

# **2nd International Workshop on Web as Corpus 2006**

**Held at EACL 2006**

**Trento, Italy  
3 April 2006**

**ISBN: 978-1-61738-484-4**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571  
[www.proceedings.com](http://www.proceedings.com)

**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2006) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2010)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

## Table of Contents

<i>Web-based frequency dictionaries for medium density languages</i> András Kornai, Péter Halácsy, Viktor Nagy, Csaba Oravecz, Viktor Trón and Dániel Varga . . . . .	1
<i>BE: A search engine for NLP research</i> Mike Cafarella and Oren Etzioni . . . . .	9
<i>A comparative study on compositional translation estimation using a domain/topic-specific corpus collected from the Web</i> Masatsugu Tonoike, Mitsuhiro Kida, Toshihiro Takagi, Yasuhiro Sasaki, Takehito Utsuro and S. Sato . . . . .	11
<i>CUCWeb: A Catalan corpus built from the Web</i> Gemma Boleda, Stefan Bott, Rodrigo Meza, Carlos Castillo, Toni Badia and Vicente López . . . . .	19
<i>Annotated Web as corpus</i> Paul Rayson, James Walkerdine, William H. Fletcher and Adam Kilgarriff . . . . .	27
<i>Web coverage of the 2004 US Presidential election</i> Arno Scharl and Albert Weichselbraun . . . . .	35
<i>Corporator: A tool for creating RSS-based specialized corpora</i> Cédric Fairon . . . . .	43
<i>The problem of ontology alignment on the Web: A first report</i> Davide Fossati, Gabriele Ghidoni, Barbara Di Eugenio, Isabel Cruz, Huiyong Xiao and Rajen Subba . . . . .	51
<i>Using the Web as a phonological corpus: A case study from Tagalog</i> Kie Zuraw . . . . .	59
<i>Web corpus mining by instance of Wikipedia</i> Rüdiger Gleim, Alexander Mehler and Matthias Dehmer . . . . .	67