

A Text Analytic Approach to Rural and Urban Legal Histories

Adam Wyner - School of Computing and Physical Sciences

Jackson Armstrong and Andrew MacKillop - School of Divinity, History, and Philosophy

Phil Astley – Aberdeen City & Aberdeenshire Archives

1 4 9 5



UNIVERSITY
OF ABERDEEN

Archivists' Meetup
January 21, 2014
Aberdeen, Scotland



ABERDEEN
CITY COUNCIL

Project Specifics

- Collaboration between Computing, History, and Aberdeen Council.
- Funded by dot.rural partnership for digital economy - about £53k for six months.
- Pays primarily for
 - Associate investigator – W. Peters, a text analyst in Computing at Univ of Sheffield.
 - Research fellow - ?, interpretive tasks on source texts in middle Scots and Latin.
- Starting soon.

Technical Objectives

- To apply automatic text analytic tools to annotate and enrich transcribed Aberdeen Burgh Records (output of RISS Project).
- To query, extract, reuse, and link to other web-based documents (dictionaries, maps, DBPedia, other archives, etc.) the machine-readable content.
- Advantages: flexible, extensible, reapplicable, fast, efficient, multi-dimensional.
- Build text analytic capacity at Aberdeen.

Historical Objectives

Digital Humanities

- To carry out deep conceptual analysis based on textual data, e.g.:
 - Identify the individuals with respect to legal roles in jurisdictions.
 - Trace social/legal networks.
 - Follow the evolution of social/legal concepts.
 - Highlight historical resource management and conflict.
- To address the questions using *semantic search* within and *across* texts, no matter the size of the text. Results are in textual context.

Council and Archivists' Objectives

- To increase public access to archival materials for tourism, education, business, and research.
- To draw out, augment, and link the content of the materials.

Text Analysis

- Different things to different people, depending on source material, purposes, tools, and resources.
- We focus on *enriching the source text with annotations in order to facilitate semantic search*.
- For our historical language, the source is complex (e.g. lexical variation), the available support components very limited (electronic lexicons, parsers, and so on), and the resources constrained (funding and time).

Example

- We know that the string of characters *John* is a first name, but this is opaque to a machine.
- To make it transparent to the machine, we automatically annotate the string with machine-readable XML *tags* (computers do string processing very quickly and accurately):

```
<first_name>John</first_name>.
```
- Search for the tags and what comes between them, e.g. find all strings annotated as `first_name` in the corpus, e.g. John,...

GATE

- General Architecture for Text Engineering (GATE) - open source framework which supports plug-in NLP components to process a corpus of text.
- A GUI to work with the tools.
- A Java library to develop further applications.
- Components and sequences of processes, each process feeding the next in a “pipeline”.
- Annotated text output inline or offset XMLish formats.
- Word lists are the basis of the analysis. Can construct complex annotations from basic annotations. Queries are used to explore and extract complex patterns.

Annotations - Human View

Printed Source: E. Gemmill, ed., Aberdeen Guild Court Records (2005), pp. 88-91
MS source: Aberdeen Council Register V(2), pages as numbered 689 – 691
[689]
Curia gilde huius burgi tenta per Johannem de Marr', **prepositum**, vijmo die mensis Februarii, anno Domini, etc., xlij [=7 Feb 1442/3, session of the guild court]
Quo die Johannes Bertlotson' primo die vocatus ad intrandum coram **preposito** non comparauit, etc.
Robertus Blyndseil secundo die vocatus ad intrandum Willelmum Bird, hominem suum, etc.
Gilbert Meignes, William Scherar', Thomas Blyndseil and Joh[n]n[e] Gray, arbitra[irs] chosin betuex
Willame Voket, eldar' and Gilbert Coly apon the debate movit betuex thaim because of a certane pak of wal[d] thai have ordanit that how mykil is sald of the wald the mo[ne] of it sal be partit evynly amangis thaim, and the remanand of the wald to be partit right swa and at this be done under gude fai[th]
Item, because that Matheu Fichet said it was cum til his understand[ing] that Maroiune of Crondane has complengnit to the lord of Erole that he had al the wite and wilfully put hir fra hir' watt[eris] the quhilkis scho had of befor, quhar'for he askit a declaracion of the **counsaile** quhethir he was cause of the forsaid mater or nocht, and than he removit and the **counsaile** habely decretit that he had made na sik cause na he had na wite tharof.
Thir' ar' the names

Type	Set	Start	End	Id	Features
Office		210	220	729	{rule=Office}
Office		384	393	732	{rule=Office}
Office		1182	1191	746	{rule=Office}

- IdeologicalTerm
- LegalBody
- LegalConcept
- LegalRole
- Location
- Lookup
- Name
- Offence
- Offender
- Office
- RegisterEntry
- Original markups

Annotations - Human View

Printed Source: E. Gemmill, ed., **Aberdeen Guild Court** Records (2005), pp. 88-91

MS source: **Aberdeen** Council Register V(2), pages as numbered 689 – 691

[689]

Curia gilde huius **burgi** tenta per **Johannem de Marr'**, prepositum, vijmo die mensis Februarii, anno Domini, etc., xlij [=7 Feb 1442/3, session of the **guild court**]

Quo die **Johannes Bertlotson'** primo die vocatus ad intrandum coram **preposito** non comparauit, etc.

Robertus Blyndseil secundo die vocatus ad intrandum **Willelmum Bird**, hominem suum, etc.

Gilbert Meignes, **William Scherar'**, **Thomas Blyndseil** and **Joh[n]n[e] Gray**, arbitra[irs] chosin betuex

William Voket, **eldar'** and **Gilbert Coly** apon the debate movit betuex thaim because of a certane pak of wal[d] thai have **ordanit** that how mykil is sald of the wald the mo[ne] of it sal be partit evynly amangis thaim, and the remanand of the wald to be partit right swa and at this be done under gude fai[th]

Item, because that **Matheu Fichet** said it was cum til his understand[ing] that **Maroiune of Crondane** has complengnit to the **lord of Erole** that he had al the wite and wilfully put hir fra hir' watt[eris] the quhilkis scho had of befor, quhar'for he askit a declaracion of the **counsaile** quethir he was cause of the forsaid mater or nocht, and than he removit and the **counsaile** habely decretit that he had made na sik cause na he had na wite tharof.

Thir' ar' the names

Type	Set	Start	End	Id	Features
Location		33	41	722	{rule=Location}
RegisterEntry		42	53	723	{rule=RegisterEntry}
Location		91	99	724	{rule=Location}
LegalBody		157	162	725	{rule=LegalBody}
LegalBody		163	168	726	{rule=LegalBody}

- IdeologicalTerm
- LegalBody
- LegalConcept
- LegalRole
- Location
- Lookup
- Name
- Offence
- Offender
- Office
- RegisterEntry
- ▶ Original markings

Annotations - Machine View (XML)

```
<html><head><meta content="application/vnd.openxmlformats-officedocument.wordprocessingml.document"
  name="Content-Type"><meta content="Gemmill extract.docx" name="resourceName"><title></title></
  meta></meta></head><body><p>Printed Source: E. Gemmill, ed., <i><Location>Aberdeen</Location> <
  RegisterEntry>Guild Court</RegisterEntry> Records </i>(2005), pp. 88-91</p>
<p>MS source: <Location>Aberdeen</Location> Council Register V(2), pages as numbered 689 - 691</p>
<p><a name="_GoBack"></a></p><p>[689]</p>
<p><LegalBody>Curia</LegalBody> <LegalBody>gilde</LegalBody> huius <LegalBody>burgi</LegalBody>
  tenta per <Name>Johannem de Marr'</Name>, <Office>prepositum</Office>, vijmo die mensis
  Februarii, anno Domini, etc., xlij [=7 Feb 1442/3, session of the <RegisterEntry>guild court</
  RegisterEntry>]</p>
<p>Quo die <Name>Johannes Bertlotson'</Name> primo die vocatus ad intrandum coram
  <Office>preposito</Office> non comparavit, etc.</p>
<p><Name>Robertus Blyndseil</Name> secundo die vocatus ad intrandum <Name>Willelmum Bird</Name>,
  hominem suum, etc.</p>
<p><Name>Gilbert Meignes</Name>, <Name>William Scherar'</Name>, <Name>Thomas Blyndseil</Name> and <
  Name>Joh[n]n[e] Gray</Name>, arbitra[irs] chosin betuex <Name>Williame Voket</Name>, eldar' and
  <Name>Gilbert Coly</Name> upon the debate movit betuex thaim because of a certane pak of wal[d]
  thai have <RegisterEntry>ordanit</RegisterEntry> that how mykil is sald of the wald the mo[ne]
  of it sal be partit evynly amangis thaim, and the remanand of the wald to be partit right swa
  and at this be done under gude fai[th]</p>
<p>Item, because that <Name>Matheu Fichet</Name> said it was cum til his understand[ing] that
  <Name>Maroiune of Crondane</Name> has complengnit to the <Name>lord of Erole</Name> that he had
  al the wite and wilfully put hir fra hir' watt[eris] the quhilkis scho had of befor, quhar'for
  he askit a declaracion of the <Office>counsaile</Office> quhethir he was cause of the forsaid
  mater or noght, and than he removit and the <Office>counsaile</Office> habely decretit that he
  had made na sik cause na he had na wite tharof.</p>
```

Relations in the Register

{Name}({Token})*5{Office}

Corpus: Entire datastore Annotation set: All

Results: Context size:

Search Clear Next page of 50 results

Context of the contribucion to my Elizabeth' of Britane the hale counsaile has ordanit that it be

Token.string of the contribucion to my Elizabeth' of Britane the hale counsaile has ordanit that it be

Name Elizabeth'

Token of the contribucion to my Elizabeth' of Britane the hale counsaile has ordanit that it be

Office of Britane the hale counsaile

Page 1 (4 results) Export

Global

Left context	Match	Right context	Features
of the contribucion to my	Elizabeth' of Britane the hale counsaile	has ordanit that it be	Token.string=Elizabe
session of common council] that	Robert Blindzelle and Will' Woket, ball'	and resawaris of the Martymes	Token.string=Robert
pay the xxx lib' to	John of Mar' alderman	to wirk the commownn werk	Token.string=John, c
under[wrytyn] sal be awd[itouris] to	John' of Fiffis cwn...ice of aldermanschip	of the yer. In primis	Token.string=John', .

Annotation Type	Count
Token	33120
Lookup	143
Name	64
P	28
Office	18
Location	16
RegisterEntry	15

Relations in the Register

{Office}({Token})*5{RegisterEntry}

Corpus: Entire datastore

Results:



Search



Clear



Context neid beis / Harman / king / The bailyeis chargit roby king cordonar to pay

Token.string neid beis Harman king The bailyeis chargit roby king cordonar to pay

Office

Token

RegisterEntry

Page 1 (4 results) Export

Left context	Match	Right context	Features	Que
neid beis / Harman /	king / The bailyeis chargit	roby king cordonar to pay	Token.string=king, /...e, bailyeis, chargit	{Offic
burgi balliuos / andersoun /	King / The bailyeis Chargit	Sandris king to pay Dauid	Token.string=King, /...e, bailyeis, Chargit	{Offic
said is / forbes /	king / The bailyeis Chargit	patre forbes to Enter thome	Token.string=king, /...e, bailyeis, Chargit	{Offic
Elizabeth' of Britane the hale	counsaile has ordanit	that it be raisit and	Token.string=counsaile, has, ordanit	{Offic

Project and Beyond

- Understand better how to apply text analytic tools to develop the source material for different interests and purposes.
- Refine and add semantic annotation types (translation, part of speech, word lemmas, ...).
- Link to external web resources.
- Disseminate the methodology and materials.
- Provide a web-based interactive interface.
- Create more projects.

Thanks for your attention!

- Comments, questions, suggestions....