Third International Workshop on Paraphrasing (IWP2005)

Proceedings of the Workshop

Organizers
Mark Dras, Macquarie University, Australia
Kazuhide Yamamoto, Nagaoka University of Technology, Japan

14 October 2005
Jeju Island, Korea
Preface

This is the third in a series of international workshops dealing with paraphrasing. All have been concerned with how knowledge of paraphrases can be useful in the field of natural language processing, and each has had a theme focusing on one aspect of how such knowledge of paraphrases can be acquired, codified and used.

Because of its broad potential applicability, the subfield of paraphrase studies has occasioned interest from researchers from many institutions around the world. For this third workshop in the series, we received submissions from Asia, Australasia, the US and Europe. The accepted papers reflect this wide distribution of interest.

All papers were peer reviewed in full by at least three members of the Program Committee. We would like to thank Program Committee members for their efforts in this regard. We would also like to thank the organisers of the Second International Joint Conference on Natural Language Processing (IJCNLP-05), with which this workshop is associated, for their assistance.

Mark Dras and Kazuhide Yamamoto

October 2005
Excerpt from Call for Papers

Background and Goals
Paraphrases are alternative ways to convey the same information. As having been claimed by an increasing number of researchers, technology for generating and recognizing paraphrases can potentially benefit a broad range of NLP tasks including machine translation, reading assistance, multi-document summarization, information retrieval, and question answering.

Motivated by this background, we organized international workshops on automatic paraphrasing in 2001 and 2003 (IWP2003), which successfully drew the growing interest of NLP researchers. As both workshops attracted attention and successfully finished, we will hold the third workshop in conjunction with IJCNLP-05, in order to collect existing and emerging research topics on automatic paraphrasing during the recent two years.

The proposed workshop is intended to be the successor to these previous workshops. The goals of the workshop are to connect with a broader range of research activities related to automatic paraphrasing, and to place the workshops in a series with the aim of establishing a new research field.

General Topics
The workshop will be open to any research topic related to paraphrasing of any language. More specifically, topics of interest include, but are not limited to:

- typology of paraphrases
- representation of paraphrases
- automatic acquisition of paraphrases
- algorithms for recognizing and generating paraphrases
- existing and potential applications of automatic paraphrasing
- computational modeling of linguistic theories on paraphrases
- open resources for paraphrasing technology
- methods for evaluating paraphrasing technology

Given the location of the workshop, papers focusing on paraphrasing within the languages of the Asia-Pacific region are particularly encouraged.

Special Topic: constructing paraphrase-related resources
The theme of the previous workshop in the series was the automated acquisition of paraphrase. A particular topic of interest for this workshop, then, is the issue of constructing paraphrase-related resources that would follow from this automated acquisition: What should these look like? How would dictionaries and corpora of automatically acquired paraphrases be defined?
Organizers:

Mark Dras Macquarie University, Australia
Kazuhide Yamamoto Nagaoka University of Technology, Japan

Program Committee:

Caroline Brun Xerox Research Centre Europe, France
Mark Dras Macquarie University, Australia
Ulf Hermjakob USC Information Sciences Institute, USA
Kentaro Inui NAIST, Japan
Gen’ichiro Kikui ATR-SLT, Japan
Mirella Lapata University of Edinburgh, UK
Hiroshi Nakagawa University of Tokyo, Japan
Fabio Rinaldi University of Zurich, Switzerland
Satoshi Sato Nagoya University, Japan
Yusuke Shinyama New York University, USA
Noriko Tomuro DePaul University, USA
Hua Wu Toshiba China, P.R.China
Kazuhide Yamamoto Nagaoka University of Technology, Japan
Chengqing Zong Chinese Academy of Sciences, P.R.China
Table of Contents

Preface .......................................................................................................................... i
Excerpt from Call for Papers ......................................................................................... ii

Support Vector Machines for Paraphrase Identification and Corpus Construction
  Chris Brockett and William B. Dolan ........................................................................... 1

Automatically Constructing a Corpus of Sentential Paraphrases
  William B. Dolan and Chris Brockett ......................................................................... 9

Using Machine Translation Evaluation Techniques to Determine Sentence-level Semantic Equivalence
  Andrew Finch, Young-Sook Hwang and Eiichiro Sumita ......................................... 17

A Class-oriented Approach to Building a Paraphrase Corpus
  Atsushi Fujita and Kentaro Inui .................................................................................. 25

Structural Variation in Generated Health Reports
  Catalina Hallett and Donia Scott ................................................................................. 33

Transforming a Sentence End into News Headline Style
  Satoshi Ikeda and Kazuhide Yamamoto .................................................................... 41

Automated Generalization of Phrasal Paraphrases from the Web
  Weigang Li, Ting Liu, Yu Zhang, Sheng Li and Wei He ............................................. 49

Automatic Generation of Paraphrases to be Used as Translation References in Objective Evaluation Measures of Machine Translation
  Yves Lepage and Etienne Denoual .......................................................................... 57

Evaluating Contextual Dependency of Paraphrases using a Latent Variable Model
  Kiyonori Ohtake ....................................................................................................... 65

Automatic Generation of Large-Scale Paraphrases
  Richard Power and Donia Scott .................................................................................. 73

Automatic Paraphrase Discovery based on Context and Keywords between NE Pairs
  Satoshi Sekine .......................................................................................................... 80

Towards Statistical Paraphrase Generation: Preliminary Evaluations of Grammaticality
  Stephen Wan, Mark Dras, Robert Dale and Cecile Paris ............................................ 88

Author Index ................................................................................................................. 96