OKAN KOLAK

COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742
my_first_name_AT_umiacs.umd.edu • http://www.cs.umd.edu/users/okan/

RESEARCH INTERESTS

Natural language processing; in particular statistical modeling and methods, resource acquisition and transfer using parallel corpora, machine translation. Information retrieval.

EDUCATION

UNIVERSITY OF MARYLAND, College Park, MD, USA
Ph.D. candidate in *Computer Science*, December 2005 (expected)
Advisor: Prof. Philip Resnik

UNIVERSITY OF MARYLAND, College Park, MD, USA
M.S. in *Computer Science*, May 2003, GPA: 4.0/4.0
Advisor: Prof. Philip Resnik,

BILKENT UNIVERSITY, Ankara, Turkey
B.S. in *Computer Engineering and Information Science*, May 1998, GPA: 3.55/4.0
Advisor: Prof. Kemal Oflazer

RESEARCH EXPERIENCE

UNIVERSITY OF MARYLAND, College Park, MD, USA
Research Assistant, *Computational Linguistics and Information Processing (CLIP) Lab*, August 1999 – cont.
Designed a generative probabilistic optical character recognition model (OCR), developed parameter estimation methods for the model, and employed the model for OCR post-processing, resulting in significant reduction in recognition error rate. Evaluated the degradation caused by OCR errors on NLP applications, and demonstrated that post-processing has a positive impact.
Designed and implemented Direct Projection Algorithm, which transfers dependency parse trees on one side of a parallel corpus to the other using word level alignments.
Worked on use and evaluation of domain-tuned lexicons in statistical machine translation systems.
Contributed to various cross language information retrieval systems developed within the CLIP Lab.

MICROSOFT, Redmond, WA, USA
Software Design Engineer Intern, *Adaptive User Interface Group*, June – August 2000
Conducted an initial feasibility study of speech based interface support for the Mobile Controls, a unified development toolkit for creating adaptive Web interfaces accessible by a diverse range of clients from computers to cell phones. Designed and implemented a prototype that employed Microsoft VoiceXML.

COMPUTER AND COMMUNICATIONS RESEARCH LAB, NEC USA, San Jose, CA, USA
Research Assistant, June – August 1999
Worked on several components of the NetTopix Focused Search Engine Project. Introduced the concept of logical domains, which is used to cluster search results and to generate multi-granular, topic-focused Web site maps. Designed and implemented an algorithm for identifying and extracting logical domains in a physical web domain. Developed the full-text search component that allowed weighted term selection and supported several ranking and presentation schemes.

TEACHING EXPERIENCE

UNIVERSITY OF MARYLAND, College Park, MD, USA
Volunteer Teaching Assistant, KNES147N SCUBA Diving, February 2001 – July 2004

UNIVERSITY OF MARYLAND, College Park, MD, USA
Teaching Assistant, CMSC723/LING645 Intro to Computational Linguistics, February – June 2000

UNIVERSITY OF MARYLAND, College Park, MD, USA
Teaching Assistant, CMSC106 Introduction to C Programming, August 1998 – June 1999

BILKENT UNIVERSITY COMPUTER CLUB, Ankara, Turkey
Volunteer Instructor, 1996 – 1998
Designed and delivered various lectures ranging from basic computer skills to HTML.

ESENEVLER HIGH SCHOOL, Ankara, Turkey
Volunteer Instructor, September 1996 – June 1997
Taught basic computer skills and user applications such as word-processing and spreadsheets as part of the Computer Literacy Courses Program of the Bilkent University Public Services.

JOURNAL PUBLICATIONS

1. Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and **Okan Kolak**. "Bootstrapping Parsers via Syntactic Projection across Parallel Texts", In the *Special Issue of the Journal of Natural Language Engineering on Parallel Texts*, Eds. Rada Mihalcea and Michel Simard. To appear.

2. Necip Fazil Ayan, Wen-Syan Li, **Okan Kolak**, "Automating Extraction of Logical Domains in a Web Site", In *International Journal of Data and Knowledge Engineering*, 43(2), Elsevier Science, pp. 179-205, November 2002.

OTHER REFEREED PUBLICATIONS

3. **Okan Kolak** and Philip Resnik, "OCR Post-Processing for Low Density Languages", In Proceedings of the Human Language Technology Conference (HLT-EMNLP 2005), Vancouver, Canada, To appear.

4. Necip Fazil Ayan, Bonnie Dorr, **Okan Kolak**, "Domain Tuning of Bilingual Lexicons for MT", In *Proceedings of the Evaluation Workshop at the MT Summit IX*, New Orleans, Louisiana, USA, pp. 3-11, September 2003.

5. **Okan Kolak**, William Byrne, Philip Resnik, "A Generative Probabilistic OCR Model for NLP Applications", In *Proceedings of the Human Language Technology Conference (HLT-NAACL 2003)*, Edmonton, Canada, May 2003.

6. Rebecca Hwa, Philip Resnik, Amy Weinberg, and **Okan Kolak**, "Evaluating Translational Correspondence using Annotation Projection", In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, Philadelphia, Pennsylvania, USA, July 2002.

7. **Okan Kolak** and Philip Resnik, "OCR Error Correction Using a Noisy Channel Model", In *Proceedings of the Human Language Technology Conference (HLT 2002)*, San Diego, California, USA, March 2002.

8. Wen-Syan Li, Necip Fazil Ayan, **Okan Kolak**, Quoc Vu, Hajime Takano, and Hisashi Shimamura, "Constructing Multi-Granular and Topic-Focused Web Site Maps", In *Proceedings of 10th World Wide Web Conference*, Hong Kong, China, May 2001.

9. Wen-Syan Li, **Okan Kolak**, Quoc Vu, and Hajime Takano, "Defining Logical Domains in a Web Site", In *Proceedings of the 2000 ACM Hypertext Conference*, San Antonio, Texas, USA, May 2000.

10. **Okan Kolak** and Wen-Syan Li, "On Ranking and Organizing Web Query Results", In *Proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX)*, Chicago, Illinois, USA, November 8, 1999.

PATENTS

1. Wen-Syan Li, **Okan Kolak**, and Quoc Vu, "Method of Defining and Utilizing Logical Domains to Partition and to Reorganize Physical Domains" *US Patent No. 6,647,381* issued on November 11, 2003, assigned to NEC USA, Inc., Princeton, New Jersey, USA.

SOFTWARE EXPERIENCE

C++, C, C#, Perl, Java, Pascal, Delphi, Lisp, Prolog, COBOL, SQL, CGI, AT&T FSM Toolkit, UNIX, Linux, Windows.

RELEVANT COURSEWORK

Natural Language Processing, AI Planning, Neural Modeling, Programming Language Implementation—Implementing Java, Database Systems Implementation, Data Structures, Computer Networks, Compiler Design, Distributed Systems, Computational Geometry, Computer Graphics, High Performance Computing, Program Verification, Software Engineering.

HONORS

- 1998 Ranked 2nd in Computer Eng. and Information Sci. Dpt., Bilkent University
- 1993-98 Full Scholarship, including tuition, stipend, and accommodation, Bilkent University
- 1997 Certificate of Acknowledgement, Bilkent University Public Services
- 1994-98 Dean's High Honor List, Bilkent University
- 1993 Scholarship, for undergraduate study abroad, Turkish Republic Ministry of Education

PERSONAL

- Permanent U.S. Resident
- SCUBA certifications: Advanced Diver, Nitrox Diver, Rescue Diver
- Class B Commercial Driver License with Air Brake and Passenger endorsements

REFERENCES

Available upon request