A STUDY IN GEOSPATIAL DEDUCTION
AN INTERACTIVE WEB APPLICATION FOR VISUALISING FICTIONAL LITERATURE ON THE EXAMPLE OF SHERLOCK HOLMES

MASTERARBEIT
zur Erlangung des Mastergrades
an der naturwissenschaftlichen Fakultät
der Paris-Lodron-Universität Salzburg

eingereicht von
Elisabeth Binder (0720731)

Gutachter: Univ.-Prof. Dr. Josef Strobl
Interfakultärer Fachbereich Geoinformatik – Z_GIS
Salzburg, Oktober 2018
Abstract

In Literary Geography, the research focus is on fictional literature as a source of spatial information, thereby maps and visualisations serve as tools for analysing the physical geospace and the literary meta-space. Geovisualisation approaches have proved feasible for this purpose as they foster information gain and facilitate understanding of large complex datasets by integrating principles from cartography, image analysis, exploratory data analysis and GIScience. Web GIS applications relying on common standards and directives for a Spatial Data Infrastructure have thereby become a powerful alternative to desktop GIS. In Literary Geography, they are used in collaborative scholarly projects and for disseminating results to a broad audience.

In this project, Sir Arthur Conan Doyle’s Sherlock Holmes stories serve as an example for creating an interactive web application for visualising the spatial dimension of fictional literature. By leveraging open source web technologies for geovisualisation and analysis, the application provides new insights into an extremely popular work of literature by rendering visible the geographical context. The application is directed at a broad audience including scientific experts and laymen with an interest in fictional literature and digital media and shall serve as a tool for exploration and analysis.

The tripartite architecture for the web application consists of the data collection in a PostgreSQL database, an instance of Geoserver to fetch data from the database and a web server to communicate with the Geoserver and, finally, the user interface programmed in API-extended JavaScript code. The application is to a high degree interactive and dynamic with on-the-fly loading, processing and visualising of data. Functionalities of the application comprise map display with data driven symbologies, exploratory spatial analysis features, graphical charts as well as map animations. Data architecture, data model and basic application design can be considered a model and be easily applied to similar datasets from comparable literary or other sources.
# Table of Contents

1 Background..............................................................................................................................1
   1.1 Early literary geography.................................................................................................1
   1.2 From objectivism to subjectivism to the spatial turn......................................................3
   1.3 Literary cartography.......................................................................................................5
   1.4 Critical literary cartography and de-cartographisation..................................................7
      1.4.1 Critique....................................................................................................................7
      1.4.2 Ontologies.................................................................................................................8
   1.5 Digital literary cartography and re-cartographisation.....................................................12
   1.6 New approaches in cartography......................................................................................13
      1.6.1 Geovisualisation........................................................................................................13
      1.6.2 Analysis and visualisation methods in literary cartography.....................................16
      1.6.3 Dealing with uncertainty..........................................................................................17
   1.7 Web mapping, the geoweb and literary cartography.......................................................19
      1.7.1 Technological advances............................................................................................19
      1.7.2 Fictional literature in the geoweb.............................................................................22
   1.8 Literary mapping projects..............................................................................................23
      1.8.1 Mapping the Lakes....................................................................................................23
      1.8.2 The Digital Literary Atlas of Wales..........................................................................28
      1.8.3 Palimpsest: Literary Edinburgh................................................................................31
      1.8.4 The Literary Atlas of Europe....................................................................................32
      1.8.5 Projects on Sherlock Holmes..................................................................................34

2 Objective..................................................................................................................................37
   2.1 Scope...............................................................................................................................37
   2.2 Choice of data..................................................................................................................39
   2.3 Aims..................................................................................................................................40

3 The data..................................................................................................................................41
   3.1 Data acquisition...............................................................................................................41
   3.2 Ontology..........................................................................................................................43
      3.2.1 Locations and geometry type....................................................................................44
List of Figures

Figure 1: Static map of Samuel Taylor Coleridge's tour through the Lake District, 1802. (Cooper et al., n.d.) .................................................................................................................... 24

Figure 2: Smooth density surface maps of places visited by Samuel Taylor Coleridge. (Cooper et al., n.d.) .................................................................................................................... 25

Figure 3: Smooth density surface map of places not visited personally by Samuel Taylor Coleridge. (Cooper et al., n.d.) .................................................................................................................... 25

Figure 4: Comparative map of Gray's and Coleridge's tours through the Lake District. (Cooper et al., n.d.) .................................................................................................................... 26

Figure 5: Map of Thomas Gray's emotional responses to places mentioned in his journal. (Cooper et al., n.d.) .................................................................................................................... 27

Figure 6: Interactive Google Earth map depicting Gray's and Coleridge's tours as well as a historical map of the Lake District. The respective texts can be opened by clicking on the layers in the legend. (Cooper et al., n.d.) .................................................................................................................... 28

Figure 7: Map illustrating the coverage of Amy Dillwyn's The Rebecca Riots. (Anderson et al., n.d.) ........................................................................................................................................ 29

Figure 8: Plotline 'Plotting the Rebecca Riots' represented in a 'deep map'. (Anderson et al., n.d.) ............................................................................................................................................ 30

Figure 9: Map of locations of blue writer's plaques, selected the plaque for children's book author Roald Dahl. (Anderson et al., n.d.) .................................................................................................................... 30

Figure 10: Red markers for a path and contextual information on a selected point location on the right. (Loxley et al., 2018) .................................................................................................................... 32

Figure 11: Map of Zikmund Winter's 'Meister Kampanus' with a detailed interactive legend and a complex symbology. (Piatti and Weber, 2018) .................................................................................................................... 34

Figure 12: Web map 'The London of Sherlock Holmes' by Thomas Wheeler. (Wheeler, 2017) ..................................................................................................................................................... 36

Figure 13: Web map from 'The London of Sherlock Holmes ... Mapped' by Matt Brown. (Brown, 2009) ..................................................................................................................................................... 36

Figure 14: Highlighting geographical references in the text ..................................................................................................................... 41

Figure 15: Conceptual model applied to geographical references in the Sherlock Holmes stories ..................................................................................................................... 43
Figure 16: Churches (red) around Edgware Road, London (blue). None of them is St. Monica’s mentioned in A Scandal in Bohemia.

Figure 17: Data architecture consisting of three main tiers for data processing and retrieval and additional secondary sources for external WFS, after Fu & Sun (Fu and Sun, 2011).

Figure 18: Database model illustrating main tables, attributes and keys.

Figure 19: Popup showing information for 7 stacked map features.

Figure 20: Control panel for requesting a data subset.

Figure 21: Control panels for switching basemaps.

Figure 22: Control panel for requesting an administrative boundary layer from a WFS.

Figure 23: Control panel for manipulating the map symbology according to the four attributes ‘certainty’, ‘reality’, ‘accuracy’ and ‘locationtype’.

Figure 24: Legend for the multifactorial symbology.

Figure 25: Combined symbology for ‘certainty’, ‘reality’, ‘accuracy’ and ‘locationtype’ shown on London locations from The Red-Headed League.

Figure 26: Results of convex hull and mean centre analysis for a subset of all point features in The Hound of the Baskervilles.

Figure 27: Result of distance measurement. The two points chosen for distance measurements are connected by a light green line. A summary of all pairwise distances within the dataset is given in the panel on the left.

Figure 28: Result of spatial frequency assessment. The darker the colour of the polygon, the more storylocations lie within the boundaries. A popup reveals the number of contained locations on click.

Figure 29: End product of map animation: the animation sequence started with the darkest lines and ended with the lightest.

Figure 30: Information on begin and end point as well as frequency of connection delivered on click.

Figure 31: Control panel for the map animation.

Figure 32: Bar chart of character appearances within the selected subset.

Figure 33: Co-occurrence matrix juxtaposing ‘location type’ and ‘country’ for a subset of feature from The Adventure of Silver Blaze.

Figure 34: Crude polygon simplifications owed to performance reasons.
List of Tables

Table 1: Juxtaposition of the storymap and the web app approach to Sherlock Holmes........38
Table 2: Colour scheme used in text highlighting.................................................................42
Table 3: Geometry type of map features according to the reference scale.........................45
Table 4: Description of ontological categories ‘real’ and ‘fictional’ applied to the project......46
Table 5: Description of ‘historic’ categories for storylocations........................................46
Table 6: Categories for ‘certainty’..........................................................................................47
Table 7: Categories for ‘accuracy’........................................................................................48
Table 8: Values for ‘mappability’..........................................................................................49
Table 9: Location roles used for categorising storylocations.............................................49
Table 10: Software products used for project realisation.....................................................52
Table 11: Properties of ‘locations’.......................................................................................54
Table 12: Properties of ‘storylocations’...............................................................................55
Table 13: Properties of ‘persons’.........................................................................................56
Table 14: Properties of ‘stories’............................................................................................56
Abbreviations

AJAX – Asynchronous JavaScript and XML
API – Application Programming Interface
ASCII – American Standard Code for Information Interchange
CGI – Common Gateway Interface
CRS – Coordinate Reference System
CSS – Cascading Style Sheets
ESRI – Environmental Systems Research Institute
FOSS – Free and Open-Source Software
FOSS4G – Free and Open Source Software for Geospatial
FSF – Free Software Foundation
GIS – Geographic Information System
GIScience – Geographic Information Science
GML – Geography Markup Language
GNU – GNU’s not Unix
GRASS – Geographic Resources Analysis Support System
HTML – Hypertext Markup Language
HTTP – Hypertext Transfer Protocol
JPEG – Joint Photographic Experts Group

JSON – JavaScript Object Notation
KML – Keyhole Markup Language
OGC – Open Geospatial Consortium
OSGeo – Open Source Geospatial Foundation
OSI – Open Source Initiative
OSM – OpenStreetMap
PNG – Portable Network Graphics
SQL – Structured Query Language
SVG – Scalable Vector Graphics
TIFF – Tagged Image File Format
TOPP – The Open Planning Project
URL – Uniform Resource Locator
WCS – Web Coverage Service
WFS – Web Feature Service
WFS-T – Transactional Web Feature Service
WMS – Web Map Service
WMFS – Web Map Tile Service
WPS – Web Processing Service
XML – Extensible Markup Language


1 Background

Engaging with the spatial aspects of fictional literature has a long tradition in both literary and geographical studies that spans more than one hundred years. Throughout the decades, the emerging research subject has been responding to current academic developments with several shifts in focus and paradigms. Additionally, technological advances of the digital age have been offering new methods for analysis, visualisation and publication for geographical research in the realm of literary fiction. Today, literary cartography has been established as an interdisciplinary field of research. Modern media of the geoweb benefit not only the academic pursuit but enable also a variety of crowd-sourced projects.

This master's thesis describes the process and results of creating a prototype for an interactive web application for visualising and exploring spatial data from fictional literature. Selected stories from the Sherlock Holmes canon shall thereby serve as an example. This first chapter provides an overview of the academic and technological developments that form the basis for this project.

1.1 Early literary geography

The origins of literary geography can be traced back to the verge of the 20th century. In 1899, William Lyon Phelps published a *Literary Map of England*, a cardboard map with markings of literary places for his students at Yale University. (Mitchell, 2017; Phelps, 1899) William Sharp’s book *Literary Geography* followed in 1904. (Sharp, 1904) The title was not only adopted one year later by Virginia Woolf for one of her review essays (Woolf and McNeillie, 1989), but also anticipated the name of the research subject which was to emerge during the following decades (Piatti and Hurni, 2011). Sharp's book contained a series of maps showing the geographical areas covered by various authors' novels (Noble and Dhussa, 1990), for example “Dickens-Land”, “Scott-Land”, “The Brontë Country” and “The Literary Geography of the English Lakes”. (Alexander, 2015) Sharp’s illustrations can be viewed as early examples of maps of literature and allow the map user to “see immediately where real and fictional space overlap”. (Piatti et al., 2009)
In 1907, Siegfried Robert Nagel released his *Deutscher Literaturatlas* (*German Literary Atlas*) where he illustrated places of work and residence of various authors. (Nagel, 1907) In his atlas, Nagel intended to show how certain geographical features of German lands conditioned the lives, productivity and mobility of writers and which regions and cities played the role of intellectual centres in different epochs. (Juvan, 2015)

Similarly, John George Bartholomew’s *Literary & Historical Atlas of Europe* of 1910 contained a chapter on “Maps Illustrating Districts Connected with famous Books and their Authors”. (Bartholomew, 1910)

This list of early works on literary geography shows two main directives: on the one hand the localisation of authors and their living environment, on the other hand the zones of actions where these authors’ works are set. (Piatti and Hurni, 2009)

A slightly different approach can be found in the works of climatologist Hugh Robert Mill who, in chapter IV of his 1910 book *Guide to Geographical Books and Appliances* uses “Geographical Novels” as sources for making his students understand geographic features. (Mill, 1910) While not making direct use of maps, Mill’s work stands at the beginning of a line of literary geography research termed “objective-use” in which “the emphasis was upon how well the literary work described areas”. (Noble and Dhussa, 1990)

In the United States, basic foundations for a literary geography can be found in the works of John K. Wright, a historical geographer. (Wright, 1926, 1924) In a note in *The Geographical Review* of 1924, Wright emphasized the opportunities that imaginary literature posed for geographical research. However, this incentive was received rather reluctantly until some decades later. (Brosseau, 2009, 1994; Noble and Dhussa, 1990)

In the German-speaking countries, Josef Nadler’s *Literaturgeschichte der Deutschen Stämme und Landschaften* (Nadler, 1912) and *Literaturgeschichte des Deutschen Volkes* (Nadler, 1938) are among the most influential works on literary geography, albeit in a negative sense. Nadler sought to explain “the historical development of regional differences in German literature by the interplay of the tribal origins of writers and their habitat (Lebensraum)”, conveying “Blut-und-Boden underpinnings”. (Juvan, 2015) Due to Nadler’s proximity to ideologies of the Third Reich, any academic discussion on the spatialities of
literature has been discredited in Europe after World War II, leaving behind what Barbara Piatti and Lorenz Hurni call the “Nadler-Trauma”. (Piatti and Hurni, 2009)

A collection of more than 230 examples published by the Library of Congress in 1999 shows, however, that Nadler’s infamous works did not affect the popularity of literary maps in the United States. A great number of maps in the collection originate from the 1940s and 1950s and represent the literary canon of the time. (Hopkins and Buscher, 1999)

1.2 From objectivism to subjectivism to the spatial turn

Up to and during the 1960s and early 1970s, regionalism arose as a major research topic in geography and with it new interest in literary geography. Coming from the tradition of historical geography, literary texts, especially realistic literature such as travel accounts, were considered as a complementary source for regional geography. Fictional literary texts were taken into account only very reluctantly for fear that they might be too subjective and contain too little reliable information. (Brosseau, 2009; Ridanpää, 2010) In two major works, Henry Clifford Darby (Darby, 1948) and Edmund William Gilbert (Gilbert, 1960) did direct their attention to imaginative literature by investigating the documentary value of the English regional novel. Darby and Gilbert followed the general assumption that, depending on the local knowledge of the author, there is indeed reliable information in literary texts. (Brosseau, 1994)

In the later 1970s finally emerged a named discipline literary geography. (Mitchell, 2017) In 1974, Christoph Salter initiated a special session termed Landscape in Literature at the Association of American Geographers annual meeting in Seattle and, together with William J. Lloyd, published the first extensive review of the new literary geography in 1977. (Noble and Dhussa, 1990; Salter and Lloyd, 1977) During the following years, the humanistic turn brought new concepts for literary geographical research. Instead of investigating the documentary nature of literary fiction and trying to find objective truths about landscapes, literary works now began to be appreciated as documents of personal experience, subjective meanings of space and place as well as symbolic interpretations of landscape. (Brosseau, 2009, 1994; Ridanpää, 2010)
1.2 From objectivism to subjectivism to the spatial turn

In their *Review of Literary Geography* of 1990, Allen G. Noble and Ramesh Dhusa identified six categories of literary geographical studies in the 1970s:

“(1) those which discuss methodological and theoretical aspects, (2) research which focuses upon the subjective use of literature, (3) studies that continue the earlier exploration of objective and descriptive information provided in literature, (4) the use of literary works in geography classroom instruction, (5) the analysis of landscape use by fiction writers, and (6) the portrayal of imaginary landscapes in literature.” (Noble and Dhussa, 1990)

During the following decades, two major movements in geographical research at the one hand, and in humanistic research at the other hand, gravely influenced further development in literary geography. The cultural turn of the 1980s and the socially critical approach to literary geography were both “a criticism of the scientific objectivism characteristic of human geography, and of the excessive subjectivism of the humanistic approach”. (Ridanpää, 2010) Analysis of literary works now focused on their role as products of the human mind embedded in their specific sociocultural context. (Alexander, 2015) Parallel to these developments in the geographic research area, the “spatial turn”, a term coined in 1989 by Edward Soja (Soja, 1989), or “topographical turn” showed itself in the humanities and brought with it new concepts of space.

“[It] developed an innovative concept of spatiality that dispenses with concepts grounded in Euclidean geometry, Cartesian subjectivism, and the Newtonian idea of a three-dimensional, absolute container of the material world; it rather draws on Leibniz’s relational concept of space, Spinoza’s monadology, Kant’s understanding of space as a category, and Einstein’s theory of the space-time continuum” (Juvan, 2015).

The spatial turn had directed researchers’ attention to the spatialities of human life worlds and encouraged research on the geographical settings and manifold spatial aspects of fictional literature. (Piatti, 2012; Piatti et al., 2009) With the spatial turn space gained importance over time as the main category of analysis in fictional literature. (Peraldo, 2016)
1.3 Literary cartography

The 1990s saw the emergence of literary cartography, a subdiscipline of literary geography which specifically engages with literary maps. According to Barbara Piatti and her team, the hierarchical relationship between literary geography and literary cartography can be described as follows: while literary geography is the overall topic, literary cartography is a tool or one possible approach to carry out research in this area by making use of a symbolic language. (Piatti et al., 2011; Piatti and Hurni, 2011)

Literary maps occur in two variants: on the one hand as maps in literature, on the other hand as maps of literature (Döring, 2013) or, in other words, as authorial maps or as post-authorial maps (Bushell, 2012). A literary map that is a map in literature is, using the words of Sally Bushell, “a graphic representation of spatial relations among places or objects (real or imagined) that is presented alongside the literary work at the time of first publication and is authorial or authorially approved” (Bushell, 2012), for example J.R.R. Tolkien’s maps of Middle-earth (Tolkien, 1937) or the maps in Jonathan Swift’s Gulliver’s Travels (Swift, 1726). As such, these maps are literary “peritexts” (Döring, 2013; Juvan, 2015) and can, in some cases, also be considered as integral part of the narrative (Peraldo, 2016).

Maps of literature, on the other hand, have been created by literary scholars (Döring, 2013) and “have been constructed to geovisualise the narrative locations of one or more literary texts and that are not themselves a part of the literary text or its peri- or para-texts” (Mitchell, 2017).

This distinction is similar to the “two principle categories” in literary geography given by David Cooper and Ian N. Gregory. Cooper and Gregory distinguish between “writerly mapping” and “readerly mapping” (Cooper and Gregory, 2011), later changed by Cooper to “authorial” and “reader-generated mapping” to avoid “post-structuralist connotations” (Cooper, 2012). Writerly or authorial mappings are “cartographic representations of geographical space which are embedded within literary texts”. (Cooper, 2012) In other words, writerly mappings are maps created by the author of a text. As such, engagement with writerly mappings is centred on the possible influence that maps exert on a writer’s
work and on the interplay or concurrence between cartographical and textual representation of space within an author’s work. (Cooper and Gregory, 2011) Moreover, writerly mappings should, according to Andrew Thacker, be read with attention to historical and cultural geography as well as to their material context within the published text. (Thacker, 2005) Readerly mapping is, on the other hand, concerned with the reader and how he or she engages with and switches between textual and cartographical representation of space, if and how the reader consults other maps while reading and, most importantly, how the reader creates his or her own maps while reading. (Cooper and Gregory, 2011) In other words, readerly mapping describes how the and by which means the reader of a text achieves an understanding of the geographical setting of the plot.

While the overview given above of the early phase of literary geography already contained examples of post-authorial maps (Phelps, 1899; Sharp, 1904), it was not before the 1990s that literary maps have been considered with systematic scientific endeavour. In 1989, Graham Huggan included a section termed First Principles for a Literary Cartography in his dissertation at the University of British Columbia. In this chapter, Huggan explains how literary maps are embedded within the cartographic discourse concerning spatial representation and the historical and political significance of maps and how maps function either as icons, motifs or metaphors in literary texts. (Huggan, 1994)

While earlier literary cartography is primarily concerned with maps in literature, Malcolm Bradbury’s Atlas of Literature of 1996 holds a collection of maps made by literary scholars. (Bradbury, 1996) The truly innovative approach was, however, delivered two years later by Franco Moretti. Even if it was the source of much critique, it influenced decades of literary cartographic research. (Bushell, 2012; Mitchell, 2017) In his Atlas of the European Novel of 1998 (Moretti, 1998), Moretti uses maps no longer as “simply ornamental, illustrative, supplementary or pedagogical” devices (Mitchell, 2017), but as tools for literary analysis and text interpretation (Piatti and Hurni, 2009). Moretti’s effort is to “dissect the text in an unusual way, bringing to light relations that would otherwise remain hidden” (Moretti, 1998). He explores both “space in literature” by creating maps that explore plot and narrative as well as “literature in space” by mapping “the spread of a certain genre or generic convention in relation to actual places in the world” (Bushell, 2012). (Juvan, 2015)
1.4 Critical literary cartography and de-cartographisation

1.4.1 Critique

As shown in the developments in literary geography since the 1960s, research on the spatial dimension of literature has been subject to criticism that was expressed in multiple waves and was directed both against the objectivism of the 1960s and early 1970s as well as the subjectivism of the later 1970s and 1980s. Similarly, literary cartography was affected by contemporary discussions on the reliability, hidden meanings and justification of maps summarized under the term “critical cartography”. (Crampton and Krygier, 2005)

However influential Moretti’s work may have been, it was also targeted by critiques. Even to this day, literary cartography is viewed critically by some literary scholars. For example, scholars criticised that Moretti’s maps show only what can obviously be mapped but fail to convey the “genius loci”, meaning that the impression of a space in a three-dimensional sense and complex social spaces are not respected. (Cooper and Gregory, 2011) Similarly, Jörg Döring notes that Moretti’s ambitions of mapping the work of Jane Austen reduce spatiality to a single marker in the map that does not convey any meaning or information on the role of the location. (Döring, 2013) With its reductive tendencies, Moretti’s work was seen as a return to “unfashionable positivist methods” (Bushell, 2012) and as disregarding critical cartography’s doubts in the objectivity of maps (Thacker, 2005).

Furthermore, while Moretti’s maps claim to emphasise “complex and fluid spatialities and mobilities”, they are themselves fixed, stable and rigid and leave the reader in an exclusively passive role. (Cooper and Gregory, 2011) Very similar to this argument is Döring’s criticism of Barbara Piatti’s Literary Atlas of Europe (see also below in section 1.8.4). Piatti’s project brings 500 years of literature to a map in a complex cartographic system, but while the dataset is diachronic, the background map is not. Instead, it is a very detailed map of 2004. (Döring, 2013) Last, Moretti presents his maps separated from the textual representation of literary space and fails to pay attention to their interconnections. (Cooper and Gregory, 2011)

More general criticism of the work of literary cartographers has been issued by Döring who states that mapping the origins of authors and the circulation of their work can only lead to
very vague assumptions on how space influenced the emergence of literary texts and conveys strong environmentally deterministic values. (Döring, 2013)

In total, Barbara Piatti identified six sources of critique to literary cartography:

a) references between literary and physical space are per se invalid, the literary space is independent from the physical;

b) literary maps are reductionist and omit complex contents;

c) literary maps depict geographies that are not suggested by the text;

d) literary maps are redundant and provide no additional information to the text;

e) literary maps are the result of subjective readings and therefore not intersubjective or even objective;

f) the choice of texts for literary maps is arbitrary and primary literature is never respected as a complete corpus. (Piatti et al., 2009)

1.4.2 Ontologies

A very basic question arising in discussions related to the spatial turn and critical literary cartography is how and if at all fiction refers to the ‘real’ world and is, therefore, mappable. According to Barbara Piatti and her team, fiction does include references to the physical world. (Piatti, 2012) The physical “geo-space”, characterised by its geographical, geological and geomorphological attributes, is thereby overlaid by the literary “meta-space”. The literary “meta-space” consists of projections, imaginations, knowledge and phantasies and references the underlying “geo-space” by means of topography and toponyms. (Piatti and Hurni, 2009)

Fictional space is, according to Anne-Kathrin Reuschel and Lorenz Hurni, characterised by the following five aspects:

a) it is fragmentary, completed and developed through the reader’s imagination;

b) it has uncertain, vague boundaries that are neither physical nor administrative;

c) localisation is at times difficult or not possible at all;

d) time may have altered the topography if it refers to a historical place;
1.4 Critical literary cartography and de-cartographisation

e) transformation and remodelling processes may be at work. (Reuschel and Hurni, 2011)

In other words, this means that descriptions of fictional settings may highlight certain details and omit others so that the image of space constructed in the reader's mind may differ from individual to individual. Furthermore, references to places in a fictional text may make it difficult to define where it begins and where it ends. In some cases, it may not be possible to connect the geographical reference to any ‘real’ place in the world and the place will remain ‘unmappable’. (Piatti and Hurni, 2011) Transformation processes that are in place when ‘real’ spaces are made into ‘fictional’ spaces include renaming of existing locations, creating a synthetic place by fusing to different places, relocating places and referring to places without a name. (Piatti, 2012)

Piatti’s ontology is roughly binary and differentiates between the “fictional”, represented by the “meta-space” and the “physical”, represented by the “geo-space”. Piatti further localises the fictional “meta-space” in Edward Soja’s “secondspace”. (Piatti and Hurni, 2009; Reuschel et al., 2013) However, Soja’s (Bedford, 1998; Soja, 1996) theory of spaces is indeed tripartite: While the “firstspace” is the “real” space, the “concrete materiality of spatial forms”, the “secondspace” is the “imagined” space and contains “mental or cognitive forms of spatiality”. (Bedford, 1998) The “thirdspace” is, finally, where everything comes together, “subjectivity and objectivity, the abstract and the concrete, the real and the imagined, the knowable and the unimaginable”; it is “a mode of thinking about space that draws upon material and mental spaces of the traditional dualism but extends well beyond them in scope, substance and meaning”. (Soja, 1996)

Soja’s theory of spaces is reminiscent of Karl Popper’s discussion of three worlds: while “World 1” contains all material elements of the physical world, “World 2” is made up of the human consciousness and all elements of the psychical world. “World 3” emerges from “World 2” and consists of theories, ideas and beliefs about “World 1”. (Popper, 1994) Juha Ridanpää compares Popper’s “World 3” with the main perspective of the cultural turn in geography – instead of viewing space as purely physical or purely subjective, the new concept saw space as a social construct. (Ridanpää, 2010) With regard to literary geography, a purely physical view of literary space would reduce it to “a simple reflection of the real
1.4 Critical literary cartography and de-cartographisation

space”. In a purely subjective view of literary space, on the other hand, the space of reference is rejected and neglected entirely, leading to, for example, a reading of Balzac’s Paris without considering the historical Paris of the 1930s. (Joliveau, 2009)

The notion of if and how fictional literature references physical, ‘real’ spaces determines whether fictional spaces are considered ‘mappable’ at all – a view that is refuted by several literary scholars including Robert Stockhammer. Following Stockhammer’s line of thought, texts do never contain affirmations or convey a logical value in the sense that an author ever claims that what he or she says is ‘true’ in the sense of ‘real’. Therefore, fiction is, according to Stockhammer, “by definition unmappable” and only produces “the fictional effect” of “internal” and “referential mappability”. (Stockhammer, 2013)

“Internal mappability” means that the description of space in the literary text is consistent and complies with the rules of Euclidian geometry, “so that the world created by this text can be unequivocally depicted in a map to be attached to the book”. (Stockhammer, 2013)

With “referential mappability” Stockhammer means that geographical features mentioned in a literary text correspond to those found on a world or cadastral map; it is comparable to geotagging pictures in Facebook. (Stockhammer, 2013) Stockhammer deliberately avoids the term ‘reality’ or ‘real’ space because the notion of ‘reality’ can change. However, it seems that, contrary to critical cartographers, in his argumentation the map exclusively represents things that are ‘real’ and is therefore incompatible with fiction.

Instead of the antithetic notion of ‘map’ vs. ‘text’, in Bertrand Westphal’s geocritical approach, the place is at the centre of the analysis. (Westphal and Tally, 2011) According to the geocritical view, both texts and maps, together with “photos, paintings, maps, travelogues, urban plans, films, and other non-fictional documentation” (Juvan, 2015) are different forms of representing “the multi-sensory nature of space, made up of colours, textures, sounds, smells and material objects” (Joliveau, 2009). Similarly, for Christina Ljungberg cartography and writing are representations that use “different semiotic systems”: “on a map, the three-dimensional world is viewed orthogonally from above, whereas in a text, it is transformed into a one-dimensional linear one”. Narration is, for Ljungberg, also a form of mapping because it is a process that “populate[s] space, creating
1.4 Critical literary cartography and de-cartographisation

new ‘realities’ and transforming them into a ‘real’ world” and stories are therefore “narrative maps”. (Ljungberg, 2003)

This excursion into literary theory illustrates the different notions of ‘mapping’ between cartography and literary scholars. Sébastien Caquard notes that “While in cartography mapping refers to a specific set of conventions and techniques, in the arts and humanities it refers to a broader range of concepts and practices” (Caquard, 2011) and Cooper and Gregory remark that “‘mapping’ in literary studies, has frequently become synonymous with a way of reading rather than cartography” (Cooper and Gregory, 2011). In the course of the spatial turn, ‘mapping’ has in places become a metaphor, disintegrated from its cartographic roots and leading to a “de-cartographisation” of literary cartography. (Cooper, 2012; Rossetto, 2014) Cooper emphasizes that “genuinely interdisciplinary geohumanities research needs to be predicated upon a self-reflexive engagement with geographical thinking and practices rather than an uncritically imprecise reliance on spatial vocabularies and discourses” and promotes a sophisticated critical literary cartography that is informed by the critical works of Brian Harley (Harley, 1989), Denis Wood (Wood and Fels, 1992), and others (Crampton and Krygier, 2005). (Cooper, 2012)

This lack of communication between the disciplines and the reluctance of scholars to engage intensely with principles and vocabularies of ‘the other’ discipline has been a critical issue in literary geography since its emergence. Neal Alexander remarks how humanist geographers of the 1970s and 1980s “ignore[d] contemporaneous developments in literary theory”, while scholars of cultural studies made “no reference at all to relevant studies of rural, urban, and regional geography”. (Alexander, 2015) In this respect, literary geography or literary cartography mirrors the struggles between “GISers” and “non-GISers” outlined by Nadine Schuurman in the influential paper Trouble in the Heartland. (Schuurman, 2000) Within the past 20 years, however, Alexander notes that research in literary geography has developed more and more towards true interdisciplinarity – “geographers have become increasingly sophisticated readers of literary texts” and literary critics are aware of how important it is “to read the work of actual geographers”. (Alexander, 2015) While Sheila Hones acknowledges that “literary critics do work with geographical theory”, they seem reluctant to engage with “substantive work on literary texts produced by geographers”. (Hones, 2008)
Since 2015, Sheila Hones is, together with Juha Ridanpää and Angharad Saunders, one of the editors of *Literary Geographies*, an open-access e-journal that provides a platform for interdisciplinary, collaborative research in the field of literary geography. The editorial board of *Literary Geographies* furthermore includes several scholars whose works have been cited in this thesis: Neal Alexander, Marc Brosseau, Sébastien Caquard, David Cooper, Ramesh C. Dhussa, Peta Mitchell and Barbara Piatti. (“Literary Geographies,” n.d.)

### 1.5 Digital literary cartography and re-cartographisation

Two major factors have managed to overcome the “academic mapphobia” (Rossetto, 2014) and initiate a “re-cartographisation” of literary geography (Cooper, 2012) On the one hand, a new critical literary cartography led to rethinking the ontological status of the map: Instead of viewing maps as finished representations, mapping is considered a process that is influenced by social and technical practices. (Kitchin and Dodge, 2007) This new “post-representational cartography” can be considered as an epistemological shift in cartography. (Rossetto, 2014)

On the other hand, new digital mapping techniques coming with the proliferation of Geographic Information Systems (GIS) and their use in quantitative and also qualitative research in the *spatial humanities* (Bodenhamer, 2010) may help to overcome the limitations of ‘traditional’ cartography which were the target of earlier critique. (Cooper, 2012) By going beyond a purely quantitative approach in mapping, digital literary cartography is also going beyond the borders laid out by Moretti’s use of maps as analytical tools for literary studies. (Rossetto, 2014)

Closely related to the concept of *spatial humanities* is that of the *digital humanities* which is sometimes even considered as an equivalent. (Crang, 2015) The digital humanities leverage computer technologies to store, manage, visualise and analyse large amounts of data, mostly texts. (Jänicke et al., 2015) Charles Travis identifies three waves of the digital humanities with the third being currently in progress:

“In the first wave the digitization of collections and movement of disciplinary activity surged online; the second wave witnessed quantification exercises and investigations considering the parsing, analysis and visualization potential of
Digital tools. With the third wave, crosspollination between the arts, humanities, computing science, social media and Big-Data will facilitate the birth of truly hybrid digital methodologies and technologies borne by the fruit of research, experimentation, innovation, and development over the past few decades. “ (Travis, 2015)

Two important analysis methods for texts in digital humanities research are close reading and distant reading and each has their specific visualisation techniques: Close reading means engaging with the text with respect to details such as individuals, events and ideas, wording, structure and style. Visualisation techniques centre on annotating of highlighting the text itself and include colouring, font size as well as glyphs (symbols) and connections (connecting lines). (Jänicke et al., 2017, 2015)

Distant reading is a concept that goes back to Franco Moretti (Moretti, 2005) and aims to “generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s)” (Jänicke et al., 2017, 2015). The three most prominent visualisation techniques going back to Moretti’s work are graphs, used by Moretti for analysing genre changes in historical novels, maps, used by Moretti for illustrating geographical aspects of novels, and trees, used by Moretti for classifying different types of detective stories. (Moretti, 2005) Other visualisation techniques for distant reading approaches include structure overviews to illustrate the hierarchy and structural elements of an individual text or an entire corpus; heat maps or block matrices to illustrate for example textual patterns, relationships between texts of a corpus or interpersonal relationships between characters; tag clouds for highlighting the frequency of recurring words or topics; maps either in the form of thematic maps, circle glyphs or symbolised glyphs for visualising spatial information in the text. (Jänicke et al., 2017, 2015)

1.6 New approaches in cartography

1.6.1 Geovisualisation

As mentioned above, with critical literary cartography the ontological status of the map was moved beyond the purely representational. (Kitchin and Dodge, 2007; Rossetto, 2014) With the rise of GIS, the representational role of the map was also modified from a technological
perspective. In a GIS, not only powerful databases and analytical tools are added to the cartographic representation of geographic phenomena, but also the means for representation are extended by various visualisation techniques that go beyond mapping.

Additionally, during the 1980s and 1990s, a shift of paradigms occurred concerning the analytical role of the map. Maps were no longer used as the end product of analysis and as a medium to communicate results, but as analytical tools and a means to produce results. (Çöltekin et al., 2018) Similarly, in literary cartography maps are, according to Barbara Piatti, to be considered only as intermediary products, as a means for the literary scholar to derive results. (Piatti and Hurni, 2009) Maps are never meant to replace scholarly text analysis, but rather as a supplement. (Piatti et al., 2009)

**geovisualisation** has developed as a multidisciplinary approach encompassing principles of cartography, image analysis, scientific and information visualisation, exploratory data analysis and GIScience for visual exploration, analysis, synthesis and presentation of geospatial data. (Dykes et al., 2007; MacEachren et al., 2004; MacEachren and Kraak, 2001) Geovisualisation offers multiple advantages: Innovative visualisation methods added to traditional maps allow for representing and exploring the same dataset in multiple ways and for comparing them by dynamically linked views. (Kraak, 2003) As an intuitive exploration method, geovisualisation requires little complex understanding, is specifically suitable for large, multivariate, spatio-temporal datasets, and is also applicable to non-homogeneous and noisy data. (MacEachren and Kraak, 2001)

The basic working principle of geovisualisation is illustrated in Alan MacEachren’s “cartography cube”. The sides of the cube represent three dimensions of a geovisualisation product: the task at hand, the level of user interaction, and the kind of users acting as recipients of the geovisualisation product. (MacEachren et al., 2004) Users may vary from the public or non-experts to specialists like researchers. With a varying degree of interaction, the geovisualisation application can be used to communicate information or to enable the user to explore the data and discover patterns. The central axis of the cube suggests that specialists tend to benefit from a higher degree of interaction to gain new knowledge by exploration while communication of synthesized results to the public requires fewer possibilities for interaction. (Çöltekin et al., 2018)
Two major factors that distinguish geovisualisation approaches from ‘traditional’ mapping are *interactivity* and *animation*. (Krisp, 2006)

**Interactivity**

According to Alan MacEachren, “geovisualisation is not a passive process of either seeing or reading maps. It is an active process in which an individual engages in sorting, highlighting, filtering, and otherwise transforming data in a search for patterns and relationships.” (MacEachren and Kraak, 2001) Methods for interacting with the application include in a general sense clicking buttons, panning, zooming and turning (Krisp, 2006), but also more specialised methods like focusing and brushing. Focusing in this respect means narrowing down the dataset for more thorough exploration of a subset; brushing refers to the visual highlighting of selected data throughout different forms of visualisation. (MacEachren and Kraak, 2001) Brushing requires multiple dynamically linked views and a “brush” - either a bounding-box, a lasso or single clicks – with which the user can select data. The actions set by the user by brushing are subsequently seen simultaneously on each data plot, usually by visual highlighting. (Roberts and Wright, 2006) Highlighting operations include

- a) highlighting: changing the visual appearance of one or more elements in all views;
- b) shadow highlighting: like highlighting plus additionally removing non-highlighted elements from the view to isolate the selected data;
- c) deleting: removing the selected elements;
- d) labelling: showing additional information in a textual label. (Becker and Cleveland, 1987)

Visual highlighting through brushing is usually achieved by a change in colour, but other possibilities exist that exploit more visual variables as proposed by Jacques Bertin (Bertin, 1983): Visual highlighting can thus also be done by chancing hue, value, saturation, focus, transparency, resolution, shape, arrangement, texture, orientation, size or location of the selected elements. (Robinson, 2011)
1.6 New approaches in cartography

**Animation and the time dimension**

Map animations represent one, but not the only possibility for mapping the time dimension. Conversely, animations can also be used for non-temporal data. While time can also be represented statically with small multiples where the temporal succession is represented by the spatial succession – usually from left to right – (Kraak, 2003), animations can also be used for non-temporal, sequential or uncertain data (Fabrikant, 2005). Dynamic displays for uncertain data use mechanisms like symbols fading in and out of the map at random locations. (Fisher, 1993; Reuschel and Hurni, 2011) Non-temporal animations include fly-bys and animated sequences from high to low values. (Slocum et al., 2001)

Interactivity and related usability issues play a major role when working with animated map displays. Providing the user with a means to control the sequence of pacing of an animation helps to prevent overload of the user’s short-term memory as well as difficulties in detecting differences between two frames, so-called ‘change-blindness’. Therefore, interactivity is, according to Sara Irina Fabrikant, not only useful, but “necessary to alleviate perceptual problems of non-interactive animations”. (Fabrikant, 2005)

In accordance with Jacques Bertin’s visual variables for maps, new *dynamic variables* were defined for animated maps. (Fabrikant, 2005) David DiBiase et al. first proposed three dynamic variables: *duration* (how long a scene is displayed), *rate of change* (how much change is visible between two sequential scenes) and *order* (in which sequence the scenes are displayed). (DiBiase et al., 1992) Later, MacEachran added three more dynamic variables to this list: *display date* (when a dynamic event becomes visible), *frequency* (the frame rate; how fast a scene follows another) and *synchronization* (how two concurrent events are visualised on the same dynamic display). (MacEachren, 2004)

1.6.2 Analysis and visualisation methods in literary cartography

New approaches to literary cartography aim to synthesize literary as well as cartographic and other visualisation methods and are, for example, concerned with different levels of scale. The distinction between *micro-mapping* and *macro-mapping* given by Cooper and Gregory (Cooper and Gregory, 2011) is reminiscent to the differentiation of *close reading* and *distant reading* made by Franco Moretti in *Graphs, Maps and Trees* (Moretti, 2005). There
is, naturally, a notable tension between generalisation and detail. Thus, it seems useful that macroanalysis, that is visualising an overview of the data, should lead the reader to the detail by “drilling down” on noteworthy patterns. (Jänicke et al., 2015; Taylor et al., 2018)

**Macro-mapping** pays respect to the wider scale: mapping works of more authors or several works of the same author alongside each other can be used to provide the big literary picture of a landscape or region and to highlight places that are frequented more often than others. (Cooper and Gregory, 2011) It is worth noting that this approach to mapping literature is not per se new or tied exclusively to digital mapping. Maps of this kind have been around since interest in literary mapping arose in the early 20th century and can already be found in the pioneer work by William Sharp (Sharp, 1904).

**Micro-mapping** requires thorough engagement with the primary text to grasp how the writer’s understanding and representation of a landscape is reflected in localised geographies. Visualising these aspects requires the creative and sometimes experimental use of cartographic techniques, for example mood maps. (Cooper and Gregory, 2011)

**Deep mapping** is an approach where multiple layers of geospatial data, including texts, images or even data from biological research, are accumulated to achieve a “transdisciplinary understanding of a landscape”. (Taylor et al., 2018) Joliveau suggests a similar form of mapping where the imaginary and the real space, that is, the text and the map, are not only connected with each other but result in a hybrid form of representation that allows for movement in both directions – from the map to the text and from the text to the map. Additionally, media such as static images or animated images may accompany the text and the map. (Joliveau, 2009) These concepts of mapping are closely related to Mark Graham’s theoretical framework of *palimpsests of place*. Following Graham, all places are palimpsests in the sense that they consist of a multitude of physical, cultural, historical, personal or political layers. With the geoweb (see below section 1.7), a virtual palimpsest consisting of photographs, descriptions, blogs, stories etc. is created. (Graham, 2010)

### 1.6.3 Dealing with uncertainty

A major issue with mapping in literary cartography and other digital humanities is uncertainty in the geographic information. (Caquard, 2011) Uncertainty, vagueness of
localisation and borders, as well as temporal vagueness are inherent characteristics of fictional spaces (see above section 1.4.2). In other words, “[t]exts may contain placenames of varying granularity (e.g., country, region, city) or type (e.g., points for cities, polygons for areas, polylines for rivers) or even fictional placenames, which are hard to represent.” (Jänicke et al., 2015)

For the *Literary Atlas of Europe*, Anne-Kathrin Reuschel and Lorenz Hurni developed a systematic approach to uncertainty in literary fiction (see below section 1.8.4). Reuschel first defines five concepts of uncertainty:

a) ambiguity in linguistic terms;

b) averaging due to merges in respect to scale or time;

c) continuousness, i.e. lack of a distinct delineation or borders;

c) subjectivity in deferring localisation from implicit rather than implicit references;

d) vagueness, i.e. lack of a clear distinction of geometry data. (Reuschel et al., 2013)

Generally spoken, vagueness arises if an object or a class of objects lacks a clear definition. (Fisher et al., 2006) In geographical terms, vagueness means that there is no definite consense on ‘where’ a named location is or where its borders are. Ambiguity, on the other hand, means that an object can be element of more than one class, depending on the classification system used. (Fisher et al., 2006) In respect to geography and fictional literature, ambiguity applies to place names that can be connected with more than one ‘real’ location. ‘Oxford Street’, for example, may either refer to a street in London, or to a street in Norfolk.

Reuschel and Hurni identify five sources of uncertainty in fictional literary texts:

a) uncertainty through artistic freedom;

b) uncertainty in linguistic concepts;

c) uncertainty in geographical concepts;

d) uncertainty of the interpreters/reader;
1.6 New approaches in cartography

e) uncertainty in visualisation. (Reuschel and Hurni, 2011)

Uncertainty in geospatial data is, consequently, a challenge for cartographic visualisation. Approaches to visualising uncertainty include fuzzy shapes or grey scales (MacEachren, 2001) as well as animated map symbols (Fisher, 1993) (see also above section 1.6.1 on animation). A sophisticated model for visualising uncertainty has been developed for the Literary Atlas of Europe (Reuschel et al., 2013) that will be discussed in more detail below (see section 1.8.4).

1.7 Web mapping, the geoweb and literary cartography

1.7.1 Technological advances

Literary cartography did not only develop with the proliferation of GIS and digital mapping in general, but was and is still also influenced by the possibilities provided by web mapping technologies. In the early stage, web mapping was mainly in the hands of specialists. (Graham, 2010) One of the first web mapping products was the Xerox PARC Map Viewer that was introduced in 1993 and relied primarily on the use of Hypertext Markup Language (HTML). A request for a map including rendering options was sent to the server via Hypertext Transfer Protocol (HTTP) where the information in the Uniform Resource Locator (URL) was processed by a Common Gateway Interface (CGI). The resulting HTML document was then sent back to the client and rendered in the web browser. (Haklay et al., 2008; Putz, 1994) Due to technical knowledge and financial expenses required for producing web maps, the user’s role was mainly limited to that of the consumer with little possibilities for customisation. (Graham, 2010; Haklay et al., 2008)

The introduction of extensible and portable technologies such as Extensible Markup Language (XML), Cascading Style Sheets (CSS) and Asynchronous JavaScript and XML (AJAX) as well as related formats such as Geography Markup Language (GML) and Keyhole Markup Language (KML) simplified the process of web mapping and played an important role for the development of the so-called Web 2.0. (Graham, 2010; Haklay et al., 2008; Turner, 2006) Sophisticated system architectures going beyond the simple client – server communication in the Xerox PARC Map Viewer facilitated the development of powerful Web GIS applications. Modern architectures exhibit a three-tier structure consisting of a data
Web mapping, the geoweb and literary cartography

tier (GIS database), a logical tier (a GIS server and a web server) and the presentation tier (client). Currently, web GIS architectures are developing towards an n-tier structure with multiple resources combined to mashups. (Fu and Sun, 2011) Web services standards such as Web Feature Service (WFS), Web Map Service (WMS) and Web Processing Service (WPS) laid out by the Open Geospatial Consortium (OGC) further facilitated publishing and retrieving of geographic information. (Haklay et al., 2008)

A WMS allows for requesting georeferenced map images from one or more GIS databases. Two requests are mandatory for a WMS to fulfil the OGC standard: GetCapabilities and GetMap. GetCapabilities returns an XML document containing the service specifications. These comprise general service metadata, a list of available layers and their metadata, as well as format specifiers. GetMap is used for requesting a map from the service. The request includes a definition of layers, styles, extent, and format and returns a georeferenced map image. (“Web Map Service,” 2018, “Web Map Service | OGC,” n.d.)

A WFS provides an interface for querying and retrieving map features from a data source. Transactional Web Feature Services (WFS-T) additionally allow for creating, deleting and updating map features. Mandatory operations for a WFS include GetCapabilities, DescribeFeatureType and GetFeature. GetCapabilities returns XML encoded service metadata including a list of available feature types and DescribeFeatureType the feature definition for a feature type in XML. GetFeature finally retrieves the map features according to the chosen feature type, Coordinate Reference System (CRS), extent and output format. (“Web Feature Service,” 2017, “Web Feature Service | OGC,” n.d.)

The WPS standard defines requests and responses for calling geospatial processing services. Mandatory operations include GetCapabilities, DescribeProcess and Execute. As above, GetCapabilities returns the service metadata. With DescribeProcess, the XML encoded description of a specific process, the ProcessOfferings, can be requested by providing a process identifier. An Execute request is then used to send the data that needs to be processed as input to the service – either by including the data directly in the request or by referencing external sources and to retrieve the result of the process in the requested format. (“Web Processing Service,” 2018, “Web Processing Service | OGC,” n.d.)
Another major driving force for the emergence of the closely related concept Neogeography was the publication of Application Programming Interfaces (APIs), especially that of Google Maps launched in 2005, that facilitated the creation of web mapping application and made it more accessible to non-expert users. (Haklay et al., 2008; Mitchell, 2017; Warf and Sui, 2010)

Other APIs for geospatial web applications include OpenLayers (“OpenLayers - Welcome,” n.d.) and Leaflet (“Leaflet — an open-source JavaScript library for interactive maps,” 2017) which are both JavaScript libraries. Additionally, these APIs represent a special case not only of web GIS technology but of geospatial technology in general. Both products are FOSS, which is *Free Open Source Software* and encompasses software that is both *Free Software* and *Open-source Software*. While there are ample similarities between the two licenses, the main difference between the *Free* and *Open-source* approach is rather philosophical. (“Free and open-source software,” 2018) The Free Software movement was initiated by Richard Stallman when he launched the GNU project (“GNU’s not Unix”) in 1983. (“Free software,” 2018) Today, it is represented by the Free Software Foundation (FSF) and is understood as a social movement (meaning ‘free’ as in “free speech”, not as in “free beer”) providing users with the four essential freedoms:

- freedom 0: freedom to run the program as you wish, for any purpose;
- freedom 1: freedom to study the program and change it as you wish, the prerequisite therefore is open source code;
- freedom 2: freedom to redistribute copies;
- freedom 3: freedom to distribute copies of your modified version and thus benefit to the community. (Free Software Foundation, 2018)

The aims of the Open Source movement represented by the Open Source Initiative (OSI) are of more of practical nature and focus on development of better software. The OSI’s principles which are based on the Debian Free Software Guidelines include free redistribution without royalty fee, mandatory inclusion of source code in the program, permission to redistribute modified versions of the software and prohibit discrimination against persons, groups or fields of endeavour. (Open Source Initiative, 2007) For geospatial applications of FOSS, the
term FOSS4G – Free Open Source Software for Geospatial – has been coined in 2004 and is also the name of an annual meeting which is rooted in GRASS (Geographic Resources Analysis Support System) and MapServer communities. (“FOSS4G - OSGeo,” 2018) FOSS4G is strongly promoted by the Open Source Geospatial Foundation (OSGeo), a non-profit organisation based on the four principles of Open Data, Open Education, Open Science and Open Standards. (“About OSGeo - OSGeo,” 2018) The benefits of FOSS4G platforms lies in the fact that they produce highly interoperable, sophisticated software and that access to code and algorithms guarantees reproducibility of results and facilitates quality assessment of methods. (Brovelli et al., 2012) FOSS4G does not only comprise desktop GIS software like QGIS, spatial databases like PostGIS, but also Web GIS solutions like the above mentioned JavaScript APIs or mapping servers like MapServer or GeoServer (Bandyophadyay et al., 2012; Singh Bhatia et al., 2018; Singh and Gambhir, 2014). Open Source Web GIS architectures have proven to provide a strong basis for various projects in application domains as diverse as disaster management (Kubota et al., 2014), public health (Tiwari and Jain, 2013), participatory sensing (Nakayama et al., 2017) or agriculture (De Filippis et al., 2013; Golhani et al., 2015). In this project, it will be used for literary geography. Details on the architecture will follow in section 4, a description of the project scope in section 2.

1.7.2 Fictional literature in the geoweb

As a consequence of the technical developments presented in the previous section, the new geoweb is characterised by a high degree of user participation with users creating, sharing and consuming geographic content on the web. (Graham, 2010; Turner, 2006; Warf and Sui, 2010) With the widespread use of social networks and mapping platforms such as Google Maps, OpenStreetMap and the like, maps and map products can be created without expert knowledge. Consequently, the user’s role has shifted from the pure consumer to a merged role of map user and map producer, the so-called “prosumer”. (Hoffmann, 2013)

For literary cartography, the emergence of neogeography meant that non-expert users can tell their own stories through geotagging places and pictures (Caquard, 2013; Joliveau, 2009) and also contribute to mapping fiction (Joliveau, 2009) The Atlas of Fiction for example is a crowd-sourced project for mapping places in literary fiction, including works of Agatha Christie, Arthur Conan Doyle, Charles Dickens, Victor Hugo and many others.
Associating place names with coordinates results in a simple web map with pin point markers. (“The Atlas of Fiction. real places imagined by great writers,” n.d.) Instead of manually associating mentions of place names with ‘real-world’ coordinates, automated geoparsing has facilitated the process of mapping literature. (Caquard, 2013; Joliveau, 2009) According to a blogpost dating back to 2007, Google also saw the potential of mapping literature when they announced that, beginning with a few selected works, the Google Books page would be accompanied by an interactive map created by automatically extracting placenames from the text. (Joliveau, 2009; Petrou, 2007) The map would be available under the section “About the book”, but currently none such option can be found and it is likely that the project was abandoned.

While these two approaches to literary mapping on the geoweb are of rather unsystematic nature without a sophisticated methodology or data management, neogeography did also benefit experts who leverage the new technical possibilities for analysis. (Mitchell, 2017) A number of literary web mapping projects are therefore presented in the next section.

### 1.8 Literary mapping projects

#### 1.8.1 Mapping the Lakes

**Project outline**

*Mapping the Lakes* is a research project initiated by the Wordsworth Centre for the Study of Poetry at the Department of English & Creative Writing at Lancaster University and was funded by the British Academy. The main aim of the project is “to test whether GIS technology has the potential to open up new spatial thinking about the geo-specific literature of place and space”. (Cooper et al., n.d.)

**Methods**

Two literary texts have been analysed during the project which are set in the English Lake District, a terrain that is not only geographically clearly delineated from the rest of England but also an area that has a rich history of literary reflections and travellers seeking the picturesque and the quiet. (Cooper and Gregory, 2011) The two textual accounts of journeys through the Lake District are Thomas Gray’s tour of the region in autumn 1769 and Samuel
Taylor Coleridge’s ‘circumcision’, the documentation of a 9-day walking expedition of the area, in August 1802. Both texts have been digitised and ‘tagged’ by identifying place names. The corresponding coordinates have been derived manually from the Ordnance Survey’s 1:50000 gazetteer and stored in an Microsoft Access database. Mapping was done in ArcGIS and KML exports have been generated for use in Google Earth. (Cooper et al., n.d.; Cooper and Gregory, 2011)

**Results**

During the project, four kinds of maps have been produced:

**a) base maps** (Figure 1): Static map representations of both authors’ tours are made up of point symbols corresponding to whether or not the site was actually visited that are connected by straight lines. (Cooper, 2012; Cooper and Gregory, 2011) The lines thereby represent the shortest route between to waypoints, but not necessarily the actual movement of the author. (Cooper et al., n.d.)

![Figure 1: Static map of Samuel Taylor Coleridge’s tour through the Lake District, 1802. (Cooper et al., n.d.)](image-url)
b) analytical maps (Figure 2 and Figure 3): Density smoothing was used to produce two kinds of analytical maps where in both cases colour shading indicates the frequency of placenames. First, the frequency of places actually visited by the author was compared to the density of mentioned, but not visited places. Darker areas delineate each author’s “spatial and imaginative ‘hot-spot’”. (Cooper et al., n.d.)

Figure 2: Smooth density surface maps of places visited by Samuel Taylor Coleridge. (Cooper et al., n.d.)

Figure 3: Smooth density surface map of places not visited personally by Samuel Taylor Coleridge. (Cooper et al., n.d.)

Apart from these maps were each author’s work was mapped separately, comparative maps where both authors’ maps are represented alongside each other have been produced (Figure 4). Density smoothing used with a colour scheme from red (only Coleridge) to green (only Gray) was intended to illustrate spatial intertextuality. Additionally, the mean centres and standard deviational ellipses of each author’s tour was represented in a map. (Cooper et al., n.d.; Cooper and Gregory, 2011)
1.8 Literary mapping projects

c) exploratory maps (Figure 5): These maps approach subjective, qualitative cartographies and represent the authors’ “emotional responses to named locations” in the form of mood maps. (Cooper, 2012; Cooper and Gregory, 2011) To that end, certain adjectives in the text were used as key words to derive a ten-stage sliding scale from ‘unpleasant’ (1) over ‘adequate’ (3), ‘beautiful’ (6) to ‘terrible’ and ‘fearful’ (10) spatial experiences. In the map, this scale was represented by a colour scale from white (0) to dark red (10). (Cooper et al., n.d.)
d) interactive maps (Figure 6): While all map examples so far are only published as static maps over the website, links to interactive websites that can be viewed with Google Earth are also available on the website. They contain point locations and connecting lines representing the tours of both authors as well as the authors’ texts, structured by days and with highlighted placenames. (Cooper et al., n.d.; Cooper and Gregory, 2011) This direct connection between the map and the text was intended to overcome the “principal weakness of Moretti’s literary geography in that the reader-generated GIS were dislocated, in presentational terms, from the literary texts that had been mapped”. (Cooper, 2012)
1.8 Literary mapping projects

1.8.2 The Digital Literary Atlas of Wales

Project outline

The Literary Atlas of Wales is a project funded by the Arts and Humanities Research Council by researchers from Cardiff and Swansea University. In the atlas, the research team mapped locations from twelve English language novels that are primarily set in Wales. (Anderson et al., n.d.; Smith, 2017) The literary atlas serves on the one hand as “a digital resource, providing textual and contextual information about novels and the locations in which they are set” and is on the other hand aimed at fostering a “more responsive, sensitive understandings of places and our relation to them” and, consequently, “inscribe places – and themselves – with new meaning”. (Smith, 2017)
Methods and results

For producing the interactive maps, geographical references were extracted from the novels by reading and translated into coordinates in order to map them as ‘plotpoints’. The atlas provides several kinds of maps:

a) ‘distant’ maps (Figure 7): Single plotpoints are mapped on a flat earth or on a globe in the form of pinpoints and collapsed points, respectively. On click, a popup shows the name of the location as well as the author, the novel and the context it is referenced in. Additionally, coverage maps show the geographical coverage of each novel in the form of multipolygon features. Checking layers on or off enables the user to compare the coverages of multiple novels. (Anderson et al., n.d.)

![Figure 7: Map illustrating the coverage of Amy Dillwyn’s The Rebecca Rieter. (Anderson et al., n.d.)](image)

b) ‘deep’ maps (Figure 8): For each novel, a deep map, augmented with supplementary materials such as photographs, historical maps, films, interviews, audio extracts, and excerpts from the novels is available to the user. The deep maps can be navigated similar to an Esri Storymap by scrolling through the locations within a single plotline. (Anderson et al., n.d.)
1.8 Literary mapping projects

c) **map of blue writer's plaques** (Figure 9): The locations of blue writer's plaques for famous Welsh authors are shown in a map together with each author's biography and a dynamic plot of the author's lifespans. (Anderson et al., n.d.)

![Image of map with blue writer's plaques and biography](image)

**Figure 9:** Map of locations of blue writer's plaques, selected the plaque for children's book author Roald Dahl. (Anderson et al., n.d.)
1.8.3 Palimpsest: Literary Edinburgh

Project outline

*Palimpsest: Literary Edinburgh* is supported by the Arts and Humanities Research Council and resulted in an “interactive, mobile-friendly website” featuring maps and visualisations representing the literary landscape of Edinburgh. (Loxley et al., 2017) This website called *LitLong: Edinburgh* is aimed at both experts and non-experts, literary scholars and interested laymen. (Alex et al., 2016) Creating and storing routes (‘paths’) between literary points of interest in the city of Edinburgh, the application can be used as navigation system for literary tourists. (Loxley et al., 2018, 2017) The project title refers to the concept of places as palimpsest that is also promoted by Graham (Graham, 2010) in order to “evoke the multi-layered imaginative, conceptual and historical cityscapes of our everyday settings that the resource seeks to bring to life” (Loxley et al., 2017).

Methods

For LitLong, a body of over 380,000 works in five literary document collections was processed automatically to retrieve a list of candidates having Edinburgh as main setting. After manual curation by literary scholars, the input data was reduced to 546 works of literature. Geographical references were extracted from the texts by applying natural language processing technologies, translated into coordinates using the Edinburgh Geoparser and stored in a database. (Alex et al., 2016; Loxley et al., 2018, 2017)

Results

LitLong is a map of Edinburgh with single point locations that, by clicking on them, reveal the title and author of the related work as well as the excerpt retrieved from the text by the text mining tool (Figure 10). A filter function allows for choosing a subset of the extensive amount of data by textual genre, by author, title, gender and date. Selected locations can be added to a ‘path’ that can be stored within a user profile and used as a navigation device for literary ventures through the city. (Loxley et al., 2018)
1.8.4 The Literary Atlas of Europe

Project outline

The Literary Atlas of Europe is the most extensive of the projects presented here and also the project with the most sophisticated methodology. A prototype for the Literary Atlas is currently being developed at the Institute of Cartography, ETH Zürich (CH) in cooperation with the Georg August University Goettingen (GER) and the Charles University Prague (CZ). (Piatti et al., 2009)

Three text collections comprising between 200 and 900 represent the following three model regions: Lake Lucerne/Gotthard (CH) as exemplary for an alpine landscape, North Friesland (GER) as coastal border area and Prague (CZ) as urban space (Piatti et al., 2011; Piatti and Hurni, 2009; Piatti and Weber, 2018)

Methods

In contrast to LitLong Edinburgh, data extraction was not done automatically, but manually by a team of literary scholars. To that end, a complex user interface was created where georeferencing of extracted locations as well as attribution according to toponymy,
locatability (precise, zonal or indefinite), function or process of formation is conducted. (Piatti and Hurni, 2009) Furthermore, the data comprises information on the author, the text, and the story line. Spatial objects are classified according to a sophisticated data model that is in accordance with the Simple Features Specification. (Open Geospatial Consortium Inc., 2010; “Simple Feature Access - Part 1: Common Architecture | OGC,” n.d.) The model distinguishes between five categories of locations:

a) setting: action takes places, characters are present;

b) projected place: characters not present, dreaming, remembering, longing;

c) marker: just mentioned, not part of any above;

d) zones of action: settings or projected places combined;

e) routes: connections between waypoints. (Piatti et al., 2009; Piatti and Weber, 2018; Reuschel et al., 2013)

The thus collected data is stored in a relational database (Piatti and Weber, 2018) and on this data basis, maps are generated automatically in respect to these attributes (Piatti et al., 2011; Piatti and Hurni, 2009)

Results

In the prototype, the user selects content by author, title, nationality or genre; information on the text and the author is provided with every map. (Piatti et al., 2011) Maps are available at different scales from local to global, underlaid with modern and historic reference maps. Comparing maps in a split screen allows for juxtaposing two works of the same or of different authors and it is possible to dynamically switch between single object and statistical maps. (Piatti et al., 2011; Reuschel and Hurni, 2011)

a) single object maps (Figure 11): show individual places of action, the distribution of single markers illustrates “the geography of an author, a genre of literature, or a certain time period”. (Piatti et al., 2009) Each location’s category is represented by a colour coding scheme in a 2x3 matrix: warm colours are used for settings and cold colours for projected places. Within each hue, three colour shades allow for distinguishing between ‘imported’, ‘transformed’ and ‘invented’ places. (Piatti and Weber, 2018) For visualising uncertain
localisations, fuzzy shapes have been used while animated symbols represent unclear point localisations within an area of probability. (Piatti et al., 2009)

![Figure 11: Map of Zikmund Winter’s 'Meister Kampanus' with a detailed interactive legend and a complex symbology. (Piatti and Weber, 2018)](image)

**b) statistical maps:** locations of the entire corpus are plotted on a map to visualise spatial distribution, patterns and density as well as temporal change. Thus, spatial literary centres can be identified as well as literary described spaces and blank, undescribed space. (Piatti and Weber, 2018)

### 1.8.5 Projects on Sherlock Holmes

**Literary geography publications**

In 1985, Yi-Fu Tuan dedicated a paper to the *Landscapes of Sherlock Holmes* where he stated that “the delightful adventures of Sherlock Holmes” may provide “knowledge concerning the landscapes, habits and thoughts of the late Victorian period.” (Tuan, 1985)

The short paper contains no maps but Tuan observed the spatial characteristics of the Sherlock Holmes stories. On the one hand, the adventures convey the international flair of
the Victorian empire, and on the other hand they show the detective’s explicit local knowledge of the town of London. Interesting deductions can also be made from the Sherlock Holmes adventures concerning the influence of nature and environment on the human spirit and how in Sir Arthur Conan Doyle’s stories crime scenes overlap with places of sinister appearance. (Tuan, 1985)

**Commentaries, collections and gazetteers**

Several attempts have been made to collect or visualise locations from Sherlock Holmes stories. In *The New Annotated Sherlock Holmes*, the editor Leslie Klinger provides detailed information on several locations as part of an extensive commentary on the text. (Doyle and Klinger, 2005) A *Sherlock Holmes Atlas* set up in the form of a gazetteer with information on a great number of locations contains some maps but was seemingly never finished. (Newbury, 1999) The *Arthur Conan Doyle Encyclopedia* by Alexis Barquin is a wiki-style collection of information on the author, his life and work. It contains also a list of locations for each story, for selected locations also a page with more detailed information, but no maps. (Barquin, 2017) *The Sherlock Holmes Companion* lists selected locations under the category “locations”. Entries contain references to the text and sometimes further information on the location, but apart from occasional links to a Victorian London map, no cartographic representations. (Crowley, 2011)

**Web maps**

Thomas Wheeler published a travel guide named *The New Finding Sherlock’s London*, which contains a comprehensive collection of London locations mentioned in Sherlock Holmes stories that is accompanied by contextual information. (Wheeler, 2009) Moreover, Wheeler’s travel guide is complemented by a web map. It features point mappings of locations in London in a Google Maps environment and provides information on the related actions as well as historical pictures (Figure 12). Furthermore, multiple routes taken from different stories are mapped as single line layers with contextual information on each line segment. (Wheeler, 2017)
1.8 Literary mapping projects

The London of Sherlock Holmes

… Mapped is a web map of London with map markers referring to locations in the stories (Figure 13). The author of the map, Matt Brown, has developed a colour coding scheme for the markers where green markers refer to a precise location and red markers to imprecise locations. Additionally, routes taken by Sherlock Holmes are indicated in blue. (Brown, 2009)


2 Objective

This Master's project aims at creating a prototype of an interactive web GIS application for visualising fictional literature. Selected works from the corpus of Sir Arthur Conan Doyle's Sherlock Holmes stories serve as an exemplary for developing a suitable methodology based on theoretical groundwork laid by scholars of literary geography and cartography.

2.1 Scope

The project presented in this thesis is based on but going beyond the path drawn out by a student project. The result of the former project was an ArcGIS Online Storymap featuring geographical references in the four Sherlock Holmes novels (A Study in Scarlet, The Sign of Four, The Hound of the Baskervilles and The Valley of Fear). Single point mappings as well as results of spatial analyses like density surfaces, convex hulls and distance analysis with multiple ring buffers were presented in the storymap alongside links to webpages and statistical charts. (Binder, 2017a, 2017b)

This Master's project draws on ideas developed for the storymap but exhibits several enhancements and changes of direction. The most prominent difference is that the web application relies solely on open source technologies. Furthermore, the app is developed and programmed completely from scratch without using predefined templates or components and based on an ad hoc architecture. The app features a higher degree of interactivity as the storymap: while the former enables the user to scroll through the pages and explore single locations on click, the web app allows for more user interaction concerning data selection, mode of display and analysis/visualisation methods. In the storymap, only static pre-processed datasets resulting from analyses done in a desktop GIS were presented. In the app, all processing is done on-the-fly via a WPS and the result is displayed instantaneously in the map. Not least, the web application is hosted on a university-owned web server independent from proprietary user profiles. The differences between the storymap and the web app approach are summarized below in Table 1.
2.1 Scope

Table 1: Juxtaposition of the storymap and the web app approach to Sherlock Holmes.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Storymap</th>
<th>Web app</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
<td>Proprietary</td>
<td>FOSS (Free Open Source Software)</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td>Dependent on proprietary templates</td>
<td>Independent, customisable</td>
</tr>
<tr>
<td><strong>Datasets</strong></td>
<td>Static, not changeable by user</td>
<td>Dynamic, changeable according to user-defined selection</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td>Not available (only pre-processed data)</td>
<td>Real-time display of results</td>
</tr>
<tr>
<td><strong>Interactivity</strong></td>
<td>Navigation through preselected map layers and pages</td>
<td>User decisions determine map content, analysis methods and visualisation</td>
</tr>
<tr>
<td><strong>Hosting and longevity</strong></td>
<td>Proprietary servers</td>
<td>University-owned servers</td>
</tr>
</tbody>
</table>

**Application design:**

The main characteristic of the web application is a high degree of interactivity and the freedom granted to the user to select a subset of the data by spatial and non-spatial criteria according to their own interest or curiosity. This feature makes the application highly dynamic and allows for mapping, analysing and exploring a specifically selected subset.

The application is used in three main steps:

a) select data

b) create map

c) analyse and explore.

After choosing the subset, a map visualisation of all geographical references in the text provides an overview of the selected data. Depending on the geographical scale, the references are represented by point, line or polygon geometries. Zooming allows for exploring the map at different scales from global to local.

Step by step, the user can then explore the dataset and its spatial and non-spatial characteristics in increasing detail. This includes querying detailed information on each map feature concerning plot and characters as well as manipulating the symbology to represent different qualitative categories referring to each element’s ontological status.
Furthermore, the user can explore spatial characteristics of the selection. This comprises characteristics of the first order like location and extent as well as characteristics of the second order like density and distance. Density is assessed in reference to an administrative boundary layer fetched on-the-fly from different WFS. Distance measurements can be executed pairwise between features of a dataset to determine the distance between locations A and B. At the same time, the distribution of distances between all features of the subset is assessed.

2.2 Choice of data

Sir Arthur Conan Doyle’s Sherlock Holmes stories were chosen because they contain vivid descriptions of late 19\textsuperscript{th}/early 20\textsuperscript{th} century London. The whereabouts of the detective going about his investigations as well as scenes of crime and mystery are described with great attention to detail. Furthermore, frequent references to Dr Watson’s military past and other allusions to the monarchy’s sphere of influence promised to deliver a rich and interesting body of data for mapping. The intriguing and charming particularities of the geographical aspects in the Sherlock Holmes stories have also been recognised by literary geography scholars such as Yi-Fu Tuan (see also above section 1.8.5). (Tuan, 1985)

Additionally, Sherlock Holmes and related detective stories have been and are still among the most popular works of literature since the detective made his first appearance in 1887. Since then, adaptations and hommages have been published in different media including for example numerous films starting from 1900 in the silent film era (“Conan Doyle on screen - The Arthur Conan Doyle Encyclopedia,” n.d.) over Hollywood films starring Robert Downey Jr. and Jude Law (Sherlock Holmes, 2009) and TV series such as Elementary set in New York with Lucy Liu as Dr. Joan Watson (Elementary, 2012) or Sherlock (BBC One), played by Benedict Cumberbatch (Sherlock, 2010). Furthermore, a number of PC games developed by Frogwares studio, the latest Devil’s Daughter being released in 2016. (“The Adventures of Sherlock Holmes Games List from Frogwares,” 2016) With A Slight Trick of the Mind, Mitch Cullin has created a novel about an elderly Sherlock Holmes struggling with his failing memory. (Cullin, 2005) In 2015 followed a film adaption of Cullin’s book starring Ian McKellen as Mr Holmes. (Mr. Holmes, 2015) Naturally, this list of film and TV adaptions is
not nearly exhaustive. Not listed here are furthermore numerous adaptations for radio and stage set in the Sherlock Holmes universe. (“Conan Doyle on radio - The Arthur Conan Doyle Encyclopedia,” n.d., “Conan Doyle on stage - The Arthur Conan Doyle Encyclopedia,” n.d., “Sherlock Holmes pastiches,” 2018) Due to Sherlock Holmes’s prominence in popular culture, the application is expected to raise users’ attention even if they are not familiar with the original texts.

To obtain a suitable subset of data, *The Final Problem*, last story of the collection *The Memoirs of Sherlock Holmes* was chosen as breaking point. In this story, published in 1893, Arthur Conan Doyle let his most famous fictional character die together with his arch enemy Prof. Moriarty at the Reichenbach Falls in Switzerland. As the story was originally intended by the author to end the Sherlock Holmes series (Conan Doyle and Wolfreys, 1996; “Sir Arthur Conan Doyle:Sherlock Holmes - The Arthur Conan Doyle Encyclopedia,” n.d.), data from all short stories beginning at the earliest and ending at *The Final Problem* was chosen for the application. Additionally, the data collected during the previous project on the four Sherlock Holmes novels (*A Study in Scarlet*, *The Sign of Four*, *The Hound of the Baskervilles* and *The Valley of Fear*) was added to the application.

### 2.3 Aims

Rather than illustrating the results of a spatial analysis of selected Sherlock Holmes stories, the application shall serve as a tool for exploration and analysis. The application will reveal new insights into a specific, exceedingly popular piece of literature by rendering visible the geographical context as well as the interplay between location, plot and characters. This may not only spark new interest in the original stories behind the famous fictional character, but also serve to exemplify the richness of fictional literature as a genre. Not only the interested public, but also literary scholars may benefit from the application and use it as a tool for generating hypotheses, conducting analyses and deriving new scholarly informed interpretations concerning the geography of Sherlock Holmes.

Moreover, the project shall emphasize the strengths of open source geospatial technologies in terms of interoperability, extensibility, institutional and economical independence, sustainability and accessibility.
3 The data

3.1 Data acquisition

Texts were retrieved from the platform Project Gutenberg (Hart, 2018) and downloaded as text files. All spatial and contextual information was extracted manually by working through the texts in a text editor (LibreOffice). Every occurrence of a spatial reference was highlighted in the text with a colour scheme referring to the scale of reference (see below Figure 14 and Table 2).

![Image showing text editor with color-coding of geographical references]

Figure 14: Highlighting geographical references in the text.

For each reference, a bookmark was created to enable access to the specific text passage. Most geographical references were identified by explicit naming of countries (“Holland”, “India”, “Europe”), cities (“London”, “Chilian-whalla”, “Illinois”), regions or landmarks (“California Gold Fields”, “Attica”, “The Downs”), or places (“Waterloo Bridge”, “Halliday’s...
3.1 Data acquisition

Private Hotel”, “Vermissa police station”). Some places have been identified by using a person’s or company’s name and the related private or business location as a proxy (“Madame Lesurier, Bond Street”, “Venner & Matheson, Greenwich”, “Alexander Holder’s house, Streatham”). Others were referenced only very vaguely in the text (“Cottage near Mr and Mrs Munro’s villa, Norbury”, “A private hotel in the Strand”).

Table 2: Colour scheme used in text highlighting.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Geographical scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orange</td>
<td>Continent or country</td>
</tr>
<tr>
<td>Yellow</td>
<td>City</td>
</tr>
<tr>
<td>Purple</td>
<td>Place within a city: address, house, street etc.</td>
</tr>
</tbody>
</table>

Not used in analysis

<table>
<thead>
<tr>
<th>Colour</th>
<th>Geographical scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>References to origin, adjectives</td>
</tr>
<tr>
<td>Turquoise</td>
<td>Means of transportation</td>
</tr>
</tbody>
</table>

All locations thus derived from the texts were consequently categorised according to reality, accuracy, certainty and storylocation type and transferred with all spatial and non-spatial information to the data collection. For a description of the database see below section 4.1.

Manual extraction was deliberately chosen over automated extraction by geoparsing that was employed for LitLong: Edinburgh (Loxley et al., 2018). This was done in order to maintain control over which geographical references are included in the data collection. By reading, also implicit references could be identified that would otherwise perhaps been lost in the automated process. In cases where the references are ambiguous or vague (see above section 1.6.3 on uncertainty), manual research was deemed more suitable for deciding out of context for the most suitable localisation. Automated extraction may lead to associating the reference with the wrong place or erroneous identifications of places like reading personal names as locations. (Loxley et al., 2018) However, it is worth noting that even manual, ‘human’ referencing is not immune to errors. Furthermore, instead of extracting “the sentences immediately surrounding each mention” word by word from the original text like it was done for LitLong: Edinburgh (Loxley et al., 2018), a summary of actions occurring at
the location was transferred to the data collection for use in the application. In some cases, this necessitated summarising actions over multiple paragraphs or even pages.

3.2 Ontology

The data model for this project is based on the general assumption that fictional literature consists of characters, storyline and space. (Piatti et al., 2009) The same model has already been used for the storymap project and a simplified version of its visual representation is given below in Figure 15.

![Conceptual model applied to geographical references in the Sherlock Holmes stories.](image)

According to this model, each ‘story’ (both short story and novel) from the Sherlock Holmes canon consists of multiple ‘storylocations’. Each ‘storylocation’ is a singular combination of a **location** (for example “Sherlock Holmes’s lodgings in Baker Street 221B”), the **actions** happening at or in reference to this location at this given point within the plotline (for example “When Dr Watson returns to Baker Street, he shows Sherlock Holmes the empty treasure-box.”), and the **characters** involved in these actions (for example “Dr. Watson and
3.2 Ontology

Sherlock Holmes”). One ‘character’ can appear at multiple ‘storylocations’ within one story. Moreover, as the Sherlock Holmes canon consists of multiple ‘stories’ (56 short stories and 4 novels) that are set within the same universe, one ‘character’ can also appear in different ‘stories’. This is most obviously the case with Sherlock Holmes and Dr. Watson, but also applies to Scotland Yard inspector Greg Lestrade or Sherlock Holmes’s arch enemy Prof. James Moriarty.

3.2.1 Locations and geometry type

The ‘location’ aspect of each ‘storylocation’ is represented by geometry attributes in the form of map coordinates. The choice of a geometry type depends on the scale and accuracy of the reference as well as on the dimensions of the referenced real-world entity.

References to administrative units include ‘continent’, ‘country’, ‘state’, ‘county’, ‘borough’, ‘parish’, but also ‘island’, ‘region’ and ‘kingdom’ and have been mapped as polygon features. Wherever possible, contemporary boundaries have been used for mapping administrative units (for example for ceremonial counties of England). Likewise, references to water bodies such as a ‘lake’ have been mapped as polygon features.

All references made by mentioning names of settlements including a ‘city’ or ‘village’ have been mapped as point features set approximately in the geographical centre of the settlement. References to a ‘place’ include specific locations in a settlement, for example a street address, a road junction or a shop, as well as narrowly demarcated natural features such as a landmark. ‘Place’ references have also been mapped as point features.

A reference to a clearly demarcated elongated feature such as a ‘river’, ‘bridge’ or ‘street’ has been mapped as line feature.

References to landscapes, mountain ranges, valleys, canyons and areas within an ocean have been subsumed under ‘natural feature’. As it is not possible to delineate these areas with crisp boundaries, ‘natural feature’ references have been mapped as point locations set approximately in the geographical centre of the area.
3.2 Ontology

Table 3: Geometry type of map features according to the reference scale.

<table>
<thead>
<tr>
<th>Reference scale</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative boundaries:</td>
<td></td>
</tr>
<tr>
<td>Water bodies:</td>
<td></td>
</tr>
<tr>
<td>lake (“The Serpentine”)</td>
<td>Polygon</td>
</tr>
<tr>
<td>Elongated features:</td>
<td></td>
</tr>
<tr>
<td>ridge (“Waterloo Bridge”), street (“The Strand”), river (“Severn”)</td>
<td>Line</td>
</tr>
<tr>
<td>Settlements:</td>
<td></td>
</tr>
<tr>
<td>city (“Bristol”), village (“Petersfield”)</td>
<td>Point</td>
</tr>
<tr>
<td>Specific locations:</td>
<td></td>
</tr>
<tr>
<td>place (“St. Pancras Hotel”)</td>
<td>Point</td>
</tr>
<tr>
<td>Natural features:</td>
<td></td>
</tr>
<tr>
<td>natural (“Boscombe Valley”)</td>
<td>Point (annotation)</td>
</tr>
</tbody>
</table>

3.2.2 Reality

An overview of the complex theoretical framework on the relationship between ‘real’ and ‘fictional’ spaces has been given above in section 1.4.2. While some scholars refute any connections drawn between places named in fictional literature and places in the physical world (Stockhammer, 2013), others have thoroughly investigated the processes in fictional literature that transform ‘real’ places to places of varying fictionality (Piatti and Hurni, 2011). For the purpose of this project, the ontological classification has been simplified and reduced to two categories: ‘real’ and ‘fictional’. As in the previous project, the ontological category assigned to a ‘storylocation’ refers to the smallest spatial unit that is given in the text to describe the location of the place. A fictional address (“221B Baker Street”) located in a real street (“Baker Street”) is for example categorised as ‘fictional’.

The decision whether a location is categorised as ‘real’ or ‘fictional’ has been made according to mostly web-based research. Besides modern web maps such as Google Maps and OpenStreetMap, historical maps and gazetteers proved to be of value. These included a digitised Map of City of London and its Environs published by the Ordnance Survey,
3.2 Ontology

Southampton between 1869 and 1880 which is provided as online resource by British History Online. (“Map of City of London and its Environs (1869-1880),” 2017) The same website offers also digitised versions of a series of topographical dictionaries for England, Scotland and Wales, published between 1846 and 1848 by Samuel Lewis. (“A Topographical Dictionary of England (1848),” 2015) The National Library of Scotland has published georeferenced historical maps of various date and scale that can be freely accessed in a web map. (National Library of Scotland, 2016) Additionally, the commentaries, gazetteers and collections that have been described above in section 1.8.5 have been used as sources for localising places mentioned in the texts and finding out whether they refer to ‘real’ locations or not. A description of the ontological categories ‘real’ and ‘fictional’ that have been applied to this project is given below in Table 4.

Table 4: Description of ontological categories ‘real’ and ‘fictional’ applied to the project.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td>a place of the given name does exist or has at any time existed at this location</td>
</tr>
<tr>
<td>fictional</td>
<td>a place of the given name does not exist and has never existed at this location</td>
</tr>
</tbody>
</table>

In many cases research showed that locations named in the text do not exist anymore in the present topography of a town or country (“St. James’s Hall” in London was demolished in 1905) or that a place has in the meantime be renamed (“Broad Street” in London is today known as “Black Prince Road”). To reflect these findings in the data model, an attribute ‘historic’ has been defined. A description of the categories is given below in Table 5.

Table 5: Description of ‘historic’ categories for storylocations.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y (yes)</td>
<td>does not exist today or does exist under another name</td>
</tr>
<tr>
<td>n (no)</td>
<td>does still exist today under the given name</td>
</tr>
<tr>
<td>u (unknown)</td>
<td>is fictional</td>
</tr>
</tbody>
</table>

3.2.3 Certainty

‘Certainty’ can be understood as a measure of probability. It refers to the probability of having identified the ‘correct’ location that was ‘meant’ with the place name in the text. The
level of certainty decreases with increasing ambiguity of place names. “St. Monica’s church in Edgware Road” is a location of ‘poor’ certainty, because there are and have been several churches dedicated to St. Monica in London, but none of them in Edgware Road. There are, on the other hand, several churches alongside Edgware Road, but none of them is dedicated to St. Monica (see below, Figure 16).

Fictional locations are often of only medium or poor certainty while real locations are commonly easier and more securely identified and therefore often of ‘good’ certainty. A description of categories used for ‘certainty’ is given below in Table 6.

Table 6: Categories for ‘certainty’.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>no reason to doubt that this is the location meant in the text</td>
</tr>
<tr>
<td>medium</td>
<td>some doubts concerning the identification of the location, but the chosen location can be favoured from the context</td>
</tr>
<tr>
<td>poor</td>
<td>unclear which location is meant in the text and whether it is the one chosen for visualisation</td>
</tr>
</tbody>
</table>

3.2.4 Accuracy
‘Accuracy’ refers to the precision of the localisation. The smaller the area of probability for localising a named place, the higher the accuracy and vice versa. All polygon features representing administrative units have crisp, clearly defined boundaries that are well-known from cartographic products, so accuracy for them is always ‘high’. Likewise, most cities or towns are referenced with ‘high’ accuracy, if name and location of the settlement are well-known. More generally spoken, the higher the probability of having identified the right ‘real-world’ location over the name given in the text, i.e. the higher the ‘certainty’, the higher is in most cases the ‘accuracy’ of the localisation. When ‘certainty’ is at its lowest, meaning that there are major doubts about which ‘real-world’ location is meant and if there is, at all, a ‘real-world’ counterpart, then ‘accuracy’ is also at its lowest and the location is not mappable.

‘Accuracy’ of location mapping does also decrease with an increasing vagueness of the localisation given in the text. “Mycroft Holmes’s rooms in Pall Mall” for example cannot be localised more precisely than being located somewhere in Pall Mall. As listed in Table 7, ‘accuracy’ is in this case ‘medium’ because the location of Pall Mall is known, but the exact location of Mycroft Holmes’s rooms in Pall Mall is not.

*Table 7: Categories for ‘accuracy’.*

<table>
<thead>
<tr>
<th>Category</th>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>urban</td>
<td>localised at an address or road junction</td>
</tr>
<tr>
<td></td>
<td>regional</td>
<td>localised at a city or landmark</td>
</tr>
<tr>
<td>medium</td>
<td>urban</td>
<td>localised within a street or district but the exact address is unknown</td>
</tr>
<tr>
<td></td>
<td>regional</td>
<td>localised within a region but the exact location is unknown</td>
</tr>
<tr>
<td>low</td>
<td>urban</td>
<td>localised anywhere within a city or urban district</td>
</tr>
<tr>
<td></td>
<td>regional</td>
<td>localised very vaguely in relation to another location</td>
</tr>
<tr>
<td>none</td>
<td></td>
<td>cannot be localised, not mappable</td>
</tr>
</tbody>
</table>

### 3.2.5 Mappability

The possibilities for localising and, consequently, mapping places mentioned in the Sherlock Holmes stories naturally depend on the attributes ‘reality’, ‘historicity’ and ‘certainty’. Some purely fictional places can be localised and mapped quite well because an exact address is
given (“129 Camberwell Road, London”) or because the location is described in reference to known places (“a wharf at the end of Broad Street”). For other places, however, no suitable ‘real-world’ counterpart can be identified and thus, they remain ‘unmappable’. Values for ‘mappability’ are given below in Table 8.

Table 8: Values for ‘mappability’.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y (yes)</td>
<td>localisation possible with high or medium certainty</td>
</tr>
<tr>
<td>p (proxy)</td>
<td>localisation possible with low certainty</td>
</tr>
<tr>
<td>n (no)</td>
<td>localisation not possible, no real-world counterpart could be identified</td>
</tr>
</tbody>
</table>

3.2.6 Location type

The above listed geographical and ontological categories apply to the ‘location’ aspect of each storylocation. Additionally, storylocations have also been grouped into different location types according to the role that the location plays in the specific point in the narrative. Location roles represent the ‘action’ aspect of the storylocations. As Barbara Piatti and her research team have observed, places mentioned in fictional literature are not necessarily locations where the main plot ‘happens’ in the sense that characters are at one point in the plotline physically present and some kind of action takes place. Some locations are, however, places that appear in the characters’ thoughts or words. (Piatti and Weber, 2018) For this project, six location roles have been defined and are listed below in Table 9.

Table 9: Location roles used for categorising storylocations.

<table>
<thead>
<tr>
<th>Location role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>action</td>
<td>actions within the main timeline of the plot take place, characters are physically present</td>
</tr>
<tr>
<td>crime</td>
<td>special case of action where murders or other crimes happen</td>
</tr>
<tr>
<td>background</td>
<td>actions before the main timeline of the plot take place, characters are physically present</td>
</tr>
<tr>
<td>story</td>
<td>reference</td>
</tr>
<tr>
<td>reference</td>
<td>verbally referenced places, characters are not physically present</td>
</tr>
<tr>
<td>case</td>
<td>references to scenes of other criminological cases (untold stories)</td>
</tr>
<tr>
<td>clue</td>
<td>geographical references taken up as hints in a case</td>
</tr>
</tbody>
</table>
3.2 Ontology

The location roles can be summarised into two groups: ‘action’, ‘crime’ and ‘background story’ are locations where characters are physically present and either perform or experience any kind of action. The main difference between ‘action’ and ‘background story’ is that ‘actions’ take place within the main timeline of the plot, most often starting with Sherlock Holmes taking up investigations in a criminal case (“At the Alpha Inn, Sherlock Holmes asks the landlord about the geese.”). ‘Background stories’, however, are set before the main timeline (“Dr Watson’s experience of camp life in Afghanistan has made him a prompt and ready traveller.”). Especially in the Sherlock Holmes novels but also in some short stories, Arthur Conan Doyle provides the reader with insights into a character’s life story by making excursions into the past. ‘Crimes’ have been chosen as special case of ‘action’ (“The paper reports that a young man was found dead in the waters near Waterloo Bridge.”).

The second group of location roles comprises ‘reference’, ‘case’ and ‘clue’ and comprise most often locations that are referenced by characters when they speak or think. ‘References’ are mentioned in one character’s accounts to another character or as part of descriptions of the topography (“Eyford, Colonel Lysander Stark explains, is a little place near the borders of Oxfordshire.”). ‘Cases’ are Sherlock Holmes’s references to his past criminological cases. Some of these mentions emphasise intertextuality between the Sherlock Holmes stories (“Dr Watson has only once known Sherlock Holmes to fail: In the case of the King of Bohemia and of the Irene Adler photograph.”). Others, however refer to cases that have not been published by Arthur Conan Doyle (“Sherlock Holmes refers to the case of Dolsky in Odessa concerning the forceful use of poison.”) and are known among ‘Sherlockians’ as ‘untold stories’ (“Category:Untold Stories - The Arthur Conan Doyle Encyclopedia,” n.d.). ‘Clues’ are references to locations that serve as hints in a case, mostly the provenience of certain items (“Miss Mary Morstan and Dr Watson recognise the metal treasure-box of being a work from Benares.”).
4 Data architecture and used software

The data architecture is structured in three main layers: the data layer, the server layer and the client layer. The primary architecture comprises the web application, a GeoServer instance and a PostGIS data base. Secondary data sources are included for retrieving additional data such as administrative boundary layers from the web. The architecture is illustrated above in Figure 17.

Table 10 below provides an overview of the software and libraries used for this project. All of them are FOSS, that is Free Open Source Software products. PostGIS, Geoserver, OpenLayers and QGIS are part of the Open Source Geospatial Foundation (OSGeo), a not-for-profit organisation promoting „global adoption of open geospatial technology“. (“About OSGeo - OSGeo,” 2018)

Figure 17: Data architecture consisting of three main tiers for data processing and retrieval and additional secondary sources for external WFS, after Fu & Sun (Fu and Sun, 2011).
Table 10: Software products used for project realisation

<table>
<thead>
<tr>
<th>Name</th>
<th>Developer(s)</th>
<th>Written in</th>
<th>FOSS</th>
<th>License</th>
<th>Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostgreSQL</td>
<td>PostgreSQL Global Development Group</td>
<td>C</td>
<td>Yes</td>
<td>PostgreSQL License</td>
<td>Linux, Windows, MacOS</td>
</tr>
<tr>
<td>PostGIS</td>
<td>Refractions Research et al.</td>
<td>C</td>
<td>Yes</td>
<td>GNU General Public License</td>
<td>Linux, Windows, MacOS</td>
</tr>
<tr>
<td>pgAdmin</td>
<td>The pgAdmin Development Team</td>
<td>Python, jQuery</td>
<td>Yes</td>
<td>PostgreSQL License</td>
<td>Linux, Windows, MacOS</td>
</tr>
<tr>
<td>Geoserver</td>
<td>Boundless Spatial, GeoSolutions, Refractions Research</td>
<td>Java</td>
<td>Yes</td>
<td>GNU General Public License</td>
<td>Linux, Windows, MacOS</td>
</tr>
<tr>
<td>QGIS</td>
<td>QGIS Development Team</td>
<td>Qt, C++</td>
<td>Yes</td>
<td>GNU General Public License</td>
<td>Linux, Unix, Windows, MacOS, Android</td>
</tr>
<tr>
<td>OpenLayers</td>
<td>The OpenLayers Dev Team</td>
<td>JavaScript</td>
<td>Yes</td>
<td>2-clause BSD License</td>
<td>Web browser</td>
</tr>
<tr>
<td>jQuery</td>
<td>The jQuery Team</td>
<td>JavaScript</td>
<td>Yes</td>
<td>MIT License</td>
<td>Web browser</td>
</tr>
<tr>
<td>D3.js</td>
<td>Mike Bostock et al.</td>
<td>JavaScript</td>
<td>Yes</td>
<td>BSD License</td>
<td>Web browser</td>
</tr>
</tbody>
</table>

4.1 Data layer

4.1.1 Software

Spatial information derived from the Sherlock Holmes stories was collected in a PostgreSQL database (V9.5.12) using the PostGIS geospatial database extension (V2.2.1) for storing geodata. For data entry and management, pgAdminIII (V1.22.0) was used.

PostgreSQL is a powerful, open source object-relational database management system looking back at more than 30 years of active development. (“PostgreSQL: About,” n.d.)

Original concepts were developed in 1986 in the course of the POSTGRES project led by Michael Stonebreaker at the University of California at Berkeley. With the release of Postgres95, Structured Query Language (SQL) capabilities were added to the database and

To enable the integration of spatial data, PostGIS serves as a spatial database extender for PostgreSQL. It is written in C and published as free, open source software under the GNU General Public License. (“PostGIS — Development,” n.d., “PostGIS — Spatial and Geographic Objects for PostgreSQL,” n.d.) With PostGIS, spatial data types such as geometry, geography or raster along with applicable functions, operators and index enhancements are added to the PostgreSQL database. PostGIS spatial functionalities comply with the OGC’s Simple Features for SQL Specification which is the SQL implementation of the ISO 19125 Simple Feature Access Standard. (Open Geospatial Consortium Inc., 2010; “PostGIS — PostGIS Feature List,” n.d.)

pgAdmin is written in Python and serves as a management tool for PostgreSQL versions 9.2 and above. The graphical user interface enables access for creating, viewing and editing to all PostgreSQL database objects such as schemas, tables, columns, constraints and the like as well as an SQL query tool. Moreover, pgAdmin supports the use of procedural languages such as pl/pgsql for writing custom functions. (“pgAdmin - PostgreSQL Tools,” n.d.)

4.1.2 Database structure

The conceptual model presented above in section 3.2 is reflected in the database structure. A model of the main tables, attributes and keys is illustrated below in Figure 18.
4.1 Data layer

The entities 'locations', 'storylocations', 'persons' and 'stories' are represented by tables with the same name and connected with foreign key references. Separate tables each hold features of a certain geometry type including points, polygons and lines. An intermediate table holds the relations between storylocations and persons connecting them by numerical IDs. A small database function was used to populate the 'personsarray' field in the 'storylocations' table from all table records in 'personsinvolved' referring to a specific storylocationID. Detailed descriptions of table attributes are given below in Table 11, Table 12, Table 13 and Table 14.

Table 11: Properties of ‘locations’.

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>idlocation</td>
<td>numerical identifier of the location</td>
</tr>
<tr>
<td>name</td>
<td>name of the location as given in the text</td>
</tr>
<tr>
<td>historic</td>
<td>whether the location does still exist under that name today or not</td>
</tr>
</tbody>
</table>
name_today  if name is historic, current name of location
idcountry  numerical identifier of country the location is situated in (refers to idlocation)
country_contemp  contemporary name of the country the location is situated in
country_today  current name of the country the location is situated in
state_country  name of the state (USA) or country (UK) the location is situated in
county  name of the county the location is situated in
city  name of the city the location is situated in
district  name of the district the location is situated in
street  name of the street the location is situated in
reference_to  spatial scale of the geographical reference
genre  geometry type of the referencing map feature
reality  ontological status of the location
certainty  how secure the identification of the location with a real-world counterpart is
accuracy  how precise the location could be localised
mappable  whether the location can be represented by a map feature or not
location_source  information source for localisation
reasoning  evidences for localisation according to sources

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>idstorylocation</td>
<td>numerical identifier of the storylocation</td>
</tr>
<tr>
<td>idstory</td>
<td>numerical identifier of the story</td>
</tr>
<tr>
<td>story</td>
<td>string identifier of the story</td>
</tr>
<tr>
<td>idlocation</td>
<td>numerical identifier of the location the action takes place at</td>
</tr>
<tr>
<td>appearance_order</td>
<td>rank in sequential order of appearance within the story</td>
</tr>
<tr>
<td>location_role</td>
<td>role of the location in the narrative</td>
</tr>
<tr>
<td>actions</td>
<td>summary of actions related to the place</td>
</tr>
<tr>
<td>personsarray</td>
<td>list of characters associated with this storylocation instance</td>
</tr>
</tbody>
</table>
### 4.1 Data layer

**Table 13: Properties of ‘persons’.

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>idperson</td>
<td>numerical identifier of a person</td>
</tr>
<tr>
<td>lastname</td>
<td>last name of the person</td>
</tr>
<tr>
<td>firstname</td>
<td>first name of the person</td>
</tr>
<tr>
<td>initial</td>
<td>shortened name of the person (first two letters of firstname + first two letters of lastname)</td>
</tr>
</tbody>
</table>

**Table 14: Properties of ‘stories’.

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>idstory</td>
<td>numerical identifier of the story</td>
</tr>
<tr>
<td>title</td>
<td>full title of the original publication</td>
</tr>
<tr>
<td>abbrev</td>
<td>abbreviated title in four letters</td>
</tr>
<tr>
<td>story_type</td>
<td>whether it is a short story or a novel</td>
</tr>
<tr>
<td>publ_year</td>
<td>year of first publication</td>
</tr>
<tr>
<td>publ_medium</td>
<td>medium of first publication</td>
</tr>
<tr>
<td>collection</td>
<td>name of the collection the story was later published in</td>
</tr>
</tbody>
</table>

### 4.2 Logical or server layer

A Geoserver (V2.13.0) instance was implemented as a platform to host the spatial data in the PostgreSQL database and make it accessible and queryable through the user interface. To that end, the PostGIS database was added as a data store to the Geoserver instance and all vector features were published as an SQL view layer. This view is a join of all tables where each record represents one story location mapped as one vector feature. Consequently, only story locations that could be mapped as vector features can be queried from the data collection. Story locations without a spatial reference have been collected in the database but are not used in the application.

Eclipse Jetty was installed together with the GeoServer binary as a web server for communicating between the map server and the client application. Apache Tomcat can, however, be used as an alternative to Jetty. GeoServer is a free, open source software written

4.3 Presentation or client layer

4.3.1 Software

While the contextual information for each location was derived directly from the text, the geographical information was collected mostly by web search on platforms like Google Maps or OpenStreetMap, complemented by leveraging freely available Web Services, for example Open Government Data WFS. Additional information was obtained from web sites with historical information. Mapping of point, line, polygon and annotation features was done in QGIS (V2.8) by connecting to the PostGIS database.

The frontend of the web application is based on an HTML structure with styling defined in Cascading Style Sheets (CSS). All functionalities including processing of user inputs, forwarding fetch requests to the map server as well as dynamic styling and feature animation was programmed in JavaScript. For handling geographic data, mostly in JavaScript Object Notation (JSON) format, and creating map representations, the Openlayers API (V4.6.5) was used. Additionally, the JavaScript code was extended with jQuery (V3.2.1) to facilitate dynamic creation and manipulation of Document Object Model (DOM) elements such as divs, paragraphs, buttons and the like. Furthermore, the library d3.js (V5) was used for creating graphical visualisations like bar and pie charts.

QGIS is a free, open source GIS software written in Qt and C++ published under the GNU General Public License. (“QGIS - Foreword,” n.d.) With QGIS, users can view, explore,
create, edit, export and analyse geospatial data in various formats. Vector data support comprises PostGIS, SpatiaLite, Microsoft SQL Spatial, Oracle Spatial as sources as well as ESRI shapefiles, GML and other formats. Supported raster data formats include GeoTIFF, ERDAS IMG, ArcInfo ASCII GRID, JPEG and PNG, all enabled by the installed Geospatial Data Abstraction Library (GDAL). Moreover, QGIS can integrate OGC Web Services such as WMS, WMTS, WCS, WFS and WFS-T. (“QGIS - Features,” n.d.) In addition to the core functionalities, QGIS is extensible with a number of plugins. (“QGIS - Plugins,” n.d.)

OpenLayers is a free, open source JavaScript library and provides an API for geospatial web applications. It was created in 2005 by MetaCarta (“OpenLayers - Wikipedia,” n.d.) and supports vector data in GeoJSON, TopoJSON, KML, GML and other formats, as well as tiled layers, for example from OpenStreetMap (OSM), Bing, MapBox, Stamen. Additionally, with OpenLayers it is possible to integrate OGC mapping services such as WMS, WMTS and WFS. (“OpenLayers - Welcome,” n.d.)

jQuery is a free, open source JavaScript library that facilitates DOM element selection, manipulation, event handling and animation. Development of jQuery started in 2005 and the first release was done in 2006. (Jeresig, 2007) With jQuery, JavaScript code can be separated from HTML markup, it promotes brevity and clarity with chainable functions and shorthand function names and handles cross-browser incompatibilities. Furthermore, jQuery is to a high degree extensible. (“jQuery - Wikipedia,” n.d.)

D3.js (D3 stands for Data Driven Documents) is a free, open source JavaScript library used for binding data to DOM elements. With D3, the styling of document elements can be manipulated dynamically according to the underlying data by making use of web standards such as HTML, Scalable Vector Graphics (SVG) and CSS. D3 enables a wide variety of visual representations such as box plots, bar and line charts, treemaps and many more as well as map visualisations such as cartograms, symbol maps, choropleth maps and others. (“D3.js - Data-Driven Documents,” n.d.) D3.js was initially released in 2011 as a successor of the Protovis framework. (“D3.js - Wikipedia,” n.d.)
4.3.2 Implementation

Web Services use

Asynchronous JavaScript requests communicate with the WFS. To fetch data from the PostGIS database or administrative boundaries from external WFS, a getFeature request is built by using an OpenLayers `ol.format.WFS().writeGetFeature()` object.

With `XMLSerializer().serializeToString()`, the XML body for the request is built from the OpenLayers WFS `writeGetFeature()` object. Finally, a POST request is sent to the WFS by using the JavaScript `fetch` API. If the WFS responds successfully, the requested data is returned in JSON format. OpenLayers `ol.format.GeoJSON().readFeatures()` creates OpenLayers `ol.feature` vector features that are displayed in the browser map after adding them to a vector source and a map layer. For the boundary layers, the data is returned in GML format and turned into map features by using OpenLayers `ol.format.WFS().readFeatures()`.

WFS requests are also used for dynamically populating the dropdowns in the filter menu. Changes in the dropdowns are handled by an event listener in jQuery. Each time a selection is made from one of the dropdowns, a request is sent to the WFS using the current selection as filter object `ol.filter` to fetch values for all other dropdowns that match the current selection. If, for example, ‘The Hound of the Baskervilles’ is selected from the story dropdown, all other dropdowns will only show characters, countries etc. that occur in the chosen story. When fetching geodata from the database, the current selection from the dropdowns is used as `ol.filter` in the `getFeature` request.

The convex hull and spatial frequency assessment make use of Geoserver WPS operations. To that end, the XML body for the request is built according to the parameters listed under `DescribeProcess` by using the JavaScript DOM Reference. The geodata that is sent to the WPS for processing is embedded in the XML body in a CDATA section in the form of a JSON string. Text inside a CDATA section will not be parsed by a parser. Like with the WFS request, the XML request body is then sent to the WPS with the JavaScript `fetch` API and the JSON results are added to the map.
4.3 Presentation or client layer

Convex Hull

The **convex hull calculation** requires client-side data pre-processing followed by a two-step server-side process. Client-side pre-processing is necessary to assess the convex hull of a dataset that consists of different geometry types, for example points, linestrings and multipolygons. To that end, all vertices are in a first step extracted from all features to create a single multipoint feature. This is done with the `ol.feature.getGeometry().getCoordinates()` method. Multipolygons are beforehand split into their single parts with `ol.feature.getGeometry().getPolygons()`. From the resulting array of coordinates, a multipoint geometry `ol.geom.MultiPoint()` and, from that, a multipoint feature `ol.Feature()` is built. Finally, a GeoJSON object is created from the multipoint feature by using `ol.format.GeoJSON().writeFeatureObject()`. After the client-side pre-processing, the server-side processing starts. In the first step of the server-side operation, this multipoint JSON object is converted to a feature collection via the WPS function `gs:Feature`. Consequently, this feature collection is used as input for the second step of the server-side operation. The calculation of the Convex Hull is done via the WPS function `statistics:FeatureToConvexHull`.

Spatial Frequency

The **spatial frequency calculation** is done by creating GeoJSON feature collections from map features and administrative boundary features. Therefore, only map features with a point geometry are chosen from the selected dataset. Point and polygon features are then processed by the WPS function `statistics:PointStatistics`. This function returns the input polygon features in JSON format with an additional property ‘count’ carrying the number of point features within each polygon feature. This ‘count’ property is used for creating a colour symbology for the polygon features. To that end, the minimum and maximum count values are assessed. An OpenLayers style function then translates the range of count values to a linear scale of opacity values ranging from 0.2 to 0.8.

Calculation of the **mean centre feature** is done fully client-side by first assessing the rectangular extent of the feature geometries (minx, miny, maxx, maxy) and then determining the centre coordinate of the extent by using the `ol.extent.getCenter()`
function. Last, a new `ol.feature` is created from the centre coordinates and added to the map.

**Distance analysis**

**Distance measurements** are also done fully client-side. On initiating the distance assessment, an event listener is added to the map that will collect reference points from the map features that have been chosen by mouse click. For point features, the coordinates of the feature proper are used with `ol.feature.getGeometry().getCoordinates()`. For line features, the geographical centre of the feature is assessed with `ol.feature.getGeometry().getExtent()` and `ol.extent.getCenter()`. Last, for multipolygon features, the geographical centre of the greatest contiguous landmass is calculated. New `ol.feature` map features are created from the coordinates of the reference points of the two selected features and added to the map. Using the centre point of the geographical extent may cause the reference point to lie outside of the map features.

Additionally, distance lines connecting the two selected reference points are added to the map. The distance lines are constructed with the help of the Mapbox Arc.js API. To that end, the projected coordinates of the reference points are in a first step transformed into geographical coordinates with `ol.proj.toLonLat()` and then used as start and end points for a new `arc.GreatCircle().Arc()`. For displaying the distance line in the map, the line coordinates are transformed back to Web Mercator with `ol.geom.LineString().transform()`.

The distance measurement does not only calculate the distance between two selected features, but also all distances within the currently selected dataset. To obtain a distance matrix, the singular pairwise distances between each point and every other point in the subset are calculated. All distance calculations are done client-side in JavaScript and OpenLayers. To approximate geodesic line distances, a sphere `ol.Sphere` with a radius of 6378137m, equal to the semi-major axis of the WGS84 ellipsoid, is used as reference geometry. (“OpenLayers - API,” 2018) Consequently, line distances are assessed by using the haversine formula on pairs of geographical coordinates with
4.3 Presentation or client layer

ol.Sphere.haversineDistance(). The haversine formula is used to determine the great-circle distance between two points using longitude and latitude. ("Haversine formula," 2018)

In order to compare the distance measured between the two selected locations to the overall distribution of distances within the subset (the distance matrix), the following visualisation technique was used. First, statistical key figures (min/max and mean value, quartiles) are calculated in JavaScript for the distance dataset. Then, a boxplot is created from the figures by leveraging the D3.js API. To that end, the range of distance values is translated to a linear scale with d3.scaleLinear() utilising the width of the chart panel. The position of the box ends and whiskers is determined by using the min/max and quartile distance values and transforming them with the linear scale. Additionally, a circle feature visualising the currently measured distance is added on top of the boxplot.

Charts

For creating the charts, map features of the current selection are used as input in JSON format. All operations for creating the charts are done client-side. First, the JSON data needs to be parsed to obtain the desired value – frequency structure for the chart. To that end, the code loops through all properties of the JSON objects and increases the count by 1 if the respective property value has already been found in the dataset, otherwise starts the count for the new property value at 0. In parallel, all feature IDs carrying that specific property value are collected in an array. These will be used later. For creating the simple bar chart, the category selected by the user for creating the chart is derived from the selected index of the respective dropdown. For the co-occurrence matrix, both properties are derived from the respective dropdowns and an additional step is necessary to assess the frequency values for both properties. Similar to the procedure above, the code loops through the JSON objects and increases the count by 1 every time a singular combination of the two selected properties occurs.

The range of frequency values for the selected property is then translated to a linear scale with d3.scaleLinear() utilising the width of the chart panel. For creating the bars, one container element <g> is created for each property value within the SVG container by using D3.js data binding methods. For each property value entering the selection, a new <g> object
is added with `g.enter().append()` and for each property value leaving the selection, on `<g>` object is removed with `g.exit().remove()`. Within each `<g>` container, a rectangle is added to create the single bars. The width of each bar depends on the number of property values: the more property values, the narrower is each bar. The height of each bar represents the frequency value. Additionally, the property and frequency values are added as text to the bar chart.

The array of feature IDs collected during the parsing of the JSON features are used to allow for brushing in the visualisation. Clicking on a bar of the chart will determine all map features associated with the property value by querying the list of feature IDs associated with the property value. All map features with these feature IDs are then highlighted in the map by setting the attribute ‘selected’ of the corresponding map features to ‘true’ and thus applying the style for selected features from the OpenLayers style function. Additionally, the coordinates of the selected features are collected and the bounding extend is determined with `ol.extent.boundingExtent()`. This bounding extent is then used to enable zooming to the selected features with the link option ‘Zoom to selected’ in the bottom panel of the page.

Animation

An animation feature was implemented to represent the temporal sequence of events. The temporal sequence was obtained from the order in which the storylocation events occur in the texts. Therefore, the order (DiBiase et al., 1992) used for animating the map features does not represent the chronological sequence of events within each story (the timeline), but follows the narrative stream of the plot (the storyline). This includes interruptions by interjections, flashbacks, mentions of related places or events and the like. The classification of locational references in categories like ‘background story’, ‘action’ and ‘reference’ allows for filtering the dataset in a way that, for example, only background stories of one or more characters are represented in the animated sequence. If the data subset to be animated spans more than one Sherlock Holmes story, the animation will add the vector features story by story, following the chronological order of publication, in the internal order of
story location events (e.g. first all features belonging to STUD, published 1887, then all features belonging to SIGN, published 1890, and so on).

The intervals between the map events represent the dimensions *duration* (length of each scene) (DiBiase et al., 1992) as well as *frequency* (frame rate) (MacEachren, 2004). They are equally spaced and do not respect the density of action at a specific location. This means that the animation will pause at each map feature for an equal amount of time, irrespective of ‘how much’ happens in that scene at that location and of ‘how long’ the scene is described in the text. Thus, the intervals used for the time animation do neither represent the speed at which the story is told nor the speed in which the story takes place. (“Pace (narrative),” 2018)

The map animation is interactive and allows for determining the speed of the animation, i.e. setting the length of the intervals before starting the animation as well as for manipulating the sequence of animation events. The animation offers possibilities to pause and resume, forward, rewind and stop. Interactivity has been added to the animation feature to prevent overstraining the user. (Fabrikant, 2005)

As soon as the animation is started, the map features of the current selection are added one by one to the map. This is done by adding a timed function to the window object with `window.setInterval()` as soon as the start button is clicked. The timed function now calls the function for adding the map features in the beforehand determined interval. Each time a new feature is added to the map, visual highlighting is done by altering the symbology and applying the style for ‘new’ features from the OpenLayers style function. Additionally, the temporal sequence is represented in the appearance of the map features. Similarly to the processes described above, the position of each respective point within the selected subset is translated to a linear scale of lightness values. This results in ‘earlier’ features to appear darker and ‘later’ or more recent features to appear lighter. Upon adding a new feature to the map, a popup containing information on the feature appears automatically and the map view centres on the new feature.

With each feature that is added to the map during the animation, a line feature connecting the current feature with its predecessor will be added to the map, too. Like in the distance measurement, the Mapbox ARC.js API is used for creating the connecting lines. To that end,
coordinates of a reference point are extracted from each feature in the fashion described above for the distance measurement and transformed to a geographical SRS with \texttt{ol.proj.toLonLat()}. The great circle lines are created as above with \texttt{arc.GreatCircle().Arc()} and added as \texttt{ol.feature} map features to the map after being transformed back to Web Mercator with \texttt{ol.geom.LineString().transform()}.

These line features serve not only to illustrate the interconnections between map features and the sequence of places, but do also convey information on the frequency of any one interconnection between two locations. With every time that two features A and B appear as a direct sequence (both directions A to B and B to A), the connecting line increases in thickness. In doing so, the map resulting from the animation does not only represent a network of interconnected locations, but also a line cartogram illustrating the frequency of interconnections between each sequential pair of locations.

\textit{Symbology}

In addition to simple default symbologies for point, line and polygon features, a thematic symbology, representing the categories ‘certainty’, ‘reality’, ‘accuracy’ and ‘locationtype’ was implemented. Two major requirements were imposed on the thematic symbology: a) each category should be readable from a map feature individually and b) multiple categories should be readable from a map feature at the same time. To that end, one visual variable was used for each category according to its level of measurement (Bertin, 1983; Slocum et al., 2001): ‘certainty’ is an ordinal variable ranging from ‘poor’ over ‘medium’ to ‘good’. The measure is referring to a quality, so a colour range (variations in \textit{hue}) was used for displaying it in the map. ‘Accuracy’ is also an ordinal variable ranging from ‘low’ over ‘medium’ to ‘high’. Unlike ‘certainty’, ‘accuracy’ is a spatial characteristic. It was transferred to the map as a simplified approximation to a ‘perimeter of vagueness’ and therefore symbolized by \textit{size}. ‘Reality’ is a categorical measure and assumes in a very simplified manner the values ‘real’ or ‘fictional’. Fictionality was considered as a source of spatial uncertainty, therefore \textit{transparency} was used for representing it in the map. For ‘locationtype’, a categorical measure, different approaches have been chosen for each geometry type: \textit{shape} for points, \textit{arrangement} for lines and \textit{orientation} for polygon features (for the results see below section 5.2.3).
5 The application

5.1 Map control features

Tooltips

Each vector feature in the map delivers information on the mapped event in a popup appearing on click. It contains the name of the location as well as the location type, categories like certainty, reality and accuracy, as well as a short summary of events and a list of persons associated with the event. If several events overlap at the same location, the popups offer the possibility to page through the entry of each event in both directions by clicking the arrow buttons at the top of the popup (see Figure 19).

Zoom

The map allows for traditional zooming in and out by scrolling the mouse wheel or using the zoom in/zoom out buttons. Additionally, an extent zoom is implemented that can be used by holding down Ctrl and drawing a rectangle with the cursor. Likewise, a number of vector features in the map can be selected by holding down Shift and drawing a rectangle around the features. The selected features will then be highlighted visually and clicking “Zoom to selection” in the bottom panel will move the map view to the extent of the selected features. Moreover, several areas of interest have been chosen and provided to the map user as fixed view locations. For example, one click will let the user zoom in to the city of London, another to the USA or to India.

Figure 19: Popup showing information for 7 stacked map features.
5.1 Map control features

*Reset button*

The Reset button clears all layers from the map and deletes all temporal variables related to feature selection, extent, styling or analysis results. The map will be re-initialised and set back to the initial view.

5.2 Analysis and visualisation features

5.2.1 User-based dataset selection

The basis for all analysis and visualisation features in the web application is the user-based selection of a subset from all the data collected from the Sherlock Holmes stories. The application offers several possibilities to select a data subset.

*Filter by attributes*

First of all, the data can be filtered by four attributes including ‘story’, ‘country’, ‘reality’ and ‘type’. One value for each attribute can be selected from the respective dropdown menus (see Figure 20).

The dropdown menus are interdependent, which means that as soon as the user selects a value from one dropdown, all other dropdowns are updated so as to only hold values that intersect with the selected value from the first dropdown. With every selected value, the list of selectable values in the remaining dropdowns will be reduced.

*Figure 20: Control panel for requesting a data subset.*
5.2 Analysis and visualisation features

Filter by extent

The dataset can furthermore be limited geographically by a user-defined extent. To do so, a user can draw an extent on the map by holding down Shift and dragging an extent rectangle over the map. However, this will only work for point and line features.

Filter by geometry type

Not least, the user can choose a data subset by geometry type and may include for example only point and line features or only polygon features. By default, only point geometries are activated, but polygon and/or line features can be added, too.

5.2.2 Layer Selector

Basemaps

In the basemap section, users can switch between two different basemaps. The default basemap is the Stamen Terrain layer. If desired, the user can switch to OpenStreetMap as basemap.

Figure 21: Control panels for switching basemaps.

Administrative boundaries

Administrative boundaries of the United Kingdom and the United States can furthermore be added to the map as a reference layer, either as a means of orientation additional to the basemap, or to assess spatial frequency in the analysis tab (see below).

The administrative boundary layers are fetched from a WFS and can be requested for England, Scotland, Wales and London as well as for the United States.
States of America or the countries of the world. The selection for the world countries will always depend on the current subset of data. Thus, the world countries dataset will only contain boundary features of countries that are in the current data selection (see Figure 22).

5.2.3 Styling

The application offers the possibility to explore the selected dataset by qualitative categories such as ‘certainty’, ‘reality’, ‘accuracy’ and ‘locationtype’. To that end, a multifactorial symbology was used that enables to symbolize the map features according to any single attribute or any combination of the four attributes (see Figure 23).

Symbolization by ‘certainty’ will apply a colour scheme to the map features according to a traffic light system: green means high certainty, orange medium and red low certainty.

‘Reality’ is expressed by opacity: opaque features are almost certainly real locations, see-through features are more likely fictional locations.

Figure 23: Control panel for manipulating the map symbology according to the four attributes ‘certainty’, ‘reality’, ‘accuracy’ and ‘locationtype’.

Figure 24: Legend for the multifactorial symbology.
‘Accuracy’ determines the size of the map features and affects only point features. The higher the accuracy, the smaller the point feature.

Last, ‘locationtype’ is illustrated by the use of different symbols. For points, this includes different geometric shapes like circles, stars or triangles, for polygons different types of hatching and for lines dashes in different intervals (see Figure 24 and Figure 25).

By implementing this multifactorial approach to symbology, each attribute of the four categories can be read from each feature at all times. A small opaque star shape in green colour for example represents a point location with high certainty, high accuracy, reality type ‘real’ and location type ‘reference’.

5.2.4 Spatial analysis

Convex Hull

With this feature, the minimum bounding geometry of the selected dataset can be determined and added as vector feature to the map. The convex hull represents the spatial footprint of a story or selection and serves to illustrate the spatial extent. After selecting a subset of the data, the user can request calculation of the Convex Hull feature by clicking...
the ‘Calculate’ button. The Convex Hull layer can be switched on and off at will with the layer checkbox next to the Calculate button (see Figure 26).

**Mean Centre**

This functionality allows for calculating the geographical centre of the selected dataset. Similar to the Convex Hull functionality, the mean centre calculation is started with the ‘Calculate’ button and display of the resulting layer can be controlled by the layer checkbox (see Figure 26).

![Figure 26: Results of convex hull and mean centre analysis for a subset of all point features in The Hound of the Baskervilles.](image)

**Distance Analysis**

The distance analysis feature allows for assessing not only the distance between two features of a selected dataset, but determines also all other pairwise distances between features of the whole data subset (for details on the implementation see section 4.3).

Distance measurement starts with clicking the ‘Start’ button in the distance analysis section. A text box advises the user to select the first vector feature from the map by a mouse click. The selected feature is visually highlighted in the map by a specific symbol (a green star). After selecting the first feature, a text box further advises the user to select the
5.2 Analysis and visualisation features

second feature from the map. The second feature is also visually highlighted. Additionally, a line connecting the two selected locations is added to the map to visualise the distance that is to be assessed.

Once two features have been selected from the map, distance assessment is started by clicking the ‘Measure’ button. The distance between the two selected locations as well as all other pairwise distances are calculated now. The panel on the left of the page will show the results of the distance measurements in a table as well as visualised in a boxplot. The distance between the two currently selected locations is added as a visually highlighted reference point to the boxplot to allow for comparing that distance to all the other distances within the subset. Furthermore, the names of the two chosen locations and the measured distance in kms is added to the result panel (see below Figure 27).

![Figure 27: Result of distance measurement. The two points chosen for distance measurements are connected by a light green line. A summary of all pairwise distances within the dataset is given in the panel on the left.](image_url)

To assess more distances in the same subset, the ‘Start’ button can be used again to select two new features and the ‘Measure’ button to calculate the new distance. If the underlying
5.2 Analysis and visualisation features

dataset did not change, i.e. if the user did not make a new feature selection, the boxplot will remain unchanged in the results panel and only the purple reference point will move to represent the new distance measurement. If the user wishes to assess distances in a newly selected subset, the distance matrix and the boxplot will be calculated anew.

Spatial frequency – choropleth maps

To assess the spatial distribution of selected features in reference to an administrative boundary layer, the spatial frequency functionality can be used. As a prerequisite, a feature selection must have been made and an administrative boundary layer must have been added to the map. The map resulting from this operation is a choropleth map representing absolute frequency values. The more features lie within a boundary feature, the higher the opacity of the feature (see Figure 28).

![Figure 28: Result of spatial frequency assessment. The darker the colour of the polygon, the more storylocations lie within the boundaries. A popup reveals the number of contained locations on click.](image-url)
5.2.5 Animation

During the animation, each vector feature (including point, line and polygon features) of the selected subset is added sequentially to the map in the order of appearance. The temporal sequence is represented in the symbology of the map features. Vector features from the selected story subset as well as connecting lines will start at a very dark colour shade. While the animation progresses, these colour shades will lighten up gradually with each added feature until the very last feature will be added in the lightest colour shade. Thus, the sequence of events can be read from the resulting map even after the animation has finished (see Figure 29).

![Figure 29: End product of map animation: the animation sequence started with the darkest lines and ended with the lightest.](image)

Additionally, the frequency of interconnections between two points is visualised in the thickness of the connecting line: the thicker the line, the more frequent are direct interconnections between two points. The names of the connected locations as well as the frequency of the connection can be queried in a popup by clicking on the line feature (see Figure 30).
5.2 Analysis and visualisation features

User control

After selecting a subset of the data by using the filtering mechanisms on the first control page (see above), the animated sequence is controlled by the user with the help of a media control panel. Buttons for starting, stopping and pausing the animation as well as for going one step forward or back, are provided (see Figure 31).

Before starting the animation, the speed of the animated sequence can be chosen by the user. Speed settings will determine how long the animation will linger at each location before the next feature will be added to the map. Consequently, this means that the slower

---

Figure 30: Information on begin and end point as well as frequency of connection delivered on click.

Figure 31: Control panel for the map animation.
5.2 Analysis and visualisation features

the animation speed, the longer the info popup of each location that conveys a summary of the events taking place will be visible to the user. Speed settings have to be chosen before starting the animation and cannot be altered on-the-fly during the animation. The control buttons are described in more detail below:

The ‘Play’ button starts the animation. All vector features that are part of the selected subset will be added to the map one by one with the beforehand determined speed.

The ‘Forward’ button allows for going one step forward in the animation. Without awaiting the end of the interval determined by the speed control, the next feature will be added immediately upon button click and the map view will move to the next location. The forward button can be used repeatedly to go forward until the end of the animation.

The ‘Backward’ button allows for going one step back in the animation. The last feature that has been added to the map before hitting the button as well as the connecting line feature will be removed from the map. The map view will then move back to the map feature that has been added before the current feature. The ‘Backward’ button can be used repeatedly to go back until the beginning of the animation.

The ‘Pause’ button will pause the animation at the last feature that has been added to the map before hitting the button. The pause functionality can, for example, be used to stay longer at an interesting location and read the event summary more carefully. The animation can then be resumed by clicking the ‘Play’ button.

The ‘Stop’ button will stop the animation at the last feature that has been added to the map before hitting the button. The map will then zoom out to the extent of all features that have so far been added to the map. After hitting the ‘Stop’ button, the animation can be started anew from the beginning by clicking the ‘Play’ button.
5.2.6 Charts

The chart functionality offers a possibility for obtaining an overview of the frequency of specific property values within the dataset. The dropdown menu allows for selecting a category of interest, for example ‘Character’. In the chart that appears in the results panel on the left, all property values and their frequencies are illustrated (see Figure 32).

Upon clicking on a specific bar, for example the bar representing 24 occurrences of Henry Baskerville, the corresponding map features are visually highlighted and the user can zoom in to the selected features by clicking ‘zoom to selection’ in the bottom panel.

The ‘Charts’ menu does also offer a functionality for creating co-occurrence matrices. With this feature, the user can juxtapose two different aspects of the dataset and explore how they correlate. For example, by choosing ‘Country’ from the first dropdown and ‘Location role’ from the second, a co-occurrence matrix is created in the results panel. Values from the first dropdown are listed on the horizontal axis (x-axis) and values from the second dropdown on the vertical axis (y-axis). Due to restrictions in space, values are only given in abbreviated form in the matrix and the full value can be requested in a tooltip by clicking on the axis. In the example below (see Figure 33 below), there are 34 elements of type ‘action’ in the dataset and all of them are located in the United Kingdom.

Figure 32: Bar chart of character appearances within the selected subset.

Figure 33: Co-occurrence matrix of country and location role.
5.2 Analysis and visualisation features

Figure 33: Co-occurrence matrix juxtaposing 'location type' and 'country' for a subset of feature from The Adventure of Silver Blaze.
6 Summary and discussion

6.1 Strengths

The web application *A Study in Geospatial Deduction* which is the result of this master's project is the first extensive spatial compendium of Sherlock Holmes stories that offers more than single point mappings (see above section 1.8.5).

The application is highly interactive and, going beyond switching layers on and off (see above section 5.2.2), enables the user to choose his or her own subset of data according to various filtering mechanisms (see above section 5.2.1). In addition to providing an overview of the spatial framework of Sherlock Holmes by mapping single instances, the application also allows for exploring it in more detail by using analysis functions (see above section 5.2.4) and manipulating the symbology (see above section 5.2.3). The application combines multiple elements of the geovisualisation approach (see above section 1.6.1). These include map animations representing the time dimension of the plot locations which enable the user to view them one by one as if he or she was reading the text. To improve usability of the map animation, control features for setting the speed of the animation or pausing the animation have been added (see above section 5.2.5). Furthermore, graphical charts allow for statistically summarising the chosen subset in respect to one or more dimensions of the data (see above section 5.2.6).

The methodology is based on theoretical groundwork laid by literary scholars (see above section 1). Extracting the geographical references manually instead of using automated geoparsing (see above section 3.1) furthermore allowed for implementing an elaborate data model in the database (see above section 3.2).

In technological terms, the application is independent from the restrictions posed by proprietary software or servers. As the source code will be part of the end product and published together with the application, the requirements of the open source principles are met and benefits ensue (see above section 1.7.1).
6.1 Strengths

Moreover, the application is at any time extensible if more data from other Sherlock Holmes stories are added. Furthermore, the model is also applicable to other works of Arthur Conan Doyle and may even be used for works of other authors.

6.2 Weaknesses

With the benefits of manual extraction came also one major disadvantage which lay in the fact that it proved to be a very time consuming process. Due to this, the data collection had to be aborted before the desired endpoint (The Final Problem) could be reached. Therefore, the data basis underlying the application is not complete, but extensible.

Designing and developing the user interface widely without using proprietary or other templates and building blocks may have led to weaknesses in the functionality and usability of the user interface. The design is, for example, not responsive like LitLong: Edinburgh (Loxley et al., 2018), and cannot readily be used on mobile devices. In the cartographic domain, difficulties arose in finding suitable symbologies for visualising the categories ‘certainty’, ‘accuracy’, ‘reality’ and ‘locationtype’ (see above Figure 25). Hue was used for ‘certainty’, size for ‘accuracy’, transparency for ‘reality’ and shape/arrangement/orientation for ‘locationtype’ (see above section 4.3.2). These may negatively affect visibility and limit the user’s understanding.

Including polygon geometries also strained the application’s performance so that more or less crude simplifications had to be performed on the polygon features in QGIS before integrating them in the application. This led to trade-offs in the optical quality of the map representation and also to topological errors.

Figure 34: Crude polygon simplifications owed to performance reasons.
Not least, performance issues may also arise due to the fact that analysis functions like the convex hull, mean centre and spatial frequency assessment are done via online processing. Thus, performance is highly dependent on the user’s bandwidth. This problem is expected to aggravate if the data collection grows in size.

6.3 Outlook and opportunities

For future developments, the application design could be made responsive so that it can be used also on a mobile device. Furthermore, extending the user interface could be facilitated by employing specialised JavaScript libraries like Sencha ExtJS (“Sencha Ext JS,” 2015) or AngularJS (“AngularJS — Superheroic JavaScript MVW Framework,” n.d.). These libraries offer collections of components such as forms, menus, toolbars or panels and are very useful for creating user interfaces.

Conducting usability tests could further help to improve the interface design. (Haklay and Zafiri, 2008; Skarlatidou and Haklay, 2006; Slocum et al., 2001) By doing so, stumbling blocks that may inhibit the users’ interaction with the software and, consequently, hinder their getting familiar with the content could be identified. Future usability testing could, for example, be done with the help of questionnaires. Doing so would ensure that the intended purpose of the application, which is generating new knowledge in literary experts as well as laymen users, is met.

Cartographic representations which are currently restricted to simple feature mapping could be extended by experimental methods such as emotional maps. Yi-Fu Tuan has already noted the role that descriptions of gloomy weather or homely, run-down places play in the Sherlock Holmes stories. According to Tuan, Arthur Conan Doyle’s instrumentalisation of a cheerful or oppressing atmosphere thereby differs from that of Agatha Christie, also famous for her detective stories:

“In an Agatha Christie novel, for example, it is when we read about a charmingly thatched cottage surrounded by a white picket fence that we are on guard for some real horror to come. Conan Doyle himself was aware of this possibility, but he chose not to exploit it.” (Tuan, 1985)
6.3 Outlook and opportunities

Emotional mapping could prove very useful for visualising and exploring this interesting aspect of Sherlock Holmes.

Adding further information on each location as well as historical pictures or (more) illustrations from the original publications would further enhance user experience. Historical pictures and information are, for example, provided in Thomas Wheeler’s web map *The London of Sherlock Holmes*. (Wheeler, 2017) The ‘plotline maps’ in the *Literary Atlas of Wales* are good examples for literary web maps where important principles of *deep mapping* are integrated. (Anderson et al., n.d.) Adding *deep mapping* elements to *A Study in Geospatial Deduction* would not only improve user experience but also help to better meet the special requirements of literary geography in a cartographic sense.

Including literary scholars in a follow-up project as it was done, for example, for the *Literary Atlas of Europe* (see above section 1.8.4) (Piatti and Weber, 2018) or the *Literary Atlas of Wales* (see above section 1.8.2) (Anderson et al., n.d.), would furthermore serve to improve the application’s methodology. Consequently, these methodological improvements would help to better exhaust the potential of the web application as a tool for exploration and analysis.
7 References


Elementary, 2012.


Mr. Holmes, 2015.


Pace (narrative), 2018. . Wikipedia.


References


Sharp, W., 1904. Literary Geography. London.

Sherlock, 2010.


Web Map Service | OGC [WWW Document], n.d. URL  


Web Processing Service | OGC [WWW Document], n.d. URL  


