



**CompuTerm 2004**  
3rd International Workshop on  
Computational Terminology

**Proceedings of the Workshop**

Edited by

SOPHIA ANANIADOU

&

PIERRE ZWEIGENBAUM

29<sup>th</sup> August 2004  
Geneva, Switzerland

## PREFACE

Computational Terminology is becoming an increasingly important aspect in areas such as text mining, information retrieval, information extraction, summarisation, document management systems, question-answering systems, ontology building, etc. In text mining, the acquisition of new knowledge is best captured through terms as they denote new concepts. Terminological information is paramount for knowledge mining from texts for scientific discovery and competitive intelligence. Currently, scientific needs especially in fast growing domains, such as biomedicine, and the overwhelming amount of textual data published daily, demand that terminology is acquired and managed systematically and automatically.

The aim of this third CompuTerm workshop is to bring together Natural Language Processing researchers to discuss recent advances in computational terminology and its impact in many NLP applications. Topics such as term extraction, constructing and updating domain specific dictionaries and thesauri, and systematic terminology management are addressed as they are a necessary component of any NLP system dealing with domain-specific literature.

The topics addressed in this workshop are wide ranging: from term extraction in various languages using different techniques (verb co-occurrence, information theoretic approaches, etc), extracting translation pairs from bilingual corpora based on terminology, to semantic oriented approaches and theoretical aspects of terminology. Terminology structuring is also becoming an important research topic as the detection of semantic relationships between terms finds applications such as the population and update of existing knowledge bases, definition of domain specific templates in Information Extraction, disambiguation of terms, etc. The discovery of synonym terms or term clusters is equally beneficial to many applications in information retrieval, thesaurus generation, information extraction, question-answering ,etc. The integration of different terminological resources is also increasingly important as one has to address term variation between controlled vocabularies, thesauri and textual data. Terminology has to deal with the dynamic nature of terms, their variability, their acquisition from heterogeneous sources, their integration, standardisation and representation. Research and competition tracks on the evaluation of terminological tools still lag behind in terminology in comparison to other NLP areas. With the exception of the NTCIR-1 TMREC, dedicated competition tracks like MUC, TREC, for the evaluation of terminological management systems, term extraction tools, are still missing. Recently, evaluation in specialised areas e.g. named entity recognition in Biomedicine (BioCreative) proved very successful in identifying key problems in terminology such as term variation and ambiguity.

We would like to thank the authors for submitting their research and the members of the Programme Committee for their reviews and useful suggestions to the authors.

Sophia Ananiadou and Pierre Zweigenbaum

**EDITORS**

Sophia Ananiadou (National Centre for Text Mining, UK)  
Pierre Zweigenbaum (AP-HP, INSERM & INaLCO, France)

**INVITED SPEAKER**

Marie-Claude L'Homme (Université de Montréal, Canada)

**ORGANISERS**

Sophia Ananiadou (National Centre for Text Mining, UK)  
Pierre Zweigenbaum (AP-HP, INSERM & INaLCO, France)

**PROGRAM COMMITTEE**

Olivier Bodenreider (NLM, US)  
Didier Bourigault (ERSS, France)  
Teresa Cabre (University Pompeu Fabra, Spain)  
Key-Sun Choi (KORTERM, KAIST, Korea)  
Beatrice Daille (IRIN, France)  
Pascale Fung (University of Science and Technology, Hong Kong)  
Éric Gaussier (Xerox, France)  
Gregory Grefenstette (Clairvoyance Corp)  
Marie-Claude L'Homme (University of Montréal, Canada)  
Kyo Kageura (National Institute of Informatics, Japan)  
Philippe Langlais (RALI, Canada)  
John McNaught (UMIST & National Centre for Text Mining, UK)  
Goran Nenadic (UMIST & National Centre for Text Mining, UK)  
Koichi Takeuchi (Okayama University, Japan)

**WORKSHOP WEBSITE**

<http://www-new.biomath.jussieu.fr/~pz/computerm2004.html>

<http://www.cse.salford.ac.uk/TextMining/CompuTerm2004/>

# WORKSHOP PROGRAM

*and Table of Contents*

**Sunday, August 29**

<i>Preface</i> .....	3
8:55 – 9:00	<b>Welcome and opening of the workshop</b>
9:00 – 9:30	<i>Metalinguistic Information Extraction for Terminology</i> ..... 15 Carlos Rodríguez Penagos
9:30 – 10:00	<i>Extracting Low-frequency Translation Pairs from Japanese-English Bilingual Corpora</i> ..... 23 Keita Tsuji & Kyo Kageura
10:00 – 10:30	<i>Dependence Relationships between Gene Ontology Terms based on TIGR Gene Product Annotations</i> ..... 31 Anand Kumar, Barry Smith & Christian Borgelt
10:30 – 11:00	<i>Discovering Specific Semantic Relationships between Nouns and Verbs in a Specialized French Corpus</i> ..... 39 Vincent Claveau & Marie-Claude L’Homme
COFFEE BREAK 11:00–11:30	
11:30 – 12:00	<i>Automatically Inducing Ontologies from Corpora</i> ..... 47 Interjeet Mani, Ken Samuel, Kris Concepcion & David Vogel
12:00 – 12:30	<i>Detecting Semantic Relations between Terms in Definitions</i> ..... 55 Véronique Malaisé, Pierre Zweigenbaum & Bruno Bachimont
12:30 – 13:00	<i>Discovering Synonyms and Other Related Words</i> ..... 63 Krister Lindén & Jussi Piitulainen
LUNCH BREAK 13:00–14:30	
14:30 – 15:30	<b>Invited Talk</b> <i>A Lexico-semantic Approach to the Structuring of Terminology</i> ..... 7 Marie-Claude L’Homme
COFFEE BREAK 15:30–16:00	

16:00–17:30 **Poster session / demo session**

<i>Term Extraction from Korean Corpora via Japanese</i> .....	71
Atsushi Fujii, Tetsuya Ishikawa & Jong-Hyeok Lee	
<i>Quantitative Portraits of Lexical Elements</i> .....	75
Kyo Kageura	
<i>Term Extraction Using Verb Co-occurrence</i> .....	79
Teruo Koyama & Kyo Kageura	
<i>Design and Implementation of a Terminology-based Literature Mining and Knowledge Structuring System</i> .....	83
Hideki Mima, Sophia Ananiadou & Katsumori Matsushima	
<i>Determining the Specificity of Terms based on Information Theoretic Measures</i> ..	87
Pum-Mo Ryu & Key-Sun Choi	
<i>Construction of Grammar Based Term Extraction Model for Japanese</i> .....	91
Koichi Takeuchi, Kyo Kageura, Béatrice Daille & Laurent Romary	
<i>Author index</i> .....	95

## WORKSHOP PROGRAM

*and Table of Contents*

**Sunday, August 29**

<i>Preface</i> .....	3
8:55 – 9:00	<b>Welcome and opening of the workshop</b>
9:00 – 9:30	<i>Metalinguistic Information Extraction for Terminology</i> ..... 15 Carlos Rodríguez Penagos
9:30 – 10:00	<i>Extracting Low-frequency Translation Pairs from Japanese-English Bilingual Corpora</i> ..... 23 Keita Tsuji & Kyo Kageura
10:00 – 10:30	<i>Dependence Relationships between Gene Ontology Terms based on TIGR Gene Product Annotations</i> ..... 31 Anand Kumar, Barry Smith & Christian Borgelt
10:30 – 11:00	<i>Discovering Specific Semantic Relationships between Nouns and Verbs in a Specialized French Corpus</i> ..... 39 Vincent Claveau & Marie-Claude L’Homme
COFFEE BREAK 11:00–11:30	
11:30 – 12:00	<i>Automatically Inducing Ontologies from Corpora</i> ..... 47 Interjeet Mani, Ken Samuel, Kris Concepcion & David Vogel
12:00 – 12:30	<i>Detecting Semantic Relations between Terms in Definitions</i> ..... 55 Véronique Malaisé, Pierre Zweigenbaum & Bruno Bachimont
12:30 – 13:00	<i>Discovering Synonyms and Other Related Words</i> ..... 63 Krister Lindén & Jussi Piitulainen
LUNCH BREAK 13:00–14:30	
14:30 – 15:30	<b>Invited Talk</b> <i>A Lexico-semantic Approach to the Structuring of Terminology</i> ..... 7 Marie-Claude L’Homme
COFFEE BREAK 15:30–16:00	

16:00–17:30 **Poster session / demo session**

<i>Term Extraction from Korean Corpora via Japanese</i> .....	71
Atsushi Fujii, Tetsuya Ishikawa & Jong-Hyeok Lee	
<i>Quantitative Portraits of Lexical Elements</i> .....	75
Kyo Kageura	
<i>Term Extraction Using Verb Co-occurrence</i> .....	79
Teruo Koyama & Kyo Kageura	
<i>Design and Implementation of a Terminology-based Literature Mining and Knowledge Structuring System</i> .....	83
Hideki Mima, Sophia Ananiadou & Katsumori Matsushima	
<i>Determining the Specificity of Terms based on Information Theoretic Measures</i> ..	87
Pum-Mo Ryu & Key-Sun Choi	
<i>Construction of Grammar Based Term Extraction Model for Japanese</i> .....	91
Koichi Takeuchi, Kyo Kageura, Béatrice Daille & Laurent Romary	
<i>Author index</i> .....	95

## Author Index

### A

Ananiadou, Sophia ..... [83](#)

### B

Bachimont, Bruno ..... [55](#)

Borgelt, Christian ..... [31](#)

### C

Choi, Key-Sun ..... [87](#)

Claveau, Vincent ..... [39](#)

Concepcion, Kris ..... [47](#)

### D

Daille, Béatrice ..... [91](#)

### F

Fujii, Atsushi ..... [71](#)

### I

Ishikawa, Tetsuya ..... [71](#)

### K

Kageura, Kyo ..... [23](#), [75](#), [79](#), [91](#)

Koyama, Teruo ..... [79](#)

Kumar, Anand ..... [31](#)

### L

L'Homme, Marie-Claude ..... [7](#), [39](#)

Lee, Jong-Hyeok ..... [71](#)

Lindén, Krister ..... [63](#)

### M

Malaisé, Véronique ..... [55](#)

Mani, Interjeet ..... [47](#)

Matsushima, Katsumori ..... [83](#)

Mima, Hideki ..... [83](#)

### P

Piitulainen, Jussi ..... [63](#)

### R

Rodríguez Penagos, Carlos ..... [15](#)

Romary, Laurent ..... [91](#)

Ryu, Pum-Mo ..... [87](#)

### S

Samuel, Ken ..... [47](#)

Smith, Barry ..... [31](#)

### T

Takeuchi, Koichi ..... [91](#)

Tsuji, Keita ..... [23](#)

### V

Vogel, David ..... [47](#)

### Z

Zweigenbaum, Pierre ..... [55](#)