

# 1st International Workshop on Multilingual Web Access (MWA 2015)

Ben Steichen  
Santa Clara University, USA  
*bsteichen@scu.edu*

Nicola Ferro  
University of Padua, Italy  
*ferro@dei.unipd.it*

David Lewis  
Trinity College Dublin, Ireland  
*dave.lewis@cs.tcd.ie*

Ed H. Chi  
Google, USA  
*chi@acm.org*

## 1 Motivation and Goals of the MWA 2015 Workshop

Over the past 25 years, the World Wide Web has developed into a truly transnational information medium for users from across the globe. As of July 2013, Asia accounts for the largest share of online users in the world at 48%, followed by 22% from the Americas, and 19% from Europe. With this global development, the diversity of user languages on the Web has increased dramatically, leading to new challenges and opportunities for information access providers and consumers.

The Multilingual Web Access (MWA 2015)<sup>1</sup> workshop brought together researchers working on Cross-/Multilingual Search & Discovery, the Multilingual Social Web, as well as the Multilingual Semantic Web, in order to promote the exchange of complementary ideas and applicable/transferrable techniques between these areas. The goal of the workshop was to advance the current state of the art in Multilingual Web Access techniques, and, most importantly, to increase the adoption of multilingual techniques, methods, and tools in real-world Web applications.

## 2 Keynote Talk

*Roberto Navigli*, Sapienza University of Rome, Italy, discussed high-performance semantic analytics and interlinking on the Multilingual Web, and presented the details of the Babelnet and Babelfly projects. In his talk, Dr. Navigli showcased prototypes of novel applications such as multilingual analytics and semantic interlinking at the semantic level that they are developing at the Linguistic Computing Laboratory, Sapienza University of Rome. The applications leverage BabelNet 3.0, the largest multilingual semantic network, covering 14 million concepts and named entities and 271 languages, and Babelfly, a state-of-the-art joint multilingual word sense disambiguation and entity linking system.

---

<sup>1</sup><http://www.multilingualwebaccess.org/>

---

### 3 Submissions and Accepted Papers

The call for papers attracted submissions from Africa, Asia, Europe, and North America, and each submission was reviewed by three members of the workshop program committee. In total, 7 long papers and 1 short paper were accepted for presentation at the workshop (overall acceptance rate: 66%). These papers address a broad spectrum of pressing Multilingual Web Access issues, including topics such as Multilingual Web Search, Multilingual Social Networks, or accessibility issues on the Multilingual Web.

**Albano et al. [1]** proposes an innovative Word Sense Induction method based on multilingual data. Key to their approach is the idea that a multilingual context representation, where the context of the words is expanded by considering its translations in different languages, may improve the WSI results. The experiments showed a clear performance gain, and in their paper they give some preliminary ideas to exploit their multilingual Word Sense Induction method to Web search result clustering.

**Bloom et al. [2]** combines monolingual associative networks based on Wikipedia to create a larger, multilingual associative network, using the cross-lingual connections between Wikipedia articles. They prove that such multilingual associative networks perform better than monolingual associative networks in tasks related to document categorization by comparing the results of both types of associative network on a multilingual dataset.

**Brelstaff and Chessa [3]** describes a Multilingual Web prototype application that promotes appreciation of exceptional texts by non-native readers. The application allows dual column original/translation texts (in Open Office format) to be imported into the translator's browser, to be manually aligned for semantic correspondence, to be aligned with an audio reading, and then saved as HTML5 for subsequent presentation to non-native readers.

**Raghavi et al. [6]** faces Code-Mixing (CM), which is the embedding of linguistic units such as phrases, words, and morphemes of one language into an utterance of another language. They learn a basic Support Vector Machine (SVM) based question classification system for English-Hindi CM questions. Due to the inherent complexities involved in processing CM language and also the unavailability of language processing resources such as POS taggers, Chunkers, Parsers, they design their current system using only word-level resources such as language identification, transliteration and lexical translation. To reduce data sparsity and leverage resources available in a resource-rich language, instead of extracting features directly from the original CM words, they translate them into English and then perform featurization. They created an evaluation dataset for this task and their system achieves an accuracy of 63% and 45% in coarse-grained and fine-grained categories of the question taxonomy.

**Vahid et al. [4]** presents a set of experiments, which compare the impact of applying two of the best known online systems, Google and Bing translation, for query translation across multiple language pairs and for two very different CLIR tasks. Their experiments show that the MT systems perform differently on average for different tasks and language pairs, but more significantly for

---

different individual queries. They examine the differing translation behaviour of these tools and seek to draw conclusions in terms of their suitability for use in different settings.

**Papalexakis and Doğruöz [5]** using a novel method (tensor analysis), they reveal the social network structure of an online multilingual discussion forum which hosts an immigrant community in the Netherlands. In addition to the network structure, they automatically discover and categorize monolingual and bilingual sub-communities and track their formation, evolution and dissolution over a long period of time.

**Rodríguez Vázquez [7]** although research efforts made until now have led to a better understanding of visually-impaired users' browsing behavior and, hence, the definition of web design best practices for an improved user experience by this population group, the particularities of websites with multiple language versions have been mostly overlooked. They seek to shed light on the major challenges faced by visually-impaired users when accessing the multilingual web, as well as on why and how the web localization community should contribute to a more accessible web for all.

**Rózsa et al. [8]** searchers whose native language is not English may have to resort to queries in English in support of their information needs due to the lack of, or low quality of, the web content in their own language. However, when searching for information in a foreign language, users face a unique set of challenges that are not present for native language searching. They study this problem through qualitative research methods and report results from focus groups in this paper. The results reported in the paper describe typical problems foreign language searchers face, the differences in information seeking behavior in English and in the participants' native language, and advice and ideas shared by the focus group participants about how to search effectively and efficiently in English.

## Acknowledgments

Putting together MWA 2015 was a team effort. We first thank the authors for providing the content of the program. We are equally grateful to the program committee, who worked very hard in reviewing papers and providing very high-quality feedback for authors.

## References

- [1] Lorenzo Albano, Domenico Beneventano, and Sonia Bergamaschi. Multilingual word sense induction to improve web search result clustering. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 835–839, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [2] Niels Bloom, Mariët Theune, and Franciska De Jong. Document categorization using multilingual associative networks based on wikipedia. In *Proceedings of the 24th International Con-*

- 
- ference on World Wide Web*, WWW '15 Companion, pages 841–846, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [3] Gavin Brelstaff and Francesca Chessa. Exceptional texts on the multilingual web. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 847–851, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [4] Ali Hosseinzadeh Vahid, Piyush Arora, Qun Liu, and Gareth J.F. Jones. A comparative study of online translation services for cross language information retrieval. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 859–864, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [5] Evangelos Papalexakis and A. Seza Dođruöz. Understanding multilingual social networks in online immigrant communities. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 865–870, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [6] Khyathi Chandu Raghavi, Manoj Kumar Chinnakotla, and Manish Shrivastava. "answer ka type kya he?": Learning to classify questions in code-mixed language. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 853–858, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [7] Silvia Rodríguez Vázquez. Exploring current accessibility challenges in the multilingual web for visually-impaired users. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 871–873, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- [8] Gyöngyi Rózsa, Anita Komlodi, and Peng Chu. Online searching in english as a foreign language. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 875–880, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.