720430, Held, Other IRs: 4, Holdings: 13
7	رياض الرس للكتب والنشر. Beirut : 2001. شعيطا، وليد. يوسف شاهين : جناه للسينما / وليد شعيطا. الطبعة 1 Shamit, Walid. Yūsuf Shāhīn : hayāh lil-sinimā / Walid Shamit. al-Tab'ah 1. Beirut : Riyād al-Rayyis lil- Kutub wa-al-Nashr, 2001. 285 pages : illustrations ; 22 cm CatL:eng DLC OCLC: 48979915, Held, Other IRs: 6, Holdings: 34

Figure 7: The Result of Searching for a Personal Name as Subject

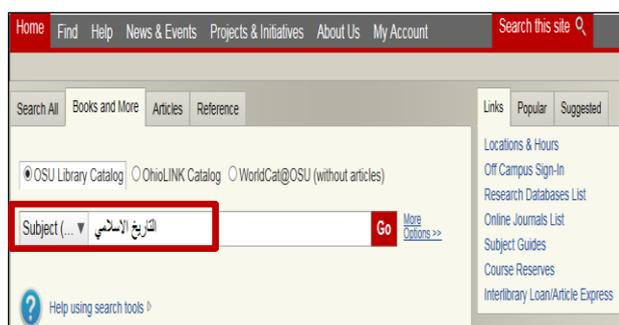


Figure 8: Type the subject query in Arabic.

The search is about Islamic History “التاريخ الإسلامي”. The search results in eleven hits about the history of Islam in the OSU online catalogue (figure 9).

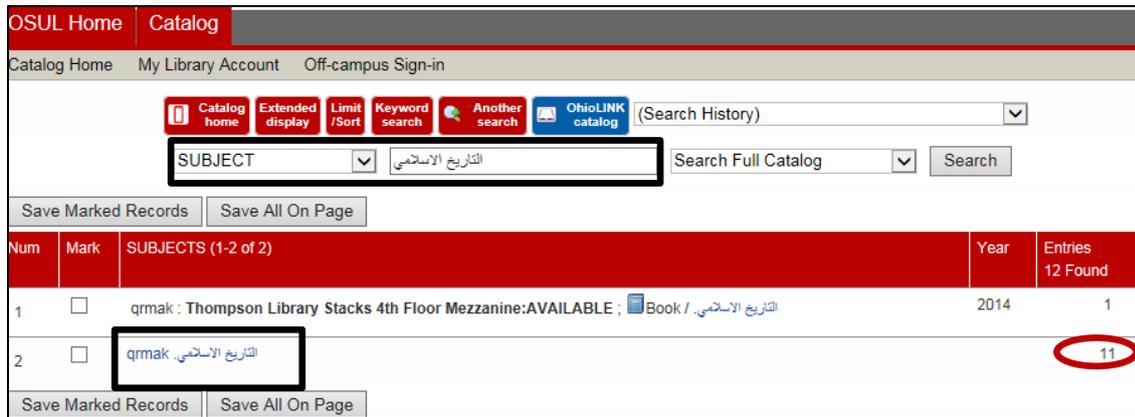


Figure 9: Result of Searching by “التاريخ الإسلامي”

When the subject term is selected, it will display the eleven hits. Figure 10 shows the first record in the list.

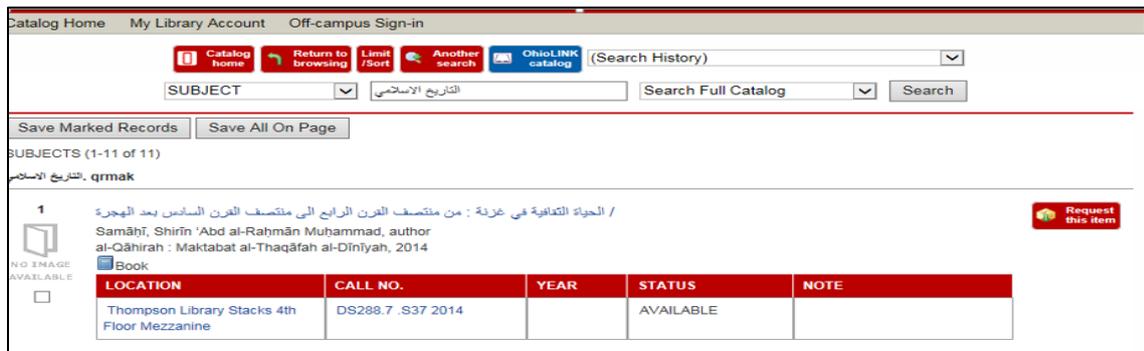


Figure 10: Sample Displaying the Records about the History of Islam

Select record number one to display the entire bibliographic record. Figure 11 shows all the elements of the record, including author, title, imprint, LCSH subject headings and other subject terms. Other subject terms show the Arabic subject terms linked to the FAST subject terms. Users can click on any of these terms to display more related titles.

SUBJECT التاريخ الإسلامي Search Full Catalog Search

Record 1 of 11
Record: [Prev](#) [Next](#)
[Permanent link to this record](#)

Author Samāhī, Shirīn 'Abd al-Rahmān Muḥammad, author
سماعي: شيرين عبد الرحمن محمد

Title al-Ḥayāh al-thaqāfiyah fī Ghaznah : min muntaṣaf al-qarn al-rābi' ilā muntaṣaf al-qarn al-sādis ba'da al-Hijrah / ta'lif Shirīn 'Abd al-Rahmān Muḥammad al-Samāhī
الحياة الثقافية في غزنة : من منتصف القرن الرابع الى منتصف القرن السادس بعد الهجرة / تأليف شيرين عبد الرحمن محمد السماحي

Imprint al-Qāhirah : Maktabat al-Thaqāfah al-Dīniyah, 2014
القاهرة : مكتبة الثقافة الدينية: 2014

LOCATION	CALL NO.	YEAR	STATUS	NOTE
Thompson Library Stacks 4th Floor Mezzanine	DS288.7 .S37 2014		AVAILABLE	
Description	630 pages : illustrations, maps ; 24 cm			
	text			
	unmediated			Other Subject
	volume			
Notes	Includes bibliographical references (pages 514-582)			
ISBN	9789773416140			
	9773416143			
OCLC #	918761769			
Author	Samāhī, Shirīn 'Abd al-Rahmān Muḥammad, author سماعي: شيرين عبد الرحمن محمد			
Subject	Ghaznī (Afghanistan) -- History Ghaznī (Afghanistan) -- Civilization Ghaznī (Afghanistan) -- Intellectual life		Genre	<ul style="list-style-type: none"> Ghaznevids -- History Ghaznevids -- Intellectual life Islamic Empire -- History Civilization. الحضارة Ghaznevids. الدولة الغزنوية. Intellectual life. لحياة الفكرية. Afghanistan -- Ghaznī. غزني (افغانستان) Islamic Empire. التاريخ الإسلامي History. تاريخ

Arabic Terms linked to FAST

Figure 11: Displaying the First Record.

As of now, there doesn't seem to be a single authorized thesaurus for Arabic subject terms that can be used universally, but this is where the opportunity presents itself. Development of a standardized, authorized Arabic subject database that could be used for the purpose of cataloguing by all libraries creates an opportunity to merge the data from Arabic and English thesauri.

Conclusion:

Linking fields is an established practice in the descriptive component of the bibliographic record. Now this can be extended to the subject component of the record. This can be accomplished by applying the Library of Congress MARC 21 Model A in the bibliographic record.

Recent developments in RDA and FAST along with the growing diversity of user populations provided incentive to internationalize the library catalogue. The added benefit is the global information sharing aspect which is now achievable through the use of multi-script subject terms. Successful implementation of this strategy could serve as a model to be applied to other non-Roman script languages, such as Chinese and Korean.

References

ALCTS (Association for Library Collection and Technical Services). 2009. "Steering Committee to Oversee the Implementation of the recommendations contained in the Report of the ACLTS Task Force on Non-English Access 2007-2009."

<http://www.ala.org/alcts/ianda/nonenglish>

ALCTS (Association for Library Collection and Technical Services). (2007). "Task Force on Non-English Access Report."

<http://www.ala.org/alcts/sites/ala.org.alcts/files/content/ianda/nonenglish/07marchrpt.pdf>

ALCTS (Association for Library Collection and Technical Services). (2009). "Non-English Access Working Group on Romanization Report."

<http://www.ala.org/alcts/sites/ala.org.alcts/files/content/ianda/nonenglish/apd15a.pdf>

Cambridge Dictionaries Online,

<http://dictionary.cambridge.org/us/dictionary/english/multilingual>

El-Sherbini, Magda, and Sherab Chen. (2001). "An assessment of the need to provide non-Roman subject access to the library online catalog." *Cataloging & Classification Quarterly* 49.6 457-483.

El-Sherbini, Magda. (2015). Multilingual Subject Retrieval: Bibliotheca Alexandrina's Subject Authority File and Linked Subject Data. In *Data Science, Learning by Latent Structures, and Knowledge Discovery* (pp.535-546). Springer Berlin Heidelberg.

Weheba, Iman. (2013). Multilingual Access to Arabic Digital Resources Using Linked Data Mechanisms and Challenges.

[http://www.mela.us/past_meetings/MELA_2012_Iman_Khairy_pres%20\(2\).pdf](http://www.mela.us/past_meetings/MELA_2012_Iman_Khairy_pres%20(2).pdf)

O'Neill, E. T., Childress, E., Dean, R., Kammerer, K., Vizine-Goetz, D., Chan, L. M., & El-Hoshy, L. (2001). [FAST: Faceted Application of Subject Terminology](#)

RDA Frequently Asked Questions. http://www.rda-rsc.org/content/rda_FAQ