

Curriculum vitae: Sumithra Velupillai

Skulptörvägen 6, 2tr
121 43 Johanneshov
Sweden
sumithra.velupillai@gmail.com

Doctoral degree

Ph. D.: 2012, Dept. of Computer and Systems Sciences, Stockholm University.
Discipline: Computer and Systems Sciences
Title: Shades of Certainty – Annotation and Classification of Swedish Medical Records
Supervisor: Prof. Hercules Dalianis, Dept. of Computer and Systems Sciences, Stockholm University.
Co-supervisor: Dr. Martin Hassel, Dept. of Computer and Systems Sciences, Stockholm University

Higher education degree

M. A.: 2005, Computational Linguistics, Uppsala University.
Title: Phrases or Words? Categorizing and Clustering Swedish Scientific Medical Text

Fil. Lic.: 2009, Computer and Systems Sciences, Stockholm University.
Title: Swedish Health Data – Information Access and Representation

Current position

Postdoctoral Research Associate: School of Computer Science and Communication, KTH, Stockholm and Biomedical Research Centre Nucleus, King's College London. Marie Curie International Career Grant. 1 Jan. 2016 – 31 Dec. 2018. 100% research.

Previous positions

2015: Postdoctoral Research Associate: Dept. Biomedical Informatics, University of Utah, Salt Lake City, US. 1 Oct. 2015 – 31 Dec. 2016. 100% research.

2013 – 2015: International postdoc: Dept. of Computer and Systems Sciences, Stockholm University. 1 Oct. 2013 – 30 Sept. 2015. 100% research, (at least) 2/3 of the time based in the US.

2012 – 2013: Lecturer (lector): Dept. of Computer and Systems Sciences, Stockholm University. 30% research.

2012: Research assistant, Dept. of Computer and Systems Sciences, Stockholm University

2007 – 2012: PhD position, Dept. of Computer and Systems Sciences, Stockholm University.

2006 – 2007: Keyword and image search, Johnér Images, Stockholm

2005: Linguistic consultant, Butler Hill Group

2005: Research assistant, Dept. of English, Justus-Liebig-Universität, Giessen, Germany

2004: Research assistant, The Linnaeus Centre for Bioinformatics, Uppsala University

Teaching:

ISBI (Internet Search Techniques and Business Intelligence) 7.5 hp, Dept. of Computer and Systems Sciences, Stockholm University: Lecturer

WEBMIN (Web Mining) 7.5 hp, Dept of Computer and Systems Sciences, Stockholm University: Participation in development of the course material, lectures and laborations, and course responsibility Spring term 2012.

ITO (It in Organizations) 7.5 hp, Dept. of Computer and Systems Sciences, Stockholm University: lab and course project mentor

OOS (Object Oriented Analysis and Design) 7.5 hp, Dept. of Computer and Systems Sciences, Stockholm University: course project mentor

Språkteknologi (Language Technology) 7.5 hp, KTH: Lecturer

Supervision: Master students, of which one has led to a publication in an international workshop, Dept. of Computer and Systems Sciences, Stockholm University.

Publications:**Peer-reviewed journal articles**

Velupillai et al., 2015. Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis. *IMIA Yearbook of Medical Informatics*, Volume 10, pp 183--193.

Velupillai et al., 2015. Louhi 2014: Special issue on health text mining and information analysis. *BMC Medical Informatics and Decision Making*, Volume 15, Issue Suppl. 2. (Editor)

Smith et al, 2014. Professional language in Swedish clinical text: Linguistic characterization and comparative studies. *Nordic Journal of Linguistics*, Volume 2, pp 297--327.

Velupillai et al, 2014. Cue-based assertion classification for Swedish clinical text -- Developing a lexicon for pyConTextSwe. *Artificial Intelligence in Medicine*, Volume 61, pp 137--144.

Allvin et al, 2011. Characteristics of Finnish and Swedish intensive care nursing narratives: a comparative analysis to support the development of clinical language technologies. *Journal of Biomedical Semantics*, Volume 2(Suppl 3):S1, doi:10.1186/2041-1480-2-S3-S1.

Dalianis & Velupillai, 2010. De-identifying Swedish Clinical Text - Refinement of a Gold Standard and Experiments with Conditional Random Fields. *Journal of Biomedical Semantics*, 1, 6.

Velupillai et al, 2009. Developing a standard for de-identifying electronic patient records written in Swedish: Precision, recall and F-measure in a manual and computerized annotation trial. *Int. J. Med. Inform.* (2009), doi:10.1016/j.ijmedinf.2009.04.005, 78, 12, pp 19--26.

Peer-reviewed conference contributions

Kalyanam et al., 2015. Facts and Fabrications about Ebola: A Twitter Based Study. In *Proc. ACM KDD Workshop on Connected Health in Big Data Era*, Sydney, Australia. Awarded best paper of ACM KDD Workshop on Connected Health in Big Data Era 2015.

Velupillai et al, 2015. BluLab: Temporal Information Extraction for the 2015 Clinical TempEval Challenge. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015, Association for Computational Linguistics, pp 814--818, Denver, Colorado.

Dalianis et al., 2015. HEALTH BANK - A Workbench for Data Science Applications in Healthcare. In Proceedings of the CAiSE-2015 Industry Track co-located with the 27th Conference on Advanced Information Systems Engineering - CAiSE 2015, CEUR, pp 1-18, Stockholm, Sweden.

Velupillai et al, 2014. Disease/Disorder Semantic Template Filling - Information Extraction Challenge in the ShARe/CLEF eHealth Evaluation Lab 2014. In Proceedings AMIA 2014, American Medical Informatics Association, Washington, DC.

Kvist & Velupillai, 2014. SCAN: A Swedish Clinical Abbreviation Normalizer. In Kanoulas, Evangelos and Lupu, Mihai and Clough, Paul and Sanderson, Mark and Hall, Mark and Hanbury, Allan and Toms, Elaine, editor, Information Access Evaluation. Multilinguality, Multimodality, and Interaction, Volume 8685, Springer International Publishing, pp 62-73.

Lövestam et al, 2014. Abbreviations in Swedish Clinical Text - use by three professions. Studies in Health Technology and Informatics, Volume 205, pp 720--724.

Velupillai, 2014. Temporal Expressions in Swedish Medical Text -- A Pilot Study. In Proceedings of BioNLP 2014, Association for Computational Linguistics, pp 88--92, Baltimore, Maryland.

Mowery et al, 2014. Generating Patient Problem Lists from the ShARe Corpus using SNOMED CT/SNOMED CT CORE Problem List. In Proceedings of BioNLP 2014, Association for Computational Linguistics, pp 54--58, Baltimore, Maryland.

Mowery et al, 2014. Task 2: ShARe/CLEF eHealth Evaluation Lab 2014. In Cappellato, L., Ferro, N., Halvey, M. and Kraaij, W., editor, CLEF2014 Working Notes, Volume 1180, CEUR-WS, pp 31--42.

Kelly et al, 2014. Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. In Kanoulas, Evangelos and Lupu, Mihai and Clough, Paul and Sanderson, Mark and Hall, Mark and Hanbury, Allan and Toms, Elaine, editor, Information Access Evaluation. Multilinguality, Multimodality, and Interaction, Volume 8685, Springer International Publishing, pp 172-191.

Grigonyte et al, 2014. Improving Readability of Swedish Electronic Health Records through Lexical Simplification: First Results. In Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations - PITR, Association for Computational Linguistics, pp 74--83, Gothenburg, Sweden.

Svee et al, 2013. Capturing and Representing Values for Requirements of Personal Health Records. In Grabis, J. and Kirikova, M. and Zdravkovic, J. and Stirna, J., editor, Short Paper Proceedings of the 6th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling (PoEM 2013, Volume 1023, CEUR-WS, pp 166--175.

Suominen et al, 2013. Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. In Forner, Pamela and Muller, Henning and Paredes, Roberto and Rosso, Paolo and Stein,

Benno, editor, Information Access Evaluation. Multilinguality, Multimodality, and Visualization. 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, September 23-26, 2013. Proceedings, Volume 8138 2013, Springer, pp 212--231.

Velupillai et al, 2013. Functions for Personal Health Records in Sweden -- Patient Perspectives. In Proceedings of the Scandinavian Conference on Health Informatics 2013, Linköping University Electronic Press, Linköpings universitet, Copenhagen, Denmark.

Kvist & Velupillai, 2013. Professional Language in Swedish Radiology Reports -- Characterization for Patient-Adapted Text Simplification. In Proceedings of the Scandinavian Conference on Health Informatics 2013, Linköping University Electronic Press, Linköpings universitet, Copenhagen, Denmark.

Chapman et al, 2013. Extending the NegEx Lexicon for Multiple Languages. Studies in Health Technology and Informatics, Volume 192, pp 677--681.

Tanushi et al, 2013. Negation Scope Delimitation in Clinical Text Using Three Approaches: NegEx, PyConTextNLP and SynNeg. In Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013) NEALT Proceedings Series 16, Oslo, Norway.

Velupillai et al, 2013. Porting a Rule-based Assertion Classifier for Clinical Text from English to Swedish. In Proceedings of the 4th International Louhi Workshop on Health Document Text Mining and Information Analysis - Louhi 2013, Sydney, Australia.

Isenius et al, 2012. Initial Results in the Development of SCAN: a Swedish Clinical Abbreviation Normalizer. In Proceedings of the CLEF 2012 Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis - CLEFeHealth2012, CLEF, Rome, Italy.

Mowery et al, 2012. Medical diagnosis lost in translation -- Analysis of uncertainty and negation expressions in English and Swedish clinical texts. In BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics, pp 56--64, Montréal, Canada.

Velupillai & Kvist, 2012. Fine-grained Certainty Level Annotations Used for Coarser-grained E-health Scenarios -- Certainty Classification of Diagnostic Statements in Swedish Clinical Text. In A. F. Gelbukh, editor, Computational Linguistics and Intelligent Text Processing. 13th International Conference, CICLing 2012, Part II, Volume 7182, Springer, pp 450--461, New Delhi, India.

Velupillai, 2011. Automatic Classification of Factuality Levels -- A Case Study on Swedish Diagnoses and the Impact of Local Context. In Proc. The Fourth International Symposium on Languages in Biology and Medicine -- LBM 2011, Singapore.

Velupillai et al, 2011. Factuality Levels of Diagnoses in Swedish Clinical Text. In Moen et al, editor, Proc. XXIII International Conference of the European Federation for Medical Informatics -- User Centred Networked Health Care, IOS Press, pp 559 -- 563, Oslo.

Kvist et al, 2011. Modeling human comprehension of Swedish medical records for intelligent access and summarization systems, a physician's perspective. In Proc. 9th Scandinavian Conference on Health Informatics, SHI, Oslo.

Hassel et al, 2011. Something Old, Something New - Applying a Pre-trained Parsing Model to Clinical Swedish. In Proc. 18th Nordic Conf. on Comp. Ling. - NODALIDA '11.

Velupillai, 2010. Semantic annotations in clinical documentation: exploring potentials for future information retrieval. In Proceedings of the third workshop on Exploiting semantic annotations in information retrieval, ACM, pp 9--10, New York, NY, USA.

Velupillai, 2010. Towards A Better Understanding of Uncertainties and Speculations in Swedish Clinical Text - Analysis of an Initial Annotation Trial. In Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, University of Antwerp, pp 14--22, Uppsala, Sweden.

Henriksson & Velupillai, 2010. Levels of Certainty in Knowledge-Intensive Corpora: An Initial Annotation Study. In Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, University of Antwerp, pp 41--45, Uppsala, Sweden.

Täckström et al, 2010. Uncertainty Detection as Approximate Max-Margin Sequence Labelling. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, Association for Computational Linguistics, pp 84--91, Uppsala, Sweden.

Allvin et al, 2010. Characteristics and Analysis of Finnish and Swedish Clinical Intensive Care Nursing Narratives. In Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, Association for Computational Linguistics, pp 53--60, Los Angeles, California, USA.

Dalianis & Velupillai, 2010. How Certain are Clinical Assessments? Annotating Swedish Clinical Text for (Un)certainties, Speculations and Negations. In Proceedings of the of the Seventh International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.

Dalianis & Velupillai, 2009. De-identifying Swedish Clinical Text - Refinement of a Gold Standard and Experiments with Conditional Random Fields. In Proceedings of The 3rd International Symposium on Languages in Biology and Medicine - LBM '09, Seogwipo-si, Jeju Island, South Korea.

Dalianis et al, 2009. Is De-identification of Electronic Health Records Possible? OR Can We Use Health Record Corpora for Research?. In Panel at Virtual Healthcare Interaction - VHI 09, Association for the Advancement of Artificial Intelligence, Technical Report FS-09-07, pp 2-3, Arlington, VA, USA.

Dalianis et al, 2009. The Stockholm EPR Corpus - Characteristics and Some Initial Findings. In Proceedings of ISHIMR 2009, Evaluation and implementation of e-health and health information initiatives: international perspectives. 14th International Symposium for Health Information Management Research, Kalmar, Sweden. Awarded best paper.

Samuelsson et al, 2008. Evaluation of local features for argument identification and classification for semantic role labeling. In Proceedings of The Second Swedish Language Technology Conference - SLTC-08, pp 2, Stockholm, Sweden.

Velupillai et al, 2008. Diagnosing Diagnoses in Swedish Clinical Records. In Proceedings of LOUHI '08 - The First Conference on Text and Data Mining of Clinical Documents, pp 110--112, Turku, Finland.

Velupillai et al, 2008. Automatic Dictionary Construction and Identification of Parallel Text Pairs. In Proceedings of UCCTS'08 - The international symposium on Using Corpora in Contrastive and Translation Studies, Hangzhou, China.

Velupillai & Dalianis, 2008. Automatic Construction of Domain-specific Dictionaries on Sparse Parallel Corpora in the Nordic languages. In Coling 2008: Proceedings of the workshop Multi-source Multilingual Information Extraction and Summarization, Coling 2008 Organizing Committee, pp 10--16, Manchester, UK.

Samuelsson et al, 2008. Mixing and blending syntactic and semantic dependencies. In Proceedings of CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning, Manchester, England. Publisher: Coling 2008 Organizing Committee. Pages: 248 - 252.

Rosell & Velupillai, 2008. Revealing Relations between Open and Closed Answers in Questionnaires through Text Clustering Evaluation. In Proc. of the 6th Int. Language Resources and Evaluation - LREC'08.

Rosell & Velupillai, 2005. The Impact of Phrases in Document Clustering for Swedish. In Proc. 15th Nordic Conf. on Comp. Ling. - NODALIDA '05.

Review articles, book chapters, books

Velupillai, 2012. Shades of Certainty -- Annotation and Classification of Swedish Medical Records. Doctoral thesis, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden.

Velupillai et al, 2010. Finding the Parallel: Automatic Dictionary Construction and Identification of Parallel Text Pairs. In Using Corpora in Contrastive and Translation Studies, Cambridge Scholars Publishing, pp 312-336, . ISBN(13): 978-1-4438-1755-4, ISBN(10): 1-4438-1755-4.

Velupillai, 2009. Swedish Health Data - Information Access and Representation. Licentiate thesis, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden.

Popular science articles/presentations:

Press about the dissertation (In Swedish)

“Hon vill råda bot på otydliga journaler”, Dagens Medicin, 2012-05-17

“Bot för otydliga journaler”, Språktidningen, nr 5 september 2012

“Nyligen avhandlat”, Tidskriften Neo, juli/augusti 2012

Research grants:

MeDESTO - Measuring Duration of Untreated Psychosis by Extraction of Symptom and Treatment Onset from mental health records using language technology. Three-year Marie Curie fellowship 2016 – 2018 funded by the Swedish Research Council/Marie Sklodowska Actions ("People"), within EU's 7th Framework Programme, for research at the Dept. Of Theoretical Computer Science, KTH, Stockholm and The National Institute for Health Research (NIHR) Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, Psychology & Neuroscience, King's College London.

De-Puzzling Time: Improving information access from Swedish medical records by modeling temporal expressions. Two-year international postdoc 2013 – 2015 funded by the Swedish Research Council for research at the University of California San Diego, US, the University

of Utah, Salt Lake City, US, and the Dept. of Computer and Systems Sciences, Stockholm University, Sweden.

An additional research grant for this postdoc project has been awarded by the Fulbright commission, encompassing the academic year 2013-14.

Interlock: Stockholm - San Diego - Inter-Language collaboration in clinical NLP. Goal: Develop inter-language data sharing and collaboration in clinical NLP with a focus on user-driven applications.

Stockholm University Academic Initiative: Travel funding for research collaboration between the research groups IT for Health, Dept. of Computer and Systems Sciences, Stockholm University and UCSD School of Medicine's Division of Biomedical Informatics, 2011 - 2013.

Miscellaneous

Review assignments:

Project proposals:

IWT - research and innovation funding agency of the Flemish (Belgian) government, 2014, 2015

Journal and conference manuscripts:

Student Research Workshop, ACL-IJCNLP, Singapore, 2009.

The Information Society (journal), 2009.

SigIR (Special Interest Group on Information Retrieval) Poster, 2010 - 2013.

ECIR (European Conference on Information Retrieval), 2010 - 2013.

NoDaLiDa (Nordic Conference on Computational Linguistics), 2011, 2013, 2015.

Artificial Intelligence in Medicine (journal), 2010, 2011, 2014

Louhi Health Documentation Text Mining and Information Analysis, 2011 - 2013.

BMC Medical Research Methodology (journal), 2012

Journal of Biomedical Informatics, 2013

Applied Clinical Informatics, 2013

American Medical Informatics Association (AMIA), 2014

Journal of the American Medical Informatics Association (JAMIA), 2014

Information Retrieval Journal, 2015

Medinfo, 2015

Natural Language Engineering, 2015

Research Community Services:

Workshop organizer:

Louhi'10 - Second Louhi Workshop on Text and Data Mining of Health Documents (held in conjunction with NAACL), Los Angeles, 2010. Organizer and local organizing chair. dsv.su.se/louhi10/

Louhi 2014: The Fifth International Workshop on Health Text Mining and Information Analysis (held in conjunction with EACL), Gothenburg, 2014. Organizer and Chair. dsv.su.se/louhi2014

Shared tasks

ShARe/CLEF eHealth challenge 2014. Task 2 co-leader. Sheffield, UK
clefehealth2014.dcu.ie/

ShARe/CLEF eHealth challenge 2013. Organizing Committee Member. Valencia, Spain.
<https://sites.google.com/site/shareclefehealth/>

Invited talks and memberships:

2014: *Clinical NLP on Swedish Medical Records and the Stockholm EPR Corpus* (joint presentation with Aron Henriksson)
Dept. of Biomedical Informatics, University of Utah, Salt Lake City, April 7 2014
Shades of Certainty -- Working with Swedish Medical Records.
AI-seminar, UCSD, US, Jan. 13 2014
Clinical data mining group, Dept. of Biomedical Informatics, Columbia University, Feb. 18 2014
Clinical Natural Language Processing Program, Childrens Hospital Boston and Harvard Medical School, Boston, Feb. 19 2014.

Member of American Medical Informatics Association (2013-2014)

Member of Association for Computational Linguistics (2014)

Travel grants:

Wenner-Gren Foundations, 2012

K A Wallenbergs stiftelse, 2010

Helge Ax:son Johnsons stiftelse, 2008, 2009, 2010, 2011

Council of European Professional Informatics Societies (CEPIS), 2007

Other appointments:

President of the PhD student board at the Department of Computer and Systems Sciences,
PhD student board member at the Faculty Board, Faculty of Social Sciences, Stockholm University, 2009-2010.

Languages:

Swedish: native speaker

English: fluent

Danish and Norwegian: very good in comprehension

German, French, Spanish, Italian, and Portuguese: basic comprehension

Referees:

Prof. Wendy Chapman

Department of Biomedical Informatics, University of Utah, Salt Lake City, USA

e-mail: wendy.chapman@utah.edu

Phone: +1 801 581 4080 (office), +1 412 654 1048 (cell)

Prof. Hercules Dalianis

Department of Computer and Systems Sciences, Stockholm University, Sweden

e-mail: hercules@dsv.su.se

Phone: +46 8 674 75 47 (office), +46 70 568 13 59 (cell)

Prof. Viggo Kann

Department of Theoretical Computer Science, KTH, Stockholm, Sweden

e-mail: viggo@nada.kth.se

Phone: +46 8 790 62 92