# **Bridging the Gap:**

## **Digital Humanities and the Arabic-Islamic Corpus**

Dafne van Kuppevelt, Patrick Bos, Melle Lyklema, Christian Lange, Umar Ryad, Janneke van der Zwaan

### Introduction

Despite some pioneering efforts in recent times, the longue durée analysis of conceptual history in the Islamic world remains largely unexplored. Researchers of Islamic intellectual history still tend to study a certain canon of texts, made available by previous Western researchers of the Islamic world. Indigenous conceptual developments and innovations are therefore insufficiently understood, particularly concerning the transition from premodern to modern thought in Islam.

What, then, are the silenced continuities, transformations and major fault lines in Arabic-Islamic discourses? The Islamic tradition offers a vast textual corpus for exploring this question from a longue durée perspective, but its very breadth poses substantial problems for the individual scholar seeking to survey the literature by traditional methods. This poster presents ongoing research to use state-of-the art Digital Humanities approaches and technologies to make pioneering forays into the vast corpus of digitized Arabic.

#### The Arabic Digital Corpus

Digitized materials

- ca. 7,900 unique titles
- ~1,1 billion words (vs. 150 million in classical Greek and Latin combined)
- ~4,000 unique titles pre-1900  $\bullet$
- <10% of what may have been written  $\bullet$



#### **Digital solution**

Combine and adapt existing tools: **nlppln** [1] for creating flexible pipelines for corpus processing, SAFAR [2] and Al-Khalil [3] for morphological analysis, **Blacklab** [4] as a search engine.

## Try the demo!

http://arabic-dh.hum.uu.nl/corpus-frontend

Per Hit Per Document			Total hits: 2553
			Total pages: 128
Course has Courts and a sum data			
Group by Century • update			Show/hide titles Export CSV
_			
Prev 1 Next			
Group	Hits		
9th/15th century		8624	
4th/10th century	8204		
5th/11th century	7531		
7th/13th century	7028		
2ndH/8thC and 3rdH/9thC century	5927		
10th/16th century	5753		
8th/14th century	4800		
6th/12th century	2553		
	« View detailed concordances in this group - Load more concordances		
	الطوابة الكلام فمرهذا الكتاب	کتاب	
	اللغيارة التعدير في هذا التعاب	الکتاب	كتاب الطفاية الكلام فمرهذا
	والثانية .	Ų LLDI	
	والمحول التحقيبة بصير مفسرا س	الكتاب	ببأدار عليهما فعله يبانا لمحمل
	بالقلم وضربت بالسيف بكتب بكار	کتیت	بدي بالمنديا . لم يكله ويقال
	بالتلم بضرب السيف فيتناول	ىكتب	وبقال كتبت بالقلم وخربت بالسيف
	اذ البيان يكون بالقول تارة	الكتاب	فعله الصلاة والسلام بيانا لمحمل
	غسا . الرحلين الحديث فيحتما . أنهما	بالكتاب	ليسه جاجة فيقي أصل الواجب
	الاختلاف الاجماع الاقتصار أسفا وكذا	کتاب	نجاسة وجكو ، ابراهيم بن جابر
	الصلاة ترك وذلك أحزأه أبو	کتاب	واحدة ومسح ها. فقد ذكر
	تيمم معه أحب الي وروي	کتاب	ووجد نبيذ توضأ ولم يتيمم
	وهؤلاء طعنوا في حديث عبد	الکتاب	النبيذ الي التراب فقد خالف
	الصوم ابتداء الوضوء فمنها النبة	کتاب	نصوص مطلقة بكره بعد للصائم
	بالمضوء لقوله آخر آية ولكن	الكتاب	النهي ينتهي اغتسال مقرون بالنبة
	بغريونيو غير بي ويني	الكتاب	والحديث حملة الآجاد تقييد مطلق
	وهو أنه انما فعا ، ذلك	الكتاب	بدلیا . وفعا . بمکن بحما . موافقة
	الآثار لأبي وشيك أصابعه لحيته	کتاب	الآداب وعند يوسف هكذا محمد
	بنا سنة نبينا يقول امرأة	کتاب	الحكم فشاور بقي فقالوا ندع
	منه فيستوى الكتابة وعلى كتاب	کتب	الدراهم التي عليها جرمة كجرمة
	وعلى كتاب التفسير يصير بمسه	الكتابة	حرمة كحرمة كتب منه فيستوى
	التفسير يصبر بمسه ماسا للقرآن	كتاب	كتب منه فيستوي الكتابة وعلى

## **Challenges Arabic digital humanities**

Many existing tools for digital humanities are **not** adapted for non-western languages. The right-to-left text direction is problematic for many web applications

Arabic is a **morphologically complex** language. 3-letter roots are combined with different prefixes, infixes and suffixes to construct semantically related words. Thus, searching on roots is valuable for research on concepts.



#### References

[1] Janneke van der Zwaan, & Dafne van Kuppevelt. nlppln http://doi.org/10.5281/zenodo.1185197 [2] Jaafar, Y. and Bouzoubaa, K. (2015) 'Arabic Natural Language Processing from Software Engineering to Complex Pipeline', in 2015 First International Conference on Arabic Computational Linguistics (ACLing). IEEE, pp. 29–36 [3] Boudchiche, M., Mazroui, A., Ould Abdallahi Ould Bebah, M., Lakhouaja, A. and Boudlal, A. (2017) 'AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer', Journal of King Saud University - Computer and Information Sciences. Elsevier, 29(2), pp. 141–146. [4] http://inl.github.io/BlackLab/

- Root search
- View local context of search results
- Aggregate results on metadata (e.g., per century)

## **Citation analysis**

- Which Quran and Hadith verses are quoted in the corpus?
- Do we see patterns in the citations? E.g., schools of thought that cite similar sources



Visualization made with Gephi – Sigma js plugin

## Want to learn more?

Contact us! d.vankuppevelt@esciencecenter.nl



*aSci* 

scie

scie



