Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

SeCo Text Annotation Services

Minna Tamper, minna.tamper@aalto.fi
Semantic Computing Research Group (SeCo)
Department of Computer Science, Aalto university

Transforming Cultural Heritage texts into a knowledge graph

Aalto University
School of Science

UNIVERSITY OF HELSINKI

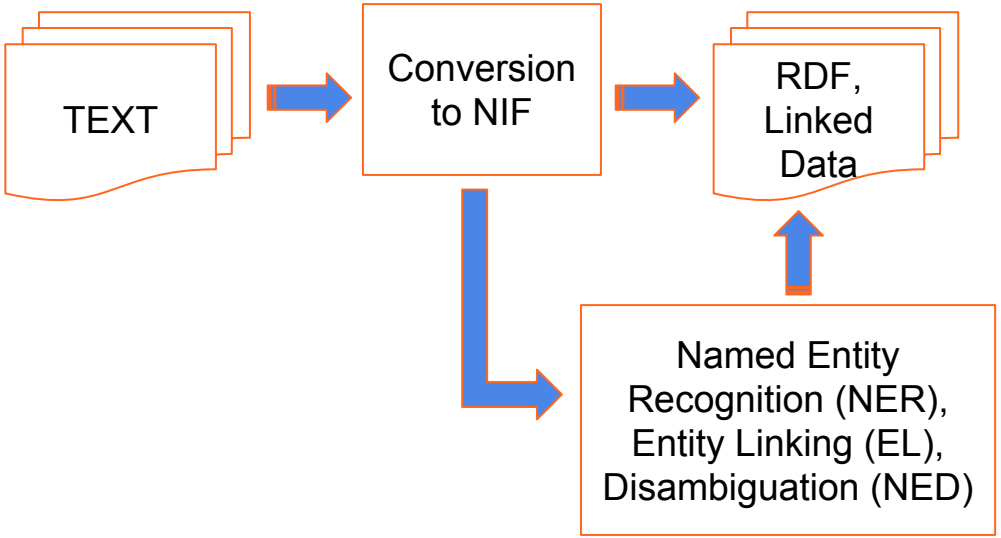
SeCo

HELDIG
Helsinki Centre for Digital Humanities

Background

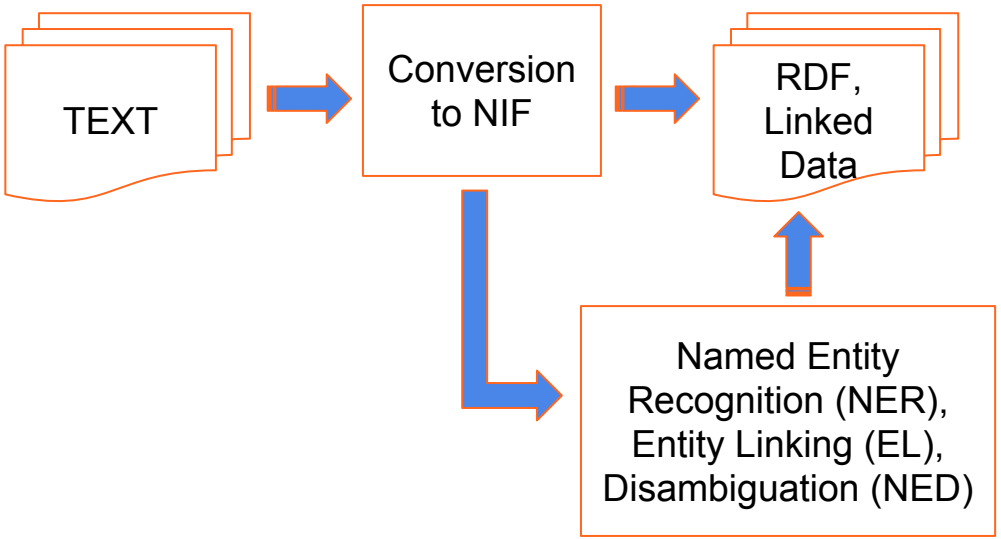
- Much of the primary data of Digital Humanities is available only in textual form, and there is an ever-growing need for structuring it for semantic analysis.
- Transforming Cultural Heritage texts into a knowledge graph and a Linked Data service provide a flexible interface to use the data for different types of analyses.

Transformation



- **Transforms text to NIF (NLP Interchange Format)**
  - Paragraphs, Titles, Sentences, Words
  - Enriched with morphological and syntactic information

Transformation



- **NIF RDF data used for named entity linking (NEL)**
  - NER, EL, and NED
- **Collecting possible tags for further usage from texts**

Application

- **Published in a SPARQL endpoint service**
- **Knowledge graph can be queried and the data can be used for different purposes**
  - Network Analysis
  - Linguistic Analysis
  - Information Retrieval
  - Contextual reader
- **Examples: biografiasampo.fi**

SeCo NLP Services

- **Tools and services for analysing texts as APIs**
- **Using multiple different tools**
  - FiNER, Finnish-dep-parser, LAS, ARPA etc.
  - Services built on top of the tools
    - *Anonymizer, automatic annotation etc.*
- **Available in near future**

Language tools in service of the humanities/social sciences

Eetu Mäkelä, D.Sc., eetu.makela@helsinki.fi
Professor (tenure track) in Humanities–Computing Interaction / University of Helsinki
Docent (Adjunct Professor) in Computer Science / Aalto University

UNIVERSITY OF HELSINKI
FACULTY OF ARTS

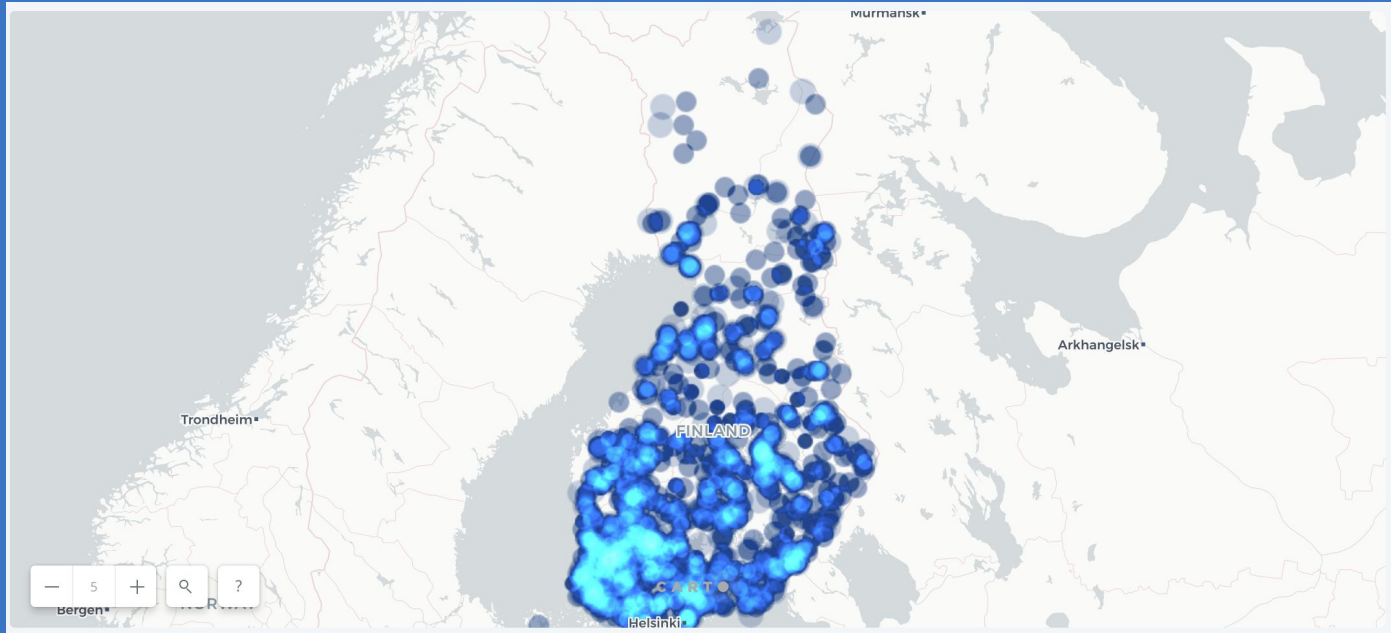
Language Analysis Tool LAS

- Need to:
  - Partition a multilingual dataset by language?
  - Lemmatize a text for e.g. topic modeling?
  - Identify all adjectives describing Juha Sipilä in a text?

Language Analysis Tool LAS

- Available as:
  - For HSS researchers, a local command line tool (http://github.com/jiemakel/las/),
  - For catalogs and search systems, an Elasticsearch analysis plugin (https://github.com/jiemakel/elasticsearch-analysis-las)
  - For developers, a web service (http://demo.seco.tkk.fi/las/ / https://hub.docker.com/r/jiemakel/las-ws/)

Need to e.g. extract all Finnish place name mentions from YLE news articles and plot them on a map?



Named Entity Linking tool ARPA

- Configurable to link against custom vocabularies and datasets

STATEMENT BY VISCOUNT BRYCE.

The final question was this :—

" As the result of this war, what hope have we of reconstruction and an altered policy in Germany ? "

Viscount Bryce answered :—

" It is to be hoped and expected that the Allies will so completely defeat Germany as to discredit the whole military system, and the ideas out of which the horrors of German war practice have developed. It is essential to inflict a defeat sufficiently decisive in the eyes of the Germa done with their milita doctrines, and it is essen themselves—discredit th thorough a manner that again. The way, then ' Frightfulness ' is by a It is our task to show t tices and total disregard must bring to pass the ju that such methods do not our British people are ur been before. The invasi committed there, and th these three series of ac people in its firm resolv complete victory. Now c isolated crimes of the German Government, like the shooting of Miss Cavell and Captain Fryatt, come these abominable deportations of Belgians into a sort of slavery."

In all communication with Lord Bryce, one feels the

13

Charles Fryatt

Charles Algernon Fryatt (2 December British mariner who was executed by t ram a U-boat in 1915. When his ship, t off the Netherlands in 1916, he was co death although he was a civilian non-c followed his execution near Bruges, Be reburied with full honours in the Unite

Laki luvan saaneista oikeudenkäyntiavustajista

Katso tekijänoikeudellinen huomautus käyttöehdoissa.

Eduskunnan päätöksen mukaisesti säädetään:

1 §
Lain soveltamisala

Tässä laissa säädetään luvasta oikeudenkäyntiasiamiehenä ja -avustajana toimimiseen sekä luvan saaneen oikeudenkäyntiavustajan velvollisuuksista ja valvonnasta.

Oikeudenkäyntiasiamiehen ja -avustajan kelpoisuusvaatimuksista oikeudenkäynnissä säädetään erikseen.

2 §
Luvan myöntämisen edellytykset

Tässä laissa tarkoitettu lupa oikeudenkäyntiasiamiehenä ja -avustajana toimimiseen myönnetään toistaiseksi henkilölle:

1) joka on suorittanut Suomessa oikeustieteen muun ylemmän korkeakoulututkinnon kuin kansainvälisen ja vertailevan oikeustieteen maisterin tutkinnon taikka ulkomailla vastaavan oikeustieteen tutkinnon, joka on tunnustettu Suomessa sen mukaan kuin erikseen säädetään;

2) joka on saavuttanut riittävän perehtyneisyyden oikeudenkäyntiasiamiehen ja -avustajan tehtäv

Konkurssi
Konkurssi eli varsinkin tilanne, jossa velallinen henkilö tai organisaatio todetaan tuomioistuimessa kyvyttömäksi maksamaan velkojaan takaisin.

3) joka ntiasiamiehen ja -avusta

4) joka ei ole konkurssissa ja jonka toimintakelpoisuutta ei ole rajoitettu.

Edellä 1 momentin 2 kohdassa tarkoitettu riittävä perehtyneisyys on henkilöllä, joka on:

1) suorittanut asianajajista annetun lain (496/1958) 3 §:n 1 momentin 4 kohdassa tarkoitetun asianajajatutkinnon;

www.thelatinlibrary.com/livy/liv.per34.shtml Edit

TITI LIVI AB URBE CONDITA LIBER

Lex Oppia, quam C. Oppius trib. pl. bello Punico de finiendis matronarum cultibus tulerat, c ea lex aboleretur.

Is in Hispaniam profectus bello, quod Emporiis orsus est, citeriorem Hispaniam pacauit.

T. Quintius Flamininus bellum aduersus Lacedaemonios et tyrannum eorum, Nabidem, prosp sub dicione tyranni erant, finiit.

Res praeterea in Hispania et ac

Senatus tunc primum secretus a

Coloniae plures deductae sunt.

M. Porcius Cato ex Hispania t

T. Quintius Flamininus, qui triumphauit.

Hannibal
Hannibal, son of Hamilcar Barca (247 – 183/182 BC) was a Punic Carthaginian military commander, generally considered one of the greatest military commanders in history. His father Hamilcar Barca, was the leading Carthaginian commander during the First Punic War, his younger brothers were Mago and Hasdrubal, and he was brother-in-law to Hasdrubal the Fair.

Legati Carthaginiensium nuntiauerunt Hannibalem, qui ad Antiochum confugerat, bellum cu litteris Carthaginem missum ad bellandum Poenos concitare.

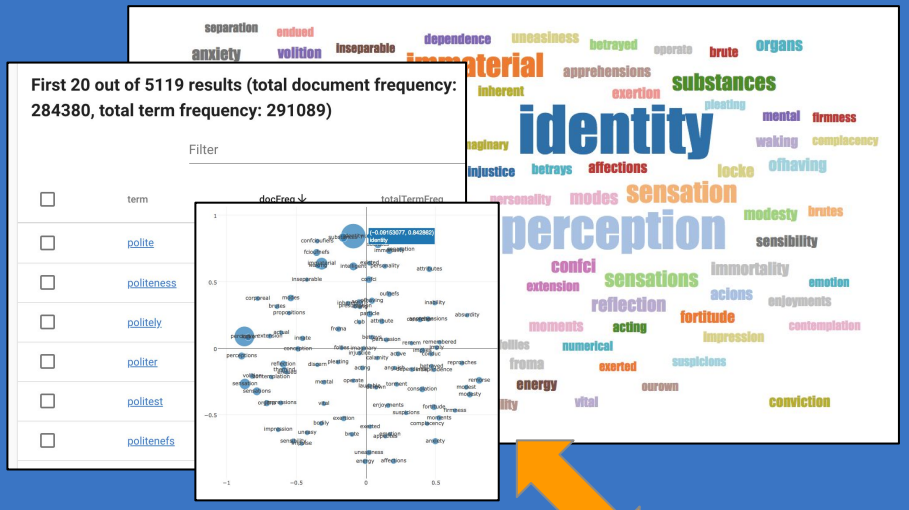
Livy The Latin Library The Clas

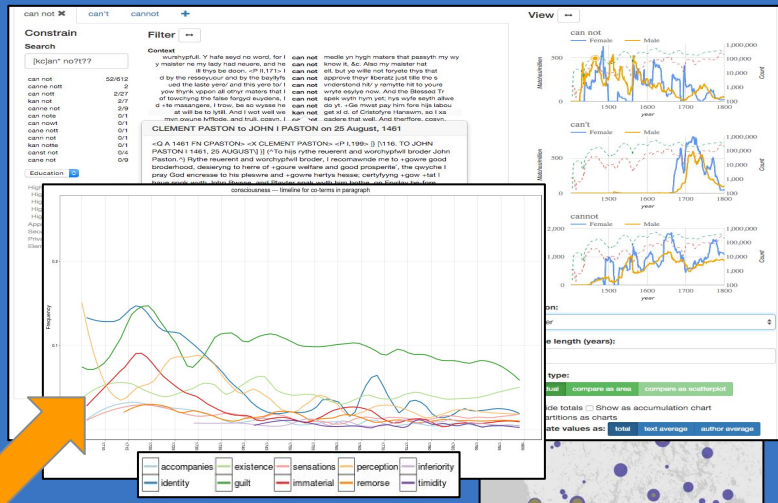
Octavo: complex queries combining structured and unstructured data

- In Dutch newspapers, find all words qualified by “national”
- In Suomi24, find all uses of the word “French” as an adjective in threads that have at least 10 comments and are not in the Travel, History or Automobiles subforums.
- In YLE news data, generate a description of how Ilkka Kanerva is depicted
- In Suomen Kansan Vanhat Runot, find all instances of the formula “small maids” and group them by time and location
- In Eighteenth Century Collections Online, compare the uses of the words “mind” and “soul” across the century, also separately between religious and profane literature

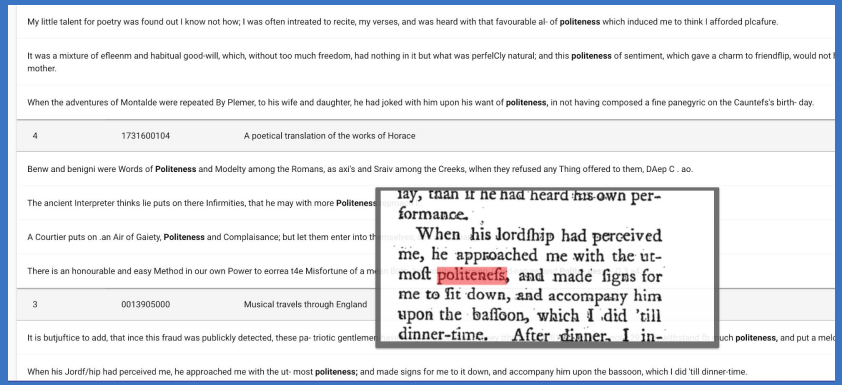
Linguistic fingerprint, either of works or of a word's neighbourhood

Temporal/geographical perspective (projected through metadata)





Close reading



Navigate between close and distant reading; linguistic, text-centric and metadata-centric analyses

Beyond keyword searches, toward complex subcorpora formulation

- Handling OCR noise
  - politeness, politenefs, politeners, …
- Linguistic complexity
  - “hopia* vaski*”~10 (any word starting with hopia within 10 words of any word starting with vaski)
  - (/pi[sk]ku.*/ OR pien*) AND /pi[ij]at.*/
  - |infl_fi_~2:Väinämöinen (any [dialectal] variation at max two edit distances from any inflected form of the Finnish name Väinämöinen)

First 20 out of 5119 results (total document frequency: 284380, total term frequency: 291089)

Filter

term	docFreq ↓	totalTermFreq
polite	155027	158986
politeness	63229	65532
politely	14568	14657
politer	7753	7789
politest	5402	5440
politenefs	2953	2976
politeffe	2806	2891
politeners	2461	2477
politesse	2140	2184