

# WWW 2008 Workshop: NLP2008 Summary

Hiroshi Nakagawa

Information Technology Center

The University of Tokyo

7-3-1 Hongou Bunkyo-ku,

Tokyo 113-0033, JAPAN

nakagawa@dl.itc.u-tokyo.ac.jp

Kentaro Torisawa

Japan Advanced Institute of

Science and Technology

1-1 Asahidai, Nomi,

Ishikawa 923-1292, JAPAN

torisawa@jaist.ac.jp

Masaru Kitsuregawa

Institute of Industrial Science

The University of Tokyo

4-6-1 Komaba Meguro-ku,

Tokyo 153-8505, JAPAN

kitsure@tkl.iis.u-tokyo.ac.jp

## ABSTRACT

The amount of information available on the Web has increased rapidly, reaching levels that few would ever have imagined possible. We live in what could be called the "information-explosion era," and this situation poses new problems for computer scientists. Users demand useful and reliable information from the Web in the shortest time possible, but the obstacles to fulfilling this demand are many including language barriers and the so-called "long tail." Even worse, users may provide only vague specifications of the information that they actually want, so that a more concrete specification must somehow be inferred by Web access tools. Natural language processing (NLP) is one of the key technologies for solving the above Web usability problems. Almost all the Web page provide with the essential information in the form of natural language texts, and the amount of these text information is huge. In order to offer solutions to these problems we must perform searching and extracting information from the Web texts using NLP technologies. The aim of this workshop: NLP Challenges in the Information Explosion Era (NLP2008) is to bring researchers and practitioners together in order to discuss our most pressing needs with respect to accessing information on the Web, and to discuss new ideas in NLP technologies that might offer viable solutions for those issues.

## Categories and Subject Descriptors

I.1.7 [DOCUMENT AND TEXT PROCESSING ] : Miscellaneous

**General Terms:** Algorithms, Experimentation, Languages, Performance, Theory.

**Keywords:** Web, Natural Language Processing

## 1. THEME AND TOPICS

As users' demands for the Web diversify, it is often the case that existing commercial search engines cannot provide useful information. This workshop focuses on diversified users' needs that have not been sufficiently satisfied by the existing search engines and other technologies, and discusses new ideas in NLP techniques that might offer viable solutions. Therefore, we invited submissions of original papers describing a wide range of applications of the NLP technologies to Web access. Possible topics are:

1) Cross-lingual and cross-cultural Web access using machine

translation, 2) Information extraction/fact retrieval from the Web, 3) Disambiguation of entities/facts on the Web, 4) Dialogue/Interactive/Exploratory systems for Web access using natural languages, 5) Automatic summarization of Web documents, 6) Combination/Discrimination of multiple document resources on the Web, 7) Trust/Credibility evaluation of Web documents using NLP, 8) Opinion extraction/Sentiment analysis on the Web, 9) Long tail/Minority/Serendipity search using NLP, 10) Entity identification on the Web e.g. people name search, 11) Knowledge Acquisition from the Web documents/query logs for effective Web access, 12) Effective search methods for mobile devices using NLP, 13) NLP-based blog analysis and 14) Other topics related to NLP and Web.

In particular, we solicited the papers that aim at fulfilling a novel type of needs in Web access and that can provide a new insight into future directions of Web access research.

## 2. ACCEPTED PAPERS

We received 15 submissions in total. They describe the original theories, methods, systems and applications. Since this is a half day workshop, we had to select six papers. The selected papers' title and authors are:

1. Refined and Incremental Centroid-based approach for Genre Categorization of Web pages. Chaker Jebari
2. Personal Name Resolution of Web People Search. Krisztian Balog, Leif Azopardi, Maarten De Rijke
3. Discriminative Dialog Analysis Using a Massive Collection of BBS comments. Eiji Aramaki, Takeshi Abekawa, Yohei Murakami, Akiyo Nadamoto
4. Building a Sentiment Summarizer for Local Service Reviews. Sasha Blair-Goldensohn, Ryan McDonald, George Reis, Tyler Neylon, Kerry Hannan, Jeff Reynar
5. Incorporate the Syntactic Knowledge in Opinion Mining in User-generated Content. Qiu, Guang
6. Towards Person Name Matching. Jakub Piskorski, Karol Wieloch

## 3. ACKNOWLEDGMENTS

We appreciate the following reviewers who did good job within very short reviewing time:

Andrew Borthwick, Julio Conzalo, Lee-Feng Chien, Atsushi Fujii, Kentaro Inui, Noriko Kando, Jussi Karlgren, Gary Geunbae Lee, Hang Li, Mun-Kew Leong, Reiner Kraft, Sadao Kurohashi, Tatsunori Mori, Massimo Poesio, Satoshi Sekine, Ralf Steinberger, Le Sun, Jun-ichi Tatemura, Takehito Utsuro and Kam-Fai Wong.

Finally our thanks to Atsushi Fujii who did the publicity task of this workshop.