

Adding Arabic script to old records

Leiden University Libraries

Tijmen Baarda | MELCOM 2019, Naples

20 June 2019



Universiteit
Leiden

The history: four(+) ways of representing Arabic

1. < 2015: PICA transcription only (>80%, 40,000 records)
2. 2015 and after: ALA-LC + Arabic script
3. Almanhal (PCI): Arabic script only
4. Legacy transcriptions

BOOK

Ṭalʿat al-šams: šarḥ šams al-uṣūl. al-Ġuz' 1.

Sālimī, ʿAbd Allāh b. Ḥumayd b. Sulūm al-, 1867-1913; Qiyyām, ʿUmar Ḥasan al-Wilāyat Badyah : Maktabat al-Imām al-Sālimī

2010

Ṭalʿat al-šams: šarḥ šams al-uṣūl

Available (more than one location) >

BOOK

الشعر الحديث في المغرب العربي / al-Shiʿr al-ḥadīth fī al-Maghrib al-ʿArabī

Nāwirī, Yūsuf ; ناوري، يوسف.

الدار البيضاء : دار توبقال للنشر ; al-Dār al-Bayḍāʿ : Dār Tūbqāl lil-Nashr;

2006-

Available (more than one location) >



BOOK

Kitab Boestân as-Salâtîn

Nūr al-Dīn al-Rānīrī, Muḥammad b. ʿAlī, -1658

1899

Available (more than one location) >



ARTICLE

المعرفة و مستقبل المغرب الكبير

فريعة، أحمد

Middle East Affairs, 04 January 2014, Vol.24(148), pp.24-30

Online access >

The history: four(+) ways of representing Arabic

1. < 2015: PICA transcription only (>80%)
2. 2015 and after: ALA-LC + Arabic script
3. Almanhal (PCI): Arabic script only
4. Legacy transcriptions

Problems:

- You always have to search at least twice (PICA and ALA-LC or PICA and Arabic script)
- You have to know the correct vocalization AND the catalogue has to be correct
- Primo does not tolerate any typos

The proposed solution: Arabic script

Solution: make sure that all records have at least Arabic script

1. < 2015: PICA transcription only (>80%) -> **add Arabic script**
2. 2015 and after: ALA-LC + Arabic script -> **do nothing**
3. Almanhal (PCI): Arabic script only -> **do nothing**
4. Legacy transcriptions -> **add Arabic script (manually)**

The proposed solution: Arabic script

Solution: make sure that all records have at least Arabic script

Advantages over correcting the transcription:

- Only the original script is obligatory according to RDA
- Users don't need knowledge of the transcription system
- Users don't need to know the correct vocalisations
- Some records will never have transcription (Almanhal, DOAB)
- Possibility to leave out transcription in the future

Is automatic conversion possible?


- Our transcription is precise and of high quality, but there are ambiguities
- The operation is only possible if the great majority of the records does not have to be checked and corrected manually

BOOK

Ṭalʿat al-šams: šarḥ šams al-uṣūl. al-Ġuz' 1.

Sālimī, ʿAbd Allāh b. Ḥumayd b. Sulūm al-, 1867-1913; Qiyyām, ʿUmar Ḥasan al-Wilāyat Badyah : Maktabat al-Imām al-Sālimī
2010

Ṭalʿat al-šams: šarḥ šams al-uṣūl

 Available (more than one location) >

BOOK

al-Ṭabaqāt al-kubrā

Ibn Saʿd, Muḥammad, 148/764-230/845; Aṭā, Muḥammad ʿAbd al-Kāliq ʿAbd Muqallid, Yaḥyā
Bayrūt : Dār al-Kutub al-ʿIlmiyya
1410-1411/1990-1991

 Available >

Problems

Some problems cannot be overcome:

- Some special ambiguities
 - Example: *qatala* may be قتل “he killed” or قتلة “killers”. The pattern gives no answer to this question.
- Transcription errors
 - حسن incorrectly transcribed as *Hasan* instead of *Ḥasan* will yield هسن
 - On the other hand: vocalisation errors may magically disappear. Example: حنين, incorrectly transcribed as *Ḥanīn* instead of *Ḥunayn*, will automatically be corrected in Arabic script

arabic.csv - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation Sans 10

A	B	C	D
463	Hikmat al-Kaldāniyyīn /	حكمة الكلدانيين	OK
464	al-Ši'r wa-mutağayyirāt al-marḥala /	الشعر ومتغيرات المرحلة	OK
465	al-°Urwa al-wuṭqā fimā ta°umm bihi al-balwā /	العروة الوثقا فيما تعم به البلوا	OK
466	Sabīl al-mu°minīn li-taḥṣīl al-zād li-yawm al-Dīn /	سبيل الممنين لتحصيل الزاد ليوم الدين	OK
467	Kamāl al-balāga : rasā°il /	كمال البلاغة رسائل	OK
468	Nuzhat al-muštāq fī tāriḳ Yahūd al-°Irāq /	نزهة المشتاق في تاريخ يهود العراق	OK
469	al-Yazīdiyya aw-°abadat al-šayṭān /	اليزيدية اوعبدة الشيطان	OK
470	Aḡlāt al-luğawiyīn al-aqdamīn /	اغلاط اللغويين الاقدمين	OK
471	al-Gāmi° al-muḳtašar fī °unwān al-tawārīḳ wa-°uyūn al-siyar /	الجامع المختصر في عنوان التواريخ وعيون السير	OK
472	al-Manhağ al-qaḥmī al-°Arabī /	المنهج القومي العربي	OK
473	Dawr al-ğayš al-°Irāqī fī ḥukūmat al-difā° al-waṭanī wa-al-ḥarb ma°a Bariṭāniyyā,	دور الجيش العراقي في حكومة الدفاع الوطني والحرب مع بريطانيا دراسة تحليلية	OK
474	Kitāb aṣl al-Šī°a wa-ušūlihā /	كتاب اصل الشيعة واصولها	OK
475	Aqsām ḍā°i°a min Kitāb Tuḥfat al-umarā° fī tāriḳ al-wuzarā° /	اقسام ضائعة من كتاب تحفة الامراء في تاريخ الوزراء	OK
476	Mawsū°a sirriyya ḳāšša bi-al-ḥizb al-šuyū°ī al-°irāqī al-sirrī /	موسوعة سرية خاصة بالحزب الشيوعي العراقي السري	OK
477	al-Wasā°il ilā musāmarat al-awā°il /	الوسائل الى مسامرة الاوائل	OK

arabic

Find

Sheet 1 of 1

Default Dutch (Netherlands) Average: ; Sum: 0 140%

trans2arabic — Dolphin Taking Screenshots - KDE UserBase Wiki - ... maqsura — Kate Naamloos — KolourPaint trans2arabic : bash — Konsole arabic.csv - LibreOffice Calc

09:10

Other languages: Persian, Hebrew...?

- Leiden has the same problem for Persian and Hebrew
- For Persian, the old transcription system is similar and brings the same challenges
- For Hebrew, automatic addition of Hebrew script is impossible, because *matres lectionis* are not indicated



BOOK

Zabān-i Fārsī dar dānišgāh-i Istrāsbūrg

Iqdāmāt-i Mu'asasa-yi Zabān wa Tamadun-i Īrān dar Dānišgāh-i Istrāsbūrg (Farānsa) (Strasbourg).
[Strasbourg] : Iqdāmāt-i Mu'asasa-yi Zabān wa Tamadun-i Īrān dar Dānišgāh-i Istrāsbūrg (Farānsa)
1344/1966

 Available >



BOOK

Sefer ha-'aggadā : mivḥar ha-'aggadōt še-ba-Talmūd ū-va-midrašim

Rawnitski, Y.H., 1859-1944; Bialik, Chaim Nachman, 1873-1934
Tel-'Aviv
5696 [1936]

 Available >

And after?

- What is the future of transcription for non-Latin scripts? Who needs it?
 - Should we continue with transcription after this operation?
 - If not, can we at least skip transcription for our backlog?