

Apologies if you have received multiple copies of this announcement

=====
=====
ACL 2010 NEWSLETTER NO. 3
(April 19, 2010)
=====
=====

=====
:: Important Dates
=====

Registration will open in early May. Cheap and guaranteed accommodation reservation deadline is May 20, 2010.

Tutorials: July 11, 2010
Main conference: July 12-14, 2010
Workshops: July 15-16, 2010

=====
:: Table of Contents
=====

1. ACL 2010 in Uppsala, Sweden
2. Organising Committee
3. Main Conference Papers
4. Student Research Workshop
5. Calls for Student Volunteers and Travel Grants
6. Abstracts of Tutorials
7. System Demonstrations
8. Workshops and Collocated Conference
9. Call for Exhibits
10. Recommended Accommodation
11. SAS Official Airline
12. Sponsorship
13. Newsletter No. 4

=====
:: 1. ACL 2010 in Uppsala, Sweden
=====

The 48th Annual Meeting of the Association for Computational Linguistics will be held in Uppsala, Sweden, July 11-16, 2010. The conference will be organized by the Department of Linguistics and Philology at Uppsala University.

The conference will take place at Uppsala University Campus in a genuine university environment, dating back as far as 1477. The city also holds a rich history, having for long periods been the political, religious and academic center of Sweden. The proximity to the capital, Stockholm, provides additional benefits as a potential site for arranging both pre- and post-conference tours, as well as for excursions or tourism during the conference. The city of Uppsala is easy to reach by plane, train or car.

=====
:: 2. Organising Committee
=====

General Chair: Jan Hajič
Program Chairs: Sandra Carberry and Stephen Clark
Local Arrangements Chair: Joakim Nivre
Workshop Chairs: Pushpak Bhattacharya and David Weir
Tutorial Chairs: Lluís Màrquez and Haifeng Wang (王海峰)
System Demonstrations Chair: Sandra Kübler
Student Research Workshop Chairs: Seniz Demir, Jan Raab, Nils Reiter, and Marketa Lopatkova and Tomek Strzalkowski (faculty advisors)
Publication Chairs: Jing-Shin Chang (張景新) and Philipp Koehn
Mentoring Service: Björn Gambäck and Diana McCarthy
Sponsorship Chairs: Stephen Pulman (Europe), Frédérique Segond (Europe), Srinivas Bangalore (Americas), Christy Doran (Americas), Hercules Dalianis (local), and Mats Wirén (local)
Publicity Chairs: Koenraad de Smedt and Beáta Megyesi
Exhibition Chair: Jörg Tiedemann
Local Arrangement Committee: Joakim Nivre (chair), Beáta Megyesi (vice chair), Rolf Carlson, Mats Dahllöf, Marco Kuhlmann, Mattias Nilsson, Markus Saers, Anna Sågvall Hein, Per Starbäck, Jörg Tiedemann, Oscar Täckström
Local Organizing Secretariat: Academic Conferences

=====
:: 3. Main Conference Papers
=====

Program Co-Chairs:

Sandra Carberry (University of Delaware, USA)
Stephen Clark (University of Cambridge, UK)

Important Dates:

Notification of acceptance: April 20, 2010
Camera ready papers due: May 12, 2010

<http://acl2010.org/papers.html>

Submission and formatting instructions: <http://acl2010.org/authors.html>

=====
:: 4. Student Research Workshop
=====

Student Research Workshop Co-Chairs:

Seniz Demir (University of Delaware, USA)
Jan Raab (Charles University, Prague, Czech Republic)
Nils Reiter (Heidelberg University, Germany)
Marketa Lopatkova (Faculty Advisor) (Charles University, Prague, Czech Republic)
Tomek Strzalkowski (Faculty Advisor) (State University of New York, Albany, USA)

Important Dates:

Notification of acceptance (postponed): April 20, 2010
Camera ready papers due: May 10, 2010

<http://acl2010.org/srw.html>

Funding for Travel

Funding is available to assist student participants with travel to Uppsala and conference expenses. Students whose papers are accepted for presentation at the workshop are eligible for travel assistance based on need. Further details will be provided along with the acceptance letters; however, in general a flat rate stipend will be awarded with the amount depending upon the student's location.

Sponsors:

Student Research Workshop is being generously supported by National Science Foundation, ACL Walker Student Fund, and The European Chapter of the ACL (EACL).

=====
:: 5. Calls for Student Volunteers and Travel Grants
=====

Student Travel Awards

Funding is available to assist student participants with travel to Uppsala and conference expenses - all students are welcome to apply. The applicants must agree to participate in the Student Volunteer Programme of ACL 2010.

Please note that the student travel grants are to subsidize student authors' participation in ACL, and won't be able to cover all the travel expenses.

Application deadline: May 15th, 2010
Notification of awards: June 10th, 2010

Student Volunteer Programme

ACL 2010 is looking for a limited number of student volunteers. In exchange for one full day's work, student volunteers receive free registration to the main conference (not to the workshops and tutorials). The work will be divided, probably into two half-day shifts, and the shifts will be scheduled to maximize volunteer access to the conference events. We may be able to provide other amenities, and will certainly try to provide a good work environment.

Tasks will include assisting at the registration desk, stuffing delegate packs, and providing technical assistance for conference events including tutorials, main conference and workshops.

Application deadline: May 15th, 2010
Notification of accepted student volunteers: June 10th, 2010

Application

Details on both assistance programs and the application form can be found on the conference website.

Student Travel Award Coordinator, ACL 2010: Marketa Lopatkova

=====
:: 6. Abstracts of Tutorials

=====

Tutorial Co-Chairs:

Lluís Màrquez (Technical University of Catalonia, Spain)

Haifeng Wang (王海峰) (Baidu.com Inc., China)

Tutorials will be held on Sunday, July 11, 2010 and the following tutorials will be offered:

T1: Annotation

Presenter: Eduard Hovy

Abstract:

As researchers seek to apply their machine learning algorithms to new problems, corpus annotation is increasingly gaining importance in the NLP community. But since the community currently has no general paradigm, no textbook that covers all the issues (though Wilcock's book published in Dec 2009 covers some basic ones very well), and no accepted standards, setting up and performing small-, medium-, and large-scale annotation projects remain somewhat of an art.

This tutorial is intended to provide the attendee with an in-depth look at the procedures, issues, and problems in corpus annotation, and highlights the pitfalls that the annotation manager should avoid. The tutorial first discusses why annotation is becoming increasingly relevant for NLP and how it fits into the generic NLP methodology of train-evaluate-apply. It then reviews currently available resources, services, and frameworks that support someone wishing to start an annotation project easily. This includes the QDAP annotation center, Amazon's Mechanical Turk, annotation facilities in GATE, and other resources such as UIMA. It then discusses the seven major open issues at the heart of annotation for which there are as yet no standard and fully satisfactory answers or methods. Each issue is described in detail and current practice is shown. The seven issues are: 1. How does one decide what specific phenomena to annotate? How does one adequately capture the theory behind the phenomenon/a and express it in simple annotation instructions? 2. How does one obtain a balanced corpus to annotate, and when is a corpus balanced (and representative)? 3. When hiring annotators, what characteristics are important? How does one ensure that they are adequately (but not over- or under-) trained? 4. How does one establish a simple, fast, and trustworthy annotation procedure? How and when does one apply measures to ensure that the procedure remains on track? How and where can active learning help? 5. What interface(s) are best for each type of problem, and what should one know to avoid? How can one ensure that the interfaces do not influence the annotation results? 6. How does one evaluate the results? What are the appropriate agreement measures? At which cutoff points should one redesign or re-do the annotations? 7. How should one formulate and store the results? When, and to whom, should one release the corpus? How should one report the annotation effort and results for best impact?

The notes include several pages of references and suggested readings.

Participants do not need special expertise in computation or linguistics.

T2: From Structured Prediction to Inverse Reinforcement Learning

Presenter: Hal Daumé III

Abstract:

Machine learning is all about making predictions; language is full of complex rich structure. Structured prediction marries these two. However, structured prediction isn't always enough: sometimes the world throws even more complex data at us, and we need reinforcement learning techniques. This tutorial is all about the *how* and the *why* of structured prediction and inverse reinforcement learning (aka inverse optimal control): participants should walk away comfortable that they could implement many structured prediction and IRL algorithms, and have a sense of which ones might work for which problems.

The first half of the tutorial will cover the "basics" of structured prediction the structured perceptron and Magerman's incremental parsing algorithm. It will then build up to more advanced algorithms that are shockingly reminiscent of these simple approaches: maximum margin techniques and search-based structured prediction.

The second half of the tutorial will ask the question: what happens when our standard assumptions about our data are violated? This is what leads us into the world of reinforcement learning (the basics of which we'll cover) and then to inverse reinforcement learning and inverse optimal control.

Throughout the tutorial, we will see examples ranging from simple (part of speech tagging, named entity recognition, etc.) through complex (parsing, machine translation).

The tutorial does not assume attendees know anything about structured prediction or reinforcement learning (though it will hopefully be interesting even to those who know some!), but *does* assume some knowledge of simple machine learning (eg., binary classification).

T3: Wide-Coverage NLP with Linguistically Expressive Grammars

Presenters: Josef van Genabith, Julia Hockenmaier and Yusuke Miyao

Abstract:

In recent years, there has been a lot of research on wide-coverage statistical natural language processing with linguistically expressive grammars such as Combinatory Categorical Grammars (CCG), Head-driven Phrase-Structure Grammars (HPSG), Lexical-Functional Grammars (LFG) and Tree-Adjoining Grammars (TAG). But although many young researchers in natural language processing are very well trained in machine learning and statistical methods, they often lack the necessary background to understand the linguistic motivation behind these formalisms. Furthermore, in many linguistics departments, syntax is still taught from a purely Chomskian perspective. Additionally, research on these formalisms often takes place within tightly-knit, formalism-specific subcommunities. It is therefore often difficult for outsiders as well as experts to grasp the commonalities of and differences between these formalisms.

This tutorial overviews basic ideas of TAG/CCG/LFG/HPSG, and provides attendees with a comparison of these formalisms from a linguistic and computational point of view. We start from stating the motivation behind using these expressive grammar formalisms for NLP, contrasting them with shallow formalisms like context-free grammars. We introduce a common set of examples illustrating various linguistic constructions that elude context-free grammars, and reuse them when introducing each formalism: bounded and unbounded non-local dependencies that arise through extraction and coordination, scrambling, mappings to meaning representations, etc. In the second half of the tutorial, we explain two key technologies for wide-coverage NLP with these grammar formalisms: grammar acquisition and parsing models. Finally, we show NLP

applications where these expressive grammar formalisms provide additional benefits.

Who are targeted:

- Researchers, developers and PhD students with a background in machine-learning and data-driven NLP who have not been exposed to linguistically expressive computational grammars.
- Researchers, developers and PhD students who have hand-crafted linguistically expressive computational grammars and who would like to get acquainted with state-of-the-art treebank-based acquisition of wide-coverage and robust linguistically expressive grammars.
- Researchers, developers and PhD students in data-driven parsing and generation who would like to get acquainted with efficient and scalable parsing and generation models for rich linguistically expressive grammars.

T4: Semantic Parsing: The Task, the State of the Art and the Future

Presenter: Rohit J. Kate and Yuk Wah Wong

Abstract:

Semantic parsing is the task of mapping natural language sentences into complete formal meaning representations which a computer can execute for some domain-specific application. This is a challenging task and is critical for developing computing systems that can understand and process natural language input, for example, a computing system that answers natural language queries about a database, or a robot that takes commands in natural language. While the importance of semantic parsing was realized a long time ago, it is only in the past few years that the state-of-the-art in semantic parsing has been significantly advanced with more accurate and robust semantic parser learners that use a variety of statistical learning methods. Semantic parsers have also been extended to work beyond a single sentence, for example, to use discourse contexts and to learn domain-specific language from perceptual contexts. Some of the future research directions of semantic parsing with potentially large impacts include mapping entire natural language documents into machine processable form to enable automated reasoning about them and to convert natural language web pages into machine processable representations for the Semantic Web to support automated high-end web applications.

This tutorial will introduce the semantic parsing task and will bring the audience up-to-date with the current research and state-of-the-art in semantic parsing. It will also provide insights about semantic parsing and how it relates to and differs from other natural language processing tasks. It will point out research challenges and some promising future directions for semantic parsing. The target audience will be of NLP researchers and practitioners but no prior knowledge of semantic parsing will be assumed.

T5: Tree-based and Forest-based Translation

Presenters: Yang Liu and Liang Huang

Abstract:

The past several years have witnessed rapid advances in syntax-based machine translation, which exploits natural language syntax to guide translation. Depending on the type of input, most of these efforts can be divided into two broad categories: (a) string-based systems whose input is a string, which is simultaneously parsed and translated by a synchronous grammar (Wu, 1997; Chiang,

2005; Galley et al., 2006), and (b) tree-based systems whose input is already a parse tree to be directly converted into a target tree or string (Lin, 2004; Ding and Palmer, 2005; Quirk et al., 2005; Liu et al., 2006; Huang et al., 2006).

Compared with their string-based counterparts, tree-based systems offer many attractive features: they are much faster in decoding (linear time vs. cubic time), do not require sophisticated binarization (Zhang et al., 2006), and can use separate grammars for parsing and translation (e.g. a context-free grammar for the former and a tree substitution grammar for the latter).

However, despite these advantages, most tree-based systems suffer from a major drawback: they only use 1-best parse trees to direct translation, which potentially introduces translation mistakes due to parsing errors (Quirk and Corston-Oliver, 2006). This situation becomes worse for resource-poor source languages without enough Treebank data to train a high-accuracy parser.

This problem can be alleviated elegantly by using packed forests (Huang, 2008), which encodes exponentially many parse trees in a polynomial space. Forest-based systems (Mi et al., 2008; Mi and Huang, 2008) thus take a packed forest instead of a parse tree as an input. In addition, packed forests could also be used for translation rule extraction, which helps alleviate the propagation of parsing errors into rule set. Forest-based translation can be regarded as a compromise between the string-based and tree-based methods, while combining the advantages of both: decoding is still fast, yet does not commit to a single parse. Surprisingly, translating a forest of millions of trees is even faster than translating 30 individual trees, and offers significantly better translation quality. This approach has since become a popular topic.

This tutorial surveys tree-based and forest-based translation methods. For each approach, we will discuss the two fundamental tasks: decoding, which performs the actual translation, and rule extraction, which learns translation rules from real-world data automatically. Finally, we will introduce some more recent developments to tree-based and forest-based translation, such as tree sequence based models, tree-to-tree models, joint parsing and translation, and faster decoding algorithms. We will conclude our talk by pointing out some directions for future work.

T6: Discourse Structure: Theory, Practice and Use
Presenters: Bonnie Webber, Markus Egg and Valia Kordoni

Abstract:

Discourse structure concerns the ways that discourses (monologic, dialogic and multi-party) are organised and those aspects of meaning that such organisation encodes. It is a potent influence on clause-level syntax, and the meaning it encodes is as essential to communication as that conveyed in a clause. Hence no modern language technology (LT) - information extraction, machine translation, opinion mining, or summarisation - can fully succeed without taking discourse structure into account. Attendees to this tutorial should gain insight into discourse structure (discourse relations; scope of attribution, modality and negation; centering; topic structure; dialogue moves and acts; macro-structure), its relevance for LT, and methods and resources that support its use. Our target audience are researchers and practitioners in LT (not necessarily discourse) who are interested in LT tasks that involve or could benefit from considering language and communication beyond the individual sentence.

<http://acl2010.org/tutorials.html>

=====
:: 7. System Demonstrations
=====

System Demonstrations Chair:

Sandra Kübler (Indiana University, USA)

Important Dates:

Notification of acceptance: April 12, 2010

Camera ready papers due: May 16, 2010

<http://acl2010.org/demos.html>

=====
:: 8. Workshops and Collocated Conference
=====

Workshop Co-Chairs:

Pushpak Bhattacharya (Indian Institute of Technology, Mumbai, India)

David Weir (University of Sussex, United Kingdom)

The following 13 workshops will be held at ACL 2010:

WS1: SemEval-2010: 5th International Workshop on Semantic Evaluations

Date: July 15-16

Chairs: Katrin Erk, Carlo Strapparava

WS2: Fifth Workshop on Statistical Machine Translation and MetricsMATR

Date: July 15-16

Chairs: Chris Callison-Burch, Philipp Koehn, Christof Monz, Kay Peterson

WS3: The 4th Linguistic Annotation Workshop (The LAW IV)

Date: July 15-16

Chairs: Nianwen Xue, Massimo Poesio

WS4: BioNLP2010

Date: July 15

Chairs: K. Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestic, Jun'ichi Tsujii, Bonnie Webber

WS5: Cognitive Modeling and Computational Linguistics

Date: July 15

Chairs: John Hale

WS6: NLP and Linguistics: Finding the Common Ground

Date: July 16

Chairs: Lori Levin, William Lewis, Fei Xia

WS7: 11th Meeting of ACL-SIGMORPHON

Date: July 15

Chairs: Jeffrey Heinz, Lynne Cahill and Richard Wicentowski

WS8: TextGraphs-5: Graph-based Methods for Natural Language Processing
Date: July 16
Chairs: Carmen Banea, Alessandro Moschitti, Swapna Somasundaran, Fabio Massimo Zanzotto

WS9: Named Entities Workshop (NEWS 2010)
Date: July 16
Chairs: A Kumaran and Haizhou Li

WS10: Applications of Tree Automata in Natural Language Processing
Date: July 16
Chairs: Frank Drewes, Marco Kuhlmann

WS11: Domain Adaptation for Natural Language Processing (DANLP)
Date: July 15
Chairs: Hal Daume III, Tejaswini Deoskar, David McClosky, Barbara Plank, Jörg Tiedemann

WS12: Companionable Dialogue Systems
Date: July 15
Chairs: Yorick Wilks, Morena Danieli, Björn Gambäck

WS13: GEMS-2010 Geometric Models of Natural Language Semantics
Date: July 16
Chairs: Roberto Basili and Marco Pennacchiotti

Collocated conference:

CoNLL-2010: The Fourteenth Conference on Computational Natural Language Learning
Date: July 15-16
Chairs: Mirella Lapata, Anoop Sarkar

=====
:: 9. Call for Exhibits
=====

Exhibition Chair:

Jörg Tiedemann (Uppsala University, Sweden)

If you have a commercial product or service of interest to the CL and NLP community, the ACL 2010 exhibits program is the perfect way to introduce it to potential customers. Possible application areas include: mobile communications, machine translation, the semantic web, language interfaces to robots, dialogue systems, CL publications and e-journals.

The ACL 2010 exhibits space will be located in the hall of the historic main building of Uppsala University which will be used for both presentations and coffee/lunch breaks, assuring excellent "foot traffic".

Note that the exhibits program is targeted primarily at commercial products, but we also hope to present many Publishers' exhibits.

On behalf of the ACL 2010 organising committee, we would like to invite you to be part of the ACL 2010 exhibits program for a small fee. For more information, see http://acl2010.org/call_exhibits.html. If you are interested, please, contact the Exhibition Chair Jörg Tiedemann: exhibits@acl2010.org.

=====
:: 10. Recommended Accommodation
=====

Reservation of accommodation is open. Book before May 20 to take advantage of the lower prices! <http://acl2010.org/accommodation.html>

Registration will open early May.

=====
:: 11. SAS Official Airline
=====

SAS is the official airline for ACL 2010 and offers you 10% conference and event discount on published fares (except 2% on lowest economy fares). For the SAS discount, you need the event code, which you can find on the conference website.

http://acl2010.org/official_airline.htm

=====
:: 12. Sponsorship
=====

ACL 2010 very gratefully acknowledges the following commitments in sponsorship:

Riksbankens Jubileumsfond (Bank of Sweden Tercentenary Foundation) - Platinum Sponsor
GSLT - Swedish National Graduate School of Language Technology - Gold Sponsor
Textkernel - Gold Sponsor
CELI - Language & Information Technology - Silver Sponsor
City of Uppsala - Silver Sponsor
Esteam - Silver Sponsor
Google - Silver Sponsor
Voice Provider - Silver Sponsor
Yahoo! Labs - Silver Sponsor
XEROX - Xerox Research Centre Europe - Bronze Sponsor and Student Fellowship Sponsor
IBM Research - Sponsor of Best Student Paper Award
Language Weaver - Conference Bag Sponsor
SICS - Swedish Institute of Computer Science - Local Student Fellowship Sponsor

Anyone interested in becoming a sponsor or finding out more about sponsorship opportunities is encouraged to contact the sponsorship chairs:
sponsorship@acl2010.org

Stephen Pulman (Europe)
Frederique Segond (Europe)
Srinivas Bangalore (Americas)
Christy Doran (Americas)
Hercules Dalianis (local)
Mats Wirén (local)

=====
:: 13. Newsletter No. 4
=====

ACL Newsletter No. 4 will be published in May with information about registration and conference program.