COLING 2012

24th International Conference on Computational Linguistics

Proceedings of the Workshop on Speech and Language Processing Tools in Education

Workshop chairs:
Radhika Mamidi and Kishore Prahallad

15 December 2012
Mumbai, India

Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

Proceedings of the Workshop on Speech and Language Processing Tools in Education
Radhika Mamidi and Kishore Prahallad (eds.)
Revised preprint edition, 2012

Published by The COLING 2012 Organizing Committee
Indian Institute of Technology Bombay,
Powai,
Mumbai-400076
India
Phone: 91-22-25764729
Fax: 91-22-2572 0022
Email: pb@cse.iitb.ac.in

This volume © 2012 The COLING 2012 Organizing Committee.
Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Nonported* license.
http://creativecommons.org/licenses/by-nc-sa/3.0/
Some rights reserved.

Contributed content copyright the contributing authors.
Used with permission.

Also available online in the ACL Anthology at http://aclweb.org

Preface

Information and communication technology (ICT) along with conventional aids assumes a central role in today's classroom teaching. With the advent of Internet and WWW, accessing information and communication in the context of a learner-learner and a teacher-learner interaction has become more effective. The resources and tools on the WWW are considered to be one of the best available, specifically because of its 'anywhere, anytime' access. One of the most important domains in ICT is speech and language technologies. Tools and applications including text-to-speech, automatic speech recognition, spelling checkers, machine translation, and information retrieval/extraction systems have found their way into a common man's day-to-day life. However, there has been little progress in their use and application for educational purposes.

This workshop provided an opportunity for a discussion and exchange of ideas on multidisciplinary research in academia by bringing together technologies from different sub-areas of speech, language technologies, learning sciences and education.

We received a good response and based on referee reports, 12 papers were selected for presentations. There were equal number of papers from India and abroad covering a wide range of sub-areas.

We thank the members of Scientific Committee for their support and cooperation for the workshop. We also thank them for giving a thorough feedback to the authors. Finally, we thank the organizers of COLING 2012 for giving us the opportunity to conduct this workshop.

Radhika Mamidi
Kishore Prahallad

Workshop Committees

Scientific Committee

- Alan W Black, Carnegie Mellon University, USA
- Robert Elliott, University of Oregon, USA
- Maxine Eskenazi, Carnegie Mellon University, USA
- Shajith Ikbal, IBM Research, India
- Radhika Mamidi, IIIT Hyderabad, India
- Nobuaki Minematsu, The University of Tokyo, Japan
- Helen Ming, The Chinese University of Hong Kong, Hong Kong
- Kannan M Moudgalya, IIT Bombay, India
- Hema A Murthy, IIT Madras, India
- Kishore Prahallad, IIIT Hyderabad, India
- Martin Russell, University of Birmingham, UK
- Helmer Strik, Radboud University, Nijmegen, Netherlands
- Vasudeva Varma, IIIT Hyderabad, India
- Ashish Verma, IBM Research, India

Organizing Committee

- Radhika Mamidi (Chair), IIIT Hyderabad, India
- Kishore Prahallad (Co-Chair), IIIT Hyderabad, India

Table of Contents

Effective Mentor Suggestion System for Online Collaboration Platform
Advait Raut, Upasana Gaikwad, Ramakrishna Bairi and Ganesh Ramakrishnan 1

Automatically Assessing Free Texts
Yllias Chali and Sadid A. Hasan 9

Automatic pronunciation assessment for language learners with acoustic-phonetic features
Vaishali Patil and Preeti Rao 17

An Issue-oriented Syllabus Retrieval System using Terminology-based Syllabus Structuring and Visualization
Hideki Mima 25

Real-Time Tone Recognition in A Computer-Assisted Language Learning System for German Learners of Mandarin
Hussein Hussein, Hansjörg Mixdorff and Rüdiger Hoffmann 37

Textbook Construction from Lecture Transcripts
Aliabbas Petiwala, Kannan Moudgalya and Pushpak Bhattacharya 43

Enriching An Academic knowledge base using Linked Open Data
Chetana Gavankar, Ashish Kulkarni, Yuan-Fang Li and Ganesh Ramakrishnan 51

Automatic Pronunciation Scoring And Mispronunciation Detection Using CMUSphinx
Ronanki Srikanth and James Salsman 61

Content Bookmarking and Recommendation
Ananth Vyasarayamut, Satyabrata Behera and Ganesh Ramakrishnan 69

A template matching approach for detecting pronunciation mismatch
Lavanya Prahallad, Radhika Mamidi and Kishore Prahallad 79

Automatic Easy Japanese Translation for information accessibility of foreigners
Manami Moku and Hidekazu Yamamoto 85

Genre-oriented Readability Assessment: a Case Study
Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni 91

Workshop on Speech and Language Processing Tools in Education

Program

Saturday, 15 December 2012

09:30–09:50	Introduction
09:50–11:30	**Presentations**
09:50–10:00	*Effective Mentor Suggestion System for Online Collaboration Platform* Advait Raut, Upasana Gaikwad, Ramakrishna Bairi and Ganesh Ramakrishnan
10:00–10:10	*Automatically Assessing Free Texts* Yllias Chali and Sadid A. Hasan
10:10–10:20	*Automatic pronunciation assessment for language learners with acoustic-phonetic features* Vaishali Patil and Preeti Rao
10:20–10:30	*An Issue-oriented Syllabus Retrieval System using Terminology-based Syllabus Structuring and Visualization* Hideki Mima
10:30–10:40	*Real-Time Tone Recognition in A Computer-Assisted Language Learning System for German Learners of Mandarin* Hussein Hussein, Hansjörg Mixdorff and Rüdiger Hoffmann
10:40–10:50	*Textbook Construction from Lecture Transcripts* Aliabbas Petiwala, Kannan Moudgalya and Pushpak Bhattacharya
10:50–11:00	*Enriching An Academic knowledge base using Linked Open Data* Chetana Gavankar, Ashish Kulkarni, Yuan-Fang Li and Ganesh Ramakrishnan
11:00–11:10	*Automatic Pronunciation Scoring And Mispronunciation Detection Using CMUSphinx* Ronanki Srikanth and James Salsman
11:10–11:20	*Content Bookmarking and Recommendation* Ananth Vyasarayamut, Satyabrata Behera and Ganesh Ramakrishnan
11:20–11:30	*A template matching approach for detecting pronunciation mismatch* Lavanya Prahallad, Radhika Mamidi and Kishore Prahallad
11:30–12:00	Tea break
12:00–13:30	**Poster Session/Demo**
13:30–14:30	Lunch