NEW METHODS FOR THE ANALYSIS OF DIGITIZED MEDIEVAL LATIN CHARTERS

by Michael Gervers and Michael Margolin

Dramatic advances in information technology and significantly increased ease of access to the Internet have produced new and exciting research methods for historians. The DEEDS Project at the University of Toronto in Canada has built a corpus of about 9,000 securely dated medieval English charters which are accessible from the World Wide Web. This paper describes the application of advanced research and presentation methods to an analysis of scribal variations among a group of about 1,300, mostly undated, medieval Latin charters from The Cartulary of the Knights of St. John of Jerusalem in England. Research methods include a wide array of computerized textual analyses: full and segregated content searches, inline phonetic transformations, computational geometric representations, statistical processing of results, and the interactive graphical interpretations of those results.

Records of property transfer are as ancient as writing itself. Consequently, the deed, or conveyance, represents the most continuous form of legal documentation available to the historian. The current objective of the DEEDS Project, founded at the University of Toronto in 1975, is the provision of computerized access to the content of English conveyances of the twelfth and thirteenth centuries.[1] Our main priority at the moment is to

---

1 Gervers, Michael, The DEEDS Project: Towards the dating and analysis of English private charters of the twelfth and thirteenth centuries, in: Le Médiéviste et l'Ordinateur 41 (2002), S. 60-66. The internet version can be consulted at <http://www.irht.cnrs.fr/meto/mo41_07.htm> (12.10.2006). See also, Michael Gervers, "The Deeds Project and the Development of a Computerised Methodology for Dating Undated English Private Charters of the Twelfth and Thirteenth Centuries", in: Michael Gervers (Hg.), Dating Undated Medieval Charters, Ro-

develop the means to provide dates for private charters, 92 per cent of which, from the time of William the Conqueror in 1066 to the accession of King Edward II in 1307, were issued without chronological indicators.[2] The questions which can be addressed through the content of our databases are limitless, but to give a few examples of our research directions in recent years we may cite the following: 1) a study of the donations to the twelfth-century Order of the Hospital of St. John of Jerusalem which led to the conclusion that the Order became militarized in response to the fall of Edessa in 1144 and the call for the Second Crusade in 1145[3]; 2) changing forms of address appearing in grants to the Hospitallers in England during the twelfth and thirteenth centuries, which closely reflect the nature and degree of the Order's presence in, or absence from, the Holy Land[4]; 3) the relative growth of the estates of the military orders in London and its suburbs from the twelfth to fourteenth centuries[5]; and 4) evidence of social unrest in England under the Interdict (1208-1214).[6] We are presently

---

chester–Woodbridge 2000, pp. 13-35; Fiallos, Rodolfo "An Overview of the Process of Dating Undated Medieval Charters: Latest Results and Future Developments", in: idem, pp. 37-48.

2   Gervers, Michael, "The Dating of Medieval English Private Charters of the Twelfth and Thirteenth Centuries", in: Brown, Jacqueline; Stoneman, William P. (Hgg.), A Distinct Voice. Medieval Studies in Honor of Leonard E. Boyle, O.P., Notre Dame (Indiana) 1997, S. 455-504; idem, "Identifying Irregularities and Establishing Chronology in Medieval Charters", in: Keats-Rohan, Katharine S.B. (Hg.), Resourcing Sources (Prosopographica et genealogica 7), Oxford 2002, S. 164-78.

3   Gervers, Michael, "Donations to the Hospitallers in England in the Wake of the Second Crusade", in: Gervers, Michael (Hg.), The Second Crusade and the Cistercians, New York 1992, pp. 155-61.

4   Gervers, Michael, "Changing Forms of Hospitaller Address in English Private Charters of the Twelfth and Thirteenth Centuries", in: Laszlovszky, József; Hunyadi, Zsolt (Hgg.), The Crusades and the Military Orders: Expanding the Frontiers of Medieval Latin Chritsianity, Budapest 2001, pp. 395-405.

5   Gervers, Michael, "The Commandery as an Economic Unit in England", in: Luttrell, Anthony; Pressouyre, Léon, La Commanderie, Institution des orders militaries dans l'Occident médiéval, Paris 2002, pp. 245-50.

6   Gervers, Michael and Nicole Hamonic, "Pro Amore Dei: Diplomatic Evidence of Social Conflict During the Reign of King John", in: Papers in honour of James A. Brundage. Publication anticipated in 2007.

researching means to identify the medieval English scribe of the late twelfth through thirteenth centuries as an author.

Our Corpus of digitized medieval charters was started a decade ago using the Oracle version 5 database running on a 386-33 PC. At that time we had a set of seven main tables, namely DOCUMENT, PERSON, PROPERTY / COMPENSATION, LEASE, RELATION, LINKAGE and ROLE, with a variety of supporting code tables.[7] Encoding information was achieved at the data-entry level by means of a linguistically based coding 'language'. Three types of 'sentences' were used to describe connections between people, property, and people and property. Actions, tense, number and mode were similarly coded. It worked very well for the time.

However, today, the DEEDS Corpus, together with metadata, is hosted by the latest Oracle database and is widely available on the World Wide Web.[8] The Corpus presently includes two groups of Latin charters, mainly from twelfth- and thirteenth-century England, derived either from printed sources or transcribed directly from the manuscripts (Figure 1).

---

7   Gervers, Michael; Long, Gillian; McCulloch, Michael, "The DEEDS Database of Mediaeval Charters: design and coding for the RDBMS ORACLE 5", in: History and Computing, 1 no. 3 (1990), pp. 1-12.

8   <http://www.utoronto.ca/deeds/research/research.html> (##.##.2006).

Figure 1:

The first group, of about 9,500 charters, has been dated internally or by the editor of the manuscript, using internal evidence (Figure 2).

59.

[John de Haliwell grants to Reginald the Beadle land in Grove Street.]

Sciant presentes & futuri quod ego Iohannes de Halliwelle corui- C. 6. 1
sarius assensu & consensu Alicie uxoris mee & amicorum meorum Mich.
dedi & concessi & quietumclamaui³ [&c.] pro me & heredibus meis Mich.
Reginaldo seruienti Uniuersitatis Oxonie totum ius meum & clama- 1261
mium [sic] quod habui vel quod habere potui [vel]⁴ quocumque
modo possit mihi descendere de tota illa terra cum omnibus per-
tinenciis suis que est inter terram que quondam fuit magistri Hugonis
de Seydegerd ex una parte & terram que quondam fuit Iohannis de
Chelsee ex altera in Groppecuntelane in parochia beate Marie virginis
Oxonie; habendum & tenendum predicto Reginaldo & heredibus suis
vel suis assignatis libere & quiete, integre, bene & in pace, faciendo
inde capitalibus dominis illius feodi seruicium debitum & consuetum.
Pro hac autem donatione [&c.] predictus Reginaldus dedit michi pre

¹ Peter de Kyllum was rector from 1248 to the end of 1274.
² Bracton, iv. 367 (Rolls Series) speaks of an important lawsuit by the rector of St. Mary's who was next in succession after William Hardel; this was Peter de Kyllum. Bracton's language rather suggests that William Hardel had treated the land of the rector as his private property and had sold it. Bracton says the case was tried before the King at Woodstock, and as it is generally assumed that he refers to nothing later than 1254, our deed would be of 1248–1254.
³ The use of this word shows that Reginald already had a rent or some other share in the tenement.
⁴ Not in MS.

manibus centum solidos sterlingorum in gersummam in magno
negocio meo. Ut igitur [&c. sealing], hiis testibus, Adam Feteplace
tunc maiore Oxonie, Galfrido Aurifabro, Ricardo filio Nicholai tunc
balliuis, Willelmo de Wintonia espeicer, Willelmo le Speicer iuniore,
Thoma le Speicer, Willelmo de Eu, Henrico Inge, Ricardo Mares-
callo, Willelmo Russello cordewanario, Henrico Barbatore, Reginaldo
Illuminatore & aliis.¹

¹ This deed and the next refer to a property mentioned in deeds 141 and 142, lying on the east side of Bedel Hall.

Figure 2:

The second group, of about 1,300 mostly undated charters, comes from The Cartulary of the Knights of St. John of Jerusalem in England (British Library, Cotton ms. Nero E vi).[9] In the computer, they are stored independently, but both groups are available for computerized analyses. Each charter is stored in digitized form, together with a photographic image of the original printed version, or of the manuscript itself. The text and any accompanying information is extracted from the original source by scanning, using the Optical Character Recognition Program (OCR) or by transcription directly from the manuscript. We have developed a program which we call "Document Manager" to convert the text of the charter, and all external and analytical data, into an Extensible Markup Language (XML) document which is later sent to the database (Figure 3). Each document stored in the database is the source for the searchable online version of that charter in Hyper Text Markup Language (HTML).

---

9   Gervers, Michael, The Hospitaller Cartulary in the British Library (Cotton MS Nero E VI). A Study of the Manuscript and its Composition, with a Critical Edition of Two Fragments of Earlier Cartularies for Essex, Toronto 1981; idem, The Cartulary of the Knights of St. John of Jerusalem in England: Essex, 2 vols., London 1982-96.

```
- <sourceDocument dnum="00930059" created="2003-08-19" cartulary="Oriel College -
    Oxford">
    <content>Sciant presentes et futuri quod ego Iohannes de Halliwelle coruisarius
     assensu et consensu Alicie uxoris mee et amicorum meorum dedi et concessi et
     quietumclamaui etc pro me et heredibus meis Reginaldo seruienti Uniuersitatis
     Oxonie totum ius meum et clamamium quod habui vel quod habere potui [vel]
     quocumque modo possit mihi descendere de tota illa terra cum omnibus
     pertinenciis suis que est inter terram que quondam fuit magistri Hugonis de
     Seydegerd ex una parte et terram que quondam fuit Iohannis de Chelsee ex
     altera in Groppecuntelane in parochia beate Marie virginis Oxonie habend et
     tenend predicto Reginaldo et heredibus suis vel suis assignatis libere et quiete
     integre bene et in pace faciendo inde capitalibus dominis illius feodi seruicium
     debitum et consuetum Pro hac autem donatione etc predictus Reginaldus dedit
     michi pre manibus centum solidos sterlingorum in gersummam in magno
     negocio meo Ut igitur etc hiis testibus Adam Feteplace tunc maiore Oxonie
     Galfrido Aurifabro Ricardo filio Nicholai tunc balliuis Willelmo de Wintonia
     espeicer Willelmo le Speicer iuniore Thoma le Speicer Willelmo de Eu Henrico
     Inge Ricardo Marescallo Willelmo Russello cordewanario Henrico Barbatore
     Reginaldo Illuminatore et aliis</content>
  + <data>
  + <notes>
  + <markup>
  </sourceDocument>
```

Figure 3:

There are several ways to access DEEDS digitized data:

1. Remotely: over the Internet, using File Transfer Protocol (FTP), connecting directly to the DEEDS database server, using Dedicated Web Services
2. Locally: from workstations that have DEEDS applications installed on their hard drive.

Charters may be viewed over the Internet using the "Browse" option on the DEEDS Website. This will start the "Document Browser" (Figure 4) program which allows one to select the text of a particular charter, view the original source, and the title page of the printed edition. The source cartulary can be chosen by name or by its internal code. In addition, the "Document Browser" displays the diplomatic parts of the charter and formulae, marked up according to a predefined color legend.

Figure 4:

A second online program permits textual queries to be directed to the chosen collection of charters. This can be started from the DEEDS website by selecting the "Search" option. This program supports exact, fuzzy and proximity text patterns. There are also multiple options for further refining the query by specifying meta data restrictions, such as the nature of the charter, the time span, the source of the text, the principal individuals appearing in it, etc. The program spans two panels. The first displays query restrictions while the second renders a search result in one of the three available modes:

1. „Search",
2. „Context",
3. „Chart".

When the "Search" mode is selected, the lower panel will display a list of charters together with their dates. Any charter from this list can be

viewed with a highlighted query term (Figure 5). The same view also allows one to find a word or phrase in the current document.
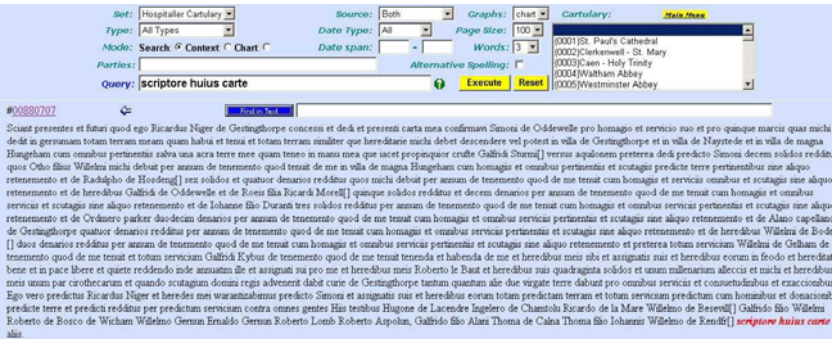
Figure 5:

When the "Context" mode is selected, each query term will be shown within its context. The context boundaries can be expanded from three words before and after the queried pattern, up to a maximum of ten words. These context words can be alphabetized independently, starting from the word nearest to the query term and moving either forwards or backwards (Figure 6).
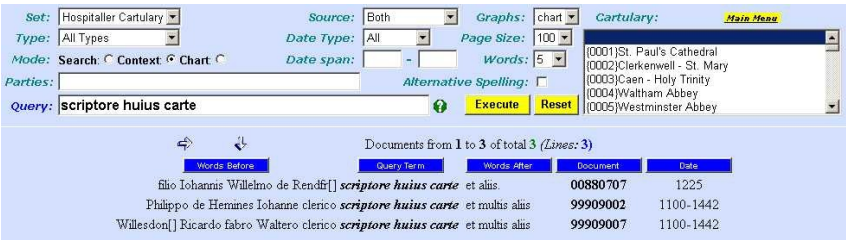
Figure 6:

When the "Chart" mode is selected the lower panel will display a graphical representation of the query using Scalable Computer Graphics (SVG)

format. SVG is an XML dialect and therefore is portable across different hardware platforms and operating systems. An additional advantage of SVG is that all graphics can be scaled to any degree without loss of image quality and direct user interaction is allowed (Figure 7). There is also one more advanced option called "Alternative Spelling". When this option is enabled some common spelling variations, like double consonants, are removed.

Figure 7:

We have recently successfully applied our online facilities to explore scribal references in The Cartulary of the Knights of St. John of Jerusalem in England, in an attempt to identify authorship, or "schools" of writing (Figure 8).

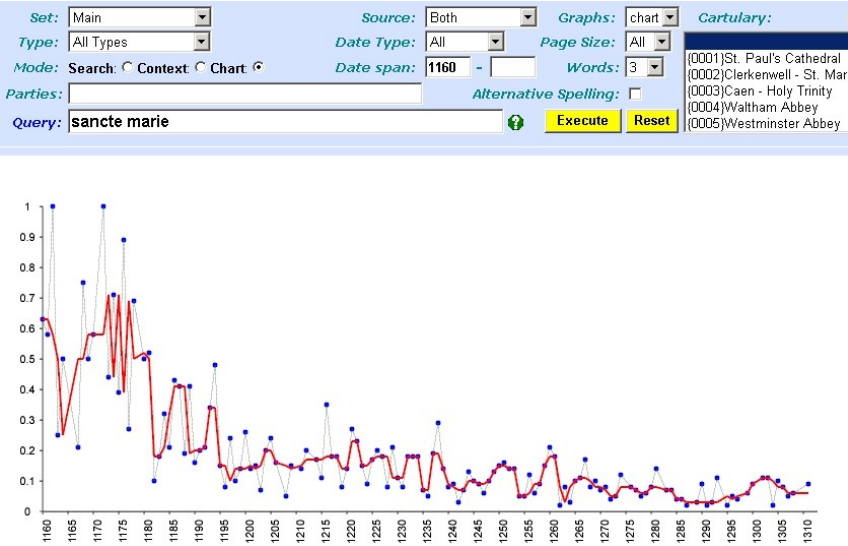|	Document	date	name	Location (titulus)	Phrase
1	Document	date	name	Location (titulus)	Phrase
2	99909001	undated	Adam diaconus de Angemare	Sumpting	presentis scripti notario
3	00880294	1180 c.	Adam scriptor	Chaureth - Chrishall	scriptore
4	99909008	undated	Edmundus	Chippenham	qui hanc cartam scripsit
5	00870019	1190 c.	Harnelinus clericus	Roydon	qui hanc cartam fecit
6	00880046	1200-1204	Hugo capellanus	Cressing-Witham: Hatfield Peve	presentis carte scriptore
7	00880394	1155 c.	Hugo scriptor	Chaureth - Henham	scriptore
8	99909004	1257-1258	Iohannes clericus	Hampton	qui hanc cartam composuit
9	99909002	undated	Iohannes clericus	Hampton	scriptore huius carte
10	99909009	undated	Iohannes de Shryneham clericus	Chippenham	qui presens scriptum composuit
11	99909000	undated	Magistro Philippo medico	Ravensthorpe	qui hanc cartam scripsit
12	00870095	1269 -	Nicholaus clericus	Rainham	huius scripti notario
13	99909003	undated	Petrus capellanus de Codham	Hampton	qui hanc cartam scripsit
14	00880541	1235-1245	Petrus capellanus de Hastede	Gestingthorpe	qui hanc cartam scripsit
15	00870103	1280 c.	Radulphus de Boughton	West Thurrock	presencium scriptore
16	00880698	1245 c.	Reginaldus clericus	Gestingthorpe	qui hanc cartam fecit
17	00880734	1230 c.	Robertus de Stistede	Gestingthorpe	qui hanc cartam scripsit
18	00880940	1255 c.	Simon clericus	Sampford	qui hanc cartam fecit
19	99900274	1219 -	Stephanus de Esseleia	London	qui hoc scripsit
20	99909006	undated	Thomas clericus	Hampton	qui hanc cartam scripsit
21	00880319	1190-1200	Walterus	Chaureth - Sawbridgeworth	qui cartam scripsit
22	00880523	1230 c.	Walterus clericus	Bumpstead Helion	huius scripti scriptore
23	99909007	undated	Walterus clericus	Hampton	scriptore huius carte
24	00880538	1185 c.	Walterus de Hull clerico magistri	Bumpstead Helion	qui hanc cartam scripsit
25	99909005	undated	Willelmus capellanus	Hampton	qui hanc cartam scripsit
26	00880707	1225 -	Willelmus de Rendfr[]	Gestingthorpe	scriptore huius carte
27	00880155	1220-1230	Willelmus de Salsetun	Little Maplestead	qui scripsit hanc cartam
28	00880512	1230-1240	Willelmus filius Derkini	Bumpstead Helion	qui hanc cartam scripsit
29	99900283	undated	Willelmus filius Derkini	Herefeld	qui hanc cartam scripsit

Figure 8:

Several more programs have been developed to work with digitized charters but, for now, they are only accessible from the DEEDS departmental workstations. We use these programs for the maintenance of digitized documents and for performing chronological and content-driven textual analysis. However, it is our intention to make those programs also accessible remotely over the Internet.

We use a set of securely dated charters to draw any quantitative conclusions with respect to chronology or content. Obviously, the number of dated charters available for each time span varies, as does the accuracy of the chronological evidence, which varies from the exact day, month and year to a range of several years. To overcome discrepancies in chronological attributes, we have developed a method of normalizing chronological information in order to facilitate computer-aided analyses. We convert the attributed date to the Julian calendar and then apply a special computer program to analyze the availability of charters over a given time span (in our case from 1050 to 1359). This program breaks up the total chronological span into periods when approximately equal numbers of dated charters are available. An index of this computer-

generated period (henceforth referred to as a "tile") is subsequently assigned to each charter. The tile value is later used as a proxy for the original date in all statistical computations. The size of the tile can vary from just one day to several years, while the number of charters referenced remains almost unchanged. The tile number assigned to the charter is automatically updated when charters are added to, or removed from, the collections. Both the actual and normalized availability of our first group of charters is shown in Figure 9.

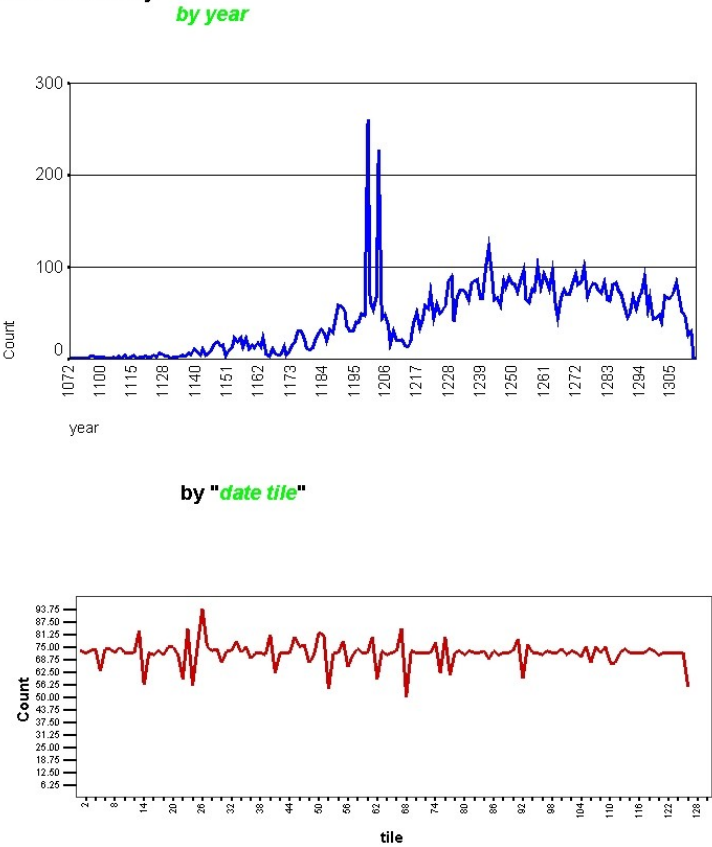**Charter availability:**
by year

by "date tile"

Figure 9:

We can apply computer-aided analysis to the charter text in order to evaluate the chronology and content. Since this is an official legal document written or issued by a religious, lay or royal institution, inevitably its vocabulary, structure and content reflect the time of its creation and also carry the "footprint" of the issuer. The core of our approach is that, by analyzing different aspects of the text, a link can be established between the given text and a set of similar charters whose attributes are known. The fact that all digitized charters are well attributed and are linked to a significant amount of meta data allows us to conduct computerized textual analyses from different angles, using the meta data for filtering information. This meta data includes elements such as: the type of legal action documented, information about the parties represented in the charter, details concerning the place of issue and locations, the name of the scribe, etc. One of its most important features is identification of the different diplomatic parts and any content-related formulae in the text of the charter. This information, which is currently available for about 50 per cent of our charters, can be accessed under the "Browse" option on our website menu.

The examination of vocabulary involves extracting combinations of two or more adjacent words, so-called "word-patterns", in consecutive order from the text of the charter and then finding occurrences of each pattern in a collection of digitized charters. The total number of valid word-patterns derived from a typical charter varies from hundreds to thousands, depending on the size of the text. All attributes of charters in which a given query term occurs are collected and later processed by the computer program. Depending on the circumstances, word-patterns can be generated in three different ways:

1. from the original text
2. from the partially normalized text
3. from the fully normalized text.

Normalization of the text (Figure 10) can include lexical transformations, phonetic transformations, or both. Partial normalization involves lexical transformation and includes the replacement of most

Roman and Arabic numerals, prepositions, definite articles and measurement units by a one-letter proxy in the text of the document. Full normalization adds phonetic transformation by further replacing words by a phonetic proxy. We tested different phonetic matching algorithms[10] and eventually decided to use the modified version of the Lawrence Philips' Double Metaphone algorithm for the phonetic transformations (Figure 11).[11]

Sciant presentes et futuri quod ego Iohannes de Halliwelle coruisarius assensu et consensu Alicie uxoris mee et amicorum meorum dedi et concessi et quietumclamaui etc pro me et heredibus meis Reginaldo seruienti Uniuersitatis Oxonie totum ius meum et clamamium quod habui vel quod habere potui [vel] quocumque modo possit mihi descendere de tota illa terra cum omnibus pertinenciis suis que est inter terram que quondam fuit magistri Hugonis de Seydegerd ex una parte et terram que quondam fuit Iohannis de Chelsee ex altera in Groppecuntelane in parochia beate Marie virginis Oxonie habend et tenend predicto Reginaldo et heredibus suis vel suis assignatis libere et quiete integre bene et in pace faciendo inde capitalibus dominis illius feodi seruicium debitum et consuetum Pro hac autem donatione etc predictus Reginaldus dedit michi pre manibus centum solidos sterlingorum in gersummam in magno negocio meo Ut igitur etc hiis testibus Adam Feteplace tunc maiore Oxonie Galfrido Aurifabro Ricardo filio Nicholai tunc balliuis Willelmo de Wintonia espeicer Willelmo le Speicer iuniore Thoma le Speicer Willelmo de Eu Henrico Inge Ricardo Marescallo Willelmo Russello cordewanario Henrico Barbatore Reginaldo Illuminatore et aliis

Figure 10:

scnt prsnt cj futr cj eg iohns pr hall corsr asns cj consn als uxrs me cj amkrm merm ded cj conss cj qutmk etk pr me cj herdp mes rejnl sernt unfrs oxn totm ius mem cj clmmm cj hap cj cj hapr pot cj qukmk mod post mih desnd pr tot il ter pr omnps pertn sus qu est pr term qu qundm fut majst hukns pr sedjr pr nm part cj term qu qundm fut lohns pr cels pr altr pr grpkn pr pars bet mar virjn oxn hapnd cj tennd prdkt rejnl cj herdp sus cj sus asnts lipr cj qut intkr ben cj pr pas fasnd ind captl domns ils fed sertm deptm cj const pr hak cj donxn etk prdkt rejnl dedt mih pr manps nm nt strln pr gersm pr man nekt me cj igtr etk his testp adm fetpl tunk mar oxn galfr aurfp rikrd fil nikl tunk balfs willm pr wintn espsr willm le spsr iunr tom le spsr willm pr eu henrk inj rikrd marsk willm rusl cordn henrk barpt rejnl ilmnt cj als

Figure 11:

We process the results that are returned by word-pattern queries in separate flows, depending on their nature, amount and quality:

10   Zobel, Justin; Dart, Philip, Phonetic String Matching: Lessons from Information Retrieval,   <http://goanna.cs.rmit.edu.au/~jz/fulltext/sigir96.pdf> (12.10.2006).
11   Philips, Lawrence, The Double Metaphone Search Algorithm, C/C++ Users Journal, June 2000,, <http://www.ddj.com/dept/cpp/184401251> (12.10.2006).

1. When a word-pattern produces more than 25 hits, each such distribution is evaluated independently.
2. When a word-pattern produces between one and 25 hits, results are first tabulated using all those of the same number and later evaluated using one for each word-pattern size.
3. All word-patterns which produce a single hit are accumulated and processed just once at the end of the process.

We accumulate intermediate results at various stages of the processing. All intermediate and final computations are performed by the integrated Statistical Engine (Figure 12).

#00930059 Gregorian date: 1260-61, Julian Date: 2181727, Tile: 79

1 hit
Total number of hits: 17
Date Tiles: 3 70 91 60 16 52 69 1 58 76 112 104 80 120 79 85 64
Statistics: 76

1..25 hits
WP size: 1: 75.7143
WP size: 2: 82
WP size: 3: 73.4022
WP size: 4: 75.1224
WP size: 5: 84.8056
WP size: 6: 95
WP size: 7: 94.5
WP size: 8: 90
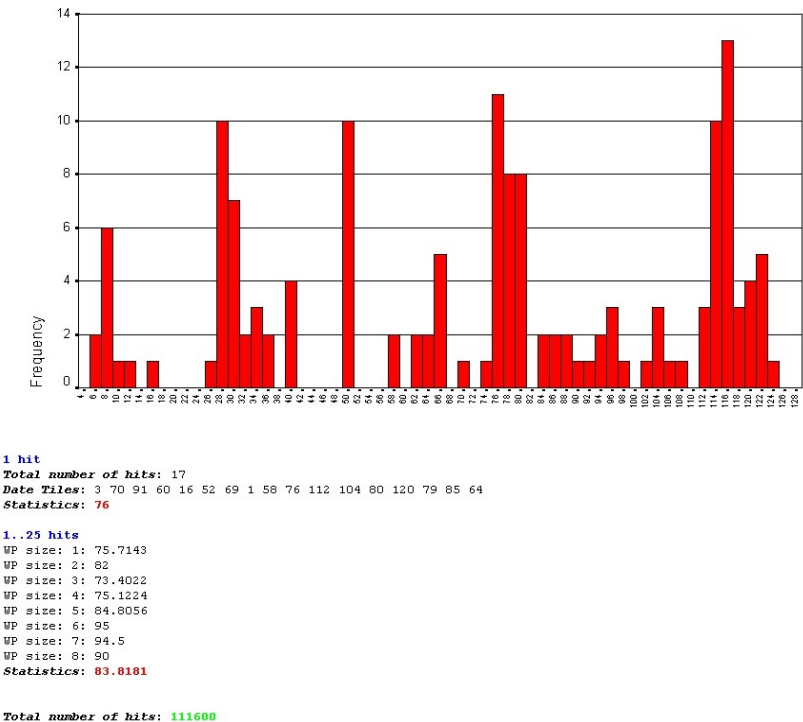Statistics: 83.8181

Total number of hits: 111600

Figure 12:

Frequently, when word-patterns generate multiple hits, they produce a distribution with peaks in multiple clusters. In this case the correct local distribution can be chosen by using results produced by the other two flows (Figure 13). The addition to the program of an industrial strength statistical engine has enabled us greatly to improve the quality of calculations through all stages of word-pattern processing. We apply Robust Statistical Routines and Fibonacci ratios base weighting to achieve our final results. An example of output (Figure 14) shows the accuracy of our fully-automated chronological evaluation of a sample of 100 charters using original word-patterns.
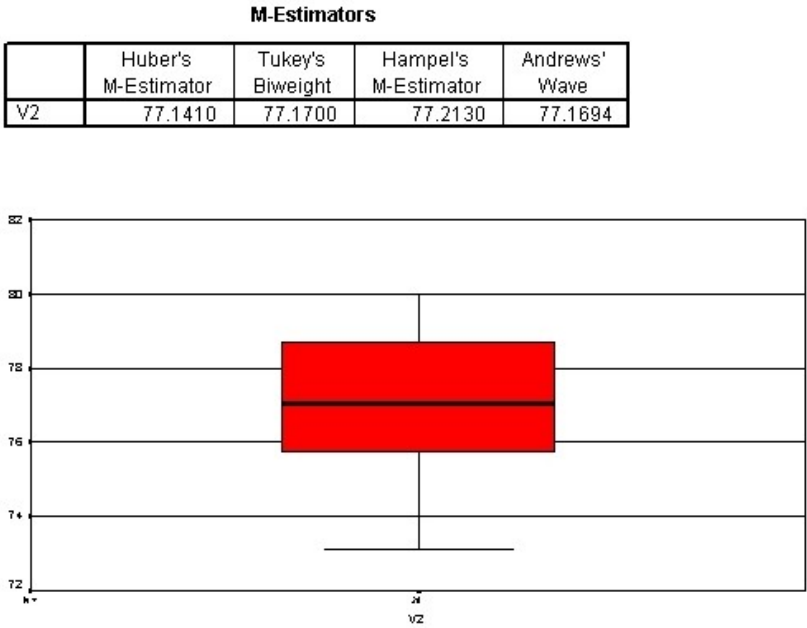
M-Estimators

	Huber's M-Estimator	Tukey's Biweight	Hampel's M-Estimator	Andrews' Wave
V2	77.1410	77.1700	77.2130	77.1694

Figure 13:

Average Error

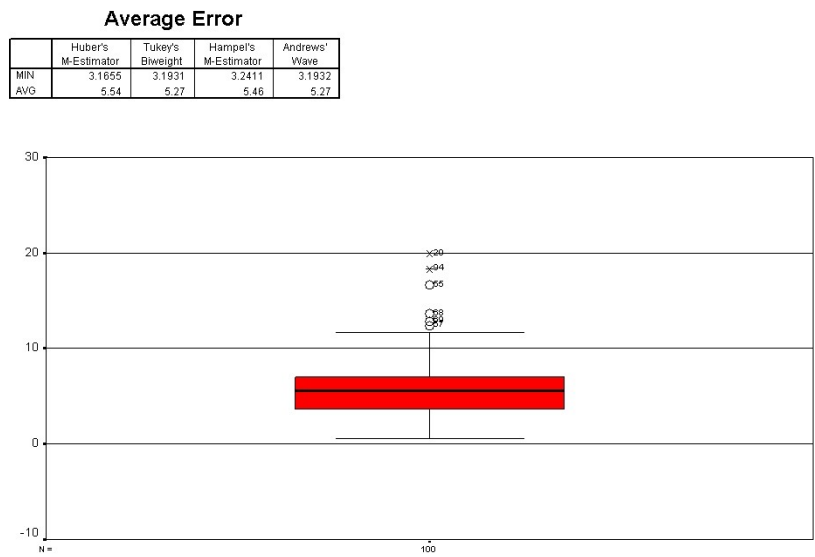|	Huber's M-Estimator	Tukey's Biweight	Hampel's M-Estimator	Andrews' Wave
MIN	3.1655	3.1931	3.2411	3.1932
AVG	5.54	5.27	5.46	5.27

<mark>Figure 14:</mark>

Using computer-generated geometrical content footprints during charter analysis appears to be a promising way of comparing charter content. According to this method, original text is converted to digital form by replacing each letter, including spaces, by their arithmetic ASCII value. After that, the program computes the frequencies of each value. At the next stage, a Computational Geometry program treats frequency/character value pairs as objects of virtual two-dimensional space. Later the program extracts a series of convex polygons (called "convex hulls") starting with the largest, using Onion Peeling Algorithms.[12] The smallest hull generated by the program is the one eventually used for a content representation of the text. Generally speaking, an overlap of hulls generated from different charters reflects the similarity of their content (Figure 15). The current

---

12 Poulus, Marious; Papavlasopoulos, Sozon; Chrissilopoulos, Vasilious, A Text Categorization Technique based on a Numerical Conversion of a Symbolic Expression and Onion Layers Algorithm, in: Journal of Digital Information, 6/1 (2004), article no. 276.

implementation of our program for content footprint evaluation uses the latest release of the Computational Geometry Algorithms Library (CGAL).[13]
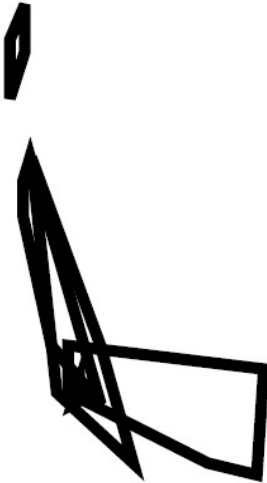
Figure 15:

To sum up, this paper has introduced the DEEDS Project approach to the textual analysis and management of a collection of digitized Medieval Latin charters. All our digitized documents are encoded using XML and, although we use our own naming rules for XML elements and attributes, the structure of encoded charter documents can easily be exported to such encoding systems as the Text Encoding Initiative (TEI) and the emerging Charter Encoding Initiative (CEI).[14] We have developed a different technical approach to encoding the metadata and physical placement of the Markup Language tags into the text. At the storage level, we store separately an original text of the charter, encoding information and meta

---

13 Computational Geometry Algorithms Library (CGAL), <http://www.cgal.org> (##.##.2006).
14 Vogeler, Georg, Towards a standard of encoding medieval charters with XML, in: Literary and Linguistic Computing 20 (2005), S. 269-280.

data information. Generally speaking we use dynamic encoding as an alternative to the more commonly used static system. This means that our system is capable of generating a variety of statically encoded documents in response to a specific client request. By employing dynamic encoding we have also eliminated the problem of overlapping markup elements.[15]

We have also discussed how we apply new methods to vocabulary and content analyses of Medieval Latin texts using "word-patterns" and Computational Geometry. Tight integration of a Statistical Engine into the process has allowed us to improve significantly the accuracy and reliability of our output results, and the use of analytical online tools like our "Browse" and "Search" programs can greatly facilitate the research. The application of Scaleable Vector Graphics as an alternative to commonly used Bitmap pictures has improved considerably the quality of graphical information available to Internet users.

For the future, we see great potential in the further development of word-pattern frequency analyses by using multiple mutually independent data flows. As we are always dealing with inexact statistical data, the search for alternative statistical methods remains one of our main priorities. We are currently working on the integration of the vocabulary and content evaluation tools into our website and we continuously seek to improve its appearance and functionality.

Michael Gervers is Professor of History at the University of Toronto. His research interests include medieval diplomatics, codicology, palaeography, chronology, charter analysis, textual editing and the dating of medieval documents. E-Mail: gervers@chass.utoronto.ca

Michael Margolin is Recearch Programmer at the DEEDS Project, University of Toronto. His research interests are quantitative analyses

---

15   Gervers, Michael, "Application of Computerized Analyses in Dating Procedures for Medieval Charters", in: Le Médiéviste et l'Ordinateur 42 (2003), S. 7-25. Co-authored with Michael Margolin

and digitization of medieval charters for the purpose of establishing chronological and authorship attribution. E-Mail: m.margolin@utoronto.ca