Reconstructing Medieval Social Networks from English and Latin Charters

ChartEx and Traces Through Time project
Arno Knobbe, Leiden University
Digital Humanities Projects

ChartEx, 2012-2013 (Digging into Data challenge)
- U. Brighton (NLP), U. Leiden (Data Mining)

Traces Through Time, 2014-2015 (Big Data, AHRC)

CATCH Links (NWO CATCH)
- IISH, U. Leiden, U. Utrecht
ChartEx, the Charter Excavator

- Over 20k charters from various regions
  - England, mostly York
  - France, Cluny
- 5 collections of charters
  - modern translations to English
  - Latin
- NLP, DM/record linkage, HCI
- Middle ages
  - 900-1400
408. Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited; paying Thomas and his heirs 1d. or [a pair of] white gloves worth 1 d. at Christmas. Warranty. Seal.
January 1252 [1252/3]

SOURCE: VC 3/Vi 326 (161 mm x 137 mm)
408. Grant by *Thomas son of Josce goldsmith and citizen of York* to his *younger son Jeremy* of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which *mag. Simon de Evesham* inhabited; paying *Thomas* and *his heirs* 1d. or [a pair of] white gloves worth 1 d. at Christmas. Warranty. Seal.


January 1252 [1252/3]
408. Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited; paying Thomas and his heirs 1d. or [a pair of] white gloves worth 1 d. at Christmas. Warranty. Seal.


January 1252 [1252/3]
408. Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited; paying Thomas and his heirs 1d. or [a pair of] white gloves worth 1 d. at Christmas. Warranty. Seal.


January 1252 [1252/3]
Record Linkage Challenges

- Fairly unstructured data, free text

- No concept of last names
  - Thomas
  - Place names, William de Gerfordby
  - Occupation, Robert le Spicer

- No street numbers (nor GPS)
  - land with buildings in Petergate, lying in length and in width between the land once of Nicholas de Bugetorp and the land once of Martin Grandeth

- Spelling variations
  - John de Eskrick, John de Eskryck, John de Escryck
Manual and Automated Annotation

- 50 charters
  - manual annotation by domain experts (ChartEx ontology)
  - training set
  - entire (digitized) collection

Natural Language Processing

- annotated collection
408. Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited; paying Thomas and his heirs 1d. or [a pair of] white gloves worth 1d. at Christmas. Warranty.
Relational representation

... although actually ‘under the hood’ it looks more like this

T1 Document 0 17 vicars-choral-408
T2 Transaction 18 23 Grant
T3 Person 27 33 Thomas
T4 Person 40 45 Josce
T5 Occupation 46 55 goldsmith

... R5 refers_to Arg1:T1 Arg2:T2
R6 is_grantor_in Arg1:T3 Arg2:T2
R7 is_son_of Arg1:T3 Arg2:T4
R8 occupation_is Arg1:T3 Arg2:T5
...
NLP layers

**Word layer**: basic information about individual words, collected ‘outside’ the main ChartEx NLP system (using external tools – part of speech tagger, stemmer, Soundex coding etc.)
NLP layers

**Token layer:** identify semantic types and intrinsic properties (e.g. gender) of known individual words (not all of these types feature in the final output). Some types are inferred – e.g. unknown capitalised words are proper nouns.
**Lexical layer:** identify simple lexical phrases – groups of tokens that act as individual units. Promote other individual tokens to lexical items (with lexical types).
NLP layers

**Syntax layer:** build (local) syntactic structure to identify basic constituents of the sentence.

---

```
Document

1 vicars-choral-408

2 Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited
```
Phrasal layer: use part-of-speech tags to build lexical items into local syntactic/semantic structures – and then ‘collapse’ all the equals relations.
How close are we?

Manual annotation

NLP output
Record Linkage

- 50 charters
- Automatically annotated collection
Relational Information (Fragmented)

Vicars Choral 408
Matching Relational Information

Vicars Choral 408

Vicars Choral 409
Thomas son of Josce, goldsmith

Statistics

\[ p(\text{Thomas}) = 0.12 \] (common name)
\[ p(\text{Josce}) = 0.0015 \] (uncommon name)
\[ p(\text{Goldsmith}) = 0.04 \] (common profession)

Dating

\( \text{vc-408} \ 1252-1253 \)
\( \text{vc-409} \ 1253-1261 \)

Final confidence

\[ \text{conf} (\text{Thomas 408, Thomas 409}) = 0.9993 \]
Thomas son of Josce matched
Vicars Choral Network (detail)
Top-ranking Matches

1. vicars-choral-408, Thomas, is_of(York), occupations_is(goldsmith), is_parent_of(Jeremy, Jeremias)
   vicars-choral-409, Thomas, is_of(York), occupations_is(goldsmith), is_parent_of(Jeremy)

2. vicars-choral-151, Ade, occupation_is(chaplain), is_child_of(Robert le hatter),
   is_landlord_of(Transaction)
   vicars-choral-152, Adam, occupation_is(chaplain), is_child_of(Robert),
   is_landlord_of(Transaction)

3. vicars-choral-445, Robert Warde, is_of(York), occupation_is(merchant),
   is_widow(Beatrice)
   vicars-choral-447, Robert Warde, is_of(York), occupation_is(merchant),
   is_widow(Beatrice)

... 

9. vicars-choral-393, Henry de Bryeland, is_of(York), occupation_is(merchant)
   is_landlord_of(Transaction)
   vicars-choral-394, Henry de Bryeland, is_of(York), occupation_is(merchant)
   is_landlord_of(Transaction)

10. vicars-choral-403, mag. Simon de Hevesham, occupation_is(precentor of York),
    is_landlord_of(Transaction)
    vicars-choral-405, Simon de Evesham, occupation_is(precentor of York),
    is_landlord_of(Transaction)
third property on Petergate
Data Mining and Workbench

- Automatically annotated collection
- Record Linkage
- Historian
- ChartEx Workbench
ChartEx Workbench

ChartEx Virtual Workbench v1.0

Search:

Thomas

Archives included in search:
- Borthwick
- TNA
- DEEDS
- Vicars Choral

Search Results

Entity S66: Petergate

Related Documents
- VC408
- VC495
- VC497
- VC503
- VC634

Related Entities
- Vicars Choral 408
  - Person
    - is granted by P99
    - is received by P98
  - Site
    - has length marker S67
    - has length marker S68
    - has breadth marker S69
    - has breadth marker S66
  - Transaction
    - Vicars Choral 495
    - Vicars Choral 497
    - Vicars Choral 503
    - Vicars Choral 634

Document Text

Lorem ipsum dolor sit amet, Petergate. Donec sit amet felis purus. Nunc ac est justo, eu tristique tortor. Sed
triginta vestibulum maurus ut velit posset. Sed in mi vitae dolor eleifend porta vel in quam. Curabitur rhoncus
tellus quis justo. Nam tristique vitae elit vel faucibus. Curabitur eu tempor odio, et ultrices non.

Cosa aptent tali sociisque ad lecta torquent per conubia nostra, per incessam himenas. Vivamus mattis nisi
sepem, nec imperdiet enim. Curabitur eu eu odio, id aliquam du. Proin nec sagittis tellus. Presertim ut risus
eus rutrum tellus metus. Vestibulum morte ultramcorper sagittis.

Show markup:
- Person
- Place
- Transaction
- Institution
- Site
- Date

Document Images

Document Details

Language: English
Date: January 1292
Held by: York Online Digital Library
Source reference: VC 3VI 326

http://www.chartex.org/docs/Chartex-Montreal-12102013-VIDEO.mov
Conclusions

- Record Linkage fairly doable
- Natural Language Processing
  - moderately successful
  - work in progress
  - Latin?
- Many more opportunities when done ‘big’
- Not sure about graphical approach
  - tabular?
- 2 years is too short!