

NAACL-HLT 2012

**Predicting and Improving Text Readability for target reader
populations (PITR 2012)**

Proceedings of the Workshop

June 7, 2012
Montréal, Canada

Production and Manufacturing by

Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2012 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 978-1-937284-20-6 / 1-937284-20-4

Foreword

The last few years have seen a resurgence of work on text simplification and readability. Examples include learning lexical and syntactic simplification operations from Simple English Wikipedia revision histories, exploring more complex lexico-syntactic simplification operations requiring morphological changes as well as constituent reordering, simplifying mathematical form, applications for target users such as deaf students, second language learners and low literacy adults, and fresh attempts at predicting readability.

The PITR 2012 workshop has been organised to provide a cross-disciplinary forum for discussing key issues related to predicting and improving text readability for target users. It will be held on June 7, 2012 in conjunction with the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, in Montréal, Québec, and is sponsored by the ACL Special Interest Group on Speech and Language Processing for Assistive Technologies (SIG-SLPAT).

These proceedings include eight papers that cover various perspectives on the topic, from machine learning to psycholinguistic methods. Three papers try to distinguish between surface level fluency and deeper comprehensibility issues (Rello et al.; Siddharthan and Katsos; Maney et al.). Other papers focus on feature engineering for better predicting and interpreting readability (Francois and Miltakaki; Tonelli et al; Ma et al). Two papers specifically address the issue of lexical difficulty (Drndarevic and Saggion; Brooke et al).

We hope this volume is a valuable addition to the literature, and look forward to an exciting Workshop.

Sandra Williams
Advaith Siddharthan
Ani Nenkova

Organizers:

Sandra Williams, The Open University, UK.
Advaith Siddharthan, University of Aberdeen, UK.
Ani Nenkova, University of Pennsylvania, USA.

Program Committee:

Gregory Aist, Iowa State University, USA.
John Carroll, University of Sussex, UK.
Kevyn Collins-Thompson, Microsoft Research (Redmond), USA.
Siobhan Devlin, University of Sunderland, UK.
Noémie Elhadad, Columbia University, USA.
Micha Elsner, University of Edinburgh, UK.
Richard Evans, University of Wolverhampton, UK.
Lijun Feng, Columbia University, USA.
Caroline Gasperin, TouchType Ltd., UK.
Albert Gatt, University of Malta, Malta.
Pablo Gervás, Universidad Complutense de Madrid, Spain.
Iryna Gurevych, Technische Universität Darmstadt, Germany.
Raquel Hervás, Universidad Complutense de Madrid, Spain.
Véronique Hoste, University College Ghent, Belgium.
Matt Huenerfauth, The City University of New York (CUNY), USA.
Iustina Ilisei, University of Wolverhampton, UK.
Tapas Kanungo, Microsoft, USA.
Mirella Lapata, University of Edinburgh, UK.
Annie Louis, University of Pennsylvania, USA.
Ruslan Mitkov, University of Wolverhampton, UK.
Hitoshi Nishikawa, NTT, Japan.
Mari Ostendorf, University of Washington, USA.
Ehud Reiter, University of Aberdeen, UK.
Lucia Specia, University of Wolverhampton, UK.
Irina Temnikova, University of Wolverhampton, UK.
Ielka van der Sluis, University of Groningen, The Netherlands.

Invited Talk: *Enriching the Web with Readability Metadata*

Dr Kevyn Collins-Thompson, Microsoft Research, USA.

Table of Contents

<i>Toward Determining the Comprehensibility of Machine Translations</i> Tucker Maney, Linda Sibert, Dennis Perzanowski, Kalyan Gupta and Astrid Schmidt-Nielsen . .	1
<i>Towards Automatic Lexical Simplification in Spanish: An Empirical Study</i> Biljana Drndarevic and Horacio Saggion	8
<i>Offline Sentence Processing Measures for testing Readability with Users</i> Advait Siddharthan and Napoleon Katsos	17
<i>Graphical Schemes May Improve Readability but Not Understandability for People with Dyslexia</i> Luz Rello, Horacio Saggion, Ricardo Baeza-Yates and Eduardo Graells	25
<i>Building Readability Lexicons with Unannotated Corpora</i> Julian Brooke, Vivian Tsang, David Jacob, Fraser Shein and Graeme Hirst	33
<i>Making Readability Indices Readable</i> Sara Tonelli, Ke Tran Manh and Emanuele Pianta	40
<i>Do NLP and machine learning improve traditional readability formulas?</i> Thomas François and Eleni Miltsakaki	49
<i>Comparing human versus automatic feature extraction for fine-grained elementary readability assessment</i> Yi Ma, Ritu Singh, Eric Fosler-Lussier and Robert Lofthus	58

Workshop Program

Thursday, June 7, 2012

Invited Talk

9:15–9:30 Opening Remarks

9:30–10:30 Invited Talk by Kevyn Collins-Thompson: Enriching the Web with Readability Metadata

Coffee

Session 1: What is readability

11:00–11:15 *Toward Determining the Comprehensibility of Machine Translations*
Tucker Maney, Linda Sibert, Dennis Perzanowski, Kalyan Gupta and Astrid Schmidt-Nielsen

11:15–11:30 *Towards Automatic Lexical Simplification in Spanish: An Empirical Study*
Biljana Drndarevic and Horacio Saggion

11:30–11:45 *Offline Sentence Processing Measures for testing Readability with Users*
Advaith Siddharthan and Napoleon Katsos

11:45–12:00 *Graphical Schemes May Improve Readability but Not Understandability for People with Dyslexia*
Luz Rello, Horacio Saggion, Ricardo Baeza-Yates and Eduardo Graells

12:00–12:30 Panel Discussion

Thursday, June 7, 2012 (continued)

Lunch

Session 2: Predicting readability

- 14:00–14:15 *Building Readability Lexicons with Unannotated Corpora*
Julian Brooke, Vivian Tsang, David Jacob, Fraser Shein and Graeme Hirst
- 14:15–14:30 *Making Readability Indices Readable*
Sara Tonelli, Ke Tran Manh and Emanuele Pianta
- 14:30–14:45 *Do NLP and machine learning improve traditional readability formulas?*
Thomas François and Eleni Miltsakaki
- 14:45–15:00 *Comparing human versus automatic feature extraction for fine-grained elementary readability assessment*
Yi Ma, Ritu Singh, Eric Fosler-Lussier and Robert Lofthus
- + 15:00–15:30 Panel Discussion

Poster Session

- 15:30–17:30 Coffee + Joint Poster Session with SLPAT

Business Meeting

- 17:30–18:30 SIG-SLPAT Business Meeting