

Today's Topics:

1. new edition of Compendium of translation software (John Hutchins)
2. ACL-IJCNLP-09: Final Call for Tutorial Proposals, Singapore (Aug, 2nd 09) (Andy Way)
3. Vacancy POSTDOC (2 years), Statistical NLP (Khalil Simaan)
4. Looking for a POSTDOC position (ghita tijani)
5. New Book: "Narrative Information" (Gian Piero Zarri)
6. New Book: "Learning Machine Translation" (Cyril Goutte)

----- Forwarded message -----

From: "John Hutchins" <John_Hutchins39@hotmail.com>
To: <mt-list@eamt.org>
Date: Mon, 19 Jan 2009 11:21:06 -0000
Subject: [Mt-list] new edition of Compendium of translation software

Please note that the latest edition of the "Compendium of translation software" is now available at:

<http://www.hutchinsweb.me.uk/Compendium.htm>

Let me know of any changes, additions, etc. for the next edition,

John Hutchins

----- Forwarded message -----

From: Andy Way <away@computing.dcu.ie>
To: mt-list@eamt.org
Date: Mon, 19 Jan 2009 16:05:22 +0000
Subject: [Mt-list] ACL-IJCNLP-09: Final Call for Tutorial Proposals, Singapore (Aug, 2nd 09)
ACL-IJCNLP-09: Final Call for Tutorial Proposals

Proposals are invited for the Tutorial Program of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. The conference is to be held from August 2 to August 7, 2009. The tutorials will be held on Sunday, August 2.

Proposals for tutorials on all topics of computational linguistics and speech processing are sought. Especially encouraged are tutorials that educate the community about advancements in cross-disciplinary areas such as machine learning and cognitive approaches as they relate to computational linguistics, as well as tutorials that span multiple areas of human language processing.

Information on the tutorial instructor payment policy can be found at

http://aclweb.org/adminwiki/index.php?title=Policy_on_tutorial_teacher_payment

PLEASE NOTE: Remuneration for Tutorial presenters is fixed according to the above policy and does not cover registration fees for the main conference.

SUBMISSION DETAILS

Proposals for tutorials should contain:

1. A title and brief description of the tutorial content and its relevance to the ACL community (not more than 2 pages).
2. A brief outline of the tutorial structure showing that the tutorial's core content can be covered in a three-hour slot (including a coffee break). In exceptional cases six-hour tutorial slots are available as well.

3. The names, postal addresses, phone numbers, and email addresses of the tutorial instructors, including a one-paragraph statement of their research interests and areas of expertise.
4. A list of previous venues and approximate audience sizes, if the same or a similar tutorial has been given elsewhere; otherwise an estimate of the audience size.
5. A description of special requirements for technical equipment (e.g., internet access).

Proposals should be submitted by electronic mail, in plain ASCII text no later than February 2, 2009 to tutorials-acl09 "at" sussex "dot" ac "dot" uk

The subject line should be: "ACL-IJCNLP 2009: TUTORIAL PROPOSAL".

PLEASE NOTE:

- 1) Proposals will not be accepted by regular mail or fax.
- 2) You will receive an email confirmation from us that your proposal has been received. If you do not receive this confirmation 24 hours after sending the proposal, please contact us personally using both emails dianam "at" sussex "dot" ac "dot" uk and cqzong "at" gmail "dot" com

TUTORIAL SPEAKER RESPONSIBILITIES

Accepted tutorial speakers will be notified by March 2, 2009, and must then provide abstracts of their tutorials for inclusion in the conference registration material by April 24, 2009. The description should be in two formats: an ASCII version that can be included in email announcements and published on the conference web site, and a PDF version for inclusion in the electronic proceedings (detailed instructions will be given). Tutorial speakers must provide tutorial materials, at least containing copies of the course slides as well as a bibliography for the material covered in the tutorial, by June 7, 2009.

IMPORTANT DATES

Submission deadline for tutorial proposals: February 2, 2009

Notification of acceptance: March 2, 2009

Tutorial descriptions due: April 24, 2009

Tutorial course material due: June 7, 2009

Tutorial date: August 2, 2009

TUTORIALS CO-CHAIRS

Diana McCarthy, University of Sussex, UK

Chengqing Zong Institute of Automation, Chinese Academy of Sciences (CASIA), China

Please send inquiries concerning ACL-09 tutorials to tutorials-acl09 "at" sussex "dot" ac "dot" uk

----- Forwarded message -----

From: Khalil Simaan <k.simaan@uva.nl>
To: "mt-list@eamt.org" <mt-list@eamt.org>, "corpora@uib.no" <corpora@uib.no>, elsnets-list@elsnets.org
Date: Sun, 25 Jan 2009 14:45:58 +0100
Subject: [Mt-list] Vacancy POSTDOC (2 years), Statistical NLP

Call for applicants.

Detail to be found on: Vacancy postdoc researcher, 2 years, computational linguistics/NLP <http://staff.science.uva.nl/%7Esimaan/postdoc_adv.html>*

Vacancy

The Institute for Logic, Language and Computation <<http://www.illc.uva.nl>>, University of Amsterdam <<http://www.uva.nl>>, has a vacancy for a

*

Postdoctoral Researcher

1.0 fte (38h per week)

For both internal and external candidates

Summary

- * Position: POSTDOC researcher
- * Duration: 2 years (full time, 38hrs per week)
- * Salary (gross per month): minimum Euro 2379 and maximum Euro 3755 in the first year.
- * The Collective Employment Agreement of the Dutch Universities is applicable.
- * Last date for application: February 15, 2009.

<http://staff.science.uva.nl/%7Esimaan/postdoc_adv.html>

Project

Priors for the Estimation of Probabilistic Grammars from Incomplete Natural Language Data
VIDI project Sima'an [2007-2011]

Abstract: The project aims at exploring the utility of available prior knowledge in specifying the model (hypothesis space) and the constraints on the objective estimation function for unsupervised estimation of probabilistic grammars for Statistical Machine Translation and Statistical Parsing. This project addresses problems of unsupervised estimation in the face of different kinds of incomplete data (parallel and monolingual corpora):

1. A raw text corpus,
2. A partially annotated corpus possibly with negative examples, and
3. A raw in-domain corpus and an out-of-domain annotated corpus.

Corpus annotation is e.g., word-alignment, syntactic or semantic structure.

Example topics studied within the project (see project related publications <http://staff.science.uva.nl/%7Esimaan/VIDI_project/Pubs_VIDI.html>):

- * Statistical machine translation with hierarchical/syntactic structure, e.g., statistical estimation [Mylonakis and Sima'an 2007 and 2008], lexical syntactic structure [Hassan, Way and Sima'an 2007 and 2008]
<http://staff.science.uva.nl/%7Esimaan/VIDI_project/Pubs_VIDI.html>
- * Statistical machine translation with morphological structure -- Work in progress 2009,
- * Inducing hidden compositional structure for machine translation (possibly using syntactic structure) -- Work in progress 2009.

For more detail see: Webpage <http://staff.science.uva.nl/%7Esimaan/postdoc_adv.html>

Requirements

We seek a cooperative yet independent researcher who is passionate about empirical research, statistical methods and NLP/CL.

The following are the main requirements:

- * obtained a Ph.D. recently (or will obtain it in the very near future, in which case testemoney from the PhD thesis Supervisor is needed)
- * a strong background, experience in and affinity with the largest subset of the following areas:

1. Computational Linguistics/Natural Language Processing working with statistical methods
2. Machine Learning Applied to NLP with good knowledge of unsupervised statistical methods.
3. Statistical Machine Translation.
4. Unsupervised statistical methods for statistical Parsing.

Appointment

The appointment will be on a temporary basis for a maximum period of 2 years. Based on a full-time appointment (38 hours per week) the gross monthly salary will range from € 2379,- to, at the most, € 3755,- , in the first year depending on expertise and experience.

The Collective Employment Agreement of the Dutch Universities is applicable.

Application

To Dr Khalil Sima'an <<mailto:simaan@science.uva.nl?SUBJECT=VIDI-POSTDOC>> by email (k.simaan@uva.nl) with subject heading: VIDI-POSTDOC).

The following documents (in pdf):

1. Letter expressing interest
2. Curriculum Vitae
3. List of publications (possibly including a downloadable pdf of your thesis)
4. Three letters of reference sent directly by the referees to Dr Khalil Sima'an <<mailto:simaan@science.uva.nl?SUBJECT=VIDI-POSTDOC>>
5. Optional: Research plan (1 page)

Last date for application: February 15, 2009.

For inquiries contact

Dr Khalil Sima'an <<mailto:simaan@science.uva.nl?SUBJECT=VIDI-POSTDOC>>.
Language and Computation <<http://www.illc.uva.nl/Research/research06b.php>>
Computer Science <<http://www.science.uva.nl/english/home.cfm>>
University of Amsterdam <<http://www.uva.nl/>>

--

Khalil Sima'an

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam
Tel#: +31-20-525 6573
Fax#: +31-20-525 5206
<http://staff.science.uva.nl/~simaan>

Address:
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands

----- Forwarded message -----
From: ghita tijani <gamor@gwu.edu>
To: mt-list@eamt.org
Date: Sun, 25 Jan 2009 11:15:20 -0500
Subject: [Mt-list] Looking for a POSTDOC position

Dear all,
I am looking for a postdoctoral position in Boston area. I have recently completed a doctorate degree in Computer Science at the George Washington University (GWU). My dissertation is entitled "Enhanced English-Arabic Cross Language Information Retrieval".
My dissertation analyses different approaches used to deal with problems related to word-by-word translation using online dictionaries. I studied the different approaches used to enhance the effectiveness of Cross Lingual Information Retrieval (CLIR) systems. I designed and improved an English-Arabic CLIR system and implemented a novel transliteration model to find Arabic equivalents of English words that are deemed to be hard to translate or not have a counterpart in other languages such as proper names. Furthermore, I did a comparative analysis of this transliteration approach to an existing statistical transliteration model. Finally, I tested and enhanced query expansion using

WordNet.

In addition to the aforementioned research activity, I gained a broad view of the field of Computer Science by participating in large scale projects in the area of software engineering and compilers. My work was fully funded by various organizations and got raving reviews for both my professional as well as personal conduct.

If anyone know of any open positions, I will appreciate if you refer me or let me know. I will gladly provide my resume and any additional information upon request.

Regards,
Ghita Amor

----- Forwarded message -----

From: Gian Piero Zarri <zarri@noos.fr>
To: mt-list@eamt.org
Date: Tue, 27 Jan 2009 20:37:34 +0100
Subject: [Mt-list] New Book: "Narrative Information"
NEW BOOK: Apologies for multiple postings.

Gian Piero ZARRI

Representation and Management of Narrative Information, Theoretical Principles and Implementation
Series: Advanced Information and Knowledge Processing
2009, X, 302 p. 55 illus., Hardcover
ISBN: 978-1-84800-077-3
Springer-Verlag London
<http://www.springer.com/computer/artificial/book/978-1-84800-077-3>

A big amount of important, economically relevant information, is buried within the huge mass of multimedia documents that correspond to some form of 'narrative' description.

Due to the ubiquity of these narrative resources, representing in a general, accurate, and effective way their semantic content - i.e., their key 'meaning' - is then both conceptually relevant and economically important. In this book, we present the main properties of NKRL ('Narrative Knowledge Representation Language'), a language expressly designed for representing and managing, in a standardised way, the 'meaning' of complex multimedia narrative documents. NKRL is also a fully implemented environment that exists in two versions: a relational database-supported version and a file-oriented one. It constitutes probably the most complete and realistic effort realised so far to deal with the huge industrial potentialities of the narrative domain.

Written from a multidisciplinary perspective, this book not only supplies an exhaustive description of NKRL and of the associated knowledge representation principles, it also constitutes a source of reference for practitioners, researchers and graduates in domains that range over narrative theories, linguistics and computational linguistics, artificial intelligence, knowledge bases, information retrieval, and languages for the ontologies and the semantic web.

Contents:

- Narratology and NKRL.
- The notion of 'event' in an NKRL context.
- Knowledge representation and NKRL.
- Architecture of NKRL, the four 'components'.
- Second order structures.
- The semantic and ontological contents.
- Ontology of 'concepts' and ontology of 'events'.
- The query and inference procedures.
- Temporal information and indexing.
- High-level inference procedures.
- Technological enhancements and theoretical enhancements.
- Appendix A: NKRL software.
- Appendix B: Plural entities in NKRL.

Professional address of the author from February 1st, 2009:

Gian Piero Zarri
University Paris-Est - LISSI Laboratory
120-122, rue Paul Armangot
94400 Vitry-sur-Seine
France
Phone: 33-1-41807383
Fax: 33-1-41807369
Email: zarri@noos.fr, gian-piero.zarri@univ-paris12.fr

----- Forwarded message -----

From: Cyril Goutte <cyril.goutte@nrc-cnrc.gc.ca>
To: <mt-list@eamt.org>
Date: Thu, 29 Jan 2009 12:22:33 -0500
Subject: [Mt-list] New Book: "Learning Machine Translation"
Dear colleagues,

We would like to announce the publication of a new book, which should be relevant to people interested in the use of Machine Learning techniques for Statistical Machine Translation.

Learning Machine Translation

C. Goutte, N. Cancedda, M. Dymetman and G. Foster (eds)

Series: Neural Information Processing

ISBN-13: 978-0-262-07297-7

<http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11753>

http://www.amazon.com/Learning-Machine-Translation-Information-Processing/dp/0262072971/ref=sr_1_1?ie=UTF8&s=books&qid=1233175113&sr=8-1

The Internet gives us access to a wealth of information in languages we don't understand. The investigation of automated or semi-automated approaches to translation has become a thriving research field with enormous commercial potential. This volume investigates how machine learning techniques can improve statistical machine translation, currently at the forefront of research in the field.

The book looks first at enabling technologies: technologies that solve problems that are not machine translation proper but are linked closely to the development of a machine translation system. These include the acquisition of bilingual sentence-aligned data from comparable corpora, automatic construction of multilingual name dictionaries, and word alignment. The book then presents new or improved statistical machine translation techniques, including a discriminative training framework for leveraging syntactic information, the use of semi-supervised and kernel-based learning methods, and the combination of multiple machine translation outputs in order to improve overall translation quality.

Contributors: Srinivas Bangalore, Nicola Cancedda, Josep M. Crego, Marc Dymetman, Jakob Elming, George Foster, Jesús Giménez, Cyril Goutte, Nizar Habash, Gholamreza Haffari, Patrick Haffner, Hitoshi Isahara, Stephan Kanthak, Alexandre Klementiev, Gregor Leusch, Pierre Mahé, Lluís Màrquez, Evgeny Matusov, I. Dan Melamed, Ion Muslea, Hermann Ney, Bruno Pouliquen, Dan Roth, Anoop Sarkar, John Shawe-Taylor, Ralf Steinberger, Joseph Turian, Nicola Ueffing, Masao Utiyama, Zhuoran Wang, Benjamin Wellington, Kenji Yamada

Neural Information Processing series

About the Editors

Cyril Goutte is a researcher in the Interactive Language Technologies Group at the Canadian National Research Council's Institute for Information Technology.

Nicola Cancedda is a researcher in the Cross-Language Technologies Research Group at the Xerox Research Centre Europe.

Marc Dymetman is a researcher in the Cross-Language Technologies Research Group at the Xerox Research Centre Europe.

George Foster is a researcher in the Interactive Language Technologies Group at the Canadian National Research Council's Institute for Information Technology.

--

Cyril.Goutte@nrc-cnrc.gc.ca <http://iit-iti.nrc-cnrc.gc.ca/>
ILT/IIT/NRC - 283, Alexandre-Taché - Gatineau, QC J8X 3X7 - Canada