MIDDLE EAST STUDIES | BROWN

Distant Reading and the Islamic Archive
Friday, October 16, 2015 | Brown University

Each year, the number of digitized books, inscriptions, images, documents, and other artifacts from the Islamic world continues to grow. As this archive expands, so too does the repertoire of digital tools for navigating and interpreting its diffuse and varied contents. Drawing upon such tools as topic modeling, context-based search, social network maps, and text reuse algorithms, the study of large-scale archives and textual corpora is undergoing significant and exciting developments.

The Middle East Studies program at Brown University is pleased to announce its third annual Digital Islamic Humanities Conference, to be held at the Joukowsky Forum on Friday, October 16, 2015. Papers shall be no longer than twenty minutes and read in English. A collection of abstracts from previous events may be found on our website (islamichumanities.org) along with recorded webcasts, a list of digital resources, and announcements for related events.

To register for the conference as a non-speaker, please contact Prof. Elias Muhanna at digitalhumanities@brown.edu. Please be advised that space is limited.

**

Program

All panels will be held at the Joukowsky Forum in the Watson Institute for International and Public Affairs, at 111 Thayer St.

8:45-9:00   *Registration*

9:00-9:15:   *Welcoming remarks*

9:15-10:30   DAVID VISHANOFF, "A Customizable Exaptive "Xap" for Charting Currents of Islamic Discourse across Multiple Bibliographic and Full Text Datasets"

                PETER VERKINDEREN, JOSÉ ANTONIO HARO PERALTA, AND HANNAH-LENA HAGEMANN, "Which Muḥammad? Computer-Based Tools for the Identification of Moving Elites in the Early Islamic Empire"

10:30-10:45   *Coffee break*

10:45-12:00   ALEXANDER MAGIDOW & YONATAN BELINKOV, "Digital Philology and the History of Written Arabic"

                ELIAS MUHANNA, "Modeling Mannerism in Classical Arabic Poetry"

12:00-1:00:   *Lunch*

1:00-2:45   MAXIM ROMANOV, "*al-Ḏahabī's Monster*: Dissecting a 50-Volume Arabic Chronicle-cum-Biographical Collection From the 14th Century CE"

SEYED MOHAMMAD BAGHER SAJADI & MOHAMMAD SADEGH RASOOLI, "Automatic Proper Names Extraction from Old Islamic Literature"

KAREN PINTO, "MIME and Other Digital Experimentations with Medieval Islamic Maps"

2:45-3:00   *Coffee break*

3:00-4:45   NIR SHAFIR, "Distant Reading the Material and Bibliographic Record of the Early Modern Islamic Archive"

ERIC VAN LIT, "A Digital Approach for Production and Transmission of Knowledge in Islamic Intellectual History"

TAIMOOR SHAHID, "Mobile Ethics: Travel and Cosmopolitanism in the Islamic Archive"

4:45-5:45   *Roundtable discussion*

6:00-8:00   All speakers are invited to a group dinner at the Brown University Faculty Club.

Paper Abstracts

Peter Verkinderen, José Antonio Haro Peralta, and Hannah-Lena Hagemann (Universität Hamburg): Which Muḥammad? Computer-Based Tools for the Identification of Moving Elites in the Early Islamic Empire

Our ERC project at Hamburg University on "The Early Islamic Empire at Work" looks at five key provinces (Fārs, Ifrīqiya, Jazīra, Khurāsān, and Syria) and investigates their geographical, economic, social and political make-up, as well as their relationship to the imperial center. Within this framework, we are particularly interested in studying the networks and dynamics of imperial and local elites. In order to conquer the huge number of sources available for the early Islamic period, we have begun to develop a series of text-mining tools aimed at harvesting relevant data in an efficient way. In this paper, we will showcase one such tool that allows us first to identify individuals belonging to specific groups, such as governors and *qāḍīs*. In a second step, it enables us to retrieve information about them from a large corpus of texts from different genres, like biographical dictionaries or chronicles. Using this information, we can analyse and visualise the geographical, social, religious or political networks these individuals were a part of. This will help us to answer questions such as whether particular families were predominant in certain offices or regions, or what factors influenced the mobility of members of the elite across the different provinces of the empire. For the purpose of this presentation, we will look at the *qāḍīs* of a specific province as a case study of how such tools can facilitate the distant reading of Arabic sources.

**

Alexander Magidow (Univ. of Rhode Island) and Yonatan Belinkov (MIT): Digital Philology and the History of Written Arabic

The availability of large collections of Arabic texts, primarily those from the classical Arabic literary and religious canon, allow for new avenues of automated and semi-automated inquiry into the history of the Arabic language. At the same time, the tools and methods to undertake this inquiry are still in development. This talk uses the question of how written Arabic can be periodized to explore the methods and tools of conducting digital philology. This research draws on a large diachronic collection of texts drawn primarily from the Maktaba Shamela text collection, totaling approximately 800 million words and covering fourteen centuries of texts. We will discuss the tools we have used to explore this question, such as lemmatization of the corpus, measures of word-lifetime, and the use of word-vectors to subtly differentiate words based on their contexts and measure changes over time. We will also discuss the challenges of eliminating redundant texts, since many of the texts in this collection quote earlier sources that are also in our collection, obscuring the changes over time.

   Preliminary evidence supports the common distinction between 'Classical' and 'Modern Standard' Arabic, but we hope to discover whether we can support further periodization. Preliminary results also suggest there is a constant rate of change in the lexicon of these texts, so that while prior to the 19th century we have not yet seen significant 'jumps' in this rate of change that are sufficient to suggest entirely new 'periods' of the language, *fusha* Arabic is less immutable than is commonly believed.

ELIAS MUHANNA (BROWN UNIVERSITY): Modeling Mannerism in Classical Arabic Poetry

As a highly patterned literary form, classical Arabic poetry represents a discourse that is particularly well suited to digital modes of analysis. The conventions of rhyme, meter, and genre conspire, together with a Semitic root-based morphology, to natively configure a literary archive in ways that invite distant reading. This talk will discuss the results of a research project that aims to map various literary embellishments associated with the so-called "innovative" (*badīʿ*) style of classical Arabic poetry, across a corpus of several thousand poems. We would like to suggest that this method opens up interesting avenues to revisit some old questions about nature and artifice, originality and derivativeness, clarity and mannerism in Arabic poetic production.

<center>**</center>

KAREN PINTO (BOISE STATE UNIVERSITY): MIME and Other Digital Experimentations with Medieval Islamic Maps

I have spent the last two decades studying medieval and early modern Islamic maps in depth. These studies involved extensive on-site visits to manuscript libraries through which I collected thousands of images of medieval Islamic maps ranging in date from the eleventh century to the nineteenth. The sheer number of these extant maps tells us, at least from the thirteenth century onwards when copies of these map-manuscripts begin to proliferate, that the world was a much-depicted place. It loomed large in the medieval Muslim imagination. It was pondered, discussed, and copied with minor and major variations again and again, and all with what seems to be a peculiar idiosyncrasy to modern eyes: the cartographers did not strive for mimesis (imitation of the real world). They did not show irregular coastlines even though some of the geographers whose work includes these maps openly acknowledge that the landmasses and their coastlines are uneven. They present instead a deliberately schematic layout of the world and the regions that comprised the Islamic empire that can be best described as "carto-ideographs."

Nine years ago thanks to a small grant from the American University of Beirut, I started a digital project called MIME—Medieval Islamic Maps Encyclopedia—to place these maps on an interactive CD and online web-based format in order to make this rich resource available to scholars, students, and the general public. The mainstay of MIME are the maps found in the manuscripts of al-Iṣṭakhrī, Ibn Ḥawqal, and al-Muqaddasī—also known as the "Islamic Atlas." The aim of MIME is to decode the place and space matrix on these carto-ideographs so that anyone, with or without Arabic, can browse them and understand how medieval Muslim cartographic artists and their patrons perceived their world.

I was able to get a pilot version of MIME started but wars, evacuations, new jobs, teaching duties, and the lack of funding prevented me from fully realizing my dream. A grant from the National Endowment for the Humanities (taken 2014-15) to work on my book project on the Mediterranean in the Islamic cartographic imagination has given me an opportunity to take MIME and other digital mapping projects related to the Mediterranean maps off the shelf. It is this digital work in progress that I seek to present at the next Digital Humanities conference at Brown.

Maxim Romanov (Universität Leipzig): *al-Ḏahabī's Monster*: Dissecting a 50-Volume Arabic Chronicle-cum-Biographical Collection From the 14th Century CE

Chronicles and biographical collections are two major genres of Arabic historical writing. Most specimens of these two forms, which often smoothly crossover into each other, are multivolume titles that aggregate and copiously reuse earlier sources, and, in their turn, get reused in later sources. Perhaps the largest specimen of these two genres is a 50-volume title, the *Taʾrīḫ al-islām* of al-Ḏahabī (d. 748/1347 CE). Dubbed "one of the most ambitious histories of the entire world of Islam," this library of a book covers 700 years of Islamic history through over 30,000 biographical records. Although nobody has ever doubted that this "History" is a compilation, we hardly have any understanding of the composition of this text: What earlier sources did al-Ḏahabī use? How and to what extent did he use them? Did he paraphrase, summarize or quote them? How is his "History" related to his other writings? What kind of understanding did al-Ḏahabī have of the historical information that he collected? How, in its turn, was al-Ḏahabī's "History" used by later biographers and chroniclers?

None of these questions can be answered convincingly with traditional methods of historical inquiry. Modern computational techniques of text analysis, however, offer a new perspective on the problem. Using an approach that combines computational algorithms for tracing similarities among texts with the use of high-power computing, my paper seeks to answer some of the above-stated questions. This approach allows to build an "x-ray image" of al-Ḏahabī's "History" that shows where exactly and to what extent earlier sources feature in this collection. Having such an "x-ray" should shed light on al-Ḏahabī's method of compilation and help us better understand his "History" in general. My preliminary results show that al-Ḏahabī patched his book together from large, practically unmodified chunks of earlier sources. Taking this approach a step further, I hope to establish connections between his "History" and his multiple writings, with my current hypothesis being that al-Ḏahabī composed his "History" as a databank for writing his other, more thematically and ideologically oriented collections. Last but not least, I will dwell on my distant reading of al-Ḏahabī's "History" in comparison with his uncharacteristically brief epistle *al-Amṣār ḏawāt al-aṯār* ("Cities and Ports for Hearing the Reports"), which shows that al-Ḏahabī had a rather solid grasp of the data that he collected.

The research has been conducted with the support of and in collaboration with David Smith (Northeastern University, Computer Science), Sarah Savant (Agha Khan University, London) and the Perseus Digital Library Project.

**

Seyed Mohammad Bagher Sajadi (Qazvin Islamic Azad University) and Mohammad Sadegh Rasooli (Columbia University): Automatic Proper Names Extraction from Old Islamic Literature

Named Entity Recognition (NER) in religious literature, specifically identifying proper names of Hadith narrators, is of great importance in Islamic studies. This significance is due to the role of narrators in Hadith validation, historical investigation of the religious literature, investigation of similarities in old religious books and their categorization. In this research a NER system, designed for extraction of the person names in old Arabic text, is introduced. A supervised learning method was used and an annotated corpus based on a historical-Islamic

book with 200K words prepared. In addition, in order to raise the performance, a rich gazetteer including special name entities focused on Hadith narrators was built using diverse Hadith sources. The recognition process, based on ensemble learning, used Boosting method by Adaboost algorithm for implementation. There are some challenges in Arabic language like lack of capital letters and morphological complexity. We tried to overcome these difficulties by tokenizing and POS tagging as pre-processing step. The model obtained an overall F-measure of 95.11% for more than 6000 persons. Finally the same model was trained and evaluated on a modern standard Arabic achieved F-measure value of 73.81%. Results show that the pattern of person names in old Arabic text is simpler than modern standard Arabic. (This paper will be presented by Mr. Rasooli on behalf of Mr. Sajadi and two other collaborators, Behrouz Minayi and Seyed Peyman Shariatpanahi).

**

NIR SHAFIR (UCLA): Distant Reading the Material and Bibliographic Record of the Early Modern Islamic Archive

Distant reading techniques are almost always directed at the body of the text itself. As valuable as such approaches can be, they often overlook one of the most valuable and distinct qualities of the early modern Islamic manuscript archive—the continued existence of millions of physical manuscripts. In this material format are two important components—1) the material aspects of the manuscript itself and 2) the relation between discrete texts grouped together in the miscellany format. This presentation suggests that analyzing the material and bibliographic record of early modern manuscripts en masse can be a form of distant reading and demonstrates how such a reading can change our understanding of various sources, especially when answering questions of reception and readership. The approach in question has been facilitated by the mass digitization of Islamic manuscripts which allows for the perusal of tens, if not hundreds, of manuscripts in a day. The presentation focuses in particular on using the miscellany structure of many early modern Islamic manuscripts to uncover a variety of larger textual connections and does so through an examination of Katib Çelebi's *The Balance of Truth* (*Mizānü'l-Ḥaḳḳ*) and other texts. It concludes by considering how the distant reading of the material record might be computationalized in the future.

**

TAIMOOR SHAHID (UNIVERSITY OF CHICAGO): Mobile Ethics: Travel and Cosmopolitanism in the Islamic Archive

This paper examines the archive of an epic romance tale called *Saif al-Mulūk*. An Egyptian-Yemeni prince's journey in search of his beloved fairy princess, *Saif al-Mulūk* has been composed in the Muslim world numerous times from the 14th to the 20th century, in both poem and prose. It is extant in disparate sociopolitical contexts (courtly, devotional, and urban) in the numerous languages found in the 7000 miles between Tunisia and Indonesia: in Arabic, Awadhi, Balochi, in Bengali, Chagatai, Dakani, in Kurdish, Malay, Persian, in Punjabi, Pushto, Seraiki, in Sindhi, Turkish, Urdu, and in Uyghur, Uzbek, and others.

Impossible for anyone to read all renditions of this text, in this paper I provide a method for interpreting this and other similar archives through a distant reading of all

renditions of *Saif al-Mulūk*. I abstract the tale to its narrative morphology—a narrative that is shared amongst all renditions inasmuch as a deviation would transform the tale into something else—and look at a *trope* structurally fundamental to all its retellings, i.e. of *travel and marriage with the Other*. A Graphical Information System (GIS) mapping of these travels within the narrative vis-à-vis the historical travels of the epic itself reveal a strange isomorphism as seen in a preliminary map below. This map is also isomorphic to historical Indian Ocean trade routes. I read this isomorphism of the distilled structural *trope* of travel with actual travels of the text and the trade routes (as revealed by the GIS) to ask why everyone is writing the *same* tale. The answer, I suggest, lies in a shared *imaginaire* of a *form*-al cosmopolitanism. *Saif al-Mulūk*, I argue, narrates an ethical comportment of social cosmopolitanism in this world of ready mobility: of merchants, courtiers, laborers, adventures, literati, and others. While the seeming limits of this cosmopolitanism are the limits of Muslim polities (re: map), a close reading of a Persian rendition commissioned by a Dutch merchant in Vengurla, Goa in 1660 reveals an even larger cosmopolitan order. Such a project—enabled only by distant reading and GIS mapping—provides new ways of concretizing the 'Islamic archive' as a discursive category, and complicates our understanding of disciplinary and area studies boundaries.

<center>**</center>

Eric van Lit (Yale University): A Digital Approach for Production and Transmission of Knowledge in Islamic Intellectual History

In this paper I provide a new theoretical framework and a practical toolkit for the study of Islamic intellectual history. Its starting point is to consider what would happen if we do not study all texts of one author, but instead all authors of one text. The success of this approach depends on including a large number of texts, which has become possible due to the digitization of an ever-larger number of texts.

This new approach allows us to break away from the focus on authorial intent, which to me is what has hampered the study of late-medieval and early-modern Islamic intellectual history. It has been standard practice to collect, read, and then paraphrase and analyze all texts of only one intellectual. But for this period, this approach often leads to the conclusion that the intellectual is 'unoriginal'. Even worse, if not unoriginal, scholars often exaggerate the originality by failing to notice influence from earlier intellectuals.

Intellectuals from this period freely copied each others works, often without acknowledgment. We ought to adapt our approach to this, rather than dismiss it. My approach thrives on unoriginality and repetitiveness, perceiving the reuse of text as a material signifier for the transmission of knowledge. Thus we are able to draw up genealogies of ideas, conceiving of Islamic intellectual discourse as a *longue durée* process of evolution rather than revolution, thereby also showing the relevance of the classical discourse for the contemporary era.

The required toolkit entails a mixed strategy of automated searches and manual labor. The former allows stringing together intellectuals in unexpected ways, making us see connections we would not have expected otherwise. The latter lays bare how Islamic intellectuals concretely go about their work. Together they allow us to proficiently study late-medieval Islamic intellectual history.

**

DAVID VISHANOFF (UNIVERSITY OF OKLAHOMA): A Customizable Exaptive "Xap" for Charting Currents of Islamic Discourse across Multiple Bibliographic and Full Text Datasets

This presentation will demonstrate a prototype of a software tool being developed by Exaptive, Inc. and the University of Oklahoma using my research on Qur'anic hermeneutics as a pilot project. This "Xap" integrates multiple data analysis and visualization tools and applies them to multiple full text and bibliographic datasets (e.g., al-Maktaba al-Shamila, the Bibliotheca Alexandrina's Digital Assets Repository, the Hathi Trust Digital Library, Google Scholar, and the researcher's own notes) to create a customizable research environment in which the distant reading of large numbers of texts assists in formulating new research projects and identifying the most promising strategies for the slow work of close reading. Starting from a few key terms such as "hermeneutics" and "Qur'an," the software will employ strategies such as topic modeling to map clusters of texts and authors that employ those terms in combination with distinctive clusters of additional concepts such as *tafsīr,* history, method, *i'jāz,* or others that the researcher might not have anticipated. It will allow the researcher to visualize those discourses regionally and chronologically, and to isolate particular authors whose works are frequently referenced by later writers, authors who straddle discourses in unique ways, or works that appear to have spawned new discourses. Given the limitations of available databases, the resulting visualizations will be suggestive rather than definitive, but they have the potential to reveal new connections between discourses, bring to light unnoticed thinkers, and prompt new research directions to which the existing scholarly literature might be blinded by its own self-referential tendency to study small canons of major texts and authors within well defined genres of Islamic literature. This software is being developed now, and by the time of the conference a second or third prototype will be ready to demonstrate using my own research questions as well as those of other conference participants.

**

Please contact Elias Muhanna (digitalhumanities@brown.edu) with any questions regarding this program.