Digital Prosopography of Renaissance Musicians

Discovery of Social and Professional Network

Susan Forscher Weiss
Musicology and German & Romance Languages and Literature
The Johns Hopkins University
Robert Lehman Visiting Professor
Harvard Center for Italian Renaissance Studies, Villa i Tatti
sweiss@jhu.edu

Ichiro Fujinaga
Music Technology Area, Department of Music Research
Schulich School of Music, McGill University
Centre for Interdisciplinary Research in Music Media and Technology
ich@music.mcgill.ca

Goals

- To create a social and professional network of Renaissance musicians
- To create a database to be able to study the network
- To make connections and discover relationships or answer questions related to dates, geographies, professions, etc.

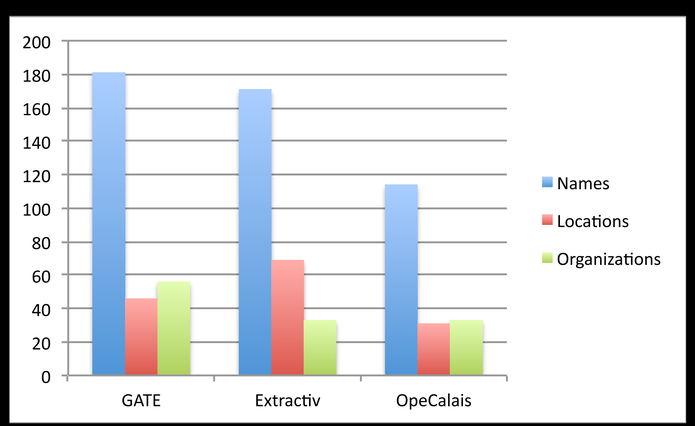
Example queries: Not easily answered by Google, Wikipedia, and other traditional methods

- Which music printers were in business in 1481 in Florence?
- Which composers were residing in Florence in 1481?
- Which composers visited Florence in 1481?
- Which trumpeters were active in Florence between 1481–86?
- What events took place in 1481 in Florence that required musical performance?
- What pieces requiring trumpets were performed in Florence in December 1481?

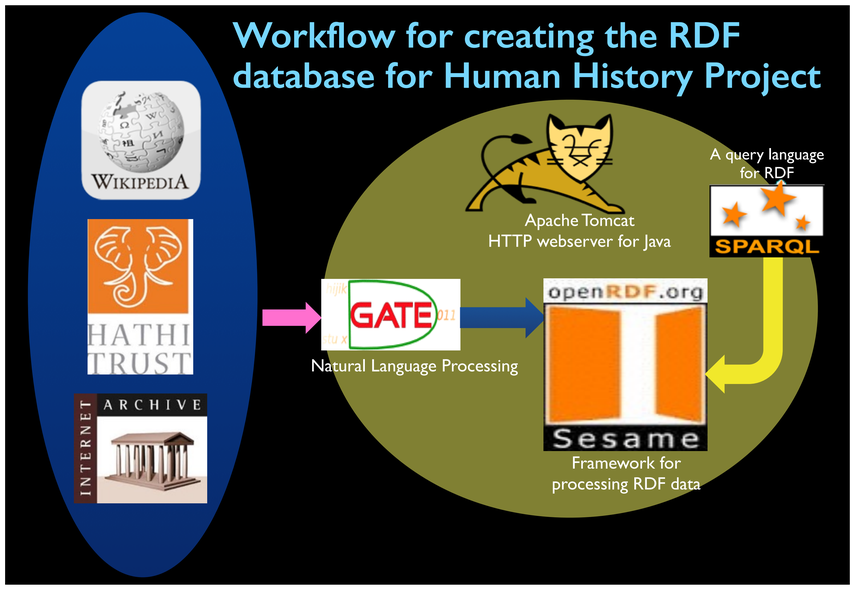
Solutions: Three major tools

- Natural Language Processing (NLP)
  - Named-entity extraction
  - Events extraction
- Linked Open Data
  - RDF (Resource Description Framework): a data model
  - SPARQL: a query language for RDF
- Crowdsourcing

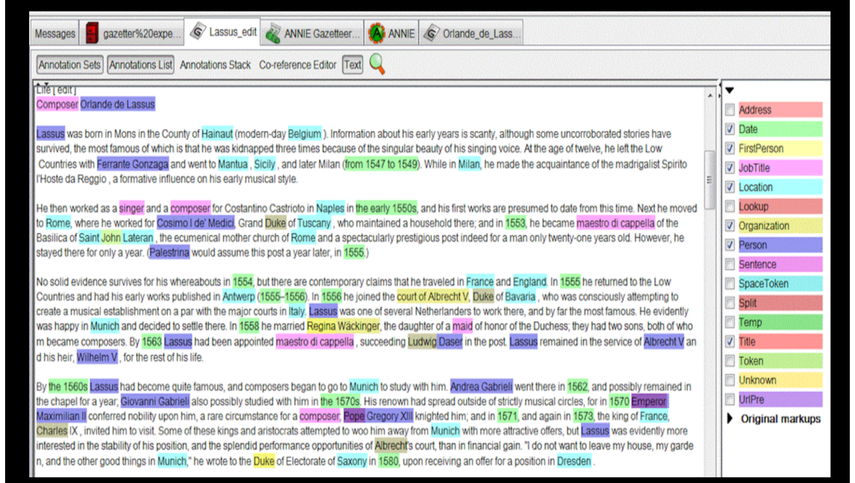
Comparison of Named-Entity Extractors



Workflow for creating the RDF database for Human History Project



An example of named-entity extraction by GATE



Named-entity extraction: Experiment I

- Four Renaissance composers' entries in Wikipedia
- Using the default Gazetteer (a dictionary)
  - 90.25% precision (the extracted entities were correctly identified)
  - 65.33% recall (the entities in the document that were found)
- Editing the Gazetteer (~15 minutes / article)
  - 98.39% precision; 91.86% recall
- Fixing the problem with plurals with the Morphological Analyzer (finding root form of words)
  - 98.45% precision; 98.45% recall

Named-entity extraction: Experiment II

- Ten composers x three sources (30) minus the four Wikipedia articles from Experiment 1: 26 articles
- Using the modified Gazetteer and Morphological Analyzer from Experiment 1
  - 99.24% precision; 98.9% recall
- Correction time of the 5,441 entities extracted
  - Average of 3 sec./entity to correct
  - 240 min. to correct 26 articles or about 10 min. / article

Next steps

- Extract relationships between named entities (e.g., REEL)
- Create web interface to correct relationships
- Create web interface to query the network of relationships

Acknowledgements

We would like to acknowledge the wonderful work by Christopher Buchanan on this project. Other team members include: Haoyuan Ji, Peng Jianxiang, Ankit Sarwal, and Chengyuan Zhang. This research has been supported by the NEH Digital Humanities Start-up Grant and the Social Sciences and Humanities Research Council of Canada.

THE PEABODY INSTITUTE OF THE JOHNS HOPKINS UNIVERSITY

NATIONAL ENDOWMENT FOR THE HUMANITIES
OFFICE OF DIGITAL HUMANITIES

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology
McGill
DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB
Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada