EACL 2012

Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra) at EACL-2012

Proceedings of the Workshop

ISBN 978-1-937284-19-0

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

Welcome to the Joint EACL Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra). This two-day workshop addresses two specific but related research problems in computational linguistics.

The ESIRMT event (1st day) aims at reducing the gap, both theoretical and practical, between information retrieval and machine translation research and applications. Although both fields have been already contributing to each other instrumentally, there is still too much work to be done in relation to solidly framing these two fields into a common ground of knowledge from both the procedural and paradigmatic perspectives.

The HyTra event (2nd day) aims at sharing ideas among researchers developing and applying statistical, example-based, or rule-based machine translation systems and who wish to enhance their systems with elements from the other approaches.

The joint workshop provides participants with the opportunity of discussing research related to technology integration and system combination strategies at both the general level of cross-language information access and the specific level of machine translation technologies.

This workshop has been supported by the Seventh Framework Programme of the European Comission through the T4ME (METANET) contract (grant agreement no.: 249119), through the TTC contract (grant agreement no.: 248005), through the Marie Curie HyghTra contract and by the Spanish Ministry of Economy and Competivity through the BUCEADOR project (TEC2009-14094-C04-01) and the Juan de la Cierva fellowship program.

We would like to thank all people who in one way or another helped in making this workshop a success. Our special thanks go to our plenary speakers, to the speakers of our invited project session, to our sponsors, to the participants of the panel discussion, to the members of the program committee who did an excellent job in reviewing the submitted papers, and to the EACL organizers, in particular the workshop general chairs Kristiina Jokinen and Alessandro Moschitti. Last but not least we would like to thank our authors and the participants of the workshop.

The Organizers
Avignon, France, April 2012

Organizers ESIRMT:

Marta R. Costa-jussà (Barcelona Media Innovation Center)
Patrik Lambert (University of Le Mans)
Rafael E. Banchs (Institute for Infocomm Research)

Organizers HyTra:

Reinhard Rapp (Universities of Mainz and Leeds)
Bogdan Babych (University of Leeds)
Kurt Eberle (Lingenio GmbH)
Tony Hartley (Toyohashi University of Technology and University of Leeds)
Serge Sharoff (University of Leeds)
Martin Thomas (University of Leeds)

Program Committee:

Jordi Atserias, Yahoo! Research , Barcelona, Spain
Sivaji Bandyopadhyay, Jadavpur University, Kolkata, India
Núria Bel, Universitat Pompeu Fabra, Barcelona, Spain
Pierrette Bouillon, ISSCO/TIM/ETI, University of Geneva, Switzerland
Chris Callison-Burch, Johns Hopkins University, Baltimore, USA
Michael Carl, Copenhagen Business School, Denmark
Oliver Culo, ICSI, University of California, Berkeley, USA
Andreas Eisele, Directorate-General for Translation, European Commission, Luxembourg
Marcello Federico, Fondazione Bruno Kessler, Trento, Italy
José A. R. Fonollosa, Universitat Politècnica de Catalunya, Barcelona, Spain
Mikel Forcada, University of Alicante, Spain
Alexander Fraser, Institute for Natural Language Processing (IMS), Stuttgart, Germany
Johanna Geiß, Lingenio GmbH, Heidelberg, Germany
Mireia Ginesti-Rosell, Lingenio GmbH, Heidelberg, Germany
Silvia Hansen-Schirra, FTSK, University of Mainz, Germany
Gareth Jones, Dublin City University, Ireland
Min-Yen Kan, National University of Singapore
Udo Kruschwitz, University of Essex, UK
Yanjun Ma, Baidu Inc. Beijing, China
Maite Melero, Barcelona Media Innovation Center, Barcelona, Spain
Haizhou Li, Institute for Infocomm Research, Singapore
Paul Schmidt, Institut for Applied Information Science, Saarbrücken, Germany
Uta Seewald-Heeg, Anhalt University of Applied Sciences, Köthen, Germany
Nasredine Semmar, CEA LIST, Fontenay-aux-Roses, France
Wade Shen, Massachusetts Institute of Technology, Cambridge, USA
Fabrizio Silvestri, Instituto de Scienza e Tecnologia del'Informazione, Pisa, Italy
Harold Somers, CNGL, Dublin City University, Ireland
Anders Søgaard, University of Copenhagen, Denmark
Jörg Tiedemann, University of Uppsala, Sweden
Zygmunt Vetulani, University of Poznan, Poland

Invited Speakers:

Christof Monz (University of Amsterdam)
Philipp Koehn (University of Edinburgh)

Speakers of the Invited Project Session:

Cristina Vertan
George Tambouratzis, Marina Vassiliou and Sokratis Sofianopoulos John Tinsley, Alexandru Ceausu and Jian Zhang
Svetla Koeva

Table of Contents

ESIRMT-HyTra Workshop Program

Monday 23rd

(9:00-9:30) Workshop Presentation

(9:30-10:30) Invited Talk Christof Monz

(10:30-11:00) Coffee break

(11:00-11:30) ESIRMT Morning Session

Semantic Web based Machine Translation
Bettina Harriehausen-Mühlbauer and Timm Heuss

(11:30-12:00)

Measuring Comparability of Documents in Non-Parallel Corpora for Efficient Extraction of (Semi-)Parallel Translation Equivalents
Fangzhong Su and Bogdan Babych

(12:00-12:30)

Full Machine Translation for Factoid Question Answering
Cristina España-Bonet and Pere R. Comas

(12:30-13:00)

An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation
Tze Yuang Chong, Rafael Banchs and Eng Siong Chng

Monday 23rd (continued)

(13:00-15:00) Lunch break

(15:00-15:30) ESIRMT Afternoon Session

Natural Language Descriptions of Visual Scenes Corpus Generation and Analysis
Muhammad Usman Ghani Khan, Rao Muhammad Adeel Nawab and Yoshihiko Gotoh

(15:30-16:00)

Combining EBMT, SMT, TM and IR Technologies for Quality and Scale
Sandipan Dandapat, Sara Morrissey, Andy Way and Josef van Genabith

(16:00-16:30) Coffee break

(16:30-17:00) Project session

Two approaches for integrating translation and retrieval in real applications
Cristina Vertan

(17:00-17:30)

PRESEMT: Pattern Recognition-based Statistically Enhanced MT
George Tambouratzis, Marina Vassiliou and Sokratis Sofianopoulos

(17:30-18:00)

PLUTO: Automated Solutions for Patent Translation
John Tinsley, Alexandru Ceausu and Jian Zhang

Monday 23rd (continued)

(18:00-18:30)

ATLAS - Human Language Technologies integrated within a Multilingual Web Content Management System
Svetla Koeva

Tuesday 24th

(9:00-9:30) HyTRA 1st Morning Session

(9:30-10:00)

Tree-based Hybrid Machine Translation
Andreas Søeborg Kirkedal

(10:00-10:30) Coffee break

(10:30-11:30) Invited Talk Philipp Koehn

(11:30-12:00) HyTRA 2nd Morning Session

Were the clocks striking or surprising? Using WSD to improve MT performance
Špela Vintar, Darja Fišer and Aljoša Vrščaj

(12:00-12:30)

Bootstrapping Method for Chunk Alignment in Phrase Based SMT
Santanu Pal and Sivaji Bandyopadhyay

Tuesday 24th (continued)

(12:30-13:00)

Design of a hybrid high quality machine translation system
Bogdan Babych, Kurt Eberle, Johanna Geiß, Mireia Ginestí-Rosell, Anthony Hartley, Reinhard Rapp, Serge Sharoff and Martin Thomas

(13:00-15:00) Lunch break

(15:00-15:30) Hytra Afternoon Session

Can Machine Learning Algorithms Improve Phrase Selection in Hybrid Machine Translation?
Christian Federmann

(15:30-16:00)

Linguistically-Augmented Bulgarian-to-English Statistical Machine Translation Model
Rui Wang, Petya Osenova and Kiril Simov

(16:30-17:00)

Using Sense-labeled Discourse Connectives for Statistical Machine Translation
Thomas Meyer and Andrei Popescu-Belis

(17:00-17:30) Coffee break

(17:30-18:30) Panel discussion