COMPUTERS IN SUPPORT OF HISTORY RESEARCH

by Stefan Gruner

History Informatics describes an emerging subdiscipline of informatics and history, following business informatics in the 1980s and biomedical informatics in the 1990s. In analogy to those sub-disciplines of Informatics, the term "History Informatics" was recently coined, derived from the German term "*Historische Fach-Informatik*". (The term "Informatics" is used in its original European sense which is different from its particular meaning here in the local South African context.)

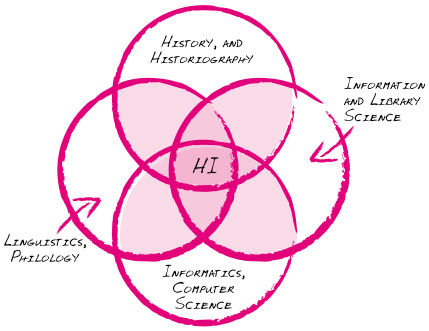
Medieval history was the first branch of the classical humanities (not taking into account modern linguistics) where computer-based quantitative and statistical research was introduced. Somewhat later, the possibilities of digital representation of the source material shifted the focus of history informatics towards the development of specific database implementations, and a discussion of electronic/digital editions began. The centre of activities in this new field is arguably Western Europe, with its rich written sources of medieval history. However, related work has recently also been found in the Oriental cultures, especially Korea and Tunisia.

The difficult relationship between text (syntax) and information (semantics), as well as the fact that the most interesting information (pragmatics) is "buried under" huge amounts of unprocessed text, still poses a major obstacle to progress in historical document engineering respectively History Informatics. Although photographic facsimiles are usually available for the most important bodies of those ancient original document collections, these facsimiles are of little use to a historian wanting to full-text-browse a large archive of medieval documents. This is especially tricky when searching for a particular phrase or text or a particular piece of information hidden in a document in the entire archive.

History informatics therefore has *two main aims*: the provision of methods, techniques and software tools that support the translation of original ancient documents into machine-accessible textual representations (syntax); and the provision of specialised (topic-specific) information retrieval methods (e.g. indexing, similarity clustering, data mining and data visualisation) to make large amounts of newly generated data automatically accessible for human perception and understanding.

With digital libraries and semantic retrieval methods becoming more available and applicable, one can reasonably expect considerable acceleration in the progress of medieval history research within the next 15 to 20 years. This has happened in biology and medicine with the support of bio-informatics and medical informatics. (Indeed, as far as the processing of *strings* is concerned, there are some methodological similarities between history informatics and bio-informatics on the one hand and history informatics and linguistic informatics –*aka* computer linguistics– on the other.)

→ *1* depicts the relationship between history informatics (HI) and other academic disciplines.



→ *1 History informatics (HI) in the centre of related disciplines*

Together with Georg Vogeler and Ben Burkard of Germany, the author presented a paper at the SAC 2007 Annual ACM Symposium on Applied Computing in Seoul, Korea in March this year. The paper discussed new software tools for XML tagging of electronic versions of medieval documents, such as meta-search methods of the arising "semantic web".

The programming of these tools was mainly done by Burkard, with additional involvement of Martin Gruner (a certified computer programmer and student of history – which illustrates the interdisciplinary nature of the field). The built-in repertoire of available XML tags is subject-specific, which allows historians with no knowledge of XML, informatics or computer science to use the tool.

→ *2* shows the interface of the context-sensitive XML tagging prototype developed for Vogeler's project in Germany. The tool supports the preparation, creation and modification of XML-tagged text files as "semantic copies" of medieval charters for the purpose of storing them in databases and accessing them via the new-generation semantic search. The lexical contents of the original charter must still be manually typed in free-text mode and cannot yet be imported via an OCR scanner. However, the selection of the XML tags to mark up the digitalised charter's text is menu-driven. (Notably, this process is still somewhat similar to the work of a medieval copyist sitting in his modestly equipped scriptorium.)

Context-sensitive means that not every tag is available in every arbitrary situation – a feature that is supposed to minimise the possibility of wrong tagging. This feature is justified by the standardised structure of the medieval charters dealt with. Thus, not only are the history-specific XML tags as such implemented into this tool as an implicit ontology, but so are their meaningful relationships between each other.

→ *3* shows a hand-written legal charter issued by the chancellery of Duke Ottokar of Lower-Austria in the 13th century AD. In this facsimile form, it cannot be searched for any contents.

After tool-supported transcription and XML tagging, a fully meta-searchable representation is yielded, as depicted in → *4*. Also note how the XML tags provide such meta information, e.g. by injecting explications to cumbersome abbreviations (much used, like nowadays, by the medieval chancellery officers), by injecting missing year numbers, or by injecting our familiar modern names to persons and places – names which usually differ considerably from the corresponding names used several hundreds of years ago.
It is hoped that this new spark will soon jump over from the Department of Computer Science at the University of Pretoria to other departments of

informatics, information science and history, as well as other universities in this country. Indeed, a 4th-year student with UP's School of Information Technology has recently started a promising software project (under the supervision of the author) in this area, and further students might possibly follow.

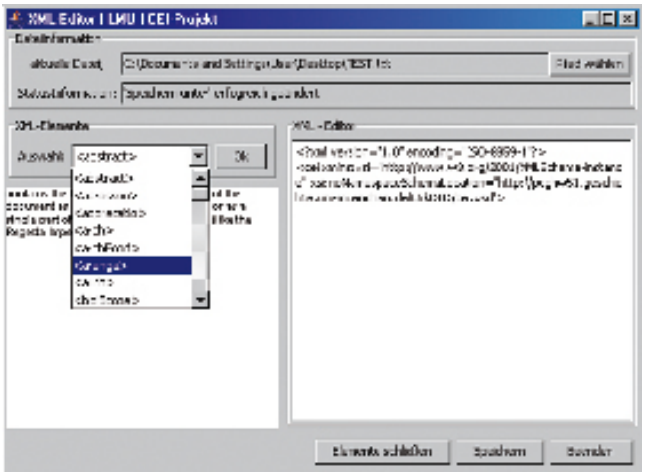
Also note that though sub-Saharan Africa, like Australia or North America, does not possess any "medieval history" in the European sense of the word, the tools and techniques of history informatics might soon be transferable to other document-based scholarly studies in the broader field of the humanities – in fact, the digitalisation of other areas of the humanities has already started as well, such that the emergence of innovative synergy effects can reasonably be expected within the next few years.

REFERENCE

Vogeler/Gruner/Burkard, "New Specialist Tools for Medieval Document XML Markup". Document Engineering Track: Proc. SAC'07 Annual ACM Symposium on Applied Computing, Vol.1, pp. 594-599, Seoul (Korea), March 2007.

***Dr Stefan Gruner** is a senior lecturer with the University of Pretoria's Department of Computer Science, (special area of interest: Software Engineering).*

sg@cs.up.ac.za



→ *2 Tool with context-sensitive tag menu (programmer: Martin Gruner)*

→ *3 Duke Ottokar's Charter, Monasterium Project, Austria*

(Source: monasterium.net, Austria)

```
<?xml version="1.0" encoding="utf-8">
<cei><text type="charter">
<tenor>
<intitulatio><persName reg="Otokar
(II.), Herzog von Österreich
und Steiermark (1251-1276),
Aussteller">Premizl, dei <expan
abbr= "gra">gracia</expan> dux <expan
abbr= "Aust"> Austrie</expan> et
<expan abbr= "Stir">Stir </expan>ie
et <expan abbr= "marcho">marchio
</expan> <expan type= "Moraw">Morawie
</expan> </persName> </intitulatio>
<inscriptio> vniu<expan abbr="er-
Haken"> er </expan>sis hanc
litt<expan abbr="er-Haken"> er </
expan>am inspecturis in perpetuum</
inscriptio>
 <publicatio>Noscat igitur tam
presens hominum etas quam futura</
publicatio>
 <arenga>quod nos retribucionis et
gracie futurum meritum exspectantes
ab eo, cuius ad sui similitudinem
prouidencia nos condidit et redemit
vt heredes faciat regni sui,</arenga>
<dispositio>huius boni non immemores
dilectis nobis fratribus <placeName
reg="Hl. Maria in Geras">claustri
ecclesie Marie in Jærus</placeName>
pro remedio anime nostre et
nostrorum remissione peccaminum
ecclesiam sitam in uilla Mislowicz
contulimus liberaliter cum uniuersis
attinenciis et prouentibus ad eandem
pertinentibus, quocunque nomine
censeantur, perpetuo possidendam</
dispositio>
<corroboratio>In cuius facti robur
memoriam et cautelam presentem
paginam ipsis dari mandauimus
sigilli nostri in testimonium
Karactere communitam. Testibus eciam
qui aderant subnotatis; qui sunt
<testis>Bosko Burgrauius Znoymensis</
testis>, <testis>Zmilo de Belcowe
et fratres sui Ratiborius</testis>,
<testis>Jerosius</testis> et <testis>
Marqvardus </testis> et alii quam
plures </corroboratio>
<datatio>Actum et datum Brune
per manum nostri prothonotarii
magistri Wilhelmi. Anno domini <num
type="römische Zahl" value="1253">M.
CC. LIII.</num> <num type= "römische
Zahl" value="3">III</num> idus
Julij <num type="römische Zahl"
value="11">XI </num>indictionis.</
datatio>
</tenor>
</cei>
```

→ *4 Tool-supported XML transcription of Duke Ottokar's Charter (provided by Georg Vogeler)*