

# **ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries**

**Held at the 50th Annual Meeting of the Association  
for Computational Linguistics 2012 (ACL 2012)**

**Jeju Island, Korea  
10 July 2012**

**ISBN: 978-1-62276-593-5**

**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© (2012) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2012)

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)

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Rediscovering ACL Discoveries Through the Lens of ACL Anthology Network Citing Sentences</i> Dragomir Radev and Amjad Abu-Jbara .....	1
<i>Towards a Computational History of the ACL: 1980-2008</i> Ashton Anderson, Dan Jurafsky and Daniel A. McFarland .....	13
<i>Discovering Factions in the Computational Linguistics Community</i> Yanchuan Sim, Noah A. Smith and David A. Smith .....	22
<i>He Said, She Said: Gender in the ACL Anthology</i> Adam Vogel and Dan Jurafsky .....	33
<i>Discourse Structure and Computation: Past, Present and Future</i> Bonnie Webber and Aravind Joshi .....	42
<i>Extracting glossary sentences from scholarly articles: A comparative evaluation of pattern bootstrapping and deep analysis</i> Melanie Reiplinger, Ulrich Schäfer and Magdalena Wolska .....	55
<i>Applying Collocation Segmentation to the ACL Anthology Reference Corpus</i> Vidas Daudaravicius .....	66
<i>Text Reuse with ACL: (Upward) Trends</i> Parth Gupta and Paolo Rosso .....	76
<i>Integrating User-Generated Content in the ACL Anthology</i> Praveen Bysani and Min-Yen Kan .....	83
<i>Towards an ACL Anthology Corpus with Logical Document Structure. An Overview of the ACL 2012 Contributed Task</i> Ulrich Schäfer, Jonathon Read and Stephan Oepen .....	88
<i>Towards High-Quality Text Stream Extraction from PDF. Technical Background to the ACL 2012 Contributed Task</i> Øyvind Raddum Berg, Stephan Oepen and Jonathon Read .....	98
<i>Combining OCR Outputs for Logical Document Structure Markup. Technical Background to the ACL 2012 Contributed Task</i> Ulrich Schäfer and Benjamin Weitz .....	104
<i>Linking Citations to their Bibliographic references</i> Huy Do Hoang Nhat and Praveen Bysani .....	110