



RECONNECT Workshop 2010

Extracting and Visualizing Information from Natural Language Text

Where: University of Southern California

When: June 6-12, 2010

About Reconnect: This CCICADA/VACCINE Summer Reconnect Workshop exposes faculty teaching undergraduates to the role of the mathematical sciences in homeland security and provides an opportunity to researchers in government or industry to learn about recent techniques in data analytics. Topics are presented in a weeklong series of lectures and activities; participants are involved in both research activities and in writing materials useful in the classroom or to share with their colleagues. Participants may develop materials for publication in either the CCICADA Technical Reports or the Educational Modules Series published by the DIMACS Center at Rutgers University.

Topic: Automatic identification and extraction of desired information from natural language text is increasingly used as a way to improve general purpose search and has a range of applications in medical informatics, business applications, and for the intelligence community. The input is one or more texts in the domain in question, and the output is a database containing just the desired fields of information, extracted from the source material and formatted appropriately. Information extraction (IE) techniques have been developed since the early 1980s, and include finite state technology, pattern-based extraction, and appropriate machine learning methods. This week-long workshop will take participants from the early, simpler, methods through the modern ones, and will include theoretical and practical topics as well as hands-on exercises using software packages. The material is very relevant to the undergraduate classroom and to many applications. The lecturers are renowned experts in the various aspects of IE and its visualization, and have a long history of giving informative, engaging, and fun lectures.

Organizers: Eduard Hovy, USC/ISI; Zornitsa Kozareva, USC/ISI; Midge Cozzens, CCICADA.

Speakers: Eduard Hovy, USC/ISI; Zornitsa Kozareva, USC/ISI; Dan Roth, UIUC; and John Stasko, GATech

Registration fees, lodging, meals and travel: *Academic participants:* registration, lodging and meals will be provided through DHS funding. *Government participants:* \$350. *For-Profit Corporation participants:* \$500 (includes all meals from Sunday dinner to Saturday lunch). Limited funds are expected to be available to waive part or the entire registration fee and/or provide partial support for travel.

Deadline for Applications is April 5, 2010 or until all slots are filled. Applications are to be submitted online at <http://ccicada.org/Reconnect/2010/ApplicationChecklistPage.html> and will be reviewed as they are received.

For more information: Christine Spassione (spassion@dimacs.rutgers.edu) or visit the Reconnect web page <http://ccicada.org/Reconnect/2010/>.

(This Reconnect Conference takes place in Los Angeles the week immediately after the Human Language Technologies (HLT-NAACL) conference of the North American Association for Computational Linguistics (June 1-6, 2010). HLT-NAACL is one of the primary conferences in this field, and participants are encouraged to consider attending it as well. See <http://naaclhlt2010.isi.edu/>).