

SDHK meets NER: Linking place names with medieval charters and historical maps

Olof Karsvall¹ and Lars Borin²

¹ The Swedish National Archives • TORA

² Språkbanken, University of Gothenburg, Sweden • SWE-CLARIN

Abstract. Mass digitization of historical text sources opens new avenues for research in the humanities and social sciences, but also presents a host of new methodological challenges. Historical text collections become more accessible, but new research tools must also be put in place in order to fully exploit the new research possibilities emerging from having access to vast document collections in digital format. This paper highlights some of the conditions to consider when place names in an older source material, in this case medieval charters, are to be matched to geographical data. The Swedish National Archives make some 43,000 medieval letters available in digital form through an online search facility. The volume of the material is such that manual markup of names will not be feasible. In this paper, we present the material, discuss the promises for research of linking, e.g., place names to other digital databases, and report on an experiment where an off-the-shelf named-entity recognition system for modern Swedish is applied to this material.

1 Introduction

Mass digitization of historical text sources opens fresh and exciting avenues for research in the humanities and social sciences (HSS), but also presents a host of new methodological challenges.

On the one hand, large-scale digitization of historical document collections certainly serves to make these collections generally more accessible to researchers (and the interested public), since they may now be browsed and searched at will on the internet, instead of by going to a physical archive with restricted opening hours. Also, unlike the physical original of a document, its digital version can be accessed by many individuals simultaneously.

On the other hand, in order to truly benefit from having large document collections in digital format, and to enable new kinds of research, such collections must become more than just digital versions of conventional physical archives.

One way of working with large-volume text collections for research purposes is often referred to as *text mining*. Although not very strictly defined, this methodology generally boils down to the application of *data mining* methods on large amounts of textual data. Data mining in its turn presupposes that the data to be processed is formally structured along the

lines of conventional databases or spreadsheets. Thus, the key step in text mining is to turn textual data into tabular data,³ i.e., in essence into *metadata* for a document or document part.

To the extent that these large digitized document collections have or can be equipped with metadata, they can be processed with data mining methods, thus providing an alternative, novel entry point for research. Further, making the data and metadata accessible on the web inevitably leads to the possibility of data linkage across collections, which in turn raises the issue of standardization of data and metadata, e.g., by spelling normalization and/or reference to authority lists.

In concert with user interfaces for visualizing, browsing and manipulating data and relations among formally structured data (Keim et al, 2010), text mining methods and linked data provide powerful – and complementary – tools for discovering new facts and generalizations from large volumes of data, since they allow us to build interfaces that combine a “distant reading”/“macroanalysis” mode of inquiry (Moretti, 2005, 2013; Jockers, 2013) with the “close reading” approach characteristic of traditional HSS research (cf. Schöch, 2013).

2 A concrete example: The Swedish medieval charters

There are several source materials from Sweden that can be said to be unusually rich and useful, even compared with what we find in other countries. The medieval charters from the 1160s and onwards and the land survey maps from the 1640s and later on are two examples of this. Most charters are legal documents, contracts between landowners concerning transactions of land and other properties. The maps are also very detailed, specifying each farm and all cultivated arable land.

These important source materials are used for addressing various historical research questions. The charters provide insights into a variety of transactions and events from the 12th century to the 16th century. The oldest maps, dating back to the 17th century, provide details of the agrarian landscape long before the major land reforms and urbanization.

Over the past 15 years, major efforts have been made to digitize these sources, and related source editions, to make them accessible on the internet. This has enabled full-text search which in some sense has revolutionized research on this material. Various events and objects can now be searched and analyzed, something that previously required time-consuming manual efforts. It is now easier than ever to formulate and explore different research questions. Digitization also opens up for quantitative analysis, as a result of simultaneous availability of larger collections of data.

However, a limitation currently is that the databases and search applications are modeled and developed as separate systems. Each source material has its own character and metadata. But each source material also contains general information that appears in other databases and registers. Names of persons and places are a typical type of information that often occurs

³ In this context, text is often characterized by computer scientists as “unstructured data”, which of course to a linguist sounds very much like turning things on their head.

in several sources. Names are therefore suitable as authority files on the semantic web, which many can use and refer to as linked data. A risk today is that research teams and developers will create their own name definitions for a lot of objects that already exist as data on the web, with duplication as a consequence.

The major point of using linked data as an approach is that it connects historical data and enables research that ‘combines’ data in a new and easy way. One example of a research question that could be highlighted concerns land ownership in the Middle Ages, where an identification of the people who buy and sell land – which is documented in the medieval letters – can be combined with the coordinates of the settlements of land ownership, which can be extracted from the historical maps.

Digitization has come a long way but we could go further. Often the databases have been designed mainly for internal use, as authoring tools for curators of the data. A great deal of effort remains in defining, labeling and linking the content of the sources. All these historical ‘knowledge bases’ taken together carry a significant potential for further research that has been little used. At present, they are not adequately combined and adapted to each other.

This paper focuses on geographic annotation methods, and more precisely the question of how place names in medieval charters can be identified and matched to geographical data from historical maps. The number of place names in all these medieval letters is not known, but very likely they hold more than one hundred thousand place names. Manual markup of these place names is thus not feasible. Instead, a method that automatically identifies place names is needed.⁴ In section 4 below, we describe such a method, but first we now turn to a more detailed technical description of our data.

2.1 Sources and data format

Since 2003, the digital edition of the medieval charters in Sweden (the Diplomatarium Main Catalogue, SDHK) has been made accessible over the internet (see figure 1). SDHK is a database with transcriptions and/or summaries of text from over 43,000 medieval charters in Sweden, from the 12th century until the early 16th century.⁵ The SDHK is a first portal for anyone looking for information on Swedish medieval charters. The web interface to the

⁴ Names of persons in the medieval charters could very likely be identified in a similar way as place names. However, personal names will not be investigated in this paper simply because it requires more. Many people named are not known, and the identification often requires access to registers which have not yet been digitized, for instance *Sveriges medeltida personnamn*: <http://www.sprakochfolkminnen.se/om-oss/om-webbplatsen/andra-sprak-an-svenska/english.html>.

⁵ The traditional editorial work of *Diplomatarium Suecanum* – the chronological national edition of Swedish medieval charters – started already in the 1820s. Having now completed the 1370s, the Diplomatarium series continues and new books are published on a regular basis.

The screenshot shows the Swedish National Archives' online search interface for SDHK (Medieval charters). The interface is organized into several sections:

- Header:** Features the Riksarkivet logo, the text "Search the collections", and user information: "Signed in as: Göteborgs universitet". There are also links for "National Archives", "Shop online", "My pages", "Sign out", and "Other languages".
- Navigation:** A horizontal menu with buttons for "Start", "SVAR, Digital Research Room", "National Archives Database", and "About the search service". Below this is a secondary menu with buttons for "Digital Research Room", "Extended search" (which is highlighted), "Archive", "Person", "Image ID", "Free text", and "Subject".
- Search Area:**
 - A dropdown menu labeled "Select extended search".
 - The title "SDHK (Medieval charters)" with a "Search tips" link.
 - A checkbox for "Digitised material only".
 - Four main search criteria: "SDHK number" (with a value of 8953), "Date", "Contents", and "Place of issue", each with an input field.
 - Two sub-sections: "Printed" and "Texts".
 - Under "Texts", there are checkboxes for "Edition", "Summery", "Document text", "Other text", and "Seals".
 - A "Search" button with a magnifying glass icon and a "Clear" button with an 'x' icon.
 - A link for "More options" with a star icon.

Fig. 1. The Swedish National Archives' online search interface to SDHK

database permits the users to search for persons, places and various subjects in the document summaries and in the full texts.⁶

It is important to emphasize that the preserved charters and documents deal above all with various property transactions, and as a consequence a large number of place names and settlements are mentioned.

The database is based on XML documents using the TEI (Text Encoding Initiative) markup language.⁷ Each XML file – corresponding to one medieval charter – contains metadata, a summary of the contents and the full literal transcription in the original source, and if applicable, references and comments. Besides the transcription of the text, some metadata has been registered concerning the original document, e.g. year of issue, issued by whom, place of issue, etc.

In addition to <header>, the element <text> is used, which consists of <front>, <body> and <back>. The full transcription of the original letter is reproduced within <body>. The full text, which is either in Old Swedish or Latin, is not normalized. Each place name can be spelled in many different ways. Hence, it is difficult to automatically identify the place names in the content of the <body> tag.⁸

The focus of this paper will therefore be on the <front> element and the editorial summary of each charter, where most known place names are listed in Swedish modern spelling.

⁶ The database is a part of the Swedish National Archives and is available at the National Archives website <http://sok.riksarkivet.se/sdhk>, see also <https://riksarkivet.se/facts-medieval-charters>.

⁷ <http://www.tei-c.org>

⁸ Tags are defined in the *TEI: P5 Guidelines* <http://www.tei-c.org/Guidelines/P5>.

SDHK No: 8953
Register of Diplomatarium Suecanum

Date of issue	Place of issue
1366 september 14	Eskilstuna

Contents

KARL ULFSSON, SVERIGES RIKES MARSK OCH LAGMAN I UPPLAND, stadfäster sina avlidna föräldrars gåvobrev till Eskilstuna kloster på 2 tön mjöl årligen från hans kvarnar i Torshälla på villkor att han och hans maka Helena Israelsdotter med arvingar och efterkommande får åtnjuta samma broderskap och privilegier, som hans far och mor haft av klostret. Utfärdaren beseglar.

Language
latin

Sources

Original: or. perg., RA 0101 [images](#)
Post-medieval copy: Hadorph: E 74, pag. 34; Peringskiöld: E 57 (reg.); Örnhielm: XI, p. 192

Document text

Vniuersis presens scriptum cernentibus . karolus wlfson regni Suecie marscalcus ac vplandiarum ' legifer salutem in domino sempiternam . Nouerint vniuersi . quod litteras patris mei . et matris mee ' felicis recordacionis . viris religiosis . priori et fratribus in æskilstunom super duabus thynionibus ' farine de molendinis meis in thorsharghe . annuatim perpetue leuandis . collatas . virtute ' presencium litterarum mearum confirmo / et donacionem huiusmodi per me et heredes meos quoscumque volo perpetuis ' temporibus iniolabiliter obseruar[i](a) . et ratam firmiter tener[i](b) . Jta tamen . quod ego . consors mea dilecta ' Helena ysraele dotter . et heredes nostri et successores . eadem confraternitate et priuilegiis ' quibus pater meus et mater mea de predictis priore et fratribus in æskilstunom . vtebantur / perfrui ' debeamus . Datum jn claustro æskilstuna Anno dominj M^oCCC^o Lx^o sexto . die Exaltacionis ' sancte Crucis meo sub sigillo

Fig. 2. SDHK 8953

Within the <front> element, place names occur within the <docDate> tag, where <date> specify when the charter was issued, and <placeName> specify the location of issue. Since all issuing locations are already marked as <placeName>, it is thus possible to match those names with authority files and geodata, for example GeoNames or – what is preferable in this case – the historical geodata registered in TORA (see below).

There are also other types of locational data related to the Swedish medieval charters. Most place names appear within a <div> tag that holds the summary of the transcription (the text under the heading “Contents” in figure 2). The summary contains one or more paragraphs <p>, without any annotations of the place names, persons etc. In order to geocode the place names mentioned in the summary – which are essentially all the names that are also present in the entire text of the text – these must first be identified as place names, i.e. <placeName> according to the TEI format. After that they could be matched with geodata.⁹ How this could be done is discussed in the following sections.

3 Geocoding historical places

A place name points to a geographical location. Regarding medieval charters most place names refer to settlement units such as towns, manors, villages, hamlets and single farms,

⁹ Writing summaries of the medieval charters is an ongoing editorial activity at the Swedish National Archives. The letters from the years 1360–1380 have the most detailed summaries. Later on, more place names will be added to the summaries of the documents from other periods.

henceforth referred as *historical settlement units*.¹⁰ There are several tools on the internet that could be used to mark a location on a modern map, as a point or area, and retrieve this data in any format, such as TEI or GeoJSON. It would thus be possible to add coordinates directly to the digital edition of a particular source material.

However, a better approach would be to match place names with existing geodata, provided by for instance GeoNames. Initiatives have also lately been taken to create open geodata for historical sites. One such initiative is the Pelagios project, that connects several sets of geodata, for instance the Digital Atlas of the Roman Empire.¹¹

Analyses of historical documents in relation to spatial data should have a great potential in several research areas. The administrative divisions of Denmark have been digitized within the DigDag project.¹² Mapping and GIS techniques are also regularly used and adapted by researchers, for example, to analyze the routes taken by individuals (Storm et al, 2017). Spatial data from historical maps has enabled identification of old cultivated hops (Strese et al, 2009). Data extracted from historical maps is the basis for an identification of late-medieval deserted farms (Karsvall, 2016).

The project TORA – Topographical Register at the National Archives in Stockholm – specifically defines the historical places in Sweden that existed from the Middle Ages until about the 18th century.¹³ The aim is to link digitized historical sources to well defined geodata. TORA includes place names and coordinates – in the form of points – for all settlement units that appears in the oldest land survey maps from the 17th century (Höglund, 2008; Tollin and Karsvall, 2010). Moreover, a large amount of settlements that appear in the Crown’s cadastres (jordeböcker) in the mid-16th century are also included.¹⁴ At present TORA holds about 22,000 spatial coordinates, which are published as linked data in RDF format.^{15,16}

The coordinates in TORA are set at the actual location of the settlement, according to the oldest map available for each site. The accuracy is specified as high, medium or low. A large

¹⁰ A historical settlement unit refers to a place mentioned in an historical source, here the medieval charters.

¹¹ <http://commons.pelagios.org>; <http://dare.ht.lu.se>

¹² <http://www.digdag.dk/>

¹³ <http://riksarkivet.se/tora>

¹⁴ Starting in the 1960s, the project *Medieval Sweden* (Det medeltida Sverige) extracts information from the medieval documents, as well as the economic provincial material of the 16th century (landskapshandlingarna). TORA includes coordinates of all settlements published in the book series of Medieval Sweden. So far 23 book volumes has been published, covering most parts of south-east Sweden: <https://riksarkivet.se/medieval-sweden-dms>

¹⁵ Resource Description Framework: <https://www.w3.org/RDF/>

¹⁶ The historical settlement units are published in RDF format (a W3C standard) using serial numbers, e.g., data.riksarkivet.se/tora/1. TORA uses the EntryStore platform developed by MetaSolutions AB (<http://entrystore.org/>). For further information see <http://riksarkivet.se/tora>.

number of abandoned settlements, not existing today, also appear in TORA, which is thus the most relevant authority file that can be linked to the place names in the medieval charters.¹⁷

3.1 An example

This relatively typical case refers to a medieval charter from 1366, issued in Eskilstuna in south-eastern Sweden (SDHK 8953; see figure 2). The original letter, written in Latin, has been fully transcribed and interpreted.¹⁸ The place of issue, Eskilstuna, that already has a <placeName> tag, can be assumed to correspond to the town of Eskilstuna. However, the town is of much later date (17th century). Most likely this charter has been issued at the monastery in Eskilstuna, which was established in the 12th century. The monastery was abandoned in the 16th century and replaced by a castle on the same spot. The corresponding settlement units and coordinates for Eskilstuna mentioned in this charter would therefore much likely be ‘Eskilstuna Slott’, as defined in TORA using the oldest maps from year 1647.¹⁹

Eskilstuna is mentioned once more in the <summary> tag in the digital edition in SDHK, as ‘Eskilstuna kloster’, which refers to the monastery. What would complicate an automatic match is the fact that ‘Eskilstuna kloster’ in more recent periods has become the name of the parish in this area. In GeoNames, for example, the coordinates linked to the name of ‘Eskilstuna kloster’ points to the parish area and not the location of the monastery.²⁰

The same charter, SDHK 8953, holds one more place name, Torshälla, just north of Eskilstuna. TORA has two registrations of Torshälla. One that refers to the town, which dates from the early 1300s, and one that refers to the neighboring settlement of a vicarage. It is not clear which of the two that is intended. In such cases it could be useful to make a temporary match that includes the other options available as alternatives.²¹

4 Towards automatic name tagging in SDHK

As mentioned above, every document in the SDHK database has a human-composed summary written in modern Swedish (labeled “Contents” in figure 2), although the language of the summaries sometimes has an archaic feel to it, especially when the original letter is in Old Swedish (and not in Latin as the example in figure 2).

These summaries form the point of departure for the experiment described in this paper. Employing state-of-the-art natural language processing and information extraction techniques, the summaries can be used as a rich source of metadata for the underlying documents,

¹⁷ The method of how to set coordinates for settlement units using historical maps is the subject of a forthcoming article by the first author.

¹⁸ https://sok.riksarkivet.se/bildvisning/Sdhk_8953.jpg

¹⁹ <https://data.riksarkivet.se/tora/2675>; <https://data.riksarkivet.se/tora/2675>; https://sok.riksarkivet.se/bildvisning/R0000151_00049

²⁰ <http://www.geonames.org/8132118/eskilstuna-kloster.html>

²¹ <https://data.riksarkivet.se/tora/12463>; <https://data.riksarkivet.se/tora/2733>

complementing the manually compiled metadata already accompanying them. At least the kinds of (legal) action, the involved individuals and other agents, and the involved locations, should be retrievable from the summaries with automatic methods.

Given the aim expressed above of linking SDHK to the historical place name register TORA, in our first experiment described here, we apply a named-entity recognition (NER) system designed for modern Swedish to the summaries and evaluate its accuracy on data which – although written in the modern language – may contain many names which are no longer current.

4.1 Experimental setup

For this experiment, 14 documents were sampled from the SDHK database and their summaries extracted, amounting to a total of approximately 1,500 words.²² All names of persons and places were manually identified and annotated in the text of the summaries, in order to have a gold standard dataset for evaluating the automated NER accuracy.²³

For the NER we use a rule-based system developed over many years in Språkbanken at the University of Gothenburg (Kokkinakis et al, 2014). It has already proven its worth in some digital humanities applications, although with textual data of more recent date than in the present case (Borin and Kokkinakis, 2010; Borin et al, 2014).

NER is one of the annotations available through Språkbanken’s *Sparv* infrastructure (Borin et al, 2016; see figure 3). *Sparv* exposes the linguistic annotation pipeline used as the first step in Språkbanken’s *Korp* corpus infrastructure (Borin et al, 2012).

Sparv also makes available a corpus upload function for offline processing of larger volumes of text. In this way the 14 SDHK summaries were linguistically annotated.

4.2 Experimental results and discussion

Table 1 shows the accuracy figures for the two types of names.²⁴ We note that the recall for personal names is much better than that for place names, i.e., most personal names in the

²² The sample comprised documents SDHK 2382, 8846, 8847, 8861, 8887, 8896, 8901, 8913, 8931, 8953, 8954, 8955, 8959, 8970.

²³ Using *gold standard* data for evaluating an automatic process is considered absolutely *de rigueur* in natural language processing, as the only way of obtaining an objective and reproducible measure of the accuracy of the result of the process. In the case at hand, this means that the automatic NER is applied to the same documents that were manually annotated but with the manual annotations hidden from the automatic process, and accuracy is then computed from a comparison of the NER annotations with the manual annotations.

²⁴ The figures have been calculated so that partial matches and full matches are counted as equal, i.e., a NER place name match for *Nöttja* is considered a full match for the manually identified *Nöttja socken* ‘Nöttja Parish’. The measures in the last two columns of the table are the standard information retrieval evaluation measures of *recall* and *precision*. Recall is defined as the share of all true instances in the dataset that were actually retrieved, and precision as the share of all the retrieved instances that are true instances. Thus if a dataset contains 50 true instances and a retrieval method returns 40 hits, out of which 35 are true instances, the recall in this case will be 0.70 (35/50), whereas the precision will be 0.88 (35/40).

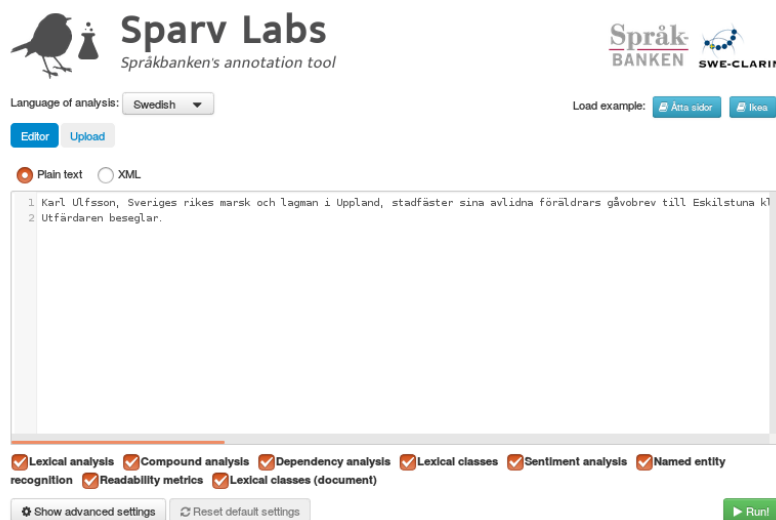


Fig. 3. Språkbanken’s Sparv corpus annotation interface

texts are in fact recovered by the system. Since the NER system relies primarily on gazetteers (name lists) for the recognition of both, a possible reason for this could be that many of the personal names used in the Middle Ages are still around, while place names have seen a higher replacement rate over the intervening centuries.

Table 1. NER accuracy figures

SDHK		NER					
name type	total	correct	false	missed	precision	recall	
Place name	114	62	4	48	0.94	0.54	
Personal name	100	92	8	8	0.92	0.92	

The precision of the NER system used in this experiment is comfortably high for both name types, i.e., there are few false positives in both place and personal name recognition. This means that even out of the box, this NER system could provide considerable added value to the digital resources of the National Archives as research data sources.

Finally, we note that the Sparv annotation pipeline is made up from independent tools of different origin. This means that the tools do not normally draw on each other’s annotations, something which is clearly reflected in the results of this experiment, if we take all annotations into account. Consequently, the named entity recognizer does not utilize the information from the part-of-speech tagger, which in fact tags 42 additional place names and 4 additional

personal names as proper nouns (tag PM), but also gives 15 false positives, i.e., non-names tagged as proper nouns. This gives a recall rate (for names) of 0.82 and a precision of 0.75 and recall rates for place names and personal names of 0.88 and 0.50, respectively. Distributing the false positives proportionally between the two classes of names, we end up with estimated precision figures for place names and personal names of 0.75 and 0.80, respectively.

The analysis, using NER, has so far been targeting the occurrences of the names. However, place names could refer to several geographical locations. Common names of settlements, which are found in our survey material, are e.g. ‘Berga’, ‘Lundby’ and ‘Söderby’. A further geographical identification of the sampled data – the 14 medieval charters with altogether 144 place names – is therefore needed. For this purpose TORA will be used. The geographic accuracy in the NER system and the National Land Survey database (Lantmäteriet, LM) will also be assessed.

The result, presented in Table 2, is based on a semi-automatic method which has two steps: firstly, the place names in the XML documents of the medieval letters have been mapped with the name lists discussed here (TORA, NER and LM); secondly, a manual check has been conducted to evaluate if the matched names refer to the geographical locations mentioned in the documents. It should be emphasized that medieval charters are difficult to interpret and that spatial localization sometimes requires expert analysis. In our sample data, however, all names (with one exception) are normalized, written in modern Swedish.

Table 2. Geocoding of place names in medieval charters

source	correct	uncertain	missed	precision	recall
NER	41	25	48	0.62	0.36
TORA	46	6	62	0.88	0.40
LM	41	24	49	0.63	0.36

As shown in Table 2, TORA, NER and LM provide an equivalent result with a recall rate between 0.36 and 0.40. All three recognize the names of major towns. Also unique and unusual place names are matched with their geographical location. Foreign place names (25 in our sample data) could not be recognized by TORA or LM, as they are limited to the current borders of Sweden.

Overall, TORA is the better recognizer. Especially since TORA has the most accurate coordinates of historical places as discussed before (see section 3). It does not cover general names of countries (as ‘Sweden’) and provinces (as ‘Uppland’). On the other hand, it includes all parishes and also other historical administrative divisions such as hundreds (härad).

The strength of TORA is above all that it recognizes most settlement units, i.e. villages and hamlets. Only six names in this experiment are uncertain giving a precision of 0.88. In comparison, NER and LM show a geographical precision of 0.62 and 0.63. As shown in Ta-

ble 1, NER successfully finds 62 place names. Of these 41 names can be assumed to point out the correct location. The uncertainty is mainly due to the occurrence of common place names as ‘Stenby’, ‘Lundby’, ‘Valla’ among others. With TORA – which includes administrative divisions, such as parishes —it is easier to determine the geographical location of a place name, even if it occurs frequently. The total amount of place names and coordinates in TORA is at present limited. About 50 percent of all historical settlement units are included so far. Names not yet included can, however, be added. As long as TORA is not complete NER and LM will be necessary complements. LM holds all official place names in Sweden today, but it does not cover historical settlements, e.g. disbanded place names and abandoned settlements.

In conclusion, the results of this small experiment are promising, and there is also clearly room for improvement of the NER system. In order to make this technology useful in the present context, the next natural step would be to match recognized names to authority lists such as TORA that give the names an identity, and makes them suitable as linked data. This would serve to add meaningful identities to place names in the medieval documents, by providing accurate coordinates that point to settlement locations specified in the oldest large-scale maps.

5 Conclusions

Since the name information in the Swedish charter today lacks clear identifiers and spatial coordinates, it is difficult to make statistical calculations and any kind of spatial comparative studies. One solution to these problems would be to relate historical data of this kind to actual physical location by using coordinates. To a large extent, matching can be done automatically based on name.

We investigated the feasibility of applying a named entity recognition system on the text of the SDHK document summaries, with quite promising results and some indications of how the NER system could be improved, and we intend to pursue this work further, including how to utilize the added metadata in user interfaces for researchers. This kind of ‘coordinate-based’ linked databases would present quite new conditions and possibilities within a number of disciplines, e.g. history, historical geography, social history, philology, onomastics and archaeology. Irrespective of administrative divisions, data can be collected freely, and previously unknown connections may appear. Large sets of data can be extracted intended for national surveys illustrating for instance regional differences. The scholars and users can devote their time to formulating research questions and making analyses instead of collecting and preparing the research material.

Acknowledgements

The research presented here is part of a project (TORA) funded by the Royal Swedish Academy of Letters, History and Antiquities, Riksbankens Jubileumsfond, and the Swedish

National Archives. The work has also drawn on the e-infrastructure Swe-Clarín, funded by the Swedish Research Council (contract 2013-2003) and by the collaborating partner institutions forming Swe-Clarín (University of Gothenburg, The Institute of Language and Folklore, KTH, Linköping University, Lund University, The National Archives, Stockholm University and Uppsala University).

References

- Borin L, Kokkinakis D (2010) Literary onomastics and language technology. In: van Peer W, Zyngier S, Viana V (eds) *Literary education and digital learning. Methods and technologies for humanities studies*, Hershey, New York, pp 53–78.
- Borin L, Forsberg M, Roxendal J (2012) Korp – the corpus infrastructure of Språkbanken. In: *Proceedings of LREC 2012, ELRA, Istanbul*, pp 474–478.
- Borin L, Dannélls D, Olsson LJ (2014) Geographic visualization of place names in Swedish literary texts. *Literary and Linguistic Computing* 29(3):400–404.
- Borin L, Forsberg M, Hammarstedt M, Rosén D, Schäfer R, Schumacher A (2016) Sparv: Språkbanken's corpus annotation pipeline infrastructure. In: *SLTC 2016. The Sixth Swedish Language Technology Conference*, Umeå University, Umeå.
- Höglund M (ed) (2008) *1600-talets jordbrukslandskap: En introduktion till de äldre geometriska kartorna*. Riksarkivet, Stockholm.
- Jockers ML (2013) *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, Urbana/Chicago/Springfield.
- Karsvall O (2016) *Utjordar och ödegårdar. En studie i retrogressiv metod*. SLU, Uppsala.
- Keim DA, Kohlhammer J, Ellis G, Mansmann F (2010) Mastering the information age – Solving problems with visual analytics. *Eurographics*, Goslar.
- Kokkinakis D, Niemi J, Hardwick S, Lindén K, Borin L (2014) HFST-SweNER: A new NER resource for Swedish. In: *Proceedings of LREC 2014, ELRA, Reykjavik*, pp 2537–2543.
- Moretti F (2005) *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, London/New York.
- Moretti F (2013) *Distant reading*. Verso, London/New York.
- Schöch C (2013) Big? Smart? Clean? Messy? Data in the humanities. *Journal of Digital Humanities* 2(3).
- Storm I, Nicol H, Broughton G, Tangherlini TR (2017) Folklore tracks: Historical GIS and folklore collection in 19th century Denmark. In: Golub K, Milrad M (eds) *Digital Humanities 2016, CEUR-WS.org, Aachen*, pp 75–98.
- Strese EM, Karsvall O, Tollin C (2009) Inventory methods for finding historically cultivated hop in Sweden (*Humulus lupulus* L.). *Genetic Resources and Crop Evolution* 57(2):219–227.
- Tollin C, Karsvall O (2010) Sveriges äldre geometriska kartor. *Bebyggelsehistorisk tidskrift* 60:94–103.