Scientific Report

First name / Family name	Dirk Ahlers
Nationality	Germany
Name of the *Host Organisation*	NTNU
First Name / family name of the *Scientific Coordinator*	Kjetil Nørvåg
Period of the fellowship	01/04/2013 to 31/03/2014

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

My fellowship was carried out with Prof. Kjetil Nørvåg at the Data and Information Management research group of the Department of Computer and Information Science (IDI) at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. My work concerned mostly the fields of information retrieval and information management. My participation in the group included contacts with fellow researchers and faculty, regular meetings of the group, collaboration within the department, and contacts with visiting researchers and ERCIM postdocs. Furthermore, I could keep ties with previous workplaces, publish some joint work, and had the opportunity to present my work at seminars or as invited talks. Within the timeframe of my ERCIM fellowship, I have focused on concluding and publishing several open threads of previous work on geographic search engines and geospatial search for developing countries. Furthermore, I have started to develop work regarding quality indicators for gazetteer data, and have looked into some data available at NTNU. Both these topics are still ongoing, while the former has already yielded some published results.

II – PUBLICATIONS DURING YOUR FELLOWSHIP

[Ahl14a] *Applying Geographic Information Retrieval — An Experience Report on Developing Local Search for a Developing Country*
Dirk Ahlers, In: Datenbank-Spektrum, Vol 14-1, 2014.
Abstract: This article reports on the efforts to establish a research project on a geospatial search engine for the Latin American country of Honduras during the author's research stay at a local university. Honduras is an interesting example of the challenges for information and knowledge management in developing countries as it combines many of the issues that are otherwise encountered more isolated or with less impact. These include low Web coverage in a low-resource country with limited Web infrastructure, and generally, work in challenging circumstances. The specific focus on geospatial information uncovers further issues that need to be addressed, such as informal addressing schemes or landmark-oriented location references, broad or incorrect locations for places, or insufficient ground truth in databases. While the tangible results of the project stay behind the original goals, several interesting results were achieved, which are condensed here as an experience report. Extended version of [Ahl13a].

[Ahl14b] *Multi-source entity-based conflation for local search*
Dirk Ahlers, In: Revista Colombiana de Computación - RCC, Volumen 13-2. 2014.
to appear.

[Ahl13d] *Assessment of the Accuracy of GeoNames Gazetteer Data*
Dirk Ahlers, In: GIR'13 - 7th Workshop on Geographic Information Retrieval, 2013.
Abstract: Gazetteers are the basis of many geospatial applications and serve an important role to collect and make available knowledge about the physical world such as place names and their coordinates. GeoNames is one of the largest and most often used gazetteer and it is generally assumed to be of sufficient quality. In this paper, we examine the quality and accuracy of the data in more detail, triggered by some anomalies encountered during its use. We present a classification of inaccuracies ranging from grid patterns, imprecise coordinates, overlaps and repetitions as well as misclassifications and visualize these for a range of countries. We finally give an outlook into potential corrections.

[Ahl13c] *Where the streets have no name - Experiences in GIR for a developing country*

Dirk Ahlers, In: GIR'13 - 7th Workshop on Geographic Information Retrieval, 2013.
Abstract: This paper gives a short overview of a project on Geographic Information Retrieval in developing countries in the form of an experience report based in Honduras. It provides some insights into encountered challenges of resource discovery, and georeferencing due to low Web coverage and informal location references as well as tested or proposed solutions to address them, including search via alternative means such as social networks.

[Ahl13b] *Towards a Development Process for Geospatial Information Retrieval and Search*
Dirk Ahlers, in: WWW 2013 - 22nd International World Wide Web Conference, 2013.
Abstract: Geospatial search as a special type of vertical search has specific requirements and challenges. While the general principle of resource discovery, extraction, indexing, and search holds, geospatial search systems are tailored to the specific use case at hand with many individual adaptations. In this short overview, we aim to collect and organize the main organizing principles for the multitude of challenges and adaptations to be considered within the development process to work towards a more formal description.

[Ahl13a] *In search of Honduras - Case report of developing local search for a developing country*
Dirk Ahlers, In: LWA 2013 - Lernen, Wissen, Adaption, Fachgruppe Information Retrieval der GI, Bamberg, Germany, 2013.
Abstract: This paper reports on the efforts to establish a research project on a geospatial search engine for the Latinamerican country of Honduras as well as establishing an encompassing research group on information retrieval during the author's stay there at a local university. Honduras is an interesting example of the challenges for information and knowledge management in developing countries as it combines many of the issues that might be encountered. These include low Web coverage in a low-resource country, cross-language information retrieval, and generally, work in challenging circumstances. The specific focus on geospatial information uncovers further issues that need to be addressed, such as informal addressing systems, broad or incorrect location references, or insufficient ground truth in databases.

[KAWB13] *Interactive Exploration of Geographic Regions with Web-based Keyword Distributions*
Chandan Kumar, Dirk Ahlers, Wilko Heuten, Susanne Boll, In: EuroHCIR 2013 - 3rd European Workshop on Human-Computer Interaction and Information Retrieval @ SIGIR2013, 2013.
Abstract: The most common and visible use of geographic information retrieval (GIR) today is the search for specific points of interest that serve an information need for places to visit. However, in some planning and decision making processes, the interest lies not in specific places, but rather in the makeup of a certain region. This may be for tourist purposes, to and a new place to live during relocation planning, or to learn more about a city in general. Geospatial Web pages contain rich spatial information content about the geo-located facilities that could characterize the atmosphere, composition, and spatial distribution of geographic regions. But the current means of Web-based GIR interfaces only support the sequential search of geo-located facilities and services individually, and limit the end users on abstracted view, analysis and comparison of urban areas. In this work we propose a system that abstracts from the places and instead generates the makeup of a region based on extracted keywords we and on the Web pages of the region. We can then use this textual fingerprint to identify and compare other suitable regions which exhibit a similar fingerprint. The developed interface allows the user to get a grid overview, but also to drill in and compare selected regions as well as adapt the list of ranked keywords.

[ASB13] *Geospatial Web Image Mining*
Dirk Ahlers, Philipp Sandhaus, Susanne Boll, In: Internet Multimedia Search and Mining, Hua, X., Worring, M., and Chua, T., Eds., Bentham Science, 2013.
Abstract: One commonly asked question when confronted with a photograph is "Where is this

place?" When talking about a place mentioned on the Web, the question arises "What does this place look like?" Today, these questions cannot reliably be answered for Web images as they typically do not reveal their relationship to an actual geographic position. Analysis of the keywords surrounding the images or the content of the images alone has not yet achieved results that would allow deriving a precise location information to select representative images. Photos that are reliably tagged with labels of place names or areas only cover a small fraction of available images and also remain at a keyword level. Results can be improved for arbitrary Web images by combining features from the Web page as image context and the images themselves as content. We propose a location-based search for Web images that allows finding images that are only implicitly related to a geographic position without having to rely on explicit tagging or metadata. Our spatial Web image search engine first crawls and identifies location-related information on Web pages to determine a geographic relation of the Web page, and then extends this geospatial reference further to assess an image's location.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

1. WWW 2013, 22nd International World Wide Web Conference, Rio de Janeiro, Brazil, May 13-17, 2013.
2. SIGIR 2013, 36th International ACM SIGIR conference on research and development in Information Retrieval, Dublin, Ireland - July 28 - August 01, 2013
3. LWA 2013, Learning, Knowledge, Adaptation, Bamberg, Germany, October 7-9, 2013.
4. ERCIM Annual Seminar, Athens, Greece, October 31-November 1, 2013.
5. SIGSPATIAL 2013, 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2013), Orlando, Florida, November 5-8, 2013.
6. GIR'13, 7th ACM SIGSPATIAL Workshop on Geographic Information Retrieval, ACM SIGSPATIAL Conference, Orlando, Florida, November 5, 2013.
7. Several seminars from NTNU IRS on Norwegian culture, language, work environment, Trondheim, Norway, 2013-2014.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

First REP: **Web Science Institute, University of Southampton**, Southampton, United Kingdom, 15.2.-4.3.2014. Host: Prof. Leslie Carr.
The interdisciplinary approach of the institute was a very good complement to my own work and served to identify possible future directions and collaborations.
Second REP: **IST - INESC-ID**, University of Lisbon, 6.3.-13.3.2014. Host: Bruno Martins.
This visit targeted the group that works in geographical information retrieval and thus was an excellent fit regarding my own work and mutual topics. It was very fruitful and served to develop and refine a joint proposal.
Additional research visit: **CWI, Amsterdam**, Netherlands, 13.3.-14.3.2014. Host: Arjen de Vries.
The visit served to strengthen previous ties and to get a better insight into the group and their work. Some overlapping fields of research could be identified.