Smart Big Data: Use of Slovenian Parliamentary Papers in Digital History

IZVLEČEK
PAMETNI MASIVNI PODATKI: UPORABA SLOVENSKIH PARLAMENTARNIH DOKUMENTOV V DIGITALNI ZGODOVINI

Avtorja v prispevku opozorita na problem velikih količin digitalnih zgodovinskih virov, s katerim se bodo srečavali raziskovalci sodobne zgodovine. Bolj nadrobno predstavita slovenske parlamentarne dokumente kot primer pametnih masivnih podatkov. Avtorja menita, da velikih količin tega digitalnega gradiva zgodovinarji ne bodo mogli obdelovati samo z uporabo klasičnih zgodovinskih metod, temveč bodo morali začeti uporabljati še metode in orodja, ki jih razvija digitalna zgodovina, digitalna humanistika in tudi jezikoslovne tehnologije.

Ključne besede: digitalna humanistika, digitalna zgodovina, Slovenija, parlament

ABSTRACT
The paper calls attention to the problem of massive amounts of digital historical sources that will eventually be faced by researchers of contemporary history. Slovenian parliamentary papers are then presented in detail as an example of smart big data. The authors believe that historians will be unable to process massive amounts of such digital materials using only standard historiographical methods and will be forced to start using methods and tools developed by digital history, digital humanities and also language technologies.

Keywords: digital humanities, digital history, Slovenia, parliament

Big Data

Big data is the buzzword of the decade. However, the very ubiquity of the term both in the industry, the media as well as in the academic community has led to big data being defined in various ways. Furthermore, big data in the humanities is not
the same as big data in natural sciences.¹ This paper uses the definition according to which historical big data is described as follows;

• large volumes of data, particularly texts that cannot be read within a reasonable time frame, and

• information that only allows us to reach new conclusions through the use of digital methods.²

Large amounts of computer-readable data are gradually becoming the reality of modern historiography. The late Roy Rosenzweig, one of the pioneers of digital history, warned as early as 2003 that historiography, instead of working within the scarcity paradigm of historical records, will have to start facing the problem of an excess of resources.³ Up to now, historiography has mostly had to deal with the lack of resources, their incompleteness and often also with high costs of acquiring additional sources. Today, on the other hand, historians can access new digitized and digital sources quickly and effectively. Although a lion’s share of analogue materials has not been and is unlikely to be digitized within reasonable time,⁴ the materials emerging today are increasingly created in the digital form. This is the reason why, for example, the Slovenian archives are currently working hard on establishing a Slovenian electronic archive – e-ARH.si.⁵ Consequently, the problem of (over)abundance of historical resources is also emerging in the studies of contemporary history,⁶ including the history of the Republic of Slovenia after 1991.⁷

The massive amount of digital materials has spurred the creation and spreading of digital humanities, which use digital methods and tools to address new research questions.⁸ In the following sections, the paper uses the example of Slovenian parlia-


mentary papers to present the advantages and drawbacks of the digital historiography methods in the analysis of big historical data. In this context, the digital history is understood as part of the digital humanities, which is primarily concerned with online distribution and presentation of historical sources using various computer tools, especially for mapping, network software and, last but not least, text analysis. The paper focuses mainly on the possibilities for the use of text analysis methods and tools.

Slovenian Parliamentary Papers

Parliamentary papers are a rich source of data used by different academic disciplines including historiography. In some European countries, a large part of these papers is already accessible in digital form, mostly in PDF format.

Both researchers and the public can avail themselves of the materials from parliamentary institutions located within today’s Slovenia and from parliamentary institutions whose members once included representatives from Slovenia. While it is true that most of the materials are currently only available in analogue form, an increasing amount has already been digitized and made available to the public:

- Austrian National Assembly (1861–1918);^11
- Styrian Provincial Assembly (1848–1914);^12
- Carniolan Provincial Assembly (1861–1869);^13
- Yugoslav legislative bodies 1919–1939, 1942–1953;^14
- People’s Assembly of the People’s Republic of Slovenia (1947–1963);^15
- Assembly of the Socialist Republic of Slovenia (1963–1990);^16
- National Assembly of the Republic of Slovenia, from 1990 until today.^17

---

With the exception of the documents from the National Assembly, which are being published in the HTML format, all documents have been published in PDF.

Chart 1 shows the number of words contained in parliamentary speeches for each parliamentary term of the Slovenian parliament which are archived in PDF at the History of Slovenia – SIstory portal (36 million words) and in HTML at the National Assembly website (69 million words). Up until 1974, minutes of the sessions also included extensive attachments (10.5 million words), which were later published in a special serial publication titled *Poročevalec*. From 2010 onward, *Poročevalec* has been regularly accessible at the National Assembly website, where visitors can also look through archived issues from 2006 to 2010. Other issues of *Poročevalec* (1974–2006) are going to be digitized within three years by the Institute of Contemporary History that manages the History of Slovenia – SIstory portal. As of this moment, the Institute has already managed to digitize issues of *Poročevalec* from 1974 to 1996, which together contain almost 37 million words (see Chart 1).

![Chart 1: Number of words in parliamentary speeches (1947–1990), attachments (1947–1974) and Poročevalec (1974–1996) at the History of Slovenia – SIstory portal and the number of words in parliamentary speeches at the National Assembly website (1990–2011); in millions of words per parliamentary term](image)

---


Searching

It is clear that no researcher is able to read that much text in its entirety. Researchers thus only read those parts that they consider relevant for their research. In doing so, they read the selected parts of the text carefully, from word to word, or quickly skim over the pages looking for relevant parts of the text. However, such research is generally based on the assumption that the researchers will find what they are looking for in the text. Researchers thus determine the text they are looking for, as well as the context of their research, in advance. In doing so, they necessarily lean on their previous comprehensive or inadequate knowledge of the area of study.\(^\text{20}\)

In history as well as in other humanities, such methods are of course completely acceptable and often yield useful results. A number of good studies has been created in such a manner using materials from the National Assembly of the Republic of Slovenia.\(^\text{21}\) Most researchers in the humanities thus primarily understand digital materials in terms of easier and quicker access to desired information.\(^\text{22}\) In the case of materials from the National Assembly, researchers can make use of a search engine that allows them to filter their results according to search modules. In a similar manner, researchers can search for older parliamentary materials at the History of Slovenia – SIstory portal.

Most researchers in the humanities first use search engines to identify sources and then do a full text keyword search. This means that studies are no longer conducted only vertically, from top to bottom, where a researcher only reads canonical texts and browses previously organized data collections. Rather, studies are being conducted in a bottom-up manner, with researchers looking for parts of text pertinent to their research that they would otherwise not have read. However, this research method has its limitations. The researcher must know the search queries in advance, and these can differ from modern thought patterns. Inasmuch as the researcher does not carefully examine every search result, the results are always lacking proper context.\(^\text{23}\)

Despite these limitations, a full text keyword search can yield very useful results. A good example of such research is the article *War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History* whose authors used carefully
selected search queries and a search engine to systematically check to which extent the Boerenpartij (Farmers’ Party) was described as “wrong” in all Dutch parliamentary debates between 1958 and 1982. The article was written as part of the War in Parliament project. The results of this project clearly showed that satisfactory research results can only be obtained if we are familiar not only with the advantages but also with the shortcomings of digital research methods.

One of the prerequisites for the War in Parliament project to be successful was the use of partly structured data in the XML format, which allowed for the search results to be filtered by speaker’s name, by party, by time period, structure of the text, etc. What was used was thus not big data in the form of plain text, but rather smart data. Smart data may be structured or partly structured, and compared to implicit big data, smart data is explicit, marked, enriched and described by metadata. The creation of smart data is often a labour-intensive process that requires human intervention.

As we have seen, parliamentary papers represent extremely extensive data collections. We thus cannot expect to be able to reveal their explicit content merely through precise manual annotation. The dilemma necessarily encountered by researchers in the use of digital parliamentary papers was succinctly stated by Christof Schöch:

“I believe the most interesting challenge for the next years when it comes to dealing with data in the humanities will be to actually transgress this opposition of smart and big data. What we need is bigger smart data or smarter big data, and to create and use it, we need to make use of new methods. So, how can we enrich big data sufficiently to make more intelligent queries possible? How can we speed up the process of creating smart data so that we can produce larger volumes of it?”

At the same time, Schöch calls our attention to two possible ways of making big data smarter: through automatic annotation and through crowdsourcing. In practice, parliamentary papers proved very suitable for automatic annotation. In particular, parliamentary debates were written down in a format that has changed very little

---

28 Ibid., 10.
with time.29 This is one of the reasons why various research projects often annotate parliamentary debates using the XML markup language. Among others, sessions of the British parliament (Hansard) from 1803,30 the Dutch parliament from 1803,31 the Spanish parliament from 1977,32 the Czech parliament from 199333 and the Polish parliament from 199334 are all available in the XML format.

The following sections of this article will present the use of Slovenian parliamentary papers, particularly the minutes of parliamentary debates, in digital history. A number of cases will be presented to illustrate the huge potential of smart big data in contemporary history studies. As an example, 2.7 million words of the minutes of parliamentary debates in the Chamber of Associated Labour of the Assembly of the Republic of Slovenia from 1990 to 1992 have been annotated using the XML format.35 In doing so, it was decided that the Text Encoding Initiative (TEI) Guidelines should be used,36 as these are the de facto standard for the encoding of texts in digital humanities.37

Automatic conversions were carried out using XSL stylesheets created specifically for the project. However, annotation was also carried out by hand, not just automatically. The reason for this was that automatic conversions can also contain annotation errors. Attempts were made to find these errors and remove them through an upgrade of XSL stylesheets. There were also some parts of text that could only be annotated manually. Using such semi-automatic annotation, brief sessions could be marked up in 30 minutes, while those of medium length usually took up to two hours and the longest (over 200,000 words) up to four hours. Speeches were marked in accordance with the TEI module for performance texts (speech, speaker, stage direction). Other annotations included the structure of the assemblies and type of

37 For example, the German Research Community (Deutsche Forschungsgemeinschaft) demands that any texts being digitized be encoded using the TEI guidelines, if at all possible. Deutsche Forschungsgemeinschaft, DFG Practical Guidelines on Digitisation (Bonn: Deutsche Forschungsgemeinschaft, 2013), 31.
sessions, individual sessions, topics of individual sessions and dates and duration of
sessions. Links were created to tables of contents and lists of speakers.38

Based on such annotated minutes of sessions, researchers can carry out various
types of fundamental analyses.39 Over 2.7 million words were thus spoken in 13,894
speeches at 54 sessions. At the longest, 36th session, which lasted eight days, the total
duration of speeches was 29 hours and 3 minutes, and the session was adjourned no
less than 21 times. On the other hand, the total duration of speeches at the briefest,
9th session was only one hour. The longest uninterrupted span was 460 minutes,
while the average (median) duration of speeches between two interruptions was 90
minutes.

Connection to External Data

However, TEI documents annotated in such a manner also have some shortcom-
ings that preclude precise analysis of the speeches. Initial analyses of the speeches of
various speakers were carried out based on their first and last names. A single person,
whose name might be written differently in other cases, is thus treated as two or
more different people. On the other hand, people with identical first and last names
are automatically considered the same person. Various historical records were thus
used to manually verify and sanitize the lists of MPs, ministers and other invited
speakers. These data are contained in a separate TEI document.

Chart 2: Number of words spoken in the Chamber of Associated Labour of the Assembly
of the Republic of Slovenia (1990/92) by organization membership; in %

<table>
<thead>
<tr>
<th>Organization</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chamber of Associated Labour</td>
<td>81%</td>
</tr>
<tr>
<td>Government</td>
<td>9%</td>
</tr>
<tr>
<td>Chamber of Municipalities</td>
<td>0%</td>
</tr>
<tr>
<td>Socio-Political Chamber</td>
<td>0%</td>
</tr>
<tr>
<td>Others</td>
<td>10%</td>
</tr>
<tr>
<td>Unknown</td>
<td>0%</td>
</tr>
</tbody>
</table>

38 Andrej Pančur, “Označevanje zbirke zapisnikov sej slovenskega parlamenta s smernicami TEI,”
in: Zbornik konference jezikovne tehnologije in digitalna humanistika, eds. Tomaž Erjavec and Darja Fišer
Federation. A feeble “ratification body” or important political decision-maker?,” Prispevki za novejšo
zgodovino 55, no. 3 (2015): 54.
Data connected in such a manner can, for example, be used to determine the provenance of speakers in the Chamber of Associated Labour (see Chart 2). In doing so, we find that almost 20% of the words were spoken by various representatives of the Government and other rapporteurs associated with the legislation that was being passed by the Assembly. As President and Vice-President of the Chamber of Associated Labour, Jože Zučančič (President, 735,166 words) and Bogo Rogina (Vice-President, 114,418 words) together spoke as much as 38.3% of all words spoken by the MPs of the Chamber of Associated Labour. Among other MPs of the Assembly, Jože Arzenšek (106,000 words), Roman Jakič (69,111 words) and Andrej Šter (67,005) were the most verbose. Then there was the silent Jože Košak who only managed to say 14 words during his term.

Speeches of the MPs can obviously also be analysed according to the parties they belonged to (see Table 1). MPs who were, at the start of their term, members of the DEMOS coalition thus spoke 21.5% of all words, opposition MPs spoke 23.3%, while the numerous independent MPs (including the President of the Chamber) spoke as much as half of all words.

Table 1: Number of words spoken by members of political party; Chamber of Associated Labour of the Assembly of the Republic of Slovenia (1990/92)

<table>
<thead>
<tr>
<th>Political parties</th>
<th>No. speakers</th>
<th>No. words</th>
<th>Percent</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDZ → DS</td>
<td>3</td>
<td>30762</td>
<td>1,4</td>
<td>21,5</td>
</tr>
<tr>
<td>SDZ → NDS</td>
<td>2</td>
<td>85246</td>
<td>3,8</td>
<td></td>
</tr>
<tr>
<td>SDSS</td>
<td>9</td>
<td>118899</td>
<td>5,4</td>
<td></td>
</tr>
<tr>
<td>SKD</td>
<td>6</td>
<td>86135</td>
<td>3,9</td>
<td></td>
</tr>
<tr>
<td>SKZ → SLS</td>
<td>8</td>
<td>88439</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ZS</td>
<td>1</td>
<td>11901</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>DEMOS</td>
<td>1</td>
<td>56858</td>
<td>2,6</td>
<td></td>
</tr>
<tr>
<td>ZKS-SDP → SDP</td>
<td>17</td>
<td>231148</td>
<td>10,4</td>
<td>23,3</td>
</tr>
<tr>
<td>ZSMS → LDS</td>
<td>9</td>
<td>236388</td>
<td>10,6</td>
<td></td>
</tr>
<tr>
<td>SZS → SSS</td>
<td>3</td>
<td>49708</td>
<td>2,2</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>19</td>
<td>1109636</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>SOPS → Independent</td>
<td>1</td>
<td>114418</td>
<td>5,2</td>
<td>5,2</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>736</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Lists of MPs, members of the Government and other speakers can also be used to formulate research questions connected to additional variables: first and last name, gender, date and place of birth, date and place of death, education, profession, residence, organization membership.

Answers to complicated research questions are possible in part thanks to the newly created TEI file that includes a thematic index of the topics dealt with by the Chamber. The creation of this table of contents used data from existing topics and tables of contents, which were then annotated according to the new scheme. The
data was first categorized in accordance with the Rules of Procedure of the National Assembly (see Chart 3).  

The largest category, Acts and Procedures, was classified in accordance with the thematic index of the Legal Information System of the Republic of Slovenia. Based

\[ \text{Table 2: Main topic categories within Acts and Procedures as per the thematic index of the Legal Information System of the Republic of Slovenia} \]

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of words</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional regime of the Republic of Slovenia</td>
<td>497575</td>
<td>21.8</td>
</tr>
<tr>
<td>Foreign affairs and international relations</td>
<td>27417</td>
<td>1.2</td>
</tr>
<tr>
<td>Interior and administrative law</td>
<td>67156</td>
<td>2.9</td>
</tr>
<tr>
<td>Civil law</td>
<td>24017</td>
<td>1.1</td>
</tr>
<tr>
<td>Criminal law</td>
<td>5925</td>
<td>0.3</td>
</tr>
<tr>
<td>Economic order</td>
<td>403880</td>
<td>17.7</td>
</tr>
<tr>
<td>Public finance</td>
<td>482764</td>
<td>21.1</td>
</tr>
<tr>
<td>Economic activities</td>
<td>165752</td>
<td>7.3</td>
</tr>
<tr>
<td>Non-economic activities</td>
<td>550466</td>
<td>24.1</td>
</tr>
<tr>
<td>Environment and spatial planning</td>
<td>43803</td>
<td>1.9</td>
</tr>
<tr>
<td>Protection against natural and other disasters</td>
<td>14960</td>
<td>0.7</td>
</tr>
</tbody>
</table>


on data annotated in such a manner, it is easily determined that over a fifth of all speeches were associated with the legislation pertaining to the constitutional arrangements in the Republic of Slovenia (see Table 2).

The greatest amount of discussion was stirred up by the Ownership Transformation of Companies Act, which resulted in 520 speeches containing over 100,000 words (see Table 3). On the other hand, debate regarding the Act Ratifying the Agreement between the Government of the Republic of Slovenia and the Federal Council of the Swiss Confederation on the Abolishment of Visas was extremely brief, consisting of only 46 words.

### Data Enrichment: Natural Language Processing (NLP)

Another extremely extensive category was the category named *Initiatives, Suggestions and Questions from the MPs* (102,858 words). However, the category is too broad to allow any conclusions about its actual content based solely on the title. Natural language processing technologies can be of some help in this regard. For example, the topic modelling method can be used to search for word patterns in the text, which can in turn assist with determining the semantic meaning of various parts of the text. One of the most popular (among historians as well as other researchers) tools used in such analyses is MALLET.\(^{42}\) Although the results were incomplete, MALLET was nonetheless successfully used to discern a number of topics within the *Initiatives, suggestions and questions* category: customs duties, healthcare, the environment, strikes, banks, etc.

A tool that searched the text for named entities (people, places and organizations)

---


yielded much better results. Named entity recognition was carried out using the Stanford NER for the Slovenian language.\textsuperscript{44} It should come as no surprise that those

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
place name & No. & Percent \\
\hline
Slovenija & 2978 & 26.2 \\
Republika Slovenija & 2153 & 18.9 \\
Jugoslavija & 579 & 5.1 \\
Evropa & 534 & 4.7 \\
Ljubljana & 270 & 2.4 \\
Maribor & 244 & 2.1 \\
Hrvaška & 241 & 2.1 \\
Nemčija & 182 & 1.6 \\
Italija & 170 & 1.5 \\
Avstrija & 167 & 1.5 \\
Poročevalca & 132 & 1.2 \\
Celje & 103 & 0.9 \\
Skupščine & 94 & 0.8 \\
Beograd & 91 & 0.8 \\
Irak & 89 & 0.8 \\
Komisija & 82 & 0.7 \\
Srbija & 82 & 0.7 \\
Koper & 55 & 0.5 \\
Lendava & 53 & 0.5 \\
Ptuj & 51 & 0.4 \\
Republika Hrvaška & 51 & 0.4 \\
Demos & 49 & 0.4 \\
Kranj & 42 & 0.4 \\
Logatca & 38 & 0.3 \\
Republiškem & 38 & 0.3 \\
Francija & 37 & 0.3 \\
Švica & 36 & 0.3 \\
Madžarska & 34 & 0.3 \\
Združene države Amerike & 33 & 0.3 \\
Piran & 31 & 0.3 \\
other named entities & 2639 & 23.2 \\
\hline
11378 & 100.0 \\
\hline
\end{tabular}
\caption{List of the 30 most common place names in speeches in the Chamber of Associated Labour of the Assembly of the Republic of Slovenia (1990/92) as identified by the Stanford NER for the Slovenian language}
\end{table}

\textsuperscript{44} Nikola Ljubešić et al., “Combining Available Datasets for Building Named Entity Recognition Models of Croatian and Slovene,” \textit{Sloveničina} 2.0 1, no. 2 (2013): 35–57.
who were most often named in the speeches of MPs were the other MPs. It is also unsurprising that the place name used most often by speakers in the Slovenian parliament was Slovenia. Table 4 thus shows that the Stanford NER for the Slovenian language recognized 11,378 place names in parliamentary speeches, 45% of which were identifiable as Slovenia or the Republic of Slovenia. However, a detailed look at the table quickly reveals that the place names identified by Stanford NER for the Slovenian language also included names of organizations (Assembly, Commission, DEMOS) and other names (Poročevalec).

These results clearly indicate that history researchers should never simply copy the results of natural language processing (NLP) technologies. At its current level of development, the technology is far from infallible. For example, the Stanford NER for the Slovenian language has 85% precision when annotating persons, while its precision is below 80% when it comes to places.45 The following warning must thus truly be taken to heart: “Historians need to be aware that, in addition to verifying reliability of sources as is common in their field, they also need to take the reliability of NLP methods into account when working with automatically extracted information.”46

The authors of this paper are certain that the most effective way to achieve this would be through close collaboration with computational linguists. The existing TEI documents that had been encoded in accordance with the TEI module used for performance texts were thus subsequently converted into TEI documents wherein the text was annotated in accordance with the speech transcription TEI module. Such documents can then be furnished with linguistic annotations at a later time. Computational linguists have thus already provided part-of-speech tagging for the text of the speeches. The corpus has been imported into the No Sketch Engine concordance base47 and all TEI documents are accessible at the CLARIN.SI repository.48

Conclusion

This article has attempted to show, based on a number of fundamental analyses of the minutes of the sessions of the Chamber of Associated Labour of the Assembly of the Republic of Slovenia (1990/92), that future historians will be unable to process the increasing amounts of digital materials using standard historiographical methods and will be forced to start supplementing these with methods and tools developed by digital humanities. In doing so, most historians will still rely on various tools developed by digital historians in order to simplify the work of their colleagues who are unfamiliar with digital humanities. At the same time, digital historians will also have to be aware of the limitations of the use of tools developed by other fields, including language technologies. On the other hand, historians who have started learning the basics of programming languages are increasingly establishing themselves in the field. The reason for this is that this is the only way to extract additional useful research results from existing digital sources. This is also the manner in which most of the analyses presented by this paper were carried out. At the same time, the authors of this article are well aware that in addition to new knowledge supplied by digital history, it remains indispensable for researchers to be very familiar with research domain. It can thus be anticipated that future research of contemporary history will take place through fruitful collaboration of experts from a multitude of different fields.

Sources and Literature

Dataset:
• Pančur, Andrej, Mojca Šorn and Tomaž Erjavec. Slovenian parliamentary corpus SlovParl 1.0 (2016). Distributed by Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1075.

Literature:
• Deutsche Forschungsgemeinschaft. DFG Practical Guidelines on Digitisation. Bonn:


PAMETNI MASIVNI PODATKI: UPORABA SLOVENSKIH PARLAMENTARNIH DOKUMENTOV V DIGITALNI ZGODOVINI

POVZETEK

Avtorja v uvodu opozoriva na dejstvo, da velika količina računalniško berljivih podatkov postaja stvarnost in neizogibno dejstvo tudi v zgodovinopisju, pri čemer poudariva, da je prav ta fenomen spodbudil nastanek in uveljavitev digitalne humanistike, ki s pomočjo digitalnih metod in orodij odgovarja na novo raziskovalna vprašanja.


Osnovne analize zapisnikov torej pokažijo, da v prihodnosti zgodovinarji vedno večjih količin digitalnega gradiva ne bodo več mogli obdelovati samo z uporabo klasičnih zgodovinskih metod, temveč bodo morali začeti uporabljati metode in orodja, ki jih ravnja digitalna humanistika. Večina zgodovinarjev se bo pri tem (še naprej) zanašala na različna orodja, ki jih ravnja digitalni zgodovinarji z namenom, da olajšajo delo svojim kolegom, ki se ne ukvarjajo z digitalno humanistiko. Pri tem pa bodo digitalni zgodovinarji morali pozoriti omejitve, ki jih prinaša uporaba orodij, razvitih v okviru drugih disciplin, med drugim tudi jezikovne tehnologije. Avtorja prispevka se dobro zaveda, da je za kakovostne analize poleg novih znanj, ki jih prinaša digitalna zgodovina, neobhodno potrebno tudi temeljito poznavanje raziskovalne domene. Zato lahko predvidimo, da bo raziskovanje sodobne zgodovine potekalo v znamenju plodnega sodelovanja strokovnjakov iz različnih področij.