

6th Web as Corpus Workshop 2010

(WAC-6)

Held at NAACL HLT 2010

**Los Angeles, California, USA
5 June 2010**

ISBN: 978-1-61738-684-8

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



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

Copyright© (2010) 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

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

| | |
|---|----|
| <i>NoWaC: a large web-based corpus for Norwegian</i> | |
| Emiliano Raul Guevara | 1 |
| <i>Building a Korean Web Corpus for Analyzing Learner Language</i> | |
| Markus Dickinson, Ross Israel and Sun-Hee Lee | 8 |
| <i>Sketching Techniques for Large Scale NLP</i> | |
| Amit Goyal, Jagadeesh Jagaralamudi, Hal Daumé III and Suresh Venkatasubramanian | 17 |
| <i>Building Webcorpora of Academic Prose with BootCaT</i> | |
| George Dillon | 26 |
| <i>Google Web 1T 5-Grams Made Easy (but not for the computer)</i> | |
| Stefan Evert | 32 |