

MultiLing 2013

**MultiLing 2013: Multilingual Multi-document
Summarization**

Proceedings of the Workshop

August 9, 2013
Sofia, Bulgaria

Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

©2013 The Association for Computational Linguistics

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

ISBN 978-1-937284-66-4

Introduction

MultiLing 2013 is a community effort, a set of research tasks and a corresponding workshop which covers three subdomains of Natural Language Processing, focused on the multilingual aspect of summarization. Each domain is allocated a separate section of the workshop. The three domains are:

- Multilingual multi-document summarization: Summarization, especially from multiple documents, has received increasing attention during the last years. This is mostly due to the increasing volume and redundancy of available online information. Recently, more and more interest arises for methods that will be able to function on a variety of languages. Multilingual multi-document summarization is the domain that researches such methods and studies their requirements and intricacies.
- Multilingual summary evaluation: Summary evaluation has been an open question for several years, even though there exist methods that correlate well to human judgment, when called upon to compare systems. In the multilingual setting, it is not obvious that these methods will perform equally well to the English language setting. In fact, some preliminary results have shown that several problems may arise in the multilingual setting [1]. This section of the workshop aims to cover and discuss these research problems and corresponding solutions.
- Multilingual summarization data collection and exploitation: The collection of multi-lingual corpora for summarization and summarization evaluation offers a challenge in itself. This section of the workshop works towards well-defined practices for the collection of such data, as well as the implementation and use of community tools for the support of the collection process. Furthermore, this section will include a discussion on how we can maximize the effect of the generated corpora in favor of the scientific community.

References

- [1] Giannakopoulos, G., El-Haj, M., Favre, B., Litvak, M., Steinberger, J., and Varma, V. (2011). TAC2011 MultiLing Pilot Overview.

Organizers:

George Giannakopoulos, NCSR “Demokritos” (Greece)
Georgios Petasis, NCSR “Demokritos” (Greece)

Program Committee:

Hoa Trang Dang, NIST (USA)
Lucy Vanderwende, Microsoft (USA)
Horacio Saggion, Universitat Pompeu Fabra (Spain)
Vangelis Karkaletsis, NCSR “Demokritos” (Greece)
Karolina Owczarzak, Oracle (USA)
John Conroy, IDA Center for Computing Sciences (USA)
George Giannakopoulos, NCSR “Demokritos” (Greece)

Invited Speaker:

Dragomir R. Radev, University of Michigan, Ann Arbor (USA)

Table of Contents

<i>Multi-document multilingual summarization corpus preparation, Part 1: Arabic, English, Greek, Chinese, Romanian</i>	
Lei Li, Corina Forascu, Mahmoud El-Haj and George Giannakopoulos	1
<i>Multi-document multilingual summarization corpus preparation, Part 2: Czech, Hebrew and Spanish</i>	
Michael Elhadad, Sabino Miranda-Jiménez, Josef Steinberger and George Giannakopoulos	13
<i>Multi-document multilingual summarization and evaluation tracks in ACL 2013 MultiLing Workshop</i>	
George Giannakopoulos	20
<i>ACL 2013 MultiLing Pilot Overview</i>	
Jeff Kubina, John Conroy and Judith Schlesinger	29
<i>CIST System Report for ACL MultiLing 2013 – Track 1: Multilingual Multi-document Summarization</i>	
Lei Li, Wei Heng, Jia Yu, Yu Liu and Shuhong Wan	39
<i>Multilingual Multi-Document Summarization with POLY2</i>	
Marina Litvak and Natalia Vanetik	45
<i>The UWB Summariser at Multiling-2013</i>	
Josef Steinberger	50
<i>Multilingual Summarization: Dimensionality Reduction and a Step Towards Optimal Term Coverage</i>	
John Conroy, Sashka T. Davis, Jeff Kubina, Yi-Kai Liu, Dianne P. O’Leary and Judith D Schlesinger	55
<i>Using a Keyness Metric for Single and Multi Document Summarisation</i>	
Mahmoud El-Haj and Paul Rayson	64
<i>Multilingual summarization system based on analyzing the discourse structure at MultiLing 2013</i>	
Daniel Anechitei and Eugen Ignat	72
<i>Multilingual Single-Document Summarization with MUSE</i>	
Marina Litvak and Mark Last	77

Conference Program

Friday August 9, 2013

09:00 Introduction by the Organizers

Session 1: (09:05) Data Contribution and Exploitation

09:05 *Multi-document multilingual summarization corpus preparation, Part 1: Arabic, English, Greek, Chinese, Romanian*

Lei Li, Corina Forascu, Mahmoud El-Haj and George Giannakopoulos

09:35 *Multi-document multilingual summarization corpus preparation, Part 2: Czech, Hebrew and Spanish*

Michael Elhadad, Sabino Miranda-Jiménez, Josef Steinberger and George Giannakopoulos

Session 2: (10:05) Multi-document Summarization

10:05 *Multi-document multilingual summarization and evaluation tracks in ACL 2013 MultiLing Workshop*

George Giannakopoulos

10:30 Coffee break

11:00 *ACL 2013 MultiLing Pilot Overview*

Jeff Kubina, John Conroy and Judith Schlesinger

11:30 *CIST System Report for ACL MultiLing 2013 – Track 1: Multilingual Multi-document Summarization*

Lei Li, Wei Heng, Jia Yu, Yu Liu and Shuhong Wan

11:50 *Multilingual Multi-Document Summarization with POLY2*

Marina Litvak and Natalia Vanetik

12:10 *The UWB Summariser at Multiling-2013*

Josef Steinberger

12:30 Lunch break

14:00 Invited Talk: Natural Language Processing for Analyzing Collective Discourse, Dragomir R. Radev, University of Michigan

15:00 *Multilingual Summarization: Dimensionality Reduction and a Step Towards Optimal Term Coverage*

John Conroy, Sashka T. Davis, Jeff Kubina, Yi-Kai Liu, Dianne P. O’Leary and Judith D Schlesinger

Friday August 9, 2013 (continued)

15:30 Coffee break

16:00 *Using a Keyness Metric for Single and Multi Document Summarisation*
Mahmoud El-Haj and Paul Rayson

Session 3: (16:20) Single-document Summarization

16:20 *Multilingual summarization system based on analyzing the discourse structure at MultiL-
ing 2013*
Daniel Anechitei and Eugen Ignat

16:40 *Multilingual Single-Document Summarization with MUSE*
Marina Litvak and Mark Last

17:00 Closing Discussion