SEMANTICS FOR MOUNTAINEERING HISTORY

Since the 18th century mountaineering has become an activity that by historians often is viewed as playing a central role in the way that much of the world has imagined conquest, human achievement and the place of wilderness in social life. Mountains and discourses on mountains and mountaineering have gained a certain significance in Europe since that time. The discourse on mountaineering is very central to the activity itself, mountaineering does not exist at all without the discourses around it – even less without descriptions of the places and people it evolves around. The go!digital project Alpenwort (2014 - 2016) has turned the Austrian Alpine Club Magazine (Zeitschrift des Deutschen und Österreichischen Alpenvereins) into the linguistically annotated Alpenwort corpus. The goal of this project is to semantically enrich the Alpenwort corpus by identifying and tagging

1. places like mountains, regions, trails or huts

2. people like mountaineers, guides or scientists, and

3. first ascent events like the first ascent of Großvenediger (by Josef Schwab in 1841)

within the corpus. This enrichment would allow for questions relevant for mountaineering research, like a list of all people mentioned in relation with the Venedigergruppe up to the year 1914, a list of places described in the years 1945 to 1950 or a list of all first ascents in the Julischen Alpen. To reach this goal we will create vocabularies for places and people important for mountaineering history. For places we prepared a very significant source that will be made publicly available through this project. It is the photo index (Lichtbilderverzeichnis) of the ÖAV from the years 1927 to 1941, containing descriptions for approx. 32.000 photos of which 85 % depict places of some kind. Additionally, the internet is a great openly available source for people as well as places, e.g. Wikipedia. Links to other sources on the internet form a network of information that can be explored (i.e. Linked Open Data). Our vocabularies will integrate these sources and thus relate to the connected information on the internet. Vocabulary entries for specific people and places will be identified in the Alpenwort corpus and thus create a relation from the places and people on the internet to the articles describing them in the Alpenwort corpus. International standards and semantic web community technologies will be used to generate our data, which will be made publicly available for research and the interested public. The developed methodology will provide possibilities to enrich and query text corpora, which exceed the possibilities of traditional full text and even corpora searches. The created vocabularies and the semantically enriched Alpenwort corpus will provide sources for mountaineering research and further enrichment of other sources that will showcase the potential of Digital Humanities research.

The actual presentation introduces the project and shows the work done so far with the applied methodologies. The first step was to transform the photo index from its OCR version into a structured data format that can be loaded in a database or a triple store. This involved differentiating between the objects documented within the photo index and the creation of unique identifiers for them. In our case these are the indexes, photos described, the places and objects depicted, regions where these places are located or topics that objects have been assigned to. In parallel we collected Place Gazetteers from national mapping institutions (from Austria and Switzerland), the Austrian Alpine Club, "regional name surveys", and internet sources like geonames, wikidata, and open street map. These are integrated into one gazetteer and used to match placenames in the photo index and the Alpenwortkorpus to automatically identify places in these sources. We manually build up a gold standard for five articles of the corpus to test the matching results for Named Entity Recognition and

Named Entity Linking. To disambiguate placenames we want to use the geographic context of the article.

Results in a nutshell:

1. Semantically enriched Alpenwort corpus

2. Publication of ÖAV photo index

3. Mountaineering place name gazetteer

4. Mountaineering people name register

5. Methodology for named entity matching / co-referencing