

Curriculum vitae

Name: Sumithra Ulrika Velupillai

Date & place of birth: October 27th, 1978, Helsingborg, Sweden

Address: c/o Lundgren
Skulptörv. 6, 2 tr
121 43 Johanneshov
Sweden

Telephone number:
(home) +46(0)850863040
(mobile) +46(0)706856121

E-mail: sumithra@dsv.su.se

Citizenship: Swedish

Education:

April 2007 - Stockholm University, Department of Computer and Systems Sciences.
PhD studies. Part of the KEA-project (Knowledge Extraction Agent, see: http://researchprojects.kth.se/index.php/kb_7795/io_9851/io.html for more information). Thesis subject: knowledge extraction through clustering and information extraction techniques.
Supervisor: Associate Professor Hercules Dalianis
Co-Supervisor: Fil. Dr Martin Hassel

Sep 2000 – 2005 Uppsala University, Language Engineering Programme.
Master of Arts in Computational Linguistics, obtained in May 2005.
The programme includes courses in for example: Algorithms and Data Structures, Phonetics and Phonology, Grammar, Introductory Course in Computational Linguistics, Mathematical Linguistics, Morphology, Corpus Linguistics, Discourse, Semantics, Language Statistics, Logic Programming, Formal Syntax, Machine Grammar, Logic, Speech Technology, Machine Translation, Object Oriented Programming, and Information Retrieval.
For more information please refer to the official webpage: <http://stp.lingfil.uu.se>

Spring 2000 Södertörn University College
Studies in Creative Writing

Sep 1997 – June 2000

Stockholm University
Studies in Humanities
Portuguese, Literature Studies, Cinema Studies, Philosophy.

Sep 1994 – June 1997

Johannes Hedbergsskolan, Helsingborg, Sweden
Senior high school. Studies in languages (Swedish, English, French, German, Spanish). Courses in English: Psychology, Philosophy, Religion, International Communication. Obligatory courses such as History, Social Sciences, Geography, Mathematics, International Marketing. 80 hours of practical study undertaken at Mexico's National Autonomous University in Mexico City.

Publications:

Sumithra Velupillai, Hercules Dalianis, Martin Hassel, Gunnar H. Nilsson: *Developing a standard for de-identifying electronic patient records written in Swedish: Precision, recall and F-measure in a manual and computerized annotation trial*
International Journal of Medical Informatics,
doi:10.1016/j.ijmedinf.2009.04.005, 2009.

Sumithra Velupillai, Hercules Dalianis, Martin Hassel: *Diagnosing Diagnoses in Swedish Clinical Records*
Proceedings of LOUHI '08 - The First Conference on Text and Data Mining of Clinical Documents, Turku, Finland, September 3-4 2008.

Sumithra Velupillai, Martin Hassel, Hercules Dalianis: *Automatic Dictionary Construction and Identification of Parallel Text Pairs*
Proceedings of UCCIS '08 - The international symposium on Using Corpora in Contrastive and Translation Studies, Hangzhou, China, September 25-27 2008.

Sumithra Velupillai, Hercules Dalianis: *Automatic Construction of Domain-specific Dictionaries on Sparse Parallel Corpora in the Nordic languages*
Coling 2008: Proceedings of the workshop Multi-source Multilingual Information Extraction and Summarization, Manchester, England, 2008.

Yvonne Samuelsson, Oscar Täckström, Sumithra Velupillai, Johan Eklund, Mark Fishel, Markus Saers: *Mixing and blending syntactic and semantic dependencies*
Proceedings of CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning, Manchester, England, August 2008.

Yvonne Samuelsson, Oscar Täckström, Sumithra Velupillai: *Evaluation of local features for argument identification and classification for semantic role labeling*

Proceedings of The Second Swedish Language Technology Conference (SLTC-08), Stockholm, Sweden, November 2008.

Magnus Rosell, Sumithra Velupillai: *Revealing Relations between Open and Closed Answers in Questionnaires through Text Clustering Evaluation*

Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08), Marrakesh, Morocco, May 28-30 2008.

Magnus Rosell, Sumithra Velupillai: *The Impact of Phrases in Document Clustering for Swedish*

Proceedings of NODALIDA'05 – 15th Nordic Conference on Computational Linguistics, Joensuu, Finland, May 20–21 2005.

Review assignment:

Student Research Workshop, ACL-IJCNLP, Singapore 2009.

Master's Thesis:

Phrases or Words? Categorizing and Clustering Swedish Scientific Medical Text

The thesis is about categorizing and clustering Swedish scientific medical text, trying to improve document clustering results. The focus lies in experimenting with word and phrase representations of documents.

Supervisor: Prof. Viggo Kann, Department of Numerical Analysis and Computer Science, Royal School of Technology, Stockholm, Sweden. The thesis is part of the project Infomat (for more information, see:

<http://www.nada.kth.se/theory/projects/infomat/>)

Certificates:

2003

Certificate for participating in the course “Applied Computational Linguistics: Information Retrieval of Multimedia Documents”, Eberhard Karls Universität, Tübingen, Germany.

1996

Certificate in Advanced English, certified by the University of Cambridge. Grade obtained: A (a).

Other Courses Attended:

Summer 2000

Södertörn University College
Publication on Internet

Summer 1999

Stockholm University.
Cypern, Fenikien och Levanten

Summer 1998

Gotland University College
Cinema Studies; Ingmar Bergman

Summer 1996 Three-week course in French in Boulogne-sur-mer.

Work experience:

April 2007 - Stockholm University, Department of Computer and Systems Sciences
Position as PhD student.

February 2006 – April 2007 Johnér Images
Work as developer and keyword writer.

November 2005 – May 2006 Butler Hill Group
Linguistic Consultant: Online news article annotation and Machine translation evaluations.
Butler Hill official webpage: <http://www.butlerhill.com>

August – October 2005 Department of English, Justus-Liebig-Universität, Giessen, Germany.
Work on corpus project.
Supervisor: Prof. Dr. Magnus Huber

November 2003 Teaching assistant for the course Computational Grammar II.
Teacher: Roussanka Loukanova
Work in LKB (see: <http://www-csli.stanford.edu/~aac/lkb.html>)

June – August 2003 The Linnaeus Centre for Bioinformatics, Uppsala University.
Summer Project, 6 weeks: “Mining gene-to-gene interactions from the MEDLINE article database”.
Supervisors: Prof. Jan Komorowski (The Linnaeus Centre for Bioinformatics, Uppsala University) and Prof. Anna Săgvall Hein (Department of Linguistics, Uppsala University).

Voluntary work:

1997, 1998 and 2005 Stockholm Film Festival. Ticket seller and hostess.

1998 Uppsala Short Film Festival. Ticket seller.

Other information:

Languages:
Swedish: speak, read and write fluently.
English: speak, read and write fluently.
French: read, write and understand without difficulties, can make myself understood.
German: read, write and understand without difficulties, can make myself understood.
Spanish: read, write and understand without difficulties, can make myself understood.
Portuguese: read and understand, can make myself understood

Danish: read and understand, can make myself understood.
Norwegian: read and understand, can make myself understood.

Other interests: Enjoy literature, film, theatre, music. Travel has been a great part of my life; I have lived in Italy (1982-4), visited USA (1987 and 1993), China (autumn 1988), and Mexico (1997), and several European countries.

References:

Associate Professor Hercules Dalianis, Department of Computer and Systems Sciences, Stockholm University, Sweden. Phone: +46 (0)8 674 75 47, e-mail: hercules@dsv.su.se

Professor Viggo Kann, CSC, Royal School of Technology, Stockholm, Sweden. Phone: +46(0)87906292, e-mail: viggo@csc.kth.se

Professor Dr. Magnus Huber, Department of English, Justus-Liebig-Universität, Giessen, Germany. e-mail: Magnus.Huber@anglistik.uni-giessen.de