Exploring Ancient Texts with a User Driven Concept Search

Muhammad Faisal Cheema, Stefan Jänicke, Christoph Weilbach, Judith Blumenstein, Gerik Scheuermann
Leipzig University, Germany

eXChange: Exploring Concept Change and Transfer in Antiquity
Outline

- Motivation
- Concept Search Pipeline
  - Concept Editor
  - Concept Search
  - Concept Search Results Viewer
- Demo
- Results & Findings
- Conclusions
Motivation

“Even with a perfect search engine, a poorly constructed search question may not lead to right answer” (Teevan et al. 2004)

- Keyword search
- Most popular method of everyday search
- Reformulate query to get better results
- Even in research paradigms
Motivation

- Data in consideration: **Ancient texts**
- Keyword search yields incomplete results “e.g. concept **Epilepsy**”
  - Full word search e.g. *'morbus comitiales'*
  - Truncated search e.g. *morb* *comiti**
  - Alternate terms e.g. *'morbus sacer' or 'epilepsia'*
  - Descriptive terms e.g. *spuma or concidere*

- Ancient texts are quite different from modern texts in terms of:
  - ancient vocabulary
  - ideas, cultures, practices that we are no more aware of
Existing techniques

- Semantic search approaches
  - Word sense disambiguation
  - Ontology based

- Statistical approaches
  - Topic Models
  - Latent Dirichlet allocation

Blackbox problem
Goal

- Ability to model a concept query that covers all dimensions of keyword search i.e. fullword, truncated, alternate and descriptive terms

- **Ability to steer the search process as desired by the user**

- **Ability to iteratively improve results, as in keyword search**

**SOLUTION**

- A visual search environment that provides the user ability to:
  - Model their ideas of concepts
  - Search the concept models and iteratively improve results
Concept Search Pipeline
Concept Search Pipeline

Concept Editor

Concept

Search Results

Concept Search Viewer

Creates/Modifies Concepts

Analyses / Filter Results

Research Scholar
Concept Search Pipeline

[Diagram showing the flow from Concept Editor to Concept Search, with arrows indicating the process of creating/modify concepts and the research scholar's involvement.]

Concept Editor -> Concept Search

Concept Search -> Concept

Concept Editor creates/modify concepts

Research Scholar
What is our Concept model

- We define our Concept Model inspired by ideas from mind maps & concept maps

Example mind map

Example concept map

Cambodia4kids.org: (https://www.flickr.com/photos/cambodia4kidsorg/6195211411)
The concept model

- Concept is modeled as a **tree** data structure
- Simpler to model for humanities scholars
- Better (re)layout options
For further on concept editor details, refer to [CJBWS15]
Concept Search Pipeline

Concept Editor -> Concept Search -> Concept Search Viewer

Concept Editor Creates/Modifies Concepts

Research Scholar Analyses / Filters Results

Concept Search Search Results
Concept Search Pipeline

Concept Editor

Concept

Search Results

Concept Search Viewer

Creates/Modifies Concepts

Research Scholar

Analyses / Filter Results
Hierarchical Keyword Search

- Word forms of all terms are searched
- Each entry is searched for with classical string lookup
- Per word, all occurrences are located, grouped and counted within groups

**Result:** Non-negative integer matrix (usually sparse) representation. Can be used to rank results according to concept hierarchy.
Assignment of Results to Concept Nodes

Goal: Show the search results in user created concept hierarchy

- Result 1
- Result 2
- Result 3
- ...
- ...
- Result N

Concept Node

Result 1, Result 2, ...
... Result N
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
### Assignment of Results to Concept Nodes

**Table:**

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

Result | Word forms
---|---
A | spuma, nervus, lumen, morbus, vinum
B | morbus, sacer, vertere, senatus
C | dies,
D | morbus, dies, senatus, lumen, tremere
E | Sacer, paupertas, vinum
F | Sacer, nervus, tremere, vium, dies
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

Result | Word forms
--- | ---
A | spuma, nervus, lumen, morbus, vinum
B | morbus, sacer, vertere, senatus
C | dies,
D | morbus, dies, senatus, lumen, tremere
E | Sacer, paupertas, vinum
F | Sacer, nervus, tremere, vium, dies
Assignment of Results to Concept Nodes

<table>
<thead>
<tr>
<th>Result</th>
<th>Word forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>spuma, nervus, lumen, morbus, vinum</td>
</tr>
<tr>
<td>B</td>
<td>morbus, sacer, vertere, senatus</td>
</tr>
<tr>
<td>C</td>
<td>dies,</td>
</tr>
<tr>
<td>D</td>
<td>morbus, dies, senatus, lumen, tremere</td>
</tr>
<tr>
<td>E</td>
<td>Sacer, paupertas, vinum</td>
</tr>
<tr>
<td>F</td>
<td>Sacer, nervus, tremere, vium, dies</td>
</tr>
</tbody>
</table>
Assignment of Results to Concept Nodes

Optional (tweakable) parameters

- Further criteria to decide when a result moves up to parent node:
  - Threshold value
  - Dark/lighter shades of concepts
  - Aggregation and intersection of results
  - Etc.

- Within a Concept node, rank texts according to
  - Term variaance
  - Word counts
  - the ratio of positive to negative words
  - Titles
  - Combinations of above
  - etc.
Concept Search Pipeline

- **Concept Editor**
  - Concept
  - Creates/Modifies Concepts

- **Concept Search**
  - Search Results
  - Analyses / Filter Results

- **Concept Search Viewer**
Concept Search Pipeline

Concept Search

Search Results

Concept Search Viewer

Analyses / Filter Results

Research Scholar
Layout calculation

- Nodes sizes expand due to search results placements
- All expanded node sizes equal
- Overlap
Layout calculation (Bottom up)

For further details refer to [CJBS16]
Concept Search Results (Aggregate View)
Concept Search Results (Word Summaries View)
Concept Search Results (Word Summaries View)

For further details refer to [JBRZS15]
Title :: Mulomedicina Chironis
Start verse :: lib. 1, § 3, p. 4, lin. 23
No. of verses :: 2

de morbo renali iumentorum, si quid enim per ignorantiam male factum vel contrarium circa animalium corpora fuerit, emendari non poesit, vincetur enim corpus animalis cedente per periculum. Primus ergo pars est medicinae causarum rationem reddere, delinde ipsa adutoria medicinae docere oportet, quemodo possint auxilia salubris corpori praestare propter distumum vitam et integritatem membrorum, quibus illeque rebus medicinae scientiae veleriae consueti omnis, ex quibus adeo aiutaria atque corpora iumentorum malam validudinem et causa liberae possint, et omne curam rationalitatem docere oportet, primum dico de sanguinis tractione. Firmamentum adulatorum est medicinae sanguinis tractatio, verum tamen si rationalitatem et locutione fiat, licet paene in omnibus validudinis fieri sanguinis tractatio praecipuit, sed se ne omnis sanguinis tractatio fieri praecipitur, cuius ergo tamen rationem protinus reddo, primum de sanguine reddenda est ratio, quacunque in corpore per ipsum vita constat, nocumentis tractato a corpore sanitate praestant, ex quo quacunque enim causa validum est, in aliquam partem corporis quae pertinet, semper indigestione fatigatus sangus corruptione, discordat per omnia membra et totus corporis languorem infert, per quem incipiet strictura et tensio nasci, per inflamationem venarum, quae strictura et tensio corporis alter laxari non potest, nisi per sanguinis retractionem. Allud simul compendium nolum est nisi sanguinis tractatio, per quern universa corporis vita et morti abstrahi possunt.
Concept Search Results

Exporting & reusing results
(An important feature)

- Save Tables
- Relevant
- Irrelevant
- Interesting

Save All session
Results, Demo and Findings
Humanities scholars in the project

Two humanities scholars have extensively used the concept search

- **Scholar A** → More Interested to explore via the distant reading aspects

- **Scholar B** → Interested to explore by reading the concerned passages directly, and not using distant reading features extensively
Concept - Bloodletting

mittere - emittere - detrhere - deductio - missio

to let/to take (resp. nouns)

bloodletting

blood...

sanguis

to disgorg/to segregate

reddere - reicere

Concept: Bloodletting

Diagram showing the relationship between various Latin terms related to bloodletting such as "mittere," "emittere," "detrhere," "deductio," and "missio." These terms are connected to "bloodletting," "blood...," and "to disgorg/to segregate." The diagram also includes "sanguis," "reddere," and "reicere."
Concept - Bloodletting
Demo
Usage Scenario

(Remodelling a concept)
Concept – Epilepsy – Version 1 (Search results)
Concept – Epilepsy – Version 2

what happens?

epilepsy labels

epilepsy-v2

persons

epilepticus

caducus

tremere

spuma

trepidatio

concidere

moral terms

vitiurn

campus

dies

political terms

epilepticus

epilepsia

comitialis

maior

morbus

Επιληψία

epilepsis

sacer
Concept – Epilepsy – Version 2 (Search results)
Concept – Epilepsy – Version 3

what happens?
epilepsy-v3

epilepsy labels

epilepsia
comitialis
epilepticus

persons
caducus
epilepticus

tremere
concidere
trepidatio

spuma
Epilepsy - Summary of approx. precision

<table>
<thead>
<tr>
<th></th>
<th>Epilepsy</th>
<th>labels</th>
<th>persons</th>
<th>Symptoms / what happens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 2</td>
<td>38</td>
<td>7053</td>
<td>85</td>
<td>800</td>
</tr>
<tr>
<td>Version 3</td>
<td>4</td>
<td>128</td>
<td>95</td>
<td>824</td>
</tr>
<tr>
<td>Version 4</td>
<td>302</td>
<td>138</td>
<td>164</td>
<td>44</td>
</tr>
<tr>
<td>Version 5</td>
<td>174</td>
<td>106</td>
<td>68</td>
<td>43</td>
</tr>
</tbody>
</table>

Total results retrieved
Findings

- Ability to **steer** the search as desired – a **prominent feature of keyword search**
- Remodelling the concept iteratively can help the scholar to increase **recall** while also keeping **precision** factor high.
- Support observations concerning of a whole phrase, not only term. Is an option that dictionaries do not provide, but only commentaries.
- Support scholars to observe different uses of phrases (**phrase reuse**) depending on time, genre and author i.e. semantic development of a phrase.
- Support scholars to verify or falsify hypothesis of their defined concepts → remodel concepts → iterative process i.e. verify scholar’s former knowledge, if the terms defined belong to the concept or not.
Usage Scenario

(Inter concept search)
Concept – **Medicines for Epilepsy**
Conclusion & Future Work
Limitations of current system

- Unable to filter time, genre, author etc due to bad metadata.

- Ability to define maximum size between two words at sentence level. Defining word distance not possible due to backend.
Visualization system turns away from being a “working black box”
Search system able to search thousands of word forms
Various word form combinations considered and ranked
Step towards extracting more and accurate results (completeness)
Iterative improvement of results (steering the search process)
Phrase reuse detection
Visual Query environment (concept level)
Applicable to other domains in future
References


Thank You for listening!

Questions / Suggestions / Feedback