COLING 2004

The 20th International Conference on Computational Linguistics

Post-Conference Workshop

Proceeding of the Workshop on Enhancing and Using Electronic Dictionaries

Editors:

Michael Zock (Limsi, CNRS)
Patrick Saint Dizier (Irit, CNRS)

August 29th, 2004
University of Geneva, Switzerland

PREFACE

A dictionary is a vital component of any natural language processing system. While the digital form has a potential which has only been partially explored, seizing the new possibilities requires some rethinking. This is the goal of this one-day workshop. Creating, maintaining and exploiting intelligently a dictionary involves many skills, some of which we still lack full understanding of. In order to get a clearer picture on these, we've invited papers addressing any of the following issues :

- type of information to be stored in the dictionary
- representation and structuring (indexing) of the data;
- management, visualization and useability of the data;
- access (aids for navigation, interface, strategies);
- acquisition of lexical data (corpus);
- possibilities for editing entries (annotation, updating);
- integration (combination of the dictionary with other ressources, writer's workbench);
- uses of dictionaries by people (learning/teaching; writing) and by machines (NLP).

While we did not get papers on all these issues, we did get a quite good number of papers on such divers issues as « design of a lexical database, standardization, indexing, word access, corpus-based lexical resources, lexical interfaces, lexical acquisition and formalization ». It was also interesting to see the variety of languages in which these issues are addressed. The proposals range from Japanese, Sanskrit, English, German, Russian, Serbo-Croatian, to French and Portuguese. In sum, the community working on dictionaries is dynamic, and there seems to be a growing awareness of the importance of some of the problems presented in our call for papers.

Hoping to enrich the discussion by giving the greatest number of people a chance to express their views, we have decided to accept next to the full papers (5) several project notes (4) and abstracts (10).

Sincere thanks to all the specialists who helped us to assure a good selection of papers, despite a very tight schedule. Their reviews were not only helpful for our decision making, but also, so we believe, for the authors to improve their papers. In the hope that their work interests you, provokes fruitful discussions and result in future collaborations.

Michael Zock and Patrick Saint Dizier

Program Chairs

- Michael Zock (LIMSI-CNRS, Orsay)
- Patrick Saint Dizier (IRIT-CNRS, Toulouse)

Programme Committee

- Antonietta Alonge (University of Perugia, Italy)
- Christian Boitet (GETA, Grenoble, France)
- Nicoletta Calzolari (ILC-CNR, Pisa, Italy)
- Christiane Fellbaum (University of Princeton, USA)
- Graeme Hirst (University of Toronto, Canada)
- Mathieu Mangeot-Nagata (NII, Tokyo)
- Rada Mihalcea (University of North Texas, USA)
- Alain Polguère (OLST, University of Montreal, Canada)
- James Pustejovsky (University of Brandeis, USA)
- Gilles Sérasset (GETA, Grenoble, France)
- Patrick Saint Dizier (IRIT-CNRS, Toulouse, France)
- Takenobu Tokunaga (Tokyo Institute of Technology, Tokyo, Japan)
- Dan Tufis (RACAI, Bucharest, Roumania)
- Jean Véronis (University of Aix en Provence, France)
- Piek Vossen (Irion Technologies, Delft, The Netherlands)
- Leo Wanner (University of Stuttgart, Germany)
- Michael Zock (Limsi-CNRS, France)

Additional reviewer

- Slaven Bilac (Tokyo Institute of Technology, Tokyo, Japan)

Provisional Program

1/2 hour for long papers,

20mn for project notes.

5mn for quick poster presentation

Note : only the first authors are mentionned

08:45	Welcome to participants, introduction
09:00-9:30	J. Breen
09:30-10:00	G. Huet
10:00-10:30	V. Pekar
10:30-11:00	coffee break
11:00-11:30	L. Romary
11:30-12:00	M. Zock
12:00-12:50	5 minutes for each poster
13:00-14:00	lunch
14:00-14:20	Apel
14:20-14:40	Bernard
14:40-15:00	Jacquemin
15:00-15:20	Neumann
15:30-16:00	coffee break
16:00-17:00	poster session
17:00-18:00	wrap up discussion
18:00	End of the workshop

Table of contents

Long Papers

James BREEN
Multiple Indexing in an Electronic Kanji Dictionary 1-7

Gérard HUET
Design of a Lexical Database for Sanskrit..... 8-14

Viktor PEKAR
Linguistic Preprocessing for Distributional Classification of Words..... 15-21

Laurent ROMARY, Susanne SALMON-ALT and Gil FRANCOPOULO
Standards going concrete : from LMF to Morphalou 22-28

Michael ZOCK and Slaven BILAC
Word Lookup on the Basis of Associations : from an Idea to a Roadmap..... 29-35

Project Notes

Ulrich APEL and Julien QUINT
Building a Graphetic Dictionary for Japanese kanji – Character Look-up Based on Brush Strokes or Stroke Groups, and the Display of Kanji as Path Data..... 36-39

Pascale BERNARD, Jacques DENDIEN and Jean-Marie PIERREL
A Computerized Dictionary : Le trésor de la langue française informatisé (TLFi) 40-43

Bernard JACQUEMIN
Dictionaries Merger for Text Expansion in Question Answering 44-47

G. NEUMANN, C. FELLBAUM, A. GEYKEN, A. HEROLD, C. HÜMMER, F. KÖRNER. U. KRAMER, K. KRELL, A., SOKIRKO, D. STANTCHEVA and E. STATHI
A Corpus-based Lexical Resource of German Idioms..... 49-52

Posters

Igor A. BOLSHAKOV and Alexander GELBUKH
A Very Large Dictionary with Paradigmatic, Syntagmatic, and Paronymic Links between Entries 53-56

Elzbieta DURA
Concordance of Snippet 57-60

Vlado KEˇSELJ, Tanja KEˇSELJ and Larisa ZLATI´C
R{j}ecnik.com : English – Serbo-Croatian Electronic Dictionary..... 61-64

Michael MAXWELL and William POSER
Morphological Interfaces to Dictionaries..... 65-68

James McCRACKEN
Rebuilding the Oxford Dictionary of English as a Semantic Network 69-72

Cristina MOTA, Paula CARVALHO and Elisabete RANCHHOD
Multiword Lexical Acquisition and Dictionary Formalization 73-76

Tom O'HARA and Janyce WIEBE
Empirical Acquisition of Differentiating Relations from Definitions 77-80

Veit REUER
Language Resources for a Network-based Dictionary 81-84

Sabine SCHULTE im WALDE
Identification, Quantitative Description, and Preliminary Distributional Analysis of German Particle Verbs..... 85-88

Svetlana SHEREMETYEVA
Application Adaptive Electronic Dictionary with Intelligent Interface 89-92

Index of authors

APEL, U.	36-39	MAXWELL, M.	65-68
BERNARD, P.	40-43	McCRACKEN, J.	66-69
BILAC, S.	29-35	MOTA, C.	73-76
BOLSHAKOV, I. A.	53-56	NEUMANN, G.	49-52
BREEN, J.	1-7	O'HARA, T.	77-80
CARVALHO, P.	73-76	PEKAR, V.	15-21
DENDIEN, J.	40-43	PIERREL, J-M.	40-43
DURA, E.	57-60	POSER, W.	65-68
FELLBAUM, C.	49-52	QUINT, J.	36-39
FRANCOPOULO, G.	22-28	RANCHHOD, E.	73-76
GELBUKH, A.	54-57	REUER, V.	81-84
GEYKEN, A.	49-52	ROMARY, L.	22-28
HEROLD, A.	49-52	SALMON-ALT, S.	22-28
HUET, G.	8-14	SCHULTE im WALDE, S.	85-88
HÜMMER, C.	49-52	SHEREMETYEVA, S.	89-92
JACQUEMIN, B.	44-47	SOKIRKO, A.	49-52
KEˇSELJ, T.	61-64	STANTCHEVA, D.	49-52
KEˇSELJ, V.	61-64	STATHI, E.	49-52
KÖRNER, F.	49-52	WIEBE, J.	77-80
KRAMER, U.	49-52	ZLATI´C, L.	61-64
KRELL, K.	49-52	ZOCK, M.	29-35