Summary

How do local people describe landscapes? This question is crucial for tackling many tasks of social relevance such as land use planning, natural resource management and crisis intervention. For all of these it is of crucial importance to have spatial information available, and in particular information that reflects how individuals conceptualize space, in order to make the appropriate decisions.

From a geographic perspective the relevance of the question as to how people describe landscape is additionally challenging since landscape descriptions are the source of numerous uncertainties. Most of these uncertainties are the result of human perception. For example, different actors may have different concepts of the same landscape and thus describe it using different words. This is particularly true for people from different cultures or language groups. Furthermore, the descriptions of the objects making up the landscape are also prone to uncertainties. Thus, for example, objects such as mountains, valleys, rivers and forests are difficult to define semantically and spatially. For instance, a mountain is neither a product of natural selection, nor is it purely artificial. The physical basis of a mountain, such as the rock from which it is formed, is natural, whereas the delineation of its extent from the earth’s surface is clearly a human, or artificial, product. Such uncertainties are often synonymously related to vagueness. We successfully deal with vagueness in everyday situations without any difficulty. Indeed, vagueness is inherent to natural language and a building block of successful communication. The statement “I spent the weekend in the mountains!” in a conversation would be unremarkable and the use of the vague concept mountain guarantees that the statement is not cluttered with irrelevant details. However, vagueness is a challenge if landscape descriptions are to be stored in a computer. Typical Geographic Information Systems are well suited for storing and analyzing precise information, with boundaries being sharp and attributes often having numeric values. Furthermore, it would not be standard practice to represent several versions of the same landscape object in order to capture vagueness in terms of variations in human perception.

Capturing information on how landscapes are described and the precise characterization of vagueness in such descriptions has long been a goal of geographic research. In ethnophysiography, for example, local people are asked to describe key landscape concepts. Such inquiries usually take place in the field, in the form of interviews or field walks. The interviewees are often indigenous people from ethnic groups distributed all over the globe. Ethnophysiographic research thus gathers information about landscape concepts at detailed local scales - at the obvious cost, however, of intensive efforts in the collection of the information and often limited spatial coverage.
In this thesis we aim to explore a new source of information for landscape descriptions and thereby address some of the limiting factors of ethnographic or field based approaches. We use written landscape descriptions contained in large compilations of digitized books. However, using geographic information from unstructured natural language sources requires us to firstly make the information explicit. Tools and approaches that are associated with this task are described in a number of disciplines, such as digital humanities, literary GIS, geographic information retrieval (GIR) and recent work with user generated content.

**Linking Landscape Descriptions to Spatial Footprints.** In a first step we aim to link some hundred volumes of text containing landscape descriptions to spatial footprints. The GIR literature offers a number of approaches for performing this task, mainly through recognizing and associating place names in text with geographic coordinates. However, landscape descriptions constitute a particular challenge to the state of the art in GIR, mainly because of the fine spatial granularity of the descriptions. Previous work in GIR has mainly focused on descriptions with place names referring to cities or communities. In order to process detailed descriptions, containing references to mountains, hills or other natural features, we introduced a new heuristic independent from the type of place name. We thus assume that particularities of place names that refer to geographic objects can be characterized using topographic information and that such information is useful for correctly recognizing and referencing place names in text. An evaluation of our heuristics shows that our final product, consisting of the spatial footprints of some 10,000 landscape descriptions, is significantly more precise compared to a state of the art baseline. Additionally, we applied our results to a spatial information retrieval task and compared it with traditional information retrieval, such as for instance performed by commercial search engines. We can thus show that for the retrieval of relevant results from detailed spatial information and for detailed queries it is crucial to use geographic intelligence. State of the art information retrieval cannot sufficiently cope with this task.

A second product from the linking of landscape descriptions to spatial footprints is a map that represents the spatial distribution and the focus of some hundred books. Literary GIS argues that such maps are an important addition to traditional close reading, since they offers insights on the content of books that cannot be reached through a close reading. Thus we can, for instance, show how the spatial footprints of landscape descriptions have changed over the last 150 years.

**Landscape Information.** In a second step we move from georeferenced landscape descriptions towards the extraction and storage of explicit landscape information. Landscape information is approximated from particular uses of geographic objects in descriptions. The recognition of geographic objects in text is
guaranteed through a preprocessing step, where a group of volunteers annotated some 1500 frequent
nouns from descriptions for filtering out geographic objects according to a set of annotation rules. Thus,
we retained a set of 94 geographic objects. The (relative) frequencies of the use of these geographic
objects in descriptions are taken as a proxy for deducing local landscape information. This methodology
for extracting and storing landscape information allows us to capture some of the vagueness in landscape
descriptions. Landscape information gathered from different landscapes can either be qualitatively or
quantitatively compared. Qualitative comparisons focus on the use of geographic objects, whereas in
quantitative comparisons numeric values from the frequency distribution of geographic objects are used
to apply statistics.

The work in this thesis is associated with contributions that relate to different scientific domains. The new
approach for linking landscape descriptions to spatial footprints can be considered a methodological
contribution to GIR and literary GIS. Previous to our approach, this task was resolved with only limited
spatial precision or it was very time consuming. In the same context, we could show that for correctly
processing spatial queries of fine spatial resolution, a search engine necessarily needs to incorporate
geographic information. This has never been shown before. A second methodological contribution is
represented by our approach for extracting and structuring geographic information from landscape
descriptions. This time the contribution is embedded in the context of compatibility of GIS for vague
human sourced information. We used the local landscape information in a series of applications and could
thus show that we contribute to the state of the art in ethnophysiographic research, in particular by
extending the spatial and temporal coverage. The retrieved landscape information is comprehensive
enough to be related to alternative sources information. We could thus show that landscape descriptions
are statistically related to local topographic characteristics. This could be relevant for local search
applications in the internet. Lacking local information could be approximated through local physical
measurements.