This is the accepted (peer-reviewed) version of the following article:

Abraham S, Mäs S, Bernard L.
Extraction of spatio-temporal data about historical events from text documents.
Transactions in GIS. 2018; 22: 677–696.

, which has been published in final form at **https://doi.org/10.1111/tgis.12448**.
This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

Extraction of Spatio-Temporal Data about Historical Events from Text Documents

Susanna Abraham; Stephan Mäs and Lars Bernard

Abstract: Often, we are faced with questions regarding past events and the answers are hidden in the historical text archives. The growing developments in Geographic Information Retrieval and Temporal Information Retrieval techniques have given new ways to explore digital text archives for spatio-temporal data. The question is how to retrieve the answers from the text documents. This work contributes to a better understanding of spatio-temporal information extraction from text documents. Natural Language Processing techniques were used to develop an information extraction approach using the GATE language processing software. The developed framework uses gazetteer matching, spatiotemporal relationship extraction and pattern based rules to recognize and annotate elements in historical text documents. The extracted spatio-temporal data is used as input for GIS studies on the time-geography context of the German-Herero war of resistance 1904 in Namibia. Related issues when analyzing the historical data in current GIS are discussed. Problematic are in particular movement data that is small scale with poor temporal density and trajectories that are short or connect very distant locations.

Keywords: spatio-temporal event extraction, natural language processing, historical visualization, geographic information retrieval

1 Introduction

Text documents have been central to the humanity sciences long before digitization. For historical studies computational history will play an increasingly important role to access the huge amounts of available data (Yeung et al. 2011). As digital text archives become more easily accessible and contain both explicit and implicit spatial temporal information, researchers in GI-Science have become aware of these new data sources for space-time analysis (Gregory, 2002).

Using unstructured text as in Web documents or historical text documents as sources of spatio-temporal information typically requires Information Extraction (IE) techniques to recognize and retrieve geographical and temporal entities. There has been a growing research interest in geographic and temporal information extraction from unstructured documents (see, e.g., Campos et al. 2014, Strötgen et al., 2010).

IE is the identification and extraction of the most relevant instances of a particular class of events or relationships in a natural language text and their transformation into a structured representation, such as a database (Campos et al. 2014, Feldman et al. 2006). Researchers in GI-Science have embraced IE techniques to explore geographical information from text documents and the capabilities to extract such information improved in recent years (Wang, 2014). IE techniques provide an opportunity to automatically extract spatiotemporal information from text document through Natural Language Processing (NLP) and Text Engineering methods. NLP techniques help to capture entities with their attributes to answer questions on what, where, when and by whom.

(Katz et al., 2013) employ a Named Entity Recognition (NER) approach that uses Toponyms Recognition (TR) and Toponyms Disambiguating (TD) as main components for their geographic IE framework. TR focusses on identifying and recognizing the occurring location names in the texts, while TD looks into associating the identified toponyms with entries stored in a database, which serves as a gazetteer for cross reference. This aims at a reduction of errors that result from the vagueness in natural language. For instance, a sentence may read "Mali is in Africa", whereby Mali is referring to a person in the document but it could be perceived as referring to the country Mali in Africa. To avoid such confusions, the TD component uses a gazetteer covering the geographic extend of the document to cross reference whether a recognized (potential) toponym can be confirmed – i.e. is contained in the gazetteer database. Temporal information retrieval concepts were not considered in their approach.

According to Strötgen et al. (2010), spatio-temporal information retrieval consists of a Temporal Information Retrieval (TIR) component and a Geographic Information Retrieval (GIR) component, both focussing on the extraction and utilization of temporal and geographic expressions from documents. They proposed an approach for combining TIR and GIR methods in order to produce spatio-temporal document profiles that allow to explore the geographical content of the document in a chronological order. The idea of combining the two methods is based on the argument that spatial and temporal expressions relating to an event usually co-exist in the document. The authors successfully applied the approach for the discovery of event location sequences from documents such as travel reports and documents related to history.

(Wang, 2014) used NLP concepts for the extraction of spatio-temporal data and semantics extraction for natural hazards from Web documents and articles. A similar but more extensive approach is proposed by Bekele (2014) for the extraction of spatio-temporal information from historical gazetteers. The advantage of this approach is that it incorporates NER concepts and rule based method for recognizing patterns in the texts. This also includes spatio-temporal relationship terms such as “2 days later”, “200 km away”. To exploit such relative spatial and temporal information spatial taggers (Leidner et al. 2011) and temporal taggers (Strötgen et al. 2016) are used.

Unlike the methods in the previously mentioned studies, the approach presented here, focusses on the extraction of spatial, temporal and attributive information depending on the linguistic and textual information contained in the documents. NLP techniques are adopted, to extract contextual information that can describe and answer questions pertaining to the movement of the German settlers and the Hereros during the anti-colonial resistance war of 1904 in Namibia. The application example illustrates the transformation of the narratives into spatio-temporal data and allows for analysis and visual representations in GIS.

The rest of the paper is structured as follows. In the next section related work in the area of historical spatio-temporal event extraction and natural language processing is discussed. Section 3 introduces our case study on the anti-colonial resistance war in Namibia and Section

4 provides details on the source data used for this study. In section 5 we describe our approach to spatio-temporal information extraction. The results are described in section 6 and section 7 provides some concluding remarks.

2 Related work

Automated text analysis is used in many different applications, generally aiming at a reduction of the efforts to analyse large amounts of texts. Application examples can be found in many areas for political texts (Grimmer and Stewart, 2013), twitter messages as a supplement to public opinion polls (O'Connor et al., 2010), biomedical literature (Cohen and Hersh, 2005) and news reports about natural hazards (Wang, 2014).

Campos et al. (2014) published an extensive survey on temporal information retrieval and its applications. Strötgen and Gertz published a series of articles focussing for example on the rule-based extraction and normalization of temporal expressions (Strötgen and Gertz 2010), the combination of TIR and GIR (Strötgen et al. 2010) and the extraction, exploration and visualization of event information (Strötgen and Gertz 2012) and temporal tagging (Strötgen and Gertz 2016).

Putting the focus also on geographic locations, Chen et al (2007) worked on mining and geolocating RSS (Really Simple Syndication) feeds to allow users so aggregation and navigation through the contents. Their GeoTracker system supports geospatial and temporal presentation for browsing and personalization of news messages. Mata et al. (2011) introduce an ontology based approach for geographic, temporal and thematic information retrieval. The ontology contains the geographical entities or events of interest. Similarly, Martins et al. (2008) extract geotemporal information from RSS feeds making use of a gazetteer for names of places and historical periods. These solutions also visualized the order of events on interactive maps including a timeline for exploration, but they did not consider moving features as it is done in this research.

Jones et al. (2008) enrich gazetteers with vague places to improve the quality of place-name based IR using knowledge harvested from the web. The inherent uncertainty of the extent of vague places is modelled by the density surface of the frequency of the co-occurrence of place names.

Closer to our application scenario Pfoser et al. (2009) work on accessing history textbooks. They implemented a European history textbook repository integrating metadata from textbooks of various countries and languages. To tag the contents, they use a gazetteer for space, time and thematic categories making use of the GATE software. Another approach is described by Yamamoto et al. (2011) to extract historical events in the from Web resources. Gupta et al. (2016a) use a probabilistic framework to analyze natural language text and determine important events. Therefore, they use semantic annotations on the texts that represent temporal expression, geographic locations and named entities. It also considers uncertainty in temporal expressions and geographic expressions. The same approach has also been used to mine historical documents (Gupta et al. 2016b).

However, most of these research focused only on the extraction of geographical and temporal entities. The movement of features is in most of the approaches not considered. Therefore, in this research we take these approaches a step further to also extract attributive information needed to produce trajectory data of moving entities from historical text documents.

3 Case Study

For the purpose of this research, we study the time-geography concept of the anti-colonial resistance war in Namibia. During the 18th century, the German settlers arrived in South West Africa currently known as Namibia in hope of settling and taking over the land. The Hereros, who were the dominant tribe in central Namibia at that time, possessed a vast amount of land and cattle. Taking advantage of the disunity, the Germans seized almost a quarter of the land and began to split the Hereros using European settlement schemes. However, the Hereros revolted in 1904 under Chief Samuel Maharero along with Hendrik Witbooi and struck out against the Germans. According to (Dierks, 2000) the Hereros besieged Okahandja and

Windhoek where the Germans had their military fort around the 12th January 1904. This marks the beginning of the Herero anti-colonial resistance war. The German empire responded to the pressure from the Hereros with overwhelming force to chase them out of the country. On the 11th of August 1904 the Germans met the Hereros at Waterberg plateau where the Hereros got defeated in a fierce battle. This marked the decisive battle for the Herero war.

For this research, we focus on the time from January to December 1904. The aim of this research is to transforming the textual historical records pertaining to movements of different troops into place and trajectory representations for a visual interactive story map that allows for GIS analysis and interactive visualizations.

4 Source Data

The primary data sources used in this research are PDFs of scanned text documents from the Namibian Environmental Information System, publications from the Namibian Scientific Society, the Namibian National Archive and online articles. The historical publications used are:

4.1 Websites and Online Articles

Chronology of the Namibian history (Dierks, 2000): The Chronology of the Namibian history is a well-researched chronology with an index delineation of the precolonial periods following World War I and the period leading up to the Namibia's independence. This work reflects the origins and migrations of many Namibian communities properly incorporated in a comprehensive index which allows researchers to follow the events in a chronological order. This publication has been the major data source on this research because of its rich coverage and exhaustive information.

The Herero Uprising 11 January 1904 (Namibia 1on1, 2013): Similar to the Chronology of the Namibian history, the Herero Uprising article from the Namibia 1on1 website provides informative descriptions on the events that transpired during this particular period in an indexed approach, which makes it suitable for the information extraction process used in this research.

4.2 Book sources

Let us die fighting (Drechsler, 1966): Let us die fighting is a quotation from a letter sent by Samuel Maharero, Supreme Chief of the Herero, to Hendrik Witbooi, Chief of the Nama, in early 1904. He renounced his early collaboration with the Germans and urged Witbooi to join hands together against the German rule. This study has been commended by the Founding Father of Namibia Dr Sam Nuuyoma in the preface, because of its impressive originality. It is covering quite extensively and exhaustively the entire period of German colonial domination in Namibia. His research work are described to entirely be based on primary sources, i.e. the archival files of the German Imperial Colonial Office in Berlin. In this book a section called "The Herero uprising 1904" explicitly outlines the unfolding events during this period and detailed explanation of the conduct of the Waterberg battle which provided more information regarding the tactics and setup of the German troops by Major General Von Trotha on the 11 August 1904.

The revolt of the Hereros (Bridgman, 1981): The revolt of the Hereros is described as an attempt to tell the story of the Herero war from the Hereros point of view. From this book, the section of the Waterberg battle was used in the research to provide more information regarding the events before, during and after the Waterberg battle.

South West Africa under German rule (Bley, 1971): Bley presents a study of the period of German colonial rule in South West Africa between 1894 and 1914. His work is described as the first study of the movements in South West Africa and it provides a composite picture of the way in which social insecurities, bureaucracy, and rigid economic thinking produced the racial extremist of the last years of German rule (Bley, 1971). In the last section of part three of the book explicitly describes the Hereros' decision to revolt in 1904 and the following section named "Military responses" discusses the revolts of 1904-1907. We found that this sources help with the provision of attributive information and supporting arguments in the events of many battles.

5 Workflow for Spatiotemporal Information Extraction

The spatiotemporal IE framework adopted in this research comprises of four main components namely:

1. Document pre-processing,
2. Creation of spatiotemporal gazetteers,
3. Contextual Information extraction and
4. Trajectory and location event extraction.

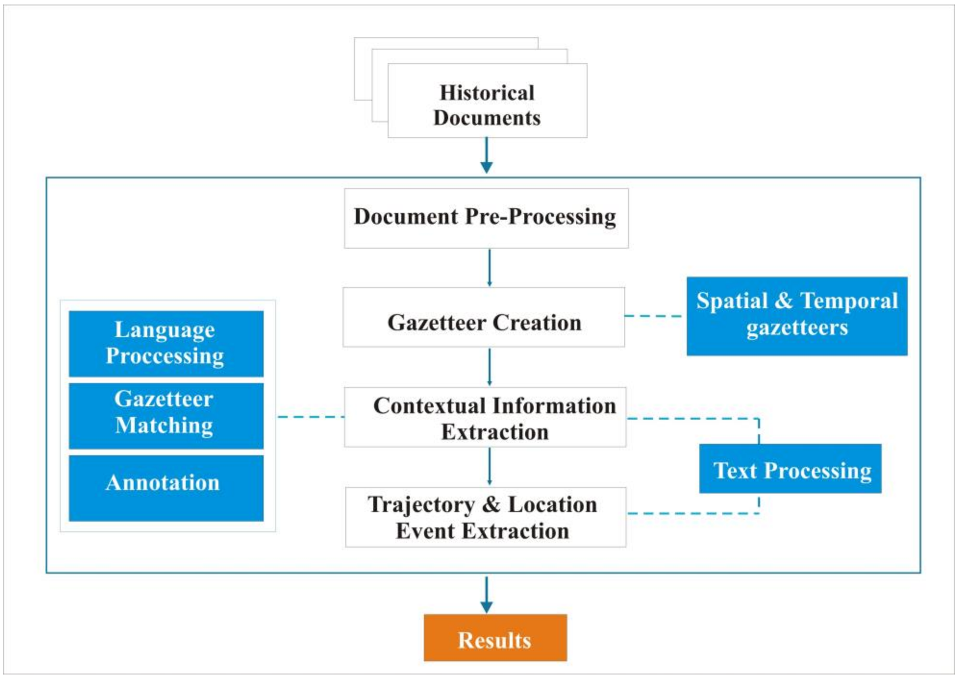


Fig. 1. : Workflow to automatically extract historical spatiotemporal information

The creation of spatiotemporal gazetteers and the extraction of contextual information have been fully implement within the GATE (General Architecture for Text Engineering) development framework[1]. GATE is an open source software for language processing and text analysis.

[1] https://gate.ac.uk

The trajectory and location event extraction is implemented in Python with the aid of the Psycopg[2] adapter that facilitates the communication between the Extensible Markup Language (XML) data source and the PostgreSQL database. The figure below illustrates the framework for automatically extracting spatiotemporal information.

5.1 Document Pre-processing

The document pre-processing handles the raw data conversions and cleaning. It provides the transformations of the raw data sources into a GATE readable format with a defined structure that allows for easy annotation and the extraction steps to follow. As described above, the acquired data sources are scanned PDF documents of chapters taken from historical books and online articles. With the proposed approach for information extraction and for the GATE development framework used in this research, scanned text pages are not recognized and therefore cannot be used in the process unless translated to Microsoft Word, Textfile, Javascript Object Notation (JSON) or XML file. For this work, the latter was chosen as the preliminary data source format for the GATE annotation framework, because of its structure that represents data with descriptive tags for reference and the easy to read tree like structure. The preparation of these XML files was part of the document pre-processing. The order of the document processing steps is depicted in the figure below:

[2] http://initd.org/psycopg/

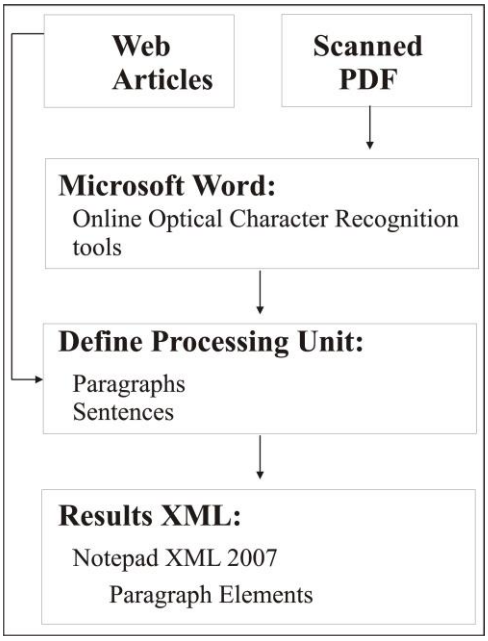
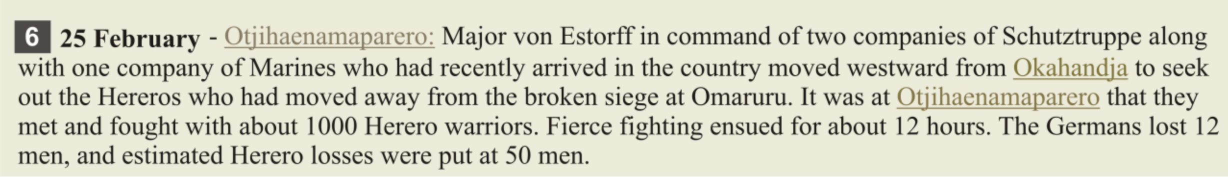


Fig. 2. : Document pre-processing workflow

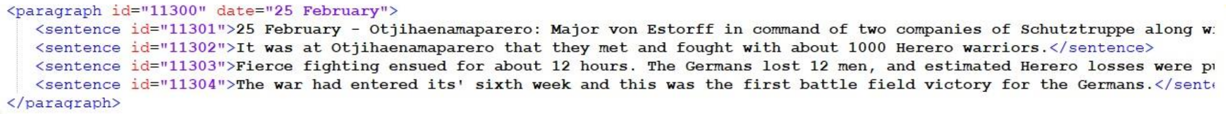
The scanned PDF images and online article contents were transformed in to XML tree like structure with paragraph elements and sentence elements using Notepad XML 2007. Therewith, the order of the events as described in the documents is kept and a processing unit, which represents a single event description, is defined. A sentence is used as a reasoning unit for an event as it contains the information needed for a location event description. It is assumed, that temporal and spatial information and attributive terms relating to one location event are most likely to be found within one sentence and the sentences following contain temporal and spatial terms for the next location event. The extraction of information from the sentence-level is commonly used, also in the literature (e.g. Strötgen et al. (2010)). However, this obviously depends on different authors and the way they express and write. Figure 3 shows an example of a paragraph describing location events that are related spatially and temporally. A paragraph in this sense contains events that happened within the same temporal and spatial boundary. The pattern in which location visit events are described in the document is clearly of crucial importance to distinguish events from each other. Here, the example in Figure 3 depicts how a paragraph contains location visit event information within sentences. The second sentence in the figure is a continuation on the first event but literally a different location visit event. Therefore, a sentence is treated as a location visit unit in this research.



6 **25 February** - Otjihaenamaparero: Major von Estorff in command of two companies of Schutztruppe along with one company of Marines who had recently arrived in the country moved westward from Okahandja to seek out the Hereros who had moved away from the broken siege at Omaruru. It was at Otjihaenamaparero that they met and fought with about 1000 Herero warriors. Fierce fighting ensued for about 12 hours. The Germans lost 12 men, and estimated Herero losses were put at 50 men.

Fig. 3. : Screenshot of the paragraph in an article from (Namibia 1on1, 2013)

In addition, the paragraph contains a bolded date at the beginning, which implies that the events described within the paragraph happened either on that date or on close dates, if no other temporal expression is found within the sentences. For this reason, the date at the beginning of the paragraph is captured in the XML as a date attribute for the paragraph element. It serves as reference date later for the location visit events from the same paragraph that do not have temporal expressions when combining the extracted spatial and temporal references. The result of the data pre-processing step is an XML with a structure that is ready for annotation in the GATE framework. Figure 4 shows an example of such XML document representing the contents of the paragraph shown in Figure 3.



```
<paragraph id="11300" date="25 February">
   <sentence id="11301">25 February - Otjihaenamaparero: Major von Estorff in command of two companies of Schutztruppe along w:
   <sentence id="11302">It was at Otjihaenamaparero that they met and fought with about 1000 Herero warriors.</sentence>
   <sentence id="11303">Fierce fighting ensued for about 12 hours. The Germans lost 12 men, and estimated Herero losses were p
   <sentence id="11304">The war had entered its' sixth week and this was the first battle field victory for the Germans.</sent
</paragraph>
```

Fig. 4. : XML example resulting from the document pre-processing

5.2 *Creation of Gazetteers*

A gazetteer is a geographic index or dictionary. The gazetteer provides a list of place names and their alternative names which is used to find matching patterns in the text documents. In this step, spatial and temporal gazetteers were built to help in entity recognition and extraction from the source documents.

Named Entities in this section refer to main elements in the texts belonging to predefined categories such as persons, organisations, locations, date expressions, etc., while NER describes the task of identifying these named entities in the texts documents. With the GATE annotation framework, there are two ways to identify entities in an input text:

1. By matching entities stored in named entity lists in the gazetteer against the input texts in the document.
2. By matching the patterns of entities with JAPE (Java Annotation Patterns Engine) grammar rule (see next section).

All entities that do not require pattern matching such as months, person names, direction indicators and place names are recognized by gazetteer matching and they need named entity lists to be created. The main gazetteers for this research are spatial gazetteers for place names and spatial expressions and temporal gazetteers for temporal expressions.

Part of the research objectives was to design self-tailored gazetteers to acquire a better understanding on the development, incorporation and usage of gazetteers in IE. On this note, GATE development framework was used with the help of JAPE transducer which provides a platform to self-tailored gazetteers.

Unlike other researchers who used existing temporal taggers and geotaggers, the historical publications used here contained narrated temporal and spatial expressions, including traditional location names which do not exist anymore or have been renamed since then, hence would not be available in an existing geo-tagger. On this note, self-tailored spatial gazetteers were deemed suitable for this research. Due to the small temporal coverage of this study and software dependency issues for the Heidel-time wrapper and GATE-Time temporal taggers, inferring of vague and normalization of temporal expression was rather done manually with the help of the sentence id incorporated in the extraction processes to facilitate the manual verification and normalization of temporal expressions. However, if possible we recommend the use of already existing temporal taggers such as GATE-TIME (Derczynski et al. 2016) or SUTime (Chang et al. 2012).

Spatial Gazetteer Creation. Spatial gazetteers play an important role in geographic information retrieval. One important point to consider when building an information extraction application using NLP and specifically the GATE software, is the ability of the gazetteers to capture the domain specific information and also the geographic extend. For the purpose of this research, the gazetteer to be used ought to recognize the traditional location names for Namibian places. Most place names mentioned in the documents for the Anti-colonial war in Namibia are historical names, which have been renamed after the Namibian Independence in 1991. The ANNIE gazetteer (A Newly New Information Extraction System) included in the GATE system is UK based and turned out to lack most place names found in the historical data sources. Therefore, these historical place names and their alternative names were collected and added into the spatial gazetteer. The local place names have been gathered from the Namibian Statistics Agency (NSA), Namibia Environmental Information Service online and from the Geonames gazetteer online.

Temporal gazetteer creation. Extracting temporal references from text documents in GATE can limit the possible amount of information to be extracted because not all temporal expressions are captured in the ANNIE gazetteer. Not all temporal terms occur in a definite and usual date formats such as DD-MM-YYYY. In order to obtain full temporal coverage of the documents as many as possible temporal expressions should be identified and extracted. For that, we had to extend the GATE's temporal gazetteer to also identify other patterns of temporal references as used in the publications, for example "11.09" or "January 13, 1904". A gazetteer for the different date formats is created using the JAPE grammar rule and stored as a JAPE transducer.

The temporal entity gazetteer created constitutes of the following date formats:

Table 1. Temporal entity gazetteer

No.	Entity	Pattern
1.	Date	June 1904
2.	Date	June 13
3.	Date	June 13, 1904
4.	Date	13 June
5.	Date	13 June 1904
6.	Date	11.06.
7.	Date	11.06.1904

5.3 Contextual Information Extraction

The spatial and temporal gazetteers and the XML documents from the document pre-processing are input for the contextual information extraction. Here, the main aim is to parse the sentences in the XML, which represents location visit events, and to match the gazetteer against the input files, annotate and tag the names of persons, as well as the spatial and the temporal information.

Contextual Information Extraction contains two main extraction pipelines namely the entity extraction pipeline and the spatio-temporal relationship extraction pipeline. Both are developed in GATE using the ANNIE gazetteer and JAPE Grammar rules. The entity extraction pipeline is run first. The order is very crucial, as the spatio-temporal relationship pipeline uses the annotation results from the entity extraction pipeline as inputs. Figure 5 illustrates the components of the Contextual Information Extraction process.

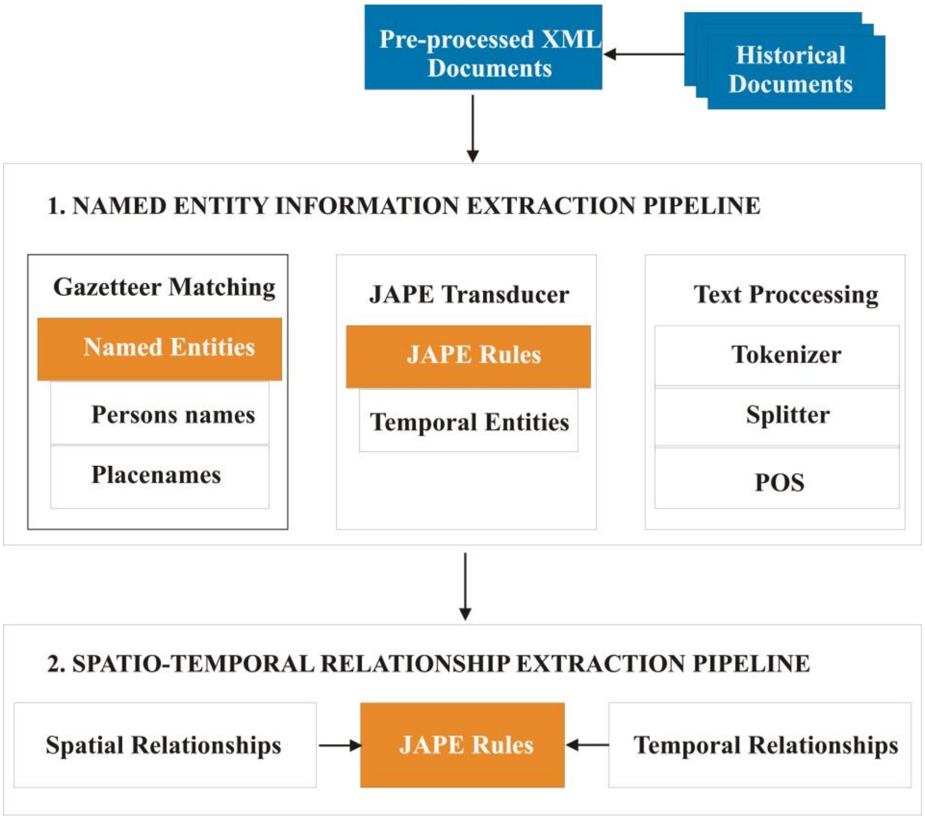


Fig. 5. : Contextual Information Extraction workflow

Named Entity Extraction pipeline. The Named Entity Extraction pipeline identifies and annotates the named elements (required are place names, date expressions and person names) in the XML documents. The Entity Extraction pipeline contains three processing sections namely: Gazetteer processing, Text Processing and Java Annotated Pattern Engine (JAPE) transducer processing. The Text Processing is done first to prepare the texts contents for further processing. Subsequently, the gazetteer matching and the JAPE transducers follow. NER is done either by the gazetteer processing or the JAPE transducer, depending on the entity being annotated. In this section, all named entities are to be annotated and assigned to an annotation class. After this stage, the person names, location names and date expressions are annotated and ready for the spatiotemporal relationship extraction pipeline.

The **Text Processing machine** acts as a temporal annotator; in this case, it does text processing using the Processing Resources (PR) available in GATE ANNIE. The reason for having the text processing machine before the NE extraction resources is that it recognizes the sentences, words, punctuations and language resources needed in the next stages. The processing resources used in this step are:

- English Tokenizer: splits the contents into token types such as words, punctuations, numbers and space tokens.
- Sentence Splitter: is responsible for segmenting texts into sentences.
- Part Of Speech Tagger (POS): is the last text processing tool that assigns POS tags to each annotation token, the tags mostly indicate a linguistic category which a token (e.g. word) belongs to, for instance Proper Nouns in singular form.

As mentioned before, NER can be done either by gazetteer matching or by the JAPE transducer. The named entities extracted for this research are place names, person names and temporal entities. The gazetteer annotator takes the words in a sentence and matches them against the words in the lists in the gazetteer. If there is a match found the word gets an annotation class tag.

JAPE[3] is a finite state transducer that operates over annotations based on regular expressions. It is useful for semantic extraction and used here because of its ability for pattern matching. A JAPE grammar consists of a set of phases, each of which consists of a pattern or rule. For instance, the temporal transducer used to match the temporal entities contains a total of seven date patterns within one transducer, of which each pattern has its own rule describing that pattern. For NER in this section, two transducers are used: the temporal transducer and the special names transducer transducer which is used to recognize special names that have a pattern containing words from other Named Entities, such as Great Waterberg or Klein Barmen.

For the seven different date formats (see Table 1) JAPE grammar rules were developed. The temporal transducer uses the annotation classes from the gazetteer-processing machine and the text processing. The annotation classes for the date patterns are as follows:

1. MonthYear: If a month is followed by a four digit number, it is annotated as MonthYear, such as July 1904.
2. MonthDate: If a month is followed by a one digit or a two digit number, it is annotated and assigned to annotation class MonthDate, such as July 13.
3. MonthDateYear: If a month is followed by a one digit or two digit number which is followed by a ‘,’ and a four digit number, it is assigned to annotation class MonthDateYear, such as July 13, 1904.
4. DateMonth: If a one digit or a two digit number is followed by a month, it is annotated and assigned to annotation class DateMonth, such as 13 July.
5. DateMonthYear: If a one digit or a two digit number is followed by a month followed by four digit number, assign it to class DateMonthYear, for instance 13 July 1904.
6. NumberDate: If a two digit number is followed by a ’,’ or ‘-’ or ‘.’, and followed by a two a digit number, it is assigned to annotation class NumberDate, for instance 13.07.
7. NumberDateYear: If a one or a two digit number is followed by a ’,’ or ‘-’ or ‘.’, it is assigned to NumberDateYear annotation class such as 13.07.1904.

[3] https://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf

The results of the Named Entity Extraction pipeline are annotated texts assigned to their annotation classes. Information such as “12km from the river” and “2 days later” are spatiotemporal relations that express the temporal period and spatial location of an event in a different way. The recognition of spatio-temporal relationship terms allows the extraction of the full spatial and temporal coverage of the input texts. This is discussed in the next section.

Spatiotemporal relationship extraction. An essential function of a language is the ability to express spatial relationships between objects and their relative location in space (Kordjamshidi et al., 2011). Such relationship expressions may also be useful as they give an approximation on where or when something happened. The spatiotemporal relationship extraction pipeline consists of two transducers, the (i) spatial relationship transducer and the (ii) temporal relationship transducer. As mentioned earlier, the spatiotemporal relationship extraction pipeline needs to follow the entity extractions, as it uses the annotation classes of the entity extraction pipeline as its inputs.

The **spatial relationships transducer** takes the annotated location entities from the entity extraction component and annotates the relationships between the location entities. The spatial relationship transducer contains 14 rules for the spatial expression patterns found in the documents. Examples of spatial relationship rules in this research are:

- Rule 1: *Between location and location* e.g. Between Windhoek and Karibib.
- Rule 2: *Directional indicator followed by the location(s)* e.g. North of Windhoek.
- Rule 3: *Distance followed by a direction indicator, followed by a location* e.g. 20km north of Omangeti.
- Rule 4: *Spatial term followed by location, followed by another location and another location* e.g. Areas of Omangeti, Windhoek and Rehoboth.
- Rule 5: *Directional indicator followed by “of”, “the” and location* e.g. South of the Great Waterberg
- Rule 6: *“from” followed by location, followed by “via” location, followed by location, followed by “and”, followed by location* e.g. from Otjimanangombe via Epata, Otjosondu and Osondema to Otjituuo

- Rule 7: *"about" followed by distance* e.g about 25 miles.

The **temporal relationship transducer** works similar to the spatial relationship transducer. Here, expressions such as "between June and July" or "the day of the battle" are extracted and assigned to the temporal relation's annotation class.

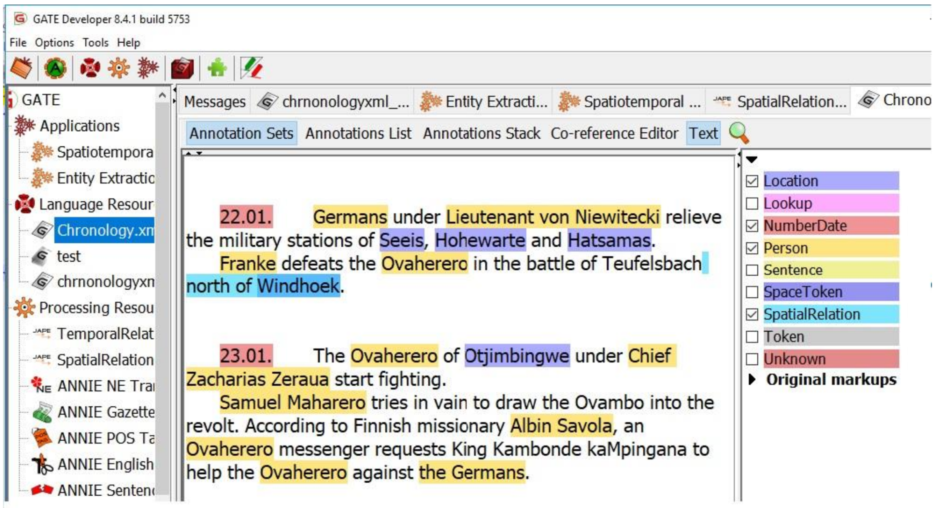
The tables and figures below summarize the annotated entities and relationships by the gazetteers or by the JAPE transducer, respectively.

Table 2. Examples of entities extracted by the JAPE Transducer

No.	Entity	Pattern	Annotation Class
1.	Date	June 1904	MonthYear
2.	Date	June 13	MonthDate
3.	Date	June 13, 1904	MonthDateYear
4.	Date	13 June	DateMonth
5.	Date	13 June 1904	DateMonthYear
6.	Date	11.06.	NumberDate
7.	Date	11.06.1904	NumberDateYear
8.	Special Name	Great Waterberg, Little Waterberg	Location
9.	Unit Names	Unit Erstoff, Unit von Trotha	Units

Table 3. Examples of relationships extracted by the JAPE Transducer

No.	Entity	Pattern	Annotation Class
1.	Spatial Relationship	20km southwest of the Great Waterberg	SpatialRelation
2.	Temporal Relationship	2 days later	TemporalRelation
3.	Temporal Relationship	The day of the battle	TemporalRelation
4.	Temporal Relationship	between June and July	TemporalRelation



22.01. Germans under Lieutenant von Niewitecki relieve the military stations of Seeis, Hohewarte and Hatsamas.
Franke defeats the Ovaherero in the battle of Teufelsbach north of Windhoek.

23.01. The Ovaherero of Otjimbingwe under Chief Zacharias Zeraua start fighting.
Samuel Maharero tries in vain to draw the Ovambo into the revolt. According to Finnish missionary Albin Savola, an Ovaherero messenger requests King Kambonde kaMpingana to help the Ovaherero against the Germans.

☑ Location
☐ Lookup
☑ NumberDate
☑ Person
☐ Sentence
☐ SpaceToken
☑ SpatialRelation
☐ Token
☐ Unknown
▸ Original markups

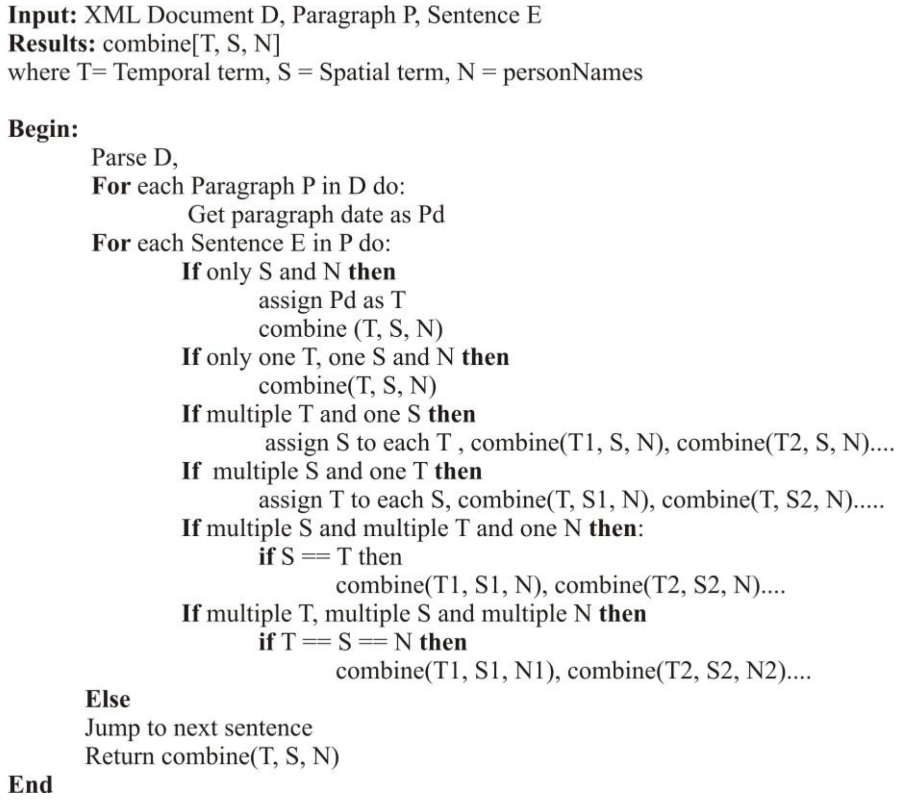
Fig. 6. : Entity recognition in GATE

Figure 6 shows the results of the entity recognition in GATE. Each annotation class is represented by a colour. The annotation list on the right contains classes of entities recognized in the document.

5.4 Trajectory and location event extraction

The fourth step in the workflow is the extraction process which produces location visit events by combining annotated spatial and temporal terms together with the attributive information such as the person names. A person name here is the name of the commander of a troop or a unit. The location visit event is described by the location name, date and the troop that was at that location. The approach of automatically extracting spatio-temporal information from the text document combines the annotated elements using the order of their appearance in the document. The extracted location visit descriptions are then automatically stored in the database (here PostgreSQL). This process is implemented in Python 3.4. In a next step it is possible to produce individual troop trajectories from the location visit events.

The extraction approach considers the textual arrangement in the sentences and assigning appropriate location, date and person's names to the appropriate location visit. As mentioned earlier, the algorithm developed takes the sentence as the processing unit for a location visit event and extract the spatial and temporal information jointly using the developed rules. The temporal term is assigned from the temporal named entity or the temporal relationship term. The same applies to the spatial term. However, not every sentence with a person name would have the spatial and temporal information and let alone the order in which the spatial and temporal expressions appears in the sentences. The following algorithm (Figure 7) is implemented to handle the cases where multiple persons, multiple temporal events and spatial locations are found in a sentence.



```
Input: XML Document D, Paragraph P, Sentence E
Results: combine[T, S, N]
where T= Temporal term, S = Spatial term, N = personNames

Begin:
        Parse D,
        For each Paragraph P in D do:
                 Get paragraph date as Pd
        For each Sentence E in P do:
                 If only S and N then
                         assign Pd as T
                         combine (T, S, N)
                 If only one T, one S and N then
                         combine(T, S, N)
                 If multiple T and one S then
                          assign S to each T , combine(T1, S, N), combine(T2, S, N)....
                 If  multiple S and one T then
                         assign T to each S, combine(T, S1, N), combine(T, S2, N).....
                 If multiple S and multiple T and one N then:
                         if S == T then
                                 combine(T1, S1, N), combine(T2, S2, N)....
                 If multiple T, multiple S and multiple N then
                         if T == S == N then
                                 combine(T1, S1, N1), combine(T2, S2, N2)....
        Else
        Jump to next sentence
        Return combine(T, S, N)
End
```

Fig. 7. : Algorithm to extract location visit events

(Hornsby, 2010) investigated five different cases in which spatio-temporal information occurs in text documents. In our case, the following possibilities are found:

1. Only spatial term and person names are present,
2. One spatial term and one temporal term,
3. Multiple temporal terms and one spatial term,
4. Multiple spatial terms and one temporal term,

5. Multiple spatial terms and multiple temporal terms.

Only spatial information. In a lot of cases were only persons and spatial information found in a sentence, for example, "*Tjetjo's group has been hiding in Okahandja, trying to escape the Germans.*" Tjetjo is a Herero Chief and Okahandja is a location name. The algorithm checks the whole sentence if there is a temporal term and if not, it checks if there is a paragraph date and assigns it as a temporal reference.

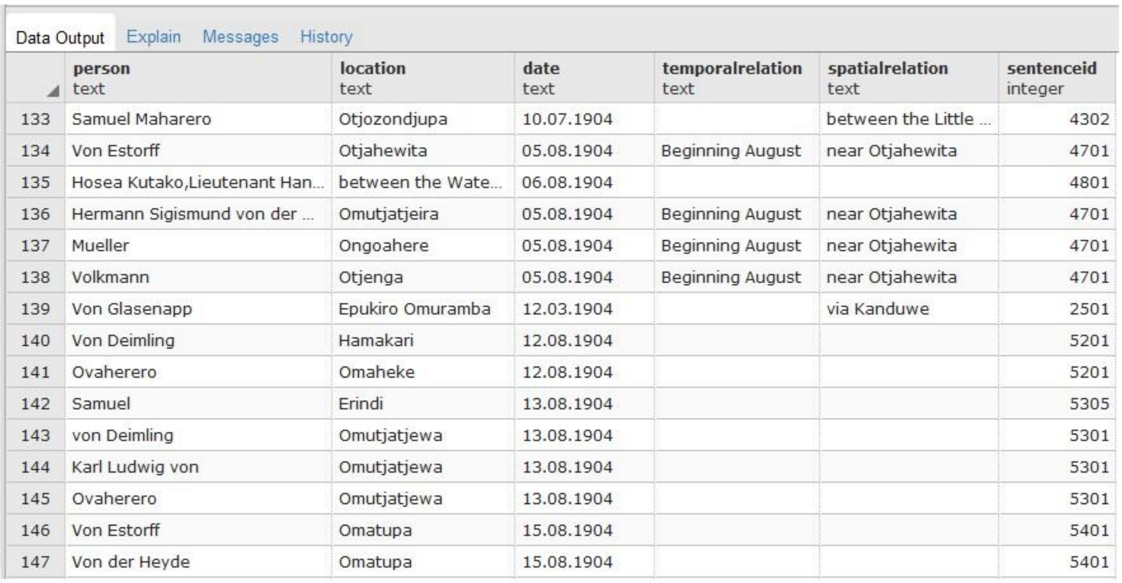
Only one spatial term and one temporal term. Some sentences have only one spatial term and one temporal term. In this case the spatial term and the location term are combined to one location event description together with the persons names found in the sentence. For instance, "*On the 06.01 Kurt Streitwolf reports on a meeting with Traugott Tjetjo in the Gobabis district*" whereby 06.01 is a temporal term, Kurt Streitwolf and Traugott Tjetjo are person's names and Gobabis a spatial term. In this example, where only one spatial term, one temporal term but multiple persons names are contained, the algorithm takes all the persons, the spatial term and temporal term and combines them into a location event description.

Multiple temporal terms and one spatial term. Sentences with multiple temporal terms and one spatial term can be a little tricky. In this case, the spatial term is assigned to each temporal term respectively. This will then constitute multiple location events on different dates. For example, "*Leutwein have been involved in skirmish with Ovahereros in Omaruru on the 12.01. and on the 17.01.*" In this example, Leutwein is a German commander, Omaruru is the spatial term while 12.01 and 17.01 are the temporal terms. The algorithm creates two location visit event description for each date, i.e. [Leutwein, Omaruru, 12.01] and [Leutwein, Omaruru, 17.01].

Multiple spatial terms and one temporal term. It is also possible for one temporal term to be mentioned with multiple spatial terms in a sentence such as, "*Von Glasenapp's unit remains defensive for the time being and is allowed to march to Otjihangwe and later to Otjihaenena arriving on 24.04.*" In this case, Von Glasenapp is a German unit, Otjihangwe and Otjihaenena are location names and 24.04 is the date. It is not clear as to when exactly they arrived in Otjihangwe, but the sentence says they arrived on the 24.04 in Otjihaenena. For this case, it is assumed that they arrived in both locations on the 24.04 and two location visits will be extracted for Otjihangwe and Otjihaenena.

Multiple temporal terms and multiple spatial terms. There are cases where multiple temporal terms and multiple spatial terms exist in a sentence. By looking into the data sources, in most cases this appears when describing the positions of different force units at different dates. And for this reason, it is assumed that the terms close to each other are of the same location visit, for example: "*The uprising is triggered off at different times: Okahandja: 12.01 by Chief Zeraua; Omaruru: 17.01 by Maharero and Otjimbingwe: 23.01 by Riraua*". In this example, Okahandja, Otjimbingwe and Omaruru are spatial terms and the temporal terms are 12.01, 17.01 and 23.01. In cases like this, where the spatial terms and temporal terms are equal to the number person names in the sentence, the algorithm assigns the first temporal term, first spatial term and first person name to one location visit description and so forth. Otherwise if the number of persons are not equal to the number of spatial terms and temporal terms, it assigns all persons to each location visit. But it is important, that a location visit can only have one spatial term.

In events where multiple expression have been found, there is high risk for errors and misplacements of information. The reference id for each sentence is also captured to be used in the data cleaning and organisation stage, which is aimed at confirming the extracted information against the data sources. All the location visit events are written to the database table (Figure 8) for geocoding and cleaning.



	person text	location text	date text	temporalrelation text	spatialrelation text	sentenceid integer
133	Samuel Maharero	Otjozondjupa	10.07.1904		between the Little ...	4302
134	Von Estorff	Otjahewita	05.08.1904	Beginning August	near Otjahewita	4701
135	Hosea Kutako,Lieutenant Han...	between the Wate...	06.08.1904			4801
136	Hermann Sigismund von der ...	Omutjatjeira	05.08.1904	Beginning August	near Otjahewita	4701
137	Mueller	Ongoahere	05.08.1904	Beginning August	near Otjahewita	4701
138	Volkmann	Otjenga	05.08.1904	Beginning August	near Otjahewita	4701
139	Von Glasenapp	Epukiro Omuramba	12.03.1904		via Kanduwe	2501
140	Von Deimling	Hamakari	12.08.1904			5201
141	Ovaherero	Omaheke	12.08.1904			5201
142	Samuel	Erindi	13.08.1904			5305
143	von Deimling	Omutjatjewa	13.08.1904			5301
144	Karl Ludwig von	Omutjatjewa	13.08.1904			5301
145	Ovaherero	Omutjatjewa	13.08.1904			5301
146	Von Estorff	Omatupa	15.08.1904			5401
147	Von der Heyde	Omatupa	15.08.1904			5401

Fig. 8. : Examples of location events in a PostgreSQL table

The cleaning of the data involves date conversions such as 13 July to 13.07.1904, date type conversion and, in cases where multiple troops are retrieved in a location visit, a cross reference check is done using the sentence id.

The location visit are geocoded making use of the spatial gazetteer built in this study, which also contains the geographic coordinates of all place names in the documents.

6 Results

A total of 263 records have been extracted from the 5 data sources. The historical data comes with temporal and spatial uncertainties that need to be recognised and treated. Problematic is the uncertain duration of events, depending on the event type this can be longer than an individual day. Some of the recognized temporal uncertainties result from ranges of dates used to identify a single event in the text, for instance "between July 13 and July 20, Chief Tjetjo met unit Volkmann in Karibib and violence erupted".

Another problem are the times between successive location stops of an object, for which no information could be found in the texts. If not treated this would results in temporal gaps where the object is not displayed and only appears on the next time step. Spatial uncertainties identified were (i) settlement names that do not exist anymore (ii) approximated locations and (iii) uncertain geographic locations. Textual descriptions always contain uncertainties and ambiguities that cannot always be dissolved. The uncertainties and errors were, as far as possible, treated with automatic or manual corrections, respectively.

The extracted data was translated to two spatio-temporal data types namely (1) moving point features in time and (2) location events in time. The trajectories of the troops were produced from the projection of the moving points on the plane by connecting locations in a temporal order. The location events give an indication of where the most movements were and possibly why. The battle information defines the location and time of individual battle events throughout the country, where by some location may have multiple battles at different dates. With discrete dynamic information in this regard, duration of change is not relevant. The moving point features are considered to exhibit dynamic and continuous dynamic temporal pattern, where the state of the point changes with time or is continuously changing with time.

The choice of tools and functions to use for temporal analysis in GIS depends largely on the scale of the data, the underlying objectives and motivation. The motivation of carrying out temporal analysis in this section is to find out the analysis methods that supports historical movement data in small scale with poor temporal density.

The question to be addressed are: Can we deduce any spatio-temporal reasoning in regards to the time and location of the historical events in our data? Are there any spatio-temporal clusters in movement data? Is there any benefit to spatio-temporal modelling and analysis of these information? Is our data sufficient and suitable for spatio-temporal analysis? To unveil the patterns and answer these questions, the following questions are used to assess the capability and functionality of ArcGIS framework to analyse spatio-temporal data in regards to our historical information:

1. At which areas were there more events?
2. During which time period did the people move more?
3. During the battle events, which were the close moving point features?
4. Are there any relationships between the battle events and movement of different groups?

While spatio-temporal clustering is a process of grouping objects based on their spatial and temporal similarity, the analysis of spatio-temporal data in context requires both temporal and spatial correlations to group events that are close in space and time. Therefore, we intend to derive and get an insight on the locations which had more events happening compared to others to answer the question “At which areas were there more events?”

The isarithmic map fin Figure 9 shows that during the period of January 1904, the collective movements were more concentrated along the central Namibia. Visualized together with the battle information during the same period, one can conclude as to why there has been more movements around those areas.

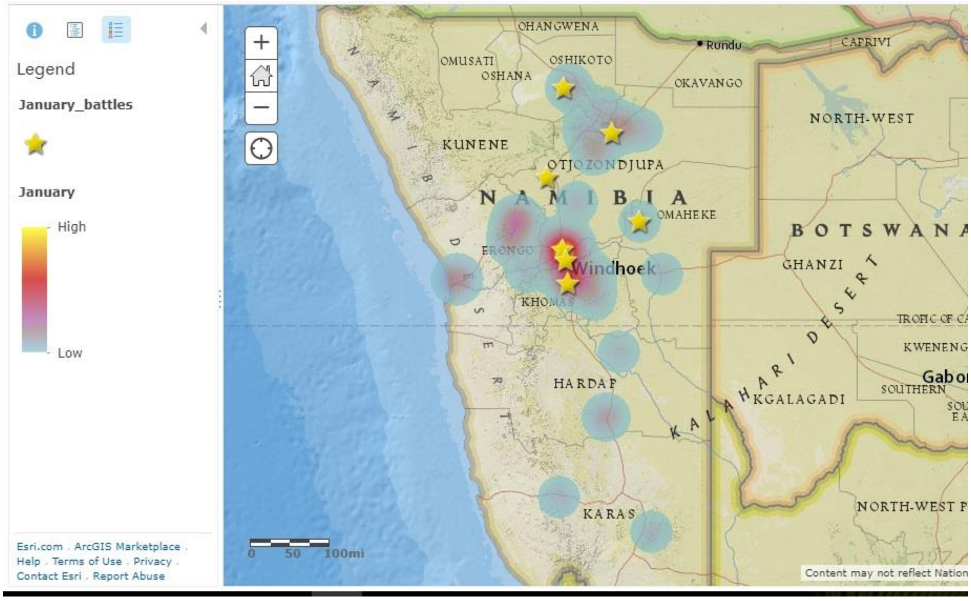


Fig. 9. : Isarithmic map of January movements

Space-time cube analysis shown in Figure 26 shows the 3D space-time cube for location events. Each layer represents location visit events per month aggregated at 50 km distance. The size and colour of the point illustrates the number of location visits at a location. The bigger and darker the point, the more the location visits have been at that position. The location along the center have had more location visits during the months of January and August, slightly significant movements during February, March, April and June around the same areas. To clearly analyse the space cube analysis, it should be visualized in an interactive environment.

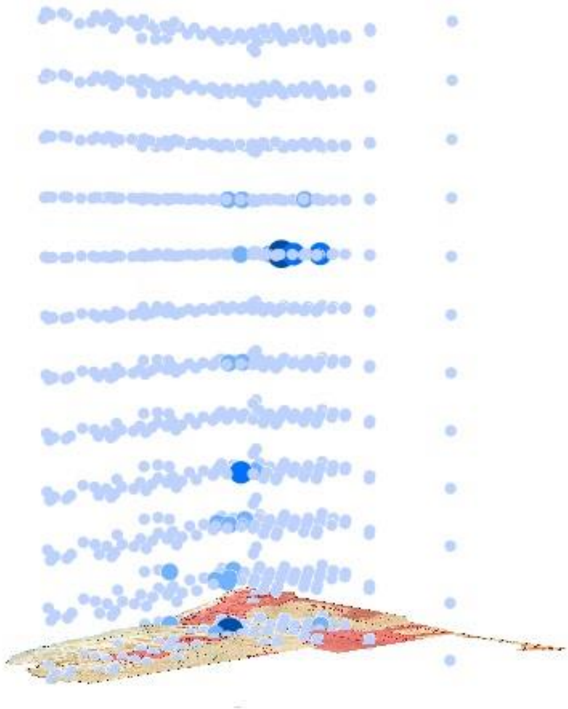


Fig. 10. : Space-time cube

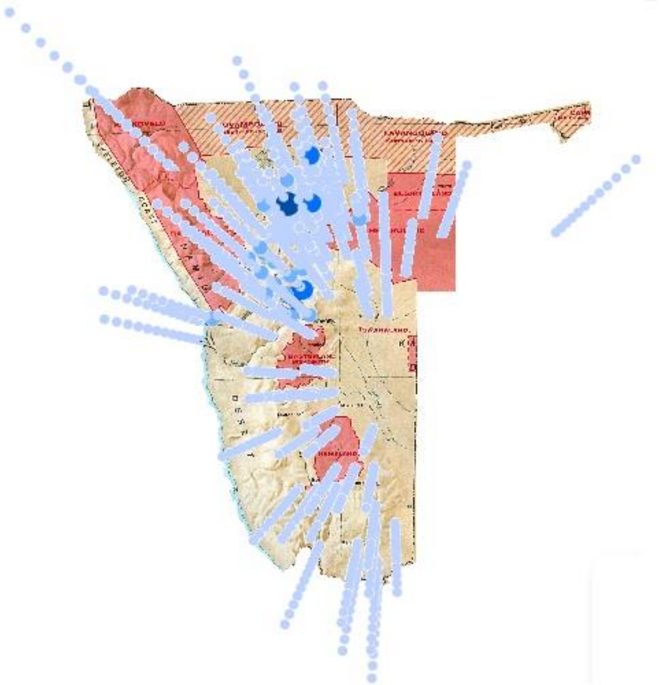
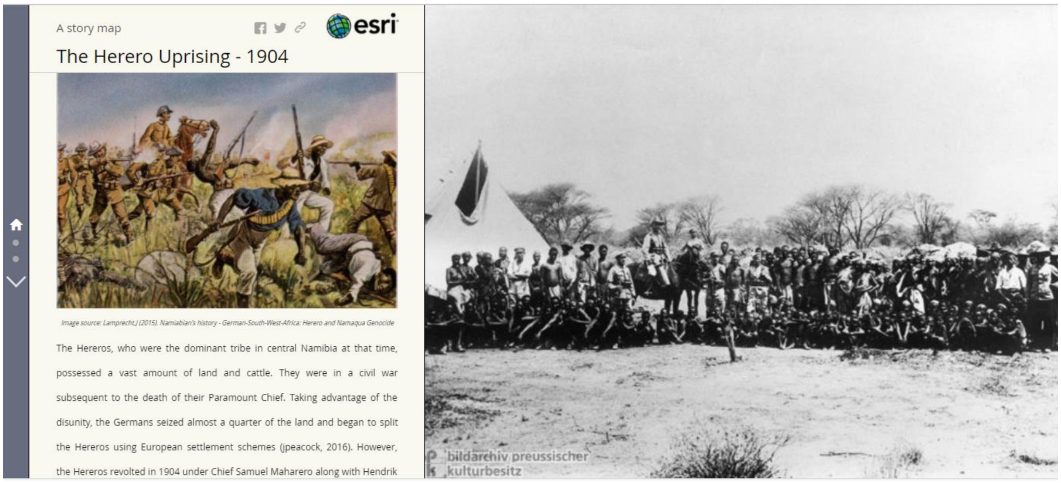


Fig. 11. : Space-time cube seen from top

Time animated Isarithmic maps were created to visually illustrate the movement density from location to locations in time. This helps in answering the question of which time period has had more movements.

Due to the poor temporal density of the data at hand, it was rather difficult to answer spatial questions in time, such as during a certain date, who was close in time and location? Such functionality is still lacking in ArcGIS and would have been of great significance in studying spatio-temporal patterns. However, time-aware maps provide the benefit of visualizing objects as they move in time and draw relations to relative objects in time and space.

For the visual representation of the extracted data ESRIs story map journal was used: A series of maps were dynamically linked together to represent the order of historical events during the Herero Uprising war in 1904. With the benefit of time representations, the time aware map is dynamically linked into the story map to illustrate the movement events during the decisive battle of Hamakari. The story map is created and hosted on ArcGIS online (Figure 12, http://namshopping.com/hererouprising/#).


A story map

The Herero Uprising - 1904

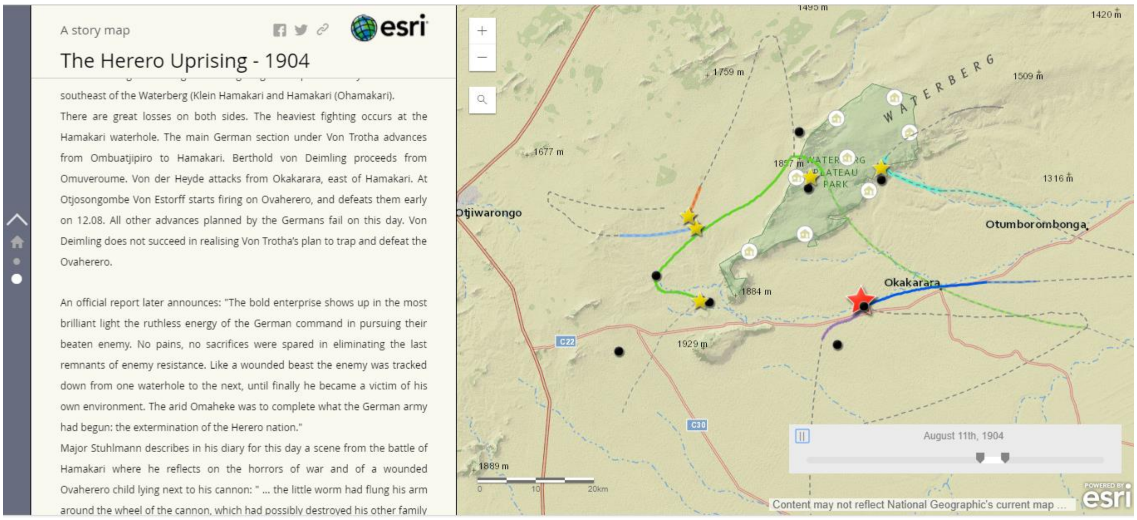
Image source: Lamprecht.J (2015). Namiabian's history - German-South-West-Africa: Herero and Namaqua Genocide

The Hereros, who were the dominant tribe in central Namibia at that time, possessed a vast amount of land and cattle. They were in a civil war subsequent to the death of their Paramount Chief. Taking advantage of the disunity, the Germans seized almost a quarter of the land and began to split the Hereros using European settlement schemes (jpeacock, 2016). However, the Hereros revolted in 1904 under Chief Samuel Maharero along with Hendrik

Fig. 12. : Screenshot of the "Herero Uprising" introductory page of the Story Map

The time aware map in Figure 13 shows the movement of the German troops before and during the battle of Hamakari on the 11 August 1904. A number of six troops had fixed positions on August 5th, surrounding the Waterberg Plateau, where the Hereros had their camps (hut icons). From the fifth, they advanced towards the Herero camp from different directions.

On the 11th August, each troop stood at a position close to the Hereros and the battle began. In the map, the coloured lines represent the trajectories of the troops and the stars represent the different battle fields. The red star is the location for the Hamakari battle where the most fighting took place. The time map contains the time slider, which provides animation controls and allows users to pause and zoom to feature of interest at that point on the time line. Along with the time map, the texts (taken from (Bridgman, 1981) and (Dierks, 2000)) on the left narrates the map situation for ease understanding and orientation.



A story map

The Herero Uprising - 1904

southeast of the Waterberg (Klein Hamakari and Hamakari (Ohamakari).

There are great losses on both sides. The heaviest fighting occurs at the Hamakari waterhole. The main German section under Von Trotha advances from Ombuatjipiro to Hamakari. Berthold von Deimling proceeds from Omuveroume. Von der Heyde attacks from Okakarara, east of Hamakari. At Otjosongombe Von Estorff starts firing on Ovaherero, and defeats them early on 12.08. All other advances planned by the Germans fail on this day. Von Deimling does not succeed in realising Von Trotha's plan to trap and defeat the Ovaherero.

An official report later announces: "The bold enterprise shows up in the most brilliant light the ruthless energy of the German command in pursuing their beaten enemy. No pains, no sacrifices were spared in eliminating the last remnants of enemy resistance. Like a wounded beast the enemy was tracked down from one waterhole to the next, until finally he became a victim of his own environment. The arid Omaheke was to complete what the German army had begun: the extermination of the Herero nation."

Major Stuhlmann describes in his diary for this day a scene from the battle of Hamakari where he reflects on the horrors of war and of a wounded Ovaherero child lying next to his cannon: " ... the little worm had flung his arm around the wheel of the cannon, which had possibly destroyed his other family

Fig. 13. : Example of a time aware story map - Battle of Hamakari 1904

7 Conclusions

The results demonstrate that automatic extraction of spatio-temporal information about historical events can provide a structured data basis that can, for example, be used for interactive visual history teaching systems. Such systems can significantly improve the knowledge transfer and the understanding of historical processes, beside of being much more engaging than traditional methods (Kossatz et al. 2017).

In this work, the trajectories are modelled as point data extracted from the location visit events. Analysing such trajectories for historical events can be challenging, because some trajectories are short or connect very distant locations and discrete times. The extracted historical movement data is relatively small scale with poor temporal density. In current GIS, it is difficult to represent such data due to the lack of functionalities to estimate time and locations between known points. However, GIS should be capable to model and represent trajectories regardless of the characteristics. In terms of visualization, the display of movement paths as lines is unnatural and does not give a realistic representation. Having functionalities that show moving progression would be useful here. Beside the representation, also a more detailed analysis of the movements would be of interest. Typical examples are queries regarding the movement of a single object, such as *"On a specific time, where was the object on the journey?"* or regarding the relation of moving objects like *"Where did the two object intersect in time and space?"*. Such analysis is hardly supported by the standard functionalities of current GIS.

The uncertainties in the extracted data partly result from heterogeneous granularity that comes with the way the information is presented in the historical texts. The modelling of such historical uncertainties is another research concept which entails in-depth studying and implementation of methods to handle the inaccuracies. In case of similarities between the detected trajectories corresponding standard methods for similarity detection can be applied. However, these things are out of the scope of this research.

References

Bekele, M. K. (2014). Spatial tracing of historic expedition: from history to trajectory. Master Thesis, University of Twente.

Bley, H. (1971) South West Africa under German rule. Northwestern University Press, ISBN 978-0810103467

Bridgman, J. M. (1981). The revolt of the Hereros. University of California Press, ISBN 0520041135

Campos, R., Dias, G., Jorge, A.M., Jatowt, A. (2014): Survey of Temporal Information Retrieval and Related Applications. ACM Comput. Surv. 47, 2, Article 15, 41 pages. DOI:http://dx.doi.org/10.1145/2619088

Chang, A. X., Manning C.D..(2012) "SUTime: A library for recognizing and normalizing time expressions." LREC.

Chen, Y-F., Fabbrizio, G. D., Gibbon, D., Jana, R., Jora, S., Renger, B., Wei, B. (2007): Geotracker: geospatial and temporal RSS navigation. In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, New York, NY, USA, 41-50. DOI: https://doi.org/10.1145/1242572.1242579

Cohen, A.M.; Hersh, W.R. (2005). A survey of current work in biomedical text mining, Briefings in Bioinformatics, Volume 6, Issue 1, 57–71, https://doi.org/10.1093/bib/6.1.57

Drechsler, H. (1966). Let us die fighting: The struggle of the Herero and Nama against German imperialism (1884-1915)

Dierks, K. (2000). Chronology of the Namibian history 38-1884. Retrieved October 5, 2017, from http://www.klausdierks.com/Chronology/38.htm

Derczynski, L., Strötgen, J., Maynard, D., Greenwood, M.A., Jung, M. (2016): GATE-Time: Extraction of Temporal Expressions and Event. In: Proceedings of the 10th Language Resources and Evaluation Conference, pp. 3702–3708

Feldman, R., Sanger, J.(2006): Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, New York, NY, USA.

Gregory, Ian (2002). A Place in History: A Guide to Using GIS in Historical Research. (AHDS Guides to Good Practice). Oxbow Books

Grimmer, J., Stewart, B. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(3), 267-297. doi:10.1093/pan/mps028

Gupta, D., Strötgen, J., & Berberich, K. EventMiner (2016a): Mining Events from Annotated Documents. In Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval (ICTIR '16). ACM, 261-270. https://doi.org/10.1145/2970398.2970411

Gupta, D., Strötgen, J., & Berberich, K. (2016b). DIGITALHISTORIAN: Search & Analytics Using Annotations. In M. Düring, A. Jatowt, J. Preiser-Kappeller, & A. van Den Bosch (Eds.),HistoInformatics 2016 (pp. 5-10).

Namibia-1on1 (2013). The Herero Uprising 11 January 1904. Retrieved October 5, 2017 from http://www.namibia-1on1.com/herero-uprising.html

Hornsby, K.; Wang. W. (2010): Representing dynamic phenomena based on spatiotemporal information extracted from web documents. In: Proceedings of the Sixth International Conference on Geographic Information Science, GI-Science, Zurich, Switzerland, 2010. Extended Abstracts.

Jones , C. B., Purves, R. S., Clough P. D., Joho H. (2008) Modelling vague places with knowledge from the Web, International Journal of Geographical Information Science, 22:10, 1045-1065, DOI: 10.1080/13658810701850547

Katz, P.; Schill, A. (2013): To Learn or to Rule: Two Approaches for Extracting Geographical Information from Unstructured Text. In Proc. Eleventh Australasian Data Mining Conference (AusDM13). 117-127

Kordjamshidi, P., Frasconi, P., Otterlo, M. V., Moens, M.-F., Raedt, L. D. (2011). Relational Learning for Spatial Relation Extraction from Natural Language. In: Inductive Logic Programming (pp. 204–220). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31951-8_20

Kossatz M., Utzig S., Schneegans S., Lauer F., Westphal T., Geelhaar J., Froehlich B. and Riehmann P. (2017). HistoGlobe - Teaching History Visually. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017), pages 201-208. DOI: 10.5220/0006103102010208

Leidner, J.L, Lieberman, M.D. (2011): Detecting geographical references in the form of place names and associated spatial natural language. SIGSPATIAL Special 3, 2 http://dx.doi.org/10.1145/2047296.2047298

Martins, B., Manguinhas, H.,Borbinha J. (2008): Extracting and Exploring the Geo-Temporal Semantics of Textual Resources. In: 2008 IEEE International Conference on Semantic Computing, Santa Clara, CA, 2008, pp. 1-9. doi: 10.1109/ICSC.2008.86

Mata F., Claramunt C. (2011) GeoST: Geographic, Thematic and Temporal Information Retrieval from Heterogeneous Web Data Sources. In: Web and Wireless Geographical Information Systems. W2GIS 2011. Lecture Notes in Computer Science, vol 6574.

O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; Smith, N. A. (2010): From tweets to polls: Linking text sentiment to public opinion time series. In Fourth International AAAI Conference on Weblogs and Social Media

Pfoser D., Efentakis A., Hadzilacos T., Karagiorgou S., Vasiliou G. (2009): Providing Universal Access to History Textbooks: A Modified GIS Case. In: Carswell J.D., Fotheringham A.S., McArdle G. (eds) Web and Wireless Geographical Information Systems. W2GIS 2009. LNCS, vol 5886. Springer

Strötgen, J., Gertz, M., Popov, P. (2010). Extraction and exploration of spatio-temporal information in documents. In Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR '10). ACM, DOI: https://doi.org/10.1145/1722080.1722101

Strötgen, J., Gertz, M. (2012) Event-centric search and exploration in document collections. In Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (JCDL '12). ACM, 223-232. http://dx.doi.org/10.1145/2232817.2232859

Strötgen, J., Gertz, M. (2010) HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 321-324.

Strötgen, J., Gertz, M. (2016): Domain-Sensitive Temporal Tagging. In: Synthesis Lectures on Human Language Technologies, 2016, Vol. 9, No. 3 , Pages 1-151

Wang, W. (2014). Automated spatiotemporal and semantic information extraction for hazards. Dissertation Thesis, The University of Iowa.

Yamamoto M., Takahashi Y., Iwasaki H., Oyama S., Ohshima H., Tanaka K. (2011) Extraction and Geographical Navigation of Important Historical Events in the Web. In: Web and Wireless Geographical Information Systems. W2GIS 2011. LNCS, vol 6574. Springer

Yeung C.A., Jatowt, A. (2011): Studying how the past is remembered: towards computational history through large scale text mining. In Conference on Information and knowledge management (CIKM '11), ACM, 1231-1240. DOI: https://doi.org/10.1145/2063576.2063755

How to cite this article:

Abraham S, Mäs S, Bernard L. Extraction of spatio-temporal data about historical events from text documents. Transactions in GIS. 2018;22:677–696. https://doi.org/10.1111/tgis.12448