Of graphs, maps, and 30,000 Muslims
Digital Humanities

Engaging in a kind of humanistic research that cannot be done without digital methods, where complexity and novelty of each research task would vary from impossible ‘without 500 monks at hand’* to otherwise unthinkable. Such research would rely on large volumes of data (e.g., corpora of full-text primary sources) and a variety of digital approaches that allow converting raw data into meaningful information and then exploring it with different visualization techniques in order to trace long-term and large-scale developments.

Method: *Distant Reading*

‘Distant reading’, I have once called this type of approach, where distance is however not an obstacle, but a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models.


[http://litlab.stanford.edu/] (Pamphlets)
Method: Algorithmic Deformation

“Algorithmic criticism is easily conceived as the form of engagement that results when imperative routines are inserted into the wider constellation of texts stipulated by critical reading. But it is also to be understood as the creation of interactive programs in which readers are forced to contend not only with deformed texts, *but with the ‘how’ of those deformations.*”


http://stephenramsay.us/
Method: Micro-, Meso-, and Macroanalysis

http://www.matthewjockers.net/
How is something like that possible in our field?
## Digital Collections of Arabic Texts

<table>
<thead>
<tr>
<th>Library</th>
<th>Media</th>
<th>Titles</th>
<th>Vols</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>al- nóng al-kabīr</td>
<td>HDD, Windows 95</td>
<td>2,400</td>
<td>5,550</td>
<td>~400 mln</td>
</tr>
<tr>
<td>al-Maktaba al-shāmila</td>
<td><a href="http://www.shamela.ws">www.shamela.ws</a></td>
<td>6,300</td>
<td>—</td>
<td>~820 mln</td>
</tr>
<tr>
<td>al-Mishkāt</td>
<td><a href="http://www.almeshkat.net">www.almeshkat.net</a></td>
<td>7,300</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ṣayd al-fawāʿid</td>
<td><a href="http://www.said.net">www.said.net</a></td>
<td>10,000</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>al-Warrāq</td>
<td><a href="http://www.alwaraq.com">www.alwaraq.com</a></td>
<td>860</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>al-Muʿjam al-fiqhī</td>
<td>DVD</td>
<td>1,130</td>
<td>3,000</td>
<td>—</td>
</tr>
<tr>
<td>al-Maktaba al-shīʿīyya</td>
<td><a href="http://www.shiaonlinelibrary.com">www.shiaonlinelibrary.com</a></td>
<td>1,970</td>
<td>4,175</td>
<td>~280 mln</td>
</tr>
</tbody>
</table>
Digital Collections: Problems

❖ Thematic/subject biases ("salafi" collections): almost no scientific, philosophical, medical and other texts
❖ Spotty chronological and geographical coverage
❖ Difficult to assess their overall coverage
Digital Collections: Coverage Issues

Distribution of Digital Arabic Sources
(by titles per century CE)

al-Maktaba al-shāmila
prior 1700 CE = 3,200 titles

al-Jāmi῾ al-kabīr
prior 1700 CE = 1,984 titles

Distribution of Digital Arabic Sources
(by titles per 50 years CE)
Digital Collections: Coverage Issues

Distribution of Digital Arabic Sources
(by words per century CE)

- **al-Maktaba al-shāmila**
  - prior 1700 CE = 456 mln

- **al-Jāmi῾ al-kabīr**
  - prior 1700 CE = 334 mln

Distribution of Digital Arabic Sources
(by words per 50 years CE)

- 800
- 1000
- 1200
- 1400
- 1600
- 1800

801-1200 CE

- 90 mln
- 80 mln
- 70 mln
- 60 mln
- 50 mln
- 40 mln
- 30 mln
- 20 mln
- 10 mln

1201-1700 CE

- 134 mln
- 200 mln

al-Maktaba al-shāmila

al-Jāmi῾ al-kabīr

prior 1700 CE = 456 mln

prior 1700 CE = 334 mln
Digital Collections of Arabic Texts

Three Collections Together:

~10,500 texts (premodern & modern)
~1.5 billion words
~4,000 unique pre-1900 titles

less than 10% of what might have been written
Digital Collections of Arabic Texts

al-Maktaba al-shāmila: 5,999 titles
al-Jāmi‘ al-kabīr: 2,364 titles
al-Maktaba al-shī‘iyya: 1,810 titles

UNIQUE: 7,895 titles (~1,1 billion words)
<table>
<thead>
<tr>
<th>Arabic Genre</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>التفسير</td>
<td>Interpretation of the Qur’ān</td>
</tr>
<tr>
<td>الحديث</td>
<td>“Prophetic sayings”</td>
</tr>
<tr>
<td>أصول الفقه ومسائله</td>
<td>Legal writings</td>
</tr>
<tr>
<td>التاريخ</td>
<td>Chronicles</td>
</tr>
<tr>
<td>التراجم والطبقات</td>
<td>Biographical collections</td>
</tr>
<tr>
<td>النحو والصرف</td>
<td>Arabic language, grammar &amp; morphology</td>
</tr>
<tr>
<td>الشعر والأدب</td>
<td>Poetry &amp; fine literature</td>
</tr>
<tr>
<td>المعاجم</td>
<td>Various dictionaries &amp; references</td>
</tr>
</tbody>
</table>
## Digital Collections: *Major Genres & Forms*

<table>
<thead>
<tr>
<th>Arabic Genres</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>التفسير</td>
<td>Interpretation of the Qur’ān</td>
</tr>
<tr>
<td>الحديث</td>
<td>“Prophetic sayings”</td>
</tr>
<tr>
<td>أصول الفقه ومسائله</td>
<td>Legal writings</td>
</tr>
<tr>
<td>التاريخ</td>
<td>Chronicles</td>
</tr>
<tr>
<td>التراجم والطبقات</td>
<td>Biographical collections</td>
</tr>
<tr>
<td>النحو والصرف</td>
<td>Arabic language, grammar &amp; morphology</td>
</tr>
<tr>
<td>الشعر والأدب</td>
<td>Poetry &amp; fine literature</td>
</tr>
<tr>
<td>المعاجم</td>
<td>Various dictionaries &amp; references</td>
</tr>
<tr>
<td><strong>Digital Collections: Biographical Collections</strong></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>~300-400 have been written</td>
<td></td>
</tr>
<tr>
<td>~250 available in digital format</td>
<td></td>
</tr>
<tr>
<td><strong>structure</strong></td>
<td>chronological, generational, alphabetical</td>
</tr>
<tr>
<td><strong>coverage</strong></td>
<td>social and religious, geographical</td>
</tr>
<tr>
<td><strong>biographies</strong></td>
<td>well-structured (linguistic formulae): onomastic section, birth [date &amp; ] place, teachers, students, contribution, miscellanea, dear: date &amp; place, reaction of the community</td>
</tr>
</tbody>
</table>
| > 400,000 biographies?
Algorithmic Analysis
Digital Collections: Structures
Abū Saʿīd Ibrāhīm b. Ṭahmān b. Šuʿayb, from the village of Bashan, a resident of Nishapur.

He traveled to Mecca and died there. He was a jurist, transmitter of Hadith. He died in 163.

... He composed The Exegesis of the Qur'an, Legal hadith, The Book of Two Celebrations, The Book of Virtues.
Algorithmic deformation (with Python)
Gradual Increase of Complexity of Research Tasks
DESCRIPTIVE NAMES

RELIGIOUS SPECIALIZATION

RELIGIOUS AFFILIATION

PHYSICAL DESCRIPTION

GEOGRAPHICAL OCCUPATIONAL

KUNYA | LAQAB

DOCTRINAL LEGAL

LOCATIONS

OTHER

BIRTH

DEATH

RESIDENCE

TRAVELS

DATES

BIRTH

DEATH

OTHER

SOURCE

VOLUME

PAGE

PATH
al-Ḏahabī and his “History of Islam”
al-Ḍahabī (d. 748/1347 CE)

- a Damascene religious scholar, Šāfīʿī jurist and historian
- “The History of Islam” (*Taʾrīḥ al-islām*)
  - “first draft” in 714/1314 CE
  - 50 volumes (here: 4-50)
  - 7 centuries (here: 41-700/661-1300 CE)
  - over 30,000 biographies (here: ~29,100)
  - ~3,2 mln. words
'Distant reading': ... *fewer elements*, hence *a sharper* sense of their overall *interconnection*. Shapes, relations, structures. Forms. Models.

—Franco Moretti. Graphs, Maps, Trees
“The History of Islam”: Distant Reading

- Death dates
- Place names / Toponyms
- “Descriptive names” (nisbats)
  - 700 unique (frequency 10 and higher)
  - 70,000 total
- ~ social profile (issue with their meaning*)
"The History of Islam": Distant Reading

Traditional Arab Name: Example

“The History of Islam”: Distant Reading

Traditional Arab Name: Example

“The History of Islam”: Distant Reading

Traditional Arab Name: Example

“The History of Islam”: Distant Reading

Traditional Arab Name: Example

"The History of Islam": Distant Reading

**Traditional Arab Name: Example**

... al-Ḳurašī al-Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-Ḥanbalī al-Wāʻīz al-Ṣaffār ...

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>al-Ḳurašī</td>
<td>member of the tribe of Quraysh (tribal)</td>
</tr>
<tr>
<td>al-Taymī</td>
<td>member of the clan of Taym (tribal)</td>
</tr>
<tr>
<td>al-Bakrī</td>
<td>descendant of Abū Bakr al-Ṣiddīq (ancestral)</td>
</tr>
<tr>
<td>al-Bağdādī</td>
<td>resident/native of Baghdād (toponymic)</td>
</tr>
<tr>
<td>al-Ḥāfiẓ</td>
<td>“Preserver” of the Tradition (religious specialization)</td>
</tr>
<tr>
<td>al-Mufassir</td>
<td>exegete of the Qurʾān (religious specialization)</td>
</tr>
<tr>
<td>al-Ḥanbalī</td>
<td>jurist of the Ḥanbali legal school (religious affiliation)</td>
</tr>
<tr>
<td>al-Wāʻīz</td>
<td>public preacher (religious specialization)</td>
</tr>
<tr>
<td>al-Ṣaffār</td>
<td>seller of copper/brass utensils (occupational)</td>
</tr>
</tbody>
</table>
Classification of *Nisbats* from the Social Perspective

Previous work on *nisbats* (mostly occupational):

- Hayyim Cohen
- Carl Petry
- Maya Shatzmiller
Social transformations

De-tribalization
Militarization
Professionalization / Institutionalization
- military (1,324; 4.55%)
- administrative (1,303; 4.48%)
- civilian (6,254; 21.49%)
- religious (11,964; 41.11%)

SRC: "The History of Islam" of al-Dhahabi
De-tribalization
Tribal and Toponymic Identities: BIO (%)

- Tribal (10,270; 35.28%)
- Toponymic (21,875; 75.15%)

% of items (20 lunar year periods)

- 0 AH 70 AH 719 CE
- 100 AH 300 AH 816 CE 913 CE
- 200 AH 300 AH 400 AH 1010 CE
- 300 AH 400 AH 500 AH 1107 CE
- 400 AH 500 AH 600 AH 1204 CE
- 500 AH 600 AH 700 AH 1301 CE

SRC: "The History of Islam" of al-Dhahabi
Militarization
Major Sectors (BIOs) (%)

- military (1,324; 4.55%)
- administrative (1,303; 4.48%)
- civilian (6,254; 21.49%)
- religious (11,964; 41.11%)
% of items (20 lunar year periods)

- military (1,324; 4.55%)
- administrative (1,303; 4.48%)
- civilian (6,254; 21.49%)
- religious (11,964; 41.11%)

Major Sectors (BIOs) (%)

SRC: "The History of Islam" of al-Dhahabi
Cumulative biographical curve (29,102; scaled down by a factor of 9.73)

Amir (796; 2.74%)
Some of the Most Prominent Madrasas

- al-Nizamiyya (185; 0.0065%)
- al-Mustansiriyya (74; 0.0026%)
- al-Zahiriyya (90; 0.0032%)
- al-Aminiyya (50; 0.0018%)
- al-Nasiriyya (50; 0.0018%)
- al-Nuriyya (48; 0.0017%)
- al-Adiliyya (42; 0.0015%)
- al-Qaymariyya (27; 0.00095%)
Patterns of Military Honorific Names: Fulān al-dawla, the most common pattern in the middle period, gets replaced by Fulān al-dīn pattern in the later period
Frequencies: CALIPH, SULTAN, AMIR

- caliph/amir al-mu’minin (1,803; 0.074%)
- sultan (1,445; 0.059%)
- amir (1,710; 0.07%)

# of items (20 lunar year periods)

0 AH - 700 AH
622 CE - 1301 CE

SRC: “The History of Islam” of al-Dhahabi
Professionalization & Institutionalization of the Learned

العلماء
Major Sectors (BIOs) (%)

- military (1,324; 4.55%)
- administrative (1,303; 4.48%)
- civilian (6,254; 21.49%)
- religious (11,964; 41.11%)

% of items (20 lunar year periods)

0 AH 622 CE 719 CE 816 CE 913 CE 1010 CE 1107 CE 1204 CE 1301 CE

SRC: “The History of Islam” of al-Dhahabi
Major Sectors (BIOs) (%)

- military (1,324; 4.55%)
- administrative (1,303; 4.48%)
- civilian (6,254; 21.49%)
- religious (11,964; 41.11%)

SRC: "The History of Islam" of al-Dhahabi

% of items (20 lunar year periods)

0 AH 100 AH 200 AH 300 AH 400 AH 500 AH 600 AH 700 AH
622 CE 719 CE 816 CE 913 CE 1010 CE 1107 CE 1204 CE 1301 CE

c. 0/622 ↘
c. 160/777 ↗
c. 270/884 ↘
c. 470/1078 ↗
c. 570/1175 ↗
Growth of Religious Specialisations

# of items (20 lunar year periods)

0 AH 100 AH 200 AH 300 AH 400 AH 500 AH 600 AH 700 AH
622 CE 719 CE 816 CE 913 CE 1010 CE 1107 CE 1204 CE 1301 CE

c. 0/622 → c. 160/777 → c. 270/884 ↓ c. 470/1078 ↑ c. 570/1175 ↑

specializations

SRC: "The History of Islam" of al-Dhahabi
Waqf Institutions

- Madrasa (545; 0.019%)
- Zawiya (123; 0.0043%)
- Maristan (110; 0.0039%)
- Khanaqa (70; 0.0025%)
- Awqaf (91; 0.0032%)

# of items (20 lunar year periods)

- c. 0/622
- c. 160/777
- c. 270/884
- c. 470/1078
- c. 570/1175

622 CE 719 CE 816 CE 913 CE 1010 CE 1107 CE 1204 CE 1301 CE

SRC: "The History of Islam" of al-Dhahabi
Lineages & Households

References to Relatives

Parents, Children, Siblings, Spouses (20,757; 0.73%)

# of items (20 lunar year periods)

0 AH
622 CE
719 CE
816 CE
913 CE
1010 CE
1107 CE
1204 CE
1301 CE
100 AH
200 AH
300 AH
400 AH
500 AH
600 AH
700 AH
C. 0/622
C. 160/777
C. 270/884
C. 470/1078
C. 570/1175

SRC: "The History of Islam" of al-Dhahabi
Lineages & Households

References to Relatives

Parents (8,966; 0.31%)
Children (3,044; 0.11%)
Siblings (4,263; 0.15%)
Spouses (387; 0.014%)

# of items (20 lunar year periods)

0 AH 100 AH 200 AH 300 AH 400 AH 500 AH 600 AH 700 AH
622 CE 719 CE 816 CE 913 CE 1010 CE 1107 CE 1204 CE 1301 CE

C. 0/622 ^v^ C. 160/777 ^→ C. 270/884 ^\downarrow^ C. 470/1078 ^v^ C. 570/1175 ^v^
Social Geography
Geo-Chronological Coverage

Andalusia (2,077)
Maghrib (310)
North African Cl. (259)
Egyptian Cl. (1,865)
East African Cl. (48)

Anatolian Cl. (276)
Syria (2,671)
Jordan (557)
Central Arabian Cl. (1,151)

Caucasian Cl. (119)
Jazira (741)
Iraq (7,845)
South Arabian Cl. (341)

Central Asian Cl. (783)
NW Iran (1,331)
NE Iran (2,688)
SW Iran (1,515)
South Eastern Cl. (136)
Modeling Geographical Connections

❖ A cartogram of a person—al-Bağdādī and al-Baṣrī—whose biography mentions Baghdad, Kufa, Basra, Medina and Mecca

❖ Such data can be grouped to show particular groups and/or periods
The Core
0632 - 0680 CE
Islamic World Changing
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 40-90AH/661-709CE

Central Asian Cluster

Andalusia

North African Cluster

Moghrib

Anatolian Cluster

Syria

Iraq

Jazira

NW Iran

NE Iran

Central Arabian Cluster

South Arabian Cluster

South Eastern Cluster

Eastern African Cluster

0632-0680 CE

Islamic World Changing

0010-0660 anno hegirae

anno domini

Maxim Romanov @ U of Michigan
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 50-100AH/670-719CE

0641 - 0690 CE
Islamic World Changing
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 150-200AH/767-816CE

0738-0787 CE

Islamic World Changing

anno domini

120-170

anno hegirae
0932–0981 CE
Islamic World Changing
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 450-500AH/1058-1107CE

Maxim Romanov @ U of Michigan

1029 - 1078 CE
Islamic World Changing
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 550-600 AH/1155-1204 CE

Maxim Romanov @ U of Michigan

1126 - 1175 CE

Islamic World Changing

anno domini

520 - 570

anno hegirae
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 650-700AH/1252-1301CE

1223-1272 CE
Islamic World Changing

anno domini 620-670 anno hegirae

Maxim Romanov © U of Michigan

Andalusia

Mazrib

North African Cluster

Anatolian Cluster

Syria

Jazira

Iraq

NW Iran

NE Iran

Central Asian Cluster

South Eastern Cluster

Central Arabian Cluster

South Arabian Cluster

East African Cluster

Caucasian Cluster
Regional Integration
Social and religious groups
Urban centers
SRC: al-Dhahabi’s Ta’rikh al-islam

Unadjusted dates: 50-700AH/670-1301CE

Maxim Romanov @ U of Michigan

2,048 Visitors
241 Migrants
2,360 Natives/Residents:
  802 - with local connections
  1,241 - with transregional connections
  1,305 - total number of transregional connections

anno domini

0641 - 1272 CE
Connections: Andalusia

anno hegirae

020 - 670
2,312 Visitors
526 Migrants
2,736 Natives/Residents:
452 - with local connections
1,398 - with transregional connections
2,077 - total number of transregional connections

641 – 1272 CE
Connections: NE Iran

anno domini

0641-1272 CE

anno hegirae

020-670
Can we actually do this kind of thing?
Reactions...
Reactions...

- Quite a few outraged “NO!”s
Reactions...

- Quite a few outraged “NO!”s
- The Ta’rīḥ al-islām is al-Ḍahabī’s vision of Islamic history from the 13th century
Reactions...

- Quite a few outraged “NO!”s
- The *Taʾrīḥ al-islām* is al-Ḍahabī’s vision of Islamic history from the 13th century
- Or is it?!
Reactions...

- Quite a few outraged “NO!”s
  - The *Ta’rīḥ al-islām* is al-Ḍahabi’s vision of Islamic history from the 13th century
  - Or is it?!

- 50 volumes
  - 700 year of Islamic history
  - 30,000 biographies
  - ~ 3.2 mln. words
  - “compilation of earlier sources”
“Compilation”?

- How does one write a book like that w/o copy-paste?!
- Did he paraphrase his sources?
  - changing their language, and thus bringing 13th century language into the descriptions of earlier periods
- Did he “stitch” the TI from his sources?
  - thus preserving the language of earlier sources
KITAB Project: kitab-project.org

Agha Khan University, U Leipzig/Tufts U, Northeastern U, Göttingen Centre for DH
KITAB Project: kitab-project.org

Sarah Savant, Associate Professor, AKU-ISMHC, and Principle Investigator of KITAB. Dr Savant is a cultural historian specialising in the Middle East and Iran ca. 600-1100. Her work on KITAB is motivated by a desire to write a cultural history of the Arabic book using digital tools. Her publications include The New Muslims of Post-Conquest Iran: Tradition, Memory, and Conversion (Cambridge University Press, 2013, winner of the Said-Sirjani book award given by the International Society for Iranian Studies on behalf of the Persian Heritage Foundation).

Ahmad Sakhi is an IT professional, specializing in Enterprise Information and Content Management space. He is currently consulting for Capgemini, a global consulting firm. Mr. Sakhi has a degree in Computer Information Systems and has completed various projects in the Finance, Telecom, Banking and Energy sectors in North America, including designing, building architecture and implementing end-to-end solutions. Mr. Sakhi is from Toronto, Canada and is involved in the KITAB project as a TKN volunteer.

Malik Merchant is a Software Consultant, specializing in Microsoft Technologies. He currently is working for Avanade, a joint venture between global consulting firm, Accenture and Microsoft. Mr. Merchant has a Bachelor's degree in Computer Science from Mumbai, India and has been involved in several projects across several industries with focus in Oil & Gas and Insurance domain. He now lives in Calgary, Canada and is involved in the KITAB project as a TKN volunteer.

Sohail Merchant is currently working as the Assistant Registrar for the Aga Khan University. Prior to working at AKU-ISMHC, he has worked with various organisations like City Hampers, Child-to-Child – University of London and Blue Systems Inc. where he has designed, managed and implemented various solutions and web applications. Sohail is a Microsoft Certified Professional and holds a Master of Science degree in Computing from London Metropolitan University.
KITAB Project: kitab-project.org

Gregory Ralph Crane
Alexander von Humboldt Professor of Digital Humanities

Greg completed his doctorate in classical philology at Harvard University and subsequently worked there as an assistant professor. From 1985, he was involved in planning the Perseus Project as a co-director and is now its Editor-in-Chief. He was associate professor at Tufts University. He has received, among others, the Google Digital Humanities Award 2010 for his work in the field.

David Smith is an Assistant Professor in the College of Computer and Information Science at Northeastern University and a founding member of the NULab for Texts, Maps, and Networks, Northeastern’s center for the digital humanities and computational social sciences. Previously, he was a research faculty member at the University of Massachusetts’ Center for Intelligent Information Retrieval, a Ph.D. student in computer science at Johns Hopkins University, and the head programmer at the Perseus Digital Library Project. His research focuses on building statistical models of human language, with applications to information retrieval, machine translation, the humanities, and social sciences. Most recently, he has been working on inference for social networks from textual evidence, in collaboration with colleagues in English, history, and political science, under the aegis of the Proteus and Viral Texts projects.

http://viraltexts.org/ Ryan Cordell, David Smith (Code), and others

Agha Khan University, U Leipzig/Tufts U, Northeastern U, Göttingen Centre for DH
KITAB Project: kitab-project.org

Gregory Ralph Crane
Alexander von Humboldt Professor of Digital Humanities

Greg completed his doctorate in classical philology at Harvard University and subsequently worked there as an assistant professor. From 1985, he was involved in planning the Perseus Project as a co-director and is now its Editor-in-Chief. He was associate professor at Tufts University. He has received, among others, the Google Digital Humanities Award 2010 for his work in the field.

David Smith
Assistant Professor in the College of Computer and Information Science at Northeastern University and a founding member of the NULab for Texts, Maps, and Networks, Northeastern's center for the digital humanities and computational social sciences. Previously, he was a research faculty member at the University of Massachusetts' Center for Intelligent Information Retrieval, a Ph.D. student in computer science at Johns Hopkins University, and the head programmer at the Perseus Digital Library Project. His research focuses on building statistical models of human language, with applications to information retrieval, machine translation, the humanities, and social sciences. Most recently, he has been working on inference for social networks from textual evidence, in collaboration with colleagues in English, history, and political science, under the aegis of the Proteus and Viral Texts projects.

http://viraltexts.org/  Ryan Cordell, David Smith (Code), and others

Agha Khan University, U Leipzig/Tufts U, Northeastern U, Göttingen Centre for DH
How to address this issue?
How to address this issue?

Text reuse: computational methods of tracing long quotations, paraphrases, allusions, etc.

al-Ta‘alibi’s *Timār al-qulūb fi-l-muḍāf wa-l-manṣūb* and one of its sources
Graphing Method
Graphing Method
Graphing Method
Graphing Method
Graphing Method

Beginning of the book

End of the book

Book A Sliced

Book B Sliced
Graphing Method

Beginning of the book

End of the book

100 words

Book A Sliced

Book B Sliced
Text Reuse in the *Taʾrīḥ al-islām*

*Taʾrīḥ al-islām* chunked and divided into periods
Text Reuse in the *Ta’rīḥ al-islām*

*Ta’rīḥ al-islām* chunked and divided into periods

*Ta’rīḥ al-islām* compared with itself
Text Reuse in the *Taʾrīḫ al-islām*

*al-Sūrat al-Nabawīyat* of Ibn Hišām (d. 213) *(28,339w; 94p; 0.872%); 50%: 30–63)*

*al-Ṭabaqāt al-kubrā* of Ibn Saʿd (d. 230) *(37,036w; 123p; 1.14%); 50%: 23–53)*
Text Reuse in the Ta’rīḥ al-islām

al-Ta’rīḥ of Ḥalifat b. Ḥayyāṭ (d. 240) (7,696 w; 25 p; 0.237%); 50%: 25–57

al-Musnad of Aḥmad b. Ḥanbal (d. 241) (35,169 w; 117 p; 1.083%); 50%: 30–63
Text Reuse in the Ta’rīḥ al-islām

al-Ta’rīḥ of al-Ṭabarī (d. 310) (37,390w; 124p; 1.151%; 50%: 26–57)

al-Ta’jārib al-umam of Miskawayh (d. 421) (9,059w; 30p; 0.279%; 50%: 23.25–63.75)
Text Reuse in the *Taʾrīḫ al-islām*

**Yatīmāt al-dahr of al-Ṭaʾālibī (d. 429) (3,192w; 10p; 0.098%; 50%: 17–33)**

**Ḥilīyat al-awliyāʾ of Abū Nuʿaym (d. 430) (30,588w; 101p; 0.942%; 50%: 21–49)**
Text Reuse in the *Taʾrīḫ al-islām*

*Dalāʾil al-nubuwwat* of al-Bayhaqi (d. 458) (111,436w; 371p; 3.431%; 50%: 28–61)

*Taʾrīḫ Baġdād* of al-Ḥaṭīb (d. 463) (74,130w; 247p; 2.282%; 50%: 21–48)
Text Reuse in the Taʾrīḥ al-islām

Taʾrīḥ Dimaṣq of Ibn ʿAsākir (d. 571) (245,161w; 817p; 7.547%); 50%: 22–48

al-Ṣilat of Ibn Baškuwāl (d. 578) (16,648w; 52p; 0.482%); 50%: 27–66
Text Reuse in the *Ta’rīḥ al-islām*

*al-Muntaẓam of Ibn al-Ǧawzī (d. 597)* (83,828w; 279p; 2.581%); 50%: 25–60

*Ḏayl mirʿāṭ al-zamān of al-Yūnīnī (d. 726)* (14,738w; 49p; 0.454%); 50%: 21–56
The Scale of Text Reuse

❖ ~11,000 pp. in the Ta’rīḥ al-islām (300 w/p)
❖ ~ 3,200 pp. from 15 major sources (100-800 pp.)
   ❖ 817 pp. > Ta’rīḥ Dimašq of Ibn ʿAsākir (d. 571)
      [passages 22-48 words]
   ❖ 413 pp > Dalāʾil al-nubuwwat of al-Bayhaqī’s (d. 458)
      [passages 28-62 words]
   ❖ 279 pp > al-Muntaẓam of Ibn al-Ǧawzī (d. 597)
      [passages 25-60 words]
   ❖ 270 pp. > Tahdīb al-kamāl of al-Mizzī (d. 742)
      [passages 21-46 words]
   ❖ 247 pp > Ta’rīḥ Baġdād of al-Ḥaṭīb (d. 463)
      [passages 21-48 words]
The Scale of Text Reuse

Counting only earliest quotations:
• 1,067,566 words
• 3,559 pages
• ~32.18%
• 50% of quotes (25-59 words)
The Scale of Text Reuse

Counting only earliest quotations:
- 1,067,566 words
- 3,559 pages
- ~32.18%
- 50% of quotes (25-59 words)

up to 74.11%
Siyar a‘lām al-nubalāʾ

30-35% from the Ta’rīḥ al-islām
50% of instances are 26-66 words
Taḏkirat al-ḥuffāż

25-26% from the Taʾrīḥ al-islām
50% of instances are 26-66 words
Maʿrifat al-qurrāʾ al-kibār ʿalā al-ṭabaqāt wa-l-aʿšār

22-24% from the Taʿrīḥ al-islām
50% of instances are 23-55 words
Working theory:

- The *Taʾrīḫ al-islām* is a databank for his other books

- Way to prove: experimentally
  - Generate his “smaller” collections computationally

>> generate and compare visualizations and various statistics (*to be continued*)
al-Ḏahabī as a historian:

Did he have any understanding what he collected?
al-Ḏahabī as a historian

❖ “Cities and Ports for Hearing the Reports”*
(al-Amṣār ḍawāt al-āṭār)

* Translation by Michael Cooperson

❖ 4-folio epistle

❖ focus on cities and provinces (~80)

❖ statements about the role of the regions of the Islamic world in Ḫadīṯ learning
[In] Mecca, the learning was sparse in the time of the Companions. Then, it became abundant at the end of their time, and then in the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in the two sacred cities (Mecca and Medina), [but] became abundant in other [cities of Islam].
[In] Mecca, the learning was sparse in the time of the Companions. Then, it flourished at the end of their time, and then at the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in both sacred cities (Mecca and Medina), [but] became abundant in other [cities of Islam].
[In] Mecca, the learning was sparse in the time of the Companions. Then, it flourished at the end of their time, and then at the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in both sacred cities (Mecca and Medina), [but] became abundant in other [cities of Islam].
Islamic early centers (%)

- Medina: al-Madani (691)
- Basra: al-Basri (1,595)
- Kufa: al-Kufi (1,432)

% of Biographies (20 AH year periods)

- c. 0/622
- c. 160/777
- c. 270/884
- c. 470/1078
- c. 570/1175

SRC: "The History of Islam" of al-Dhahabi
Classical Islamic Trivium (%)

% of items (20 lunar year periods)

- c. 0/622
- c. 160/777
- c. 270/884
- c. 470/1078
- c. 570/1175

0 AH | 100 AH | 200 AH | 300 AH | 400 AH | 500 AH | 600 AH | 700 AH
622 CE | 719 CE | 816 CE | 913 CE | 1010 CE | 1107 CE | 1204 CE | 1301 CE
al-Amṣār ḍawāt al-āṯār

- Certain statements > about large groups in the Taʾrīḥ al-islām (at least 100 people, but usually much more)
- Uncertain statements > about small groups in the Taʾrīḥ al-islām
- No statements > no identifiable groups in the Taʾrīḥ al-islām

- What was his method? Did he count?!
al-Ḏahabī’s method: How?
Collecting,
Organizing,
Categorizing
Counting Muslims

- Hadith collections
- Lexicographical dictionaries
- Onomastic dictionaries
- Terminological dictionaries
- Genealogical texts
- Biographical collections
- Geographical texts
- Chronicles

Texts with serialized data run into hundreds
Counting historians?

Keen sense of proportions
Geo-Chronological Coverage: *Comparative Perspective*
Geo-Chronological Coverage: *Comparative Perspective*
Geo-Chronological Coverage: *Comparative Perspective*
Working with data?
Movable media

Reference:
Movable media
Movable media
Movable media
Movable media
Movable media
Movable media

MAKKI (269 total; BW: mky)
Cumulative biographical curve (29,110; scaled down by a factor of 38.03)

::toponymic

<table>
<thead>
<tr>
<th>Year</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 AH</td>
<td>719</td>
</tr>
<tr>
<td>200 AH</td>
<td>816</td>
</tr>
<tr>
<td>300 AH</td>
<td>913</td>
</tr>
<tr>
<td>400 AH</td>
<td>1010</td>
</tr>
<tr>
<td>500 AH</td>
<td>1107</td>
</tr>
<tr>
<td>600 AH</td>
<td>1204</td>
</tr>
<tr>
<td>700 AH</td>
<td>1301</td>
</tr>
</tbody>
</table>
Concluding...

- Digital methods allow us to get a novel perspective on and a better understanding of:
  - our textual sources and information that they preserve
  - our authors and their methods of writings
  - and, in the long run, the entire Islamicate written tradition
There's a graph for that