Report on the Fourth Workshop on Location and the Web (LocWeb 2014)

Dirk Ahlers¹, Erik Wilde², and Bruno Martins³

 $^1 \rm NTNU-Norwegian~University~of~Science~and~Technology,~Trondheim,~Norway,$ <math display="block">dirk.ahlers@idi.ntnu.no $^2 \rm Siemens,~Berkeley,~USA,~dret@berkeley.edu$

³ULisboa, IST and INESC-ID, Lisbon, Portugal, bruno.q.martins@ist.utl.pt

Abstract

Location is a major aspect behind many approaches in Information Retrieval, naturally appearing in diverse forms and as a cross-cutting topic. To give the location topic an adequate venue, LocWeb 2014 restarted a workshop series at the intersection of geospatial search, information management, and Web architecture, with a main focus on location-aware information access. LocWeb has in the past been a 'travelling' workshop, gathering views from different communities, but always centered on the topic of location. As previously, contributions to the workshop reflected a multitude of fields that demand and use location features and geospatial information, and LocWeb 2014 featured presentations that look at the topic of location on the Web from an interdisciplinary perspective. Reflecting this interdisciplinarity, LocWeb 2014 consisted of an interesting variety of contributions from a range of topics: one keynote, four long and two short papers, and a discussion session. This report gives an overview of the workshop and summarizes its major contributions.

1 Introduction

Location has quickly moved from the next hot thing into being accepted as an important aspect of the Web and, especially, the mobile Web. Its importance is growing even more, as mobile access is surpassing other forms of Web usage and many players adopt a mobile-first strategy. Location also plays a role in the form of the explicit or implicit location of resources, locations described in content, location of users, location APIs, or mobile apps, also being used in geospatial-aware data mining or large-scale analytics. Location is thus a strong driver behind many recent innovations and research activities.

LocWeb, the Workshop on Location and the Web, has taken up location as its main topic, understanding it as a cross-cutting issue. LocWeb has been a 'travelling' workshop, having been held before at venues as diverse as WWW, CHI, and IoT. Following up on this background, the main objective of the workshop was to bring together a community of researchers at the intersection of location and the Web, for exchanging ideas and discussing and

developing the role of location. LocWeb 2014 took place on November 3rd 2014 in Shanghai, China, co-located with CIKM 2014, the 23rd ACM International Conference on Information and Knowledge Management, which was held at the Shanghai Regal International East Asia Hotel at coordinates 31.2018N, 121.4440E.

The LocWeb workshop series is aimed at establishing an integrated venue where the location aspect can be discussed in depth within an interested community. LocWeb follows the main theme of Location-Aware Information Access, with subtopics related to Search, Analytics, Mobility, Apps, Services, and Systems. The workshop is thus designed to reflect the multitude of fields that demand and utilize location features from an interdisciplinary perspective.

A short workshop introduction was available as a preparation for CIKM attendees [2]; we now aim to give a detailed report of the full workshop including its keynote and discussion session. Further details, including most of the presentations made at the workshop and previous editions, can be found on the workshop homepage¹.

2 Workshop Theme and Topics

LocWeb addresses the subject of location as a cross-cutting issue in Web research and technology, consequently examining location aspects in the domains of search, analytics, mobility, apps, services, and systems. Through LocWeb, we hope to further the integration of the geospatial dimension into the Web, by promoting research on the topic.

Our general aim is to establish a topic-specific venue where researchers from different fields and backgrounds, be it data mining, recommendation, search, systems, social media, applications, or standards, can discuss and develop the role of location. This can hopefully aid researchers in identifying common issues of geospatial information management across different fields and from different perspectives, and in learning about new approaches.

The main topics of interest, as announced in the call for papers for LocWeb 2014, are: Location-Aware Information Access, Geospatial Web Search, Location-Based Services, Geospatial Web Analytics, Geospatial Visual Analytics, Location-Aware Data Mining, Location-Aware Text Processing, Location-Based Entity Retrieval, Conflation, Merging, Integration, Location-Based Recommendation, Place Semantics, Lifecycle of Location Data, Modeling Location, Geo-Social Media and Systems, Location-Based Social Networks, Geo-Crowd-sourcing, Map-Based Interfaces and Geospatial HCIR, Geospatial Applications, Geospatial Awareness, Mobile Search and Recommendation, Mobile Apps and Mobile Context, Mobility Data, Location-Aware Web-Scale Systems, Large Ecosystems, and Location Standards.

3 Keynote

After a short introduction given by the organizers, the workshop started with a keynote by Vanessa Murdock from Microsoft, titled *Two Ways of Thinking About Where People Go* [8]. We partly paraphrase from her abstract. The keynote reflected on how classical studies in human mobility can still inspire current and highly-innovative research, particularly focusing on Ravenstein's Gravitational Model from 1885 [10] and on Stouffer's Intervening Opportunities Model from 1940 [11].

1http://dhere.de/locweb2014/

Both these authors wrote at a time when people were beginning to be mobile and affluent, a trend which nowadays continues in a way that they could not have imagined. They proposed contrasting models to explain human mobility, which are as fresh and relevant today as in the day they were originally published, due to the vast amount of data people are nowadays volunteering about where they go in their daily lives. There is an interesting aspect of granularity in the work, which was used to explain large migratory movements from rural to urban areas, yet it is general enough that it can also be applied to individual users and their local movements within a city. Vanessa Murdock presented her reflections on how these two models relate to recent work on the subject of local search and point of interest recommendation, arguing that the models naturally correspond to different reasons to why some particular places might be considered as more relevant than others (i.e., places may be considered as relevant because they possess specific properties, and their relevance is likely to decrease according to geospatial distance towards the user, and also according to the number of other relevant places that are located nearby). She argues that there is a spectrum of places ranging from merely convenient to very special. Following, the question of what makes places special to particular individuals assumes a key importance in the context of future directions for research in the area of local search.

In her presentation, she also highlighted particular engineering challenges related to data quality and large-scale information integration, given that the resources that are currently supporting Web-scale location-based services often result from the consolidation of multiple sources of heterogeneous data. Exemplarily, she demonstrated the difficulty of location detection in social media by an experiment with two native english speakers from different countries who could only reach an inter-assessor agreement of about 74% on location detection. These and other observations also fueled the discussion session.

4 Paper Presentations

The workshop accepted 6 papers, which were arranged in two sessions of two long and one short paper each. We had, on average, 3.125 reviews per paper and the overall acceptance rate was 75%. We had author groups from all over the world, namely from Australia, Germany, Japan, Taiwan, and the US. There were no cross-country authorships. Full papers had 25 minutes of presentation and 10 minutes time for discussion, while short papers were presented for 15 minutes with 5 minutes for questions, to make the workshop more interactive and facilitate discussion. The proceedings [1] are available in the ACM Digital Library².

The first session was titled *Indexing and Interfaces*, showing the first part of the broad range of topics that are currently dealing with location.

The first paper, titled Hybrid Quantized Resource Descriptions for Geospatial Source Selection and authored by Kufer and Henrich, took on the issue of representing geotagged media collections in peer-to-peer networks, so as to make them efficiently searchable. The challenge is to describe the geospatial footprint of media collections stored on one peer in a compact way, this way making querying within the distributed system efficient by selecting only promising sources. The authors proposed a hybrid approach that optimizes both selectivity and representation size to support the selection of those peers that potentially carry relevant results [5]. This work was also presented as a demo showcasing the different approaches that were implemented.

²http://dl.acm.org/citation.cfm?id=2663713

In Considering Common Data Model for Indoor Location-aware Services, Niu, Matsumoto, Saiki, and Nakamura describe their work towards the formal definition of data models for indoor positioning, abstracting from different implementations of indoor positioning systems to allow for easier generation of location-based services. This work aimed to achieve a looser coupling of components, and the authors have the long-term goal of establishing a more standardized way to model indoor locations and their relations [9].

In a short paper by Tytyk and Baldwin titled Automatic Zoom Level Prediction for Informal Location Descriptions, the authors considered the automatic selection of appropriate zoom levels in map interfaces, when presented with informal location descriptions. They presented a classification approach that uses multiple features of geospatial descriptions, evaluating it on different levels of noisy data [12].

The second paper session was titled *Processing and Understanding*, and formed the second half of the technical programme.

In Automatic Identification of Locative Expressions from Social Media Text: A Comparative Analysis, Liu, Vasardani, and Baldwin presented a comparative analysis over various geoparsers in the task of identifying locative expressions, benchmarking their performance on examples of the informal language that is commonly found in status updates in social networks. The obtained results indicate that, on informal content, the geoparsing accuracy is decidedly lower that what one would expect from the analysis of previous work on the area. The authors also discussed and analyze in detail the observed errors, as well as the process taken for the construction of the corpus supporting their experiments [7].

Chang, Fan, and Chen, in a paper titled On the Semantic Annotation of Daily Places: A Machine-Learning Approach, reported on the use of learning-based classification approaches to understand human movement data, inferring place semantics of visited locations based on temporal patterns associated to the location tracking of participants in a field study [3].

Finally, the short paper titled *HMM-based Address Parsing with Massive Synthetic Training Data Generation*, by Li, Kardes, Wang, and Sun, reported on the use of Hidden Markov Models to parse and segment addresses into their multiple components and exploring ways to generate training data along the way. This was a rare issue where the original data was too clean and training data had to be derived to add noise. Their system is used in a similarity function for record linkage [6].

5 Discussion Session

The discussion session was designed to uncover open issues in the field and lay a foundation for future work. Authors were asked to provide an overview of the open issues in their own work and bring it to the workshop as a starting point. The discussion showed a number of issues, some of which were of course focused on the individual work, and also yielded a fair number of general issues and themes.

The work presented at the workshop brings together a representative broad list of application scenarios. These include local search, indexing, location models, user support, geoparsing, and trajectories and place annotation.

The general discussions at the workshop show that there are major cultural, language, and geographical differences that have to be addressed within the location topic. For example, there is not yet a reliable unified geoparser available that would work both for broad place references and precise addresses in all countries. Individual research efforts happen for various

regions but are not yet integrated. On the other hand, commercial providers and search engines have developed their own address parsers for local search, but they are of unknown quality, with the perceived quality closely related to the perceived importance of regions to industry. This is closely related to imbalanced resources and tools for various regions are other biases, mostly population bias and technology bias, where populated regions and regions populated by technology-savvy people tend to be overrepresented in many datasources, not only in user-generated data or social media. Geoparsing and georeferencing is also an ongoing topic in TREC or the GIR workshop series [4]. We also saw that the location topic can benefit from a strong connection between different communities and for the future, also a stronger connection to W3C activities.

While we are dealing quite often with simple point-based locations, there are insufficient data models for larger places, unclear semantics of what exactly is located at the coordinates or how they were derived. In short, many places have an unknown extent, which is an issue of unknown granularity. We also see some concern towards the use of relative versus absolute positioning and a stronger demand towards location normalisation and standardisation.

One thing that we found confirmed in the papers and the discussion is the need for thorough work on error analysis and descriptions of limitations of current methods and tools. We encourage everyone working on these issues to embrace them as a good way to learn and to move the field forward.

6 Conclusion and Future Directions

The 2014 Workshop on Location and the Web had its main focus on location-aware information access. The workshop featured an interesting variety of contributions, with one keynote, four long papers, two short papers, and a final discussion session. Presenters and attendees, both from industry and academia, debated current important aspects in the field, as well as open issues and future research directions. Overall, we have seen a lot of interesting open issues. Few of them are easy, most are actually very challenging. It will be our pleasure to continue to work on them in the future.

The next edition of LocWeb, The 5th International Workshop on Location and the Web³, will be held in May 2015 at the 24th World Wide Web Conference, in Florence, Italy.

Acknowledgements

We would like to thank the authors of submitted papers for their efforts, and all attendees for making LocWeb 2014 an interesting, interactive and diverse workshop. A special thanks goes to Vanessa Murdock, for giving a very inspiring keynote. Our thanks especially include the members of the program committee for their work in reviewing the contributions:

Susanne Boll, University of Oldenburg, Germany; Keith Cheverst, Lancaster University, UK; Karen Church, Yahoo! Research, USA; Max Egenhofer, University of Maine, USA; Claudia Hauff, Delft University, Netherlands; Andreas Henrich, University of Bamberg, Germany; Christopher Jones, Cardiff University, UK; Eric Kansa, Alexandria Archive Institute, USA; Carsten Kessler, CUNY, USA; Mareike Kritzler, Siemens, USA; Chandan Kumar, OFFIS, Germany; Vanessa Murdock, Bing, USA; Kjetil Nørvåg, NTNU, Norway; Clodoveu Davis

³http://dhere.de/locweb2015/

Jr., Federal University of Minas Gerais, Brazil; Adrian Popescu, CEA LIST, France; Ross Purves, Universität Zürich, Switzerland; Daniele Quercia, Yahoo!Labs, Barcelona, Spain; Martin Raubal, ETH Zürich, Switzerland; Massimiliano Ruocco, NTNU, Norway; Johannes Schöning, Hasselt University, Belgium; Steven Schockaert, Cardiff University; Rainer Simon, AIT Austrian Institute for Technology; Arjen de Vries, CWI, Netherlands; Xing Xie, Microsoft Research Asia, China.

We are also grateful to the organizers of CIKM for hosting the workshop and to ACM and Sheridan for their support. Dirk Ahlers would like to thank the ERCIM Fellowship Program and the NTNU IDI travel fund for support. Bruno Martins acknowledges the financial support from FCT, through project grants with references EXPL/EEI-ESS/0427/2013 (KD-LBSN), EXCL/EEI-ESS/0257/2012 (DATASTORM) and PEst-OE/EEI/LA0021/2013 (INESC-ID's multi-annual funding).

References

- [1] D. Ahlers, E. Wilde, and B. Martins, editors. LocWeb '14: Proceedings of the 4th International Workshop on Location and the Web, CIKM 2014, New York, NY, USA, 2014. ACM.
- [2] D. Ahlers, E. Wilde, and B. Martins. LocWeb'14 4th International Workshop on Location and the Web: CIKM 2014 Workshop Summary. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2014. ACM.
- [3] C.-W. Chang, Y.-C. Fan, and A. Chen. On the Semantic Annotation of Daily Places: A Machine-Learning Approach. In [1].
- [4] C. B. Jones and R. S. Purves. GIR'13 Workshop Report: 7th ACM SIGSPATIAL Workshop on Geographic Information Retrieval. SIGSPATIAL Special, 6(1):14–14, 2014.
- [5] S. Kufer and A. Henrich. Hybrid Quantized Resource Descriptions for Geospatial Source Selection. In [1].
- [6] X. Li, H. Kardes, X. Wang, and A. Sun. HMM-based Address Parsing with Massive Synthetic Training Data Generation. In [1]. Short Paper.
- [7] F. Liu, M. Vasardani, and T. Baldwin. Automatic Identification of Locative Expressions from Social Media Text: A Comparative Analysis. In [1].
- [8] V. Murdock. Two ways of thinking about where people go. In [1]. Keynote.
- [9] L. Niu, S. Matsumoto, S. Saiki, and M. Nakamura. Considering Common Data Model for Indoor Location-aware Services. In [1].
- [10] E. G. Ravenstein. The Laws of Migration. Journal of the Statistical Society of London, 48(2), 1885.
- [11] S. A. Stouffer. Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, 5(6), 1940.
- [12] I. Tytyk and T. Baldwin. Automatic Zoom Level Prediction for Informal Location Descriptions. In [1]. Short Paper.