

MultiLing 2019

**SUMMARIZATION ACROSS LANGUAGES, GENRES  
AND SOURCES**

**Proceedings of the Workshop**

September 6, 2019

Varna, Bulgaria

Proceedings of the Workshop MultiLing 2019:  
Summarization Across Languages, Genres and Sources  
associated with RANLP 2019  
Varna, Bulgaria  
6 September 2019  
ISBN 978-954-452-058-8

Designed and Printed by INCOMA Ltd.  
Shoumen, BULGARIA

## Introduction

Welcome to the MultiLing 2019 Workshop, collocated with RANLP 2019. MultiLing focuses on summarization, especially in tasks related to multi-linguality, different genres and sources, as elaborated below.

**Summarization across languages, genres and sources:** Summarization has been receiving increasing attention during the last years. This is mostly due to the increasing volume and redundancy of available online information but also due to the user created content. Recently, more and more interest arises for methods that will be able to function on a variety of languages and across different types of content and genres (news, social media, transcripts). This topic of research has been mapped to different community tasks, covering different genres and source types: Multilingual single-document summarization [Kubina et al., 2013, Giannakopoulos et al., 2015]; news headline generation (new task in MultiLing 2019); financial narrative summarization (new task in MultiLing 2019, under a view of synergy and complementary to other workshops as in FNP 2018 at LREC'18 [El-Haj et al., 2018]); user-supplied comments summarization (OnForumS task [Kabadjov et al., 2015]); conversation transcripts summarization (see also [Favre et al., 2015]). The spectrum of the tasks covers many real settings, with varying intricacies, similarly to previous MultiLing endeavours [Giannakopoulos et al., 2011, Giannakopoulos, 2013, Elhadad et al., 2013, Giannakopoulos et al., 2015, Giannakopoulos et al., 2017].

**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, preliminary results have shown that several problems arise [Giannakopoulos et al., 2011], reducing the usefulness of pre-existing methods. The same challenges arise across different source types and genres, where different information needs are implied. We note that MultiLing also builds upon shared community tasks. This year these tasks concern headline generation, financial narrative summarization and summary evaluation. Based on previous experience, we estimate the number of participating systems to approximately 20 over all tasks.

Given the above, the content of the workshop includes: (a) selected system reports on the community tasks; (b) overview of the tasks and related findings; (c) selected papers within the MultiLing scope, but beyond the tasks themselves.



**Organizers:**

George Giannakopoulos - NCSR Demokritos (General Chair, Summary Evaluation Task)  
John M. Conroy - IDA Center for Computing Sciences (Headline Generation Task)  
Marina Litvak - Sami Shamoon College of Engineering (Headline Generation Task)  
Peter Rankel - Elder Research Inc. (Headline Generation Task)  
Mahmoud El-Haj - University of Lancaster (Financial Narrative Summarization Task)

**Program Committee:**

John M. Conroy - IDA Center for Computing Sciences  
Marina Litvak - Sami Shamoon College of Engineering  
Udo Kruschwitz - University of Essex  
Horacio Saggion - Universitat Pompeu Fabra  
Elena Lloret - University of Alicante  
Vangelis Karkaletsis - NCSR Demokritos  
Mark Last - Ben-Gurion University of the Negev  
Natalia Vanetik - Sami Shamoon College of Engineering, Ben-Gurion University of the Negev  
Josef Steinberger - University of West Bohemia  
Laura Plaza - UNED  
Florian Boudin - University of Nantes  
Iraklis Varlamis - Harokopio University  
George Petasis - NCSR Demokritos  
Peter Rankel - Elder Research Inc.  
George Giannakopoulos - NCSR Demokritos  
Mahmoud El-Haj - Lancaster University  
Ahmet Aker - Universität Duisburg-Essen  
Giuseppe Riccardi - University of Trento  
Corina Forascu - University "Al. I. Cuza" Iasi  
Michalis Vazirgiannis - Ecole Polytechnique; Athens University of Economics and Business

**Invited Speaker:**

Horacio Saggion, Universitat Pompeu Fabra



## Table of Contents

<i>RANLP 2019 Multilingual Headline Generation Task Overview</i>	
Marina Litvak, John M. Conroy and Peter A. Rinkel .....	1
<i>MultiLing 2019: Financial Narrative Summarisation</i>	
Mahmoud El-Haj .....	6
<i>The Summary Evaluation Task in the MultiLing - RANLP 2019 Workshop</i>	
George Giannakopoulos and Nikiforos Pittaras .....	11
<i>Multi-lingual Wikipedia Summarization and Title Generation On Low Resource Corpus</i>	
Wei Liu, Lei Li, Zuying Huang and Yinan Liu .....	17
<i>A topic-based sentence representation for extractive text summarization</i>	
Nikolaos Gialitsis, Nikiforos Pittaras and Panagiotis Stamatopoulos .....	26
<i>A Study on Game Review Summarization</i>	
George Panagiotopoulos, George Giannakopoulos and Antonios Liapis .....	35
<i>Social Web Observatory: An entity-driven, holistic information summarization platform across sources</i>	
Leonidas Tsekouras, Georgios Petasis and Aris Kosmopoulos .....	44
<i>EASY-M: Evaluation System for Multilingual Summarizers</i>	
Marina Litvak, Natalia Vanetik and Yael Veksler .....	53
<i>A study of semantic augmentation of word embeddings for extractive summarization</i>	
Nikiforos Pittaras and Vangelis Karkaletsis .....	63
<i>HEvAS: Headline Evaluation and Analysis System</i>	
Marina Litvak, Natalia Vanetik and Itzhak Eretz Kdosha .....	73





# Conference Program

**Friday, September 6th, 2019**

**09:30–10:45    Opening and Invited Talk**

09:30–09:45    *Welcome*  
George Giannakopoulos

09:45–10:45    *Invited talk: Mining and Enriching Multilingual Scientific Text Collections*  
Horacio Saggion

**10:45–11:15    Coffee Break**

**11:15–12:15    Community tasks**

11:15–11:35    *RANLP 2019 Multilingual Headline Generation Task Overview*  
Marina Litvak, John M. Conroy and Peter A. Rankel

11:35–11:55    *MultiLing 2019: Financial Narrative Summarisation*  
Mahmoud El-Haj

11:55–12:15    *The Summary Evaluation Task in the MultiLing - RANLP 2019 Workshop*  
George Giannakopoulos and Nikiforos Pittaras

**Friday, September 6th, 2019 (continued)**

**12:15–13:00 Workshop papers I**

12:15–12:30 *Multi-lingual Wikipedia Summarization and Title Generation On Low Resource Corpus*  
Wei Liu, Lei Li, Zuying Huang and Yinan Liu

12:30–12:45 *A topic-based sentence representation for extractive text summarization*  
Nikolaos Gialitsis, Nikiforos Pittaras and Panagiotis Stamatopoulos

12:45–13:00 *A Study on Game Review Summarization*  
George Panagiotopoulos, George Giannakopoulos and Antonios Liapis

**13:00–14:30 Lunch break**

**14:30–15:50 Workshop papers II**

14:30–14:50 *Social Web Observatory: An entity-driven, holistic information summarization platform across sources*  
Leonidas Tsekouras, Georgios Petasis and Aris Kosmopoulos

14:50–15:10 *EASY-M: Evaluation System for Multilingual Summarizers*  
Marina Litvak, Natalia Vanetik and Yael Veksler

15:10–15:30 *A study of semantic augmentation of word embeddings for extractive summarization*  
Nikiforos Pittaras and Vangelis Karkaletsis

15:30–15:50 *HEvAS: Headline Evaluation and Analysis System*  
Marina Litvak, Natalia Vanetik and Itzhak Eretz Kdosha

**15:50–16:00 Closing**

**16:00–16:30 Networking and final coffee break**