

Radu Soricut

Basic Information

Current Position: Ph.D. Candidate in the Department of Computer Science, University of Southern California

Address: Information Sciences Institute of University of Southern California,
4676 Admiralty Way, Suite 1001, Marina del Rey, California 90292-6601
Phone: (310) 448-8732; Fax: (310) 822-0751;
<http://www.isi.edu/~radu/>; radu@isi.edu

Education

2001– : University of Southern California

- Ph.D. Candidate in Computer Science (Expected Graduation Date: July 2006)
Thesis: Natural Language Generation for Text-to-Text Applications using an Information-Slim Representation
Adviser: Prof. Daniel Marcu

1998–2001: University of Iowa

- M.Sc. in Computer Science (July 2001)

1992–1997: University of Bucharest

- B.Sc. in Computer Science (December 1997)

Work Experience

- **August 2001 –:** Graduate Research Assistant; Intelligent Systems Division, Information Sciences Institute, University of Southern California
- **May 2003 – August 2003:** Research Intern; Machine Learning and Applied Statistics Group, Microsoft Research, Microsoft Corporation
- **May 2001 – August 2001:** Research Intern; Intelligent Systems Division, Information Sciences Institute, University of Southern California
- **May 2000 – May 2001:** Graduate Research Assistant; Image Processing Group, Department of Computer Science, University of Iowa
- **August 1999 – May 2000:** Graduate Teaching Assistant; Department of Computer Science, University of Iowa
- **August 1998 – May 1999:** Course Instructor; Department of Computer Science, University of Iowa
- **August 1996 – May 1997:** Research Fellow; International Institute for Software Technology, United Nations University

Research Interests

Natural Language Processing: statistical language generation; syntax-based language modeling; statistical machine translation; automatic question answering; automatic summarization.

Professional Activities

- Reviewer for *Journal of Computational Linguistics* (2006), *Journal of Artificial Intelligence Research* (2004)
- Reviewer for *The National Conference on Artificial Intelligence (AAAI-06)*, *The International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-06)*, *The Annual Meeting of the Association for Computational Linguistics (ACL-04)*
- Correspondent for *Computer Reviews Journal* (2000)
- Program Committee member for *Formal Methods '99 International Congress* (1999)

Academic Awards and Fellowships

- Gerard P. Weeg Scholarship (2000) at University of Iowa;
- Rockwell Collins Scholarship (1999) at University of Iowa;
- UNU/IIST Fellowship (1998) at the United Nations University;
- Undergraduate Fellowship of Merit (1992–1997) at University of Bucharest.

Publications

Refereed Journals

- **Radu Soricut** and Eric Brill (2005). Automatic Question Answering Using the Web: Beyond the Factoid. *Journal of Information Retrieval - Special Issue on Web Information Retrieval*. To appear.

Refereed Conference Proceedings

- **Radu Soricut** (2005). Natural Language Generation for Text-to-Text Applications Using an Information-Slim Representation. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, *AAAI/SIGART Doctoral Consortium*, July 9-13, Pittsburgh, PA.
- **Radu Soricut** and Daniel Marcu (2005). Towards Developing Generation Algorithms for Text-to-Text Applications. In *Proceedings of the Association for Computational Linguistics Conference (ACL-05)*, June 25-30, Ann Arbor, MI.

- **Radu Soricut** and Eric Brill (2004). A Unified Framework for Automatic Evaluation using N-gram Co-Occurrence Statistics. In *Proceedings of the Association for Computational Linguistics Conference (ACL-04)*, July 22-25, Barcelona, Spain.
- **Radu Soricut** and Eric Brill (2004). Automatic Question Answering: Beyond the Factoid. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL-04)*, May 2-5, Boston, MA.
- **Radu Soricut** and Daniel Marcu (2003). Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL-03)*, May 27-June 1, Edmonton, Canada.
- **Radu Soricut**, Kevin Knight, and Daniel Marcu (2002). Using a large monolingual corpus to improve translation accuracy. *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA-02)*, October 8-12, Tiburon, CA.

Invited/Non-Refereed Papers

- Hal Daume III, Abdessamad Echihabi, Daniel Marcu, Dragos Stefan Munteanu, and **Radu Soricut** (2002). GLEANS: A Generator of Logical Extracts and Abstracts for Nice Summaries. In *Proceedings of the Document Understanding Conference (DUC-02)*, July 11-12, Philadelphia, PA.
- Teodor Rus, Robert Kooima, **Radu Soricut**, Dragos Stefan Munteanu, and James Hunsaker (1999). A Component-Based Language Processing Environment. Midwest Society for Programming Languages and Systems Workshop. Illinois Institute of Technology, Chicago, IL.
- **Radu Soricut** (1996). Ring-buffer Protocols in Process Algebra. In *Proceedings of "Tendinte Actuale in Cibernetica si Filosofia Stiintei" (Current Trends in Cybernetics and the Philosophy of Science)*, National Conference of the Academy of Cybernetics "Stefan Odobleja", Baile Felix, Romania.

Technical Reports

- Jan A. Bergstra, Cornelius A. Middelburg, and **Radu Soricut** (1997). A Discrete Time Network Algebra for a Semantic Foundation of SDL. Research Report 98, International Institute for Software Technology, United Nations University, Macau.

Patent Applications

- **Radu Soricut** and Daniel Marcu. Natural Language Generation via Intersection of Weighted IDL Expressions (WIDL) with Ngram- and Syntax-based Language Models. Submitted to the United States Patent and Trademark Office, June 2005.
- **Radu Soricut**, Daniel Marcu and Kevin Knight. Statistical Translation Using a Large Monolingual Corpus. Submitted to the United States Patent and Trademark Office, September 2002.

Public Software

- **Radu Soricut** and Daniel Marcu (2003). Sentence-level PARSing of DiscourseE (SPADE).
 - available at: <http://www.isi.edu/licensed-sw/spade/> (released May 2003)
 - this software has approximatively 200 downloads (as of December 2005) by academia and industry researchers.

Popular Media

- **Radu Soricut** and Eric Brill (2004). Automatic Question Answering: Beyond the Factoid.
 - work featured by “The Economist” (August 28, 2004 edition), in their Science & Technology section.

Citizenship

- Romanian citizen. Studied and worked in the U.S. under F1 student visa.

References

Daniel Marcu

Information Sciences Institute of University of Southern California,
4676 Admiralty Way, Suite 1001, Marina del Rey, California 90292-6601;
Phone: (310) 448-8726; Fax: (310) 822-0751;
<http://www.isi.edu/~marcu/>;
marcu@isi.edu.

Kevin Knight

Information Sciences Institute of University of Southern California,
4676 Admiralty Way, Suite 1001, Marina del Rey, California 90292-6601;
Phone: (310) 448-8716; Fax: (310) 822-0751;
<http://www.isi.edu/~knight/>;
knight@isi.edu.

Eduard Hovy

Information Sciences Institute of University of Southern California,
4676 Admiralty Way, Suite 1001, Marina del Rey, California 90292-6601;
Phone: (310) 448-8731; Fax: (310) 823-6714;
<http://www.isi.edu/~hovy/>;
hovy@isi.edu.

Eric Brill

Microsoft Research, Microsoft Corporation,
One Microsoft Way, Redmond, WA 98052;
Phone: (425) 705-4992
<http://research.microsoft.com/~brill/>;
brill@microsoft.com.