ChartEx: Discovering spatial descriptions and relationships in medieval charters

www.chartex.org

@ChartExProject
ChartEx Project Partners

**University of York:** History and Human Computer Interaction

**University of Brighton:** Natural Language Processing

**University of Leiden:** Data Mining

**University of Washington:** History, Web Services

**University of Toronto:** History and Digital Archives

**Columbia University:** History and Digital Libraries

**Data Repositories:** The National Archives (UK), Borthwick, DEEDS Project U of Toronto, Columbia Digital Humanities

**Advisory Panel:** various cognate projects
Digital Charter Collections


The National Archives (UK) Ward 2. Abridged English translations

The Borthwick Institute, Yarburgh Muniments. Abridged English translations

DEEDS, University of Toronto. Latin charters of English provenance.

Cluny. Latin charters of French provenance
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
Chronological Sequencing and Spatial Sequencing

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.

409. Grant by Mariot widow of Thomas son of Josce goldsmith of York and by Jeremy (Jeremias) son of Thomas and Mariot to mag. Simon de Evesham canon of York of land with buildings in Steyngate, granted to them by Thomas, lying in length between Petergate and land of the prebend of Ampleford, and in breadth between Steyngate and land once of Geoffrey de Norwyz precentor of York.
ChartEx Architecture

ChartEx workbench

Charter documents

Natural language processing

Analysed individual documents

Data mining

Analysed integrated documents
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.

The NLP Task

Get from a text like this ...

... to some ‘semantic’ data like this
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.

... actually it looks more like this (BRAT markup)

(http://brat.nlplab.org/)
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.
NLP development methodology

Step 1 – find out what the mark up should be

Step 2 – ask historians to mark up some documents manually

Step 3 – use some of these examples to develop system

Step 4 – use the rest to evaluate performance
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;

**Token layer:** identify semantic types and intrinsic properties (eg gender) of known individual words (not all of these feature in final output).
NLP – layered pattern matching

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;

Lexical layer: identify simple lexical phrases – groups of tokens that act as individual units. Promote other tokens to lexical items.
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
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;

Semantic layer: build semantic relationships from syntactic phrases
Matching Relational Information

Vicars Choral 408

Vicars Choral 409
Networks of People
Charter Networks

corner
Steyngate/Petergate

third property
on Petergate
The ChartEx Virtual Workbench

Helen Petrie, Chris Power, Dave Swallow, 
*University of York*
Next Steps

**Final integration and evaluations**
Fall/Autumn 2013 (sessions in York and in New York, week of 4 – 8 November)

**Dissemination**
Not just Digital Humanities (eg novel NLP approach)

**Impact**
National and regional archives in the UK
Commercial applications?

**Follow-on funding**
Several bids building on ChartEx submitted
ChartEx: Discovering spatial descriptions and relationships in medieval charters

www.chartex.org
@ChartExProject