



	<p style="text-align: center;">SemanticMining <i>NoE 507505</i> Semantic Interoperability and Data Mining in Biomedicine</p>
---	--

**References
2005-2006**

Report Version: 1

Report Preparation Date: 2007.02.20

Contract Start Date: 2004.01.01

Duration: 3.5 years

Project Co-ordinator: Hans Åhlfeldt

Department of Biomedical Engineering / Medical Informatics

S-581 83 Linköping University, Sweden

<hans.ahlfeldt@imt.liu.se>



**Project funded by the European Community under the
FP6 programme “Integrating and Strengthening the
European Research Area” (2002-2006)**



Articles published

References 2005 (co-authored papers only – authors from more than one partner)

K. Markó, S. Schulz, U. Hahn Automatic Lexicon Acquisition for a Medical Cross-Language Information Retrieval System. Proceedings of the XIX International Congress of the European Federation for Medical Informatics (MIE '05), Geneva, Switzerland. 2005: 829-834.

K. Markó, S. Schulz, O. Medelyan, U. Hahn. Bootstrapping Dictionaries for Cross-Language Information Retrieval. Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05), Salvador, Brazil. 2005: 528-535.

K. Markó, S. Schulz, U. Hahn. MorphoSaurus - Design and Evaluation of an Interlingua-based, Cross-language Document Retrieval Engine for the Medical Domain. *Methods of Information in Medicine*. 4/2005(44): 537-545

K. Markó, S. Schulz, U. Hahn Unsupervised Multilingual Word Sense Disambiguation via an Interlingua. Proceedings of the 20th National Conference on Artificial Intelligence (AAAI '05), Pittsburgh, Pennsylvania. 2005: 1075-1080

M. Poprat, U. Hahn. Enough is Enough – Estimating Upper Bounds of the Size of Training Corpora for Unsupervised PP Attachment Disambiguation. Proceedings of Fifth International Conference on Recent Advances in Natural Language Processing (RANLP-2005)

K. Markó, S. Schulz and U. Hahn. Multilingual Lexical Acquisition by Bootstrapping Cognate Seed Lexicons. Proceedings of Fifth International Conference on Recent Advances in Natural Language Processing (RANLP-2005)

U. Hahn, P. Daumke, S. Schulz, K. Markó. Cross-Language Mining for Acronyms and their Completions from the Web. Proceedings of the 8th International Conference on Discovery Science (DS '05), Singapore. 2005.

U. Hahn, K. Markó & S. Schulz. Subword Clusters as Light-Weight Interlingua for Multilingual Document Retrieval. In: MT Summit X – Proceedings of the 10th Machine Translation. Phuket, Thailand, September 12-16, 2005. Asia-Pacific Association for Machine Translation (AAMT), 2005.

S. Schulz, K. Markó, R. L. de Andrade, E. Pacheco, P. Nohama, U. Hahn, M. Romacker. The Morphosaurus Medical Subword Lexicon. Lexicographic and Semantic Aspects. Proceedings of the 3th Workshop em Tecnologia da Informação e da Linguagem Humana (TIL '05), São Leopoldo, Brasil. 2005

Mikael Nyström, Magnus Merkel, Lars Ahrenberg, Michael Petterstedt, Håkan Petersson & Hans Åhlfeldt. Generering av ett medicinskt engelskt-svenskt lexikon med hjälp av interaktiv ordlänkning. Svenska Läkaresällskapets riksstämma 2005 (Annual Meeting of Swedish Society of Medicine).

Mikael Nyström, Magnus Merkel, Lars Ahrenberg, Pierre Zweigenbaum, Håkan Petersson & Hans Åhlfeldt. Generation of a English-Swedish medical dictionary using interactive word alignment. Submitted to BMC Medical Decision Making.

K. Marko, P. Daumke, S. Schulz, U. Hahn. Automatische Generierung einer sprachübergreifenden Akronymdatenbank. 50. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), Freiburg 11. - 15. September



2005 (Annual Meeting of the German Society of Medical Informatics, Biometry and Epidemiology)

M. Poprat, K. Markó, U. Hahn. Automatische Klassifikation medizinischer Dokumente nach Sprache und Zielgruppe für Text-Retrieval-Systeme. 50. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), Freiburg 11. - 15. September 2005 (Annual Meeting of the German Society of Medical Informatics, Biometry and Epidemiology)

R.H. Baud, M. Nyström, L. Borin, R. Evans, S. Schulz, P. Zweigenbaum. Interchanging Lexical Information for a Multilingual Dictionary. AMIA 2005 annual symposium. Washington DC. Washington, DC: AMIA. 31-35.

Stefan Schulz & Udo Hahn. Part-whole representation and reasoning in formal biomedical ontologies. *Artificial Intelligence in Medicine* (2005) 34, 179—200

Smith Barry, Ceusters Werner, Klagges Bert, Köhler Jacob, Kumar Anand, Lomax Jane, Mungall Chris, Neuhaus Fabian, Rector Alan, Rosse Cornelius. Relations in Biomedical Ontologies. *Genome Biology*, 2005, 6(5)/R 46.

Barry Smith, Jose L.V. Mejino Jr., Stefan Schulz, Anand Kumar, and Cornelius Rosse. *Anatomical Information Science*. A. G. Cohn and D. M. Mark (eds.), *Spatial Information Theory*. Proc COSIT 2005 (Lecture Notes in Computer Science) Berlin/Heidelberg/New York: Springer;:149-164.

Stefan Schulz, Philipp Daumke, Barry Smith, Udo Hahn. How to Distinguish Parthood from Location in Bioontologies. In Proc AMIA Symposium 2005, Washington DC;:669-673

Kumar Anand, Smith Barry, Pisanelli M, Gangemi Aldo, Stefanelli Mario. Clinical Guidelines as Plans: An Ontological Theory. *Methods of Information in Medicine*. In press
Anand Kumar, Barry Smith, Domenico Pisanelli, Aldo Gangemi, Mario Stefanelli. An ontological framework for the implementation of clinical guidelines in health care organizations. *Stud Health Technol Inform*. 2004;102:95-107

Stefan Schulz, Suzanne Hanser, Udo Hahn, Jeremy Rogers. Semantic Clarification of the Representation of Procedures and Diseases in SNOMED CT. *Stud Health Technol Inform*. 2005;116:773-8. (Extended version also submitted to *Meth Inf Med*)

Stefan Schulz, Anand Kumar, Thomas Bittner. Biomedical Ontologies: What part-of is and isn't. *Journal of Biomedical Informatics (Special Issue)*. In press and online at www.sciencedirect.com

Alan Rector, Jeremy Rogers, Thomas Bittner. Granularity scale and collectivity: When size does and does not matter. *Journal of Biomedical Informatics (Special Issue)*. In press and online at www.sciencedirect.com

Christiane Fellbaum, Udo Hahn, Barry Smith. Towards new information resources for public health—From WordNet to MedicalWordNet. *Journal of Biomedical Informatics (Special Issue)*. In press and online at www.sciencedirect.com

Special Issue Editors: F. Pinciroli and D.M. Pisanelli. *Ontologies in Medicine*. *Computers in Biology and Medicine*. In press and online at www.sciencedirect.com

New Journal Launch. Editors in chief: Nicola Guarino and Mark Musen. *Applied Ontology*

Forsum U, Hallander HO, Kallner A, Karlsson D, Impact of qualitative analysis in laboratory medicine. *TrAC-Trends in Analytical Chemistry*, 2005, 24: 546 -555.

Forsum U, Karlsson D, Terminology, categories and representation of examinations in laboratory medicine. *Clin. Chem. Lab. Med*. 2005;43:344-345.



-
- Rosse Cornelius, Kumar Anand, Mejino Jr. Jose L.V., Cook Daniel L., Detwiler Landon T., Smith Barry: "A Strategy for Improving and Integrating Biomedical Ontologies" (<http://ontology.buffalo.edu/bio/OBR.pdf>), in: Proceedings of AMIA Symposium 2005, Washington D.C., 2005, 639-643
- Schulz Stefan, Daumke Philipp, Smith Barry, Hahn Udo: "How to Distinguish Parthood from Location in Bioontologies" (<http://ontology.buffalo.edu/bio/Part&Location.pdf>), in: Proceedings of AMIA Symposium 2005, Washington D.C., 2005, 669-673
- Smith Barry, Ceusters Werner, Klagges Bert R. E., Köhler Jacob, Kumar Anand, Lomax Jane, Mungall Chris, Neuhaus Fabian, Rector Alan, Rosse Cornelius: "Relations in Biomedical Ontologies" (<http://genomebiology.com/2005/6/5/R46>), in: Genome Biology, 2005, R46
- Smith Barry, Mejino Jr. Jose L.V., Schulz Stefan, Kumar Anand, Rosse Cornelius. "Anatomical Information Science" (http://ontology.buffalo.edu/anatomy_GIS/FMA-AIS.pdf), in: Cohn A.G., Mark David M. (eds.): Spatial Information Theory. Proceedings of COSIT 2005 (Lecture Notes in Computer Science), Springer Verlag, Berlin/Heidelberg/New York, 2005, 149-164
- Rosse Cornelius, Kumar Anand, Leonardo Jose, Mejino V., Cook Daniel L., Detwiler Landon T., Smith Barry: "A Strategy for Improving and Integrating Biomedical Ontologies", AMIA 2005, Washington D.C., USA
- Schulz Stefan, Daumke Philipp, Smith Barry, Hahn Udo: "How to Distinguish Parthood from Location in Bioontologies", AMIA 2005, Washington D.C., USA
- Patrick Ruch, Robert Baud, Christine Chichester, Antoine Geissbühler, Frédérique Lisacek, Johann Marty, Dietrich Rebolz-Schuhmann, Imad Tbahriti, Anne-Lise Veuthey. Extracting Key Sentences with Latent Argumentative Structuring. Proceedings of MIE2005, Vol. 116, 2005, pp.1052.
- William Hersh, Jeffery Jensen, Henning Müller, Paul Gorman, Patrick Ruch (2005). Trans-Atlantic Collaboration for Evaluating Image Retrieval Systems in the ImageCLEF Biomedical Image Retrieval Task. Workshop on NSF/FP6 cooperations at the 2005 eChallenges conference, Ljubljana, Slovenia, October 2005
- M Donnelly, T Bittner, C Rosse. A formal theory for spatial representation and reasoning in biomedical ontologies. AI Medicine 2006 ; 36 : 1-28.
- Website: <http://www.morphosaurus.net>, updated in August 2005
- Website: <http://www.semanticmining.org>, updated regularly



References 2006 (co-authored papers only – authors from more than one partner)

Rector, AL, Qamar R, Marley T, Binding Ontologies & Coding Systems to Electronic Health Records and Messages. Presented at KR-Med 2006.

Towards an Upper Level Ontology for Molecular Biology, Presented at AMIA 2006

Ontological and Epistemological Aspects of the Denotation of Biological Statements, Presented at KR-MED 2006

From GENIA to BIOTOP – Towards a top-level ontology for biology. Presented at FOIS 2006

Schulz S, Hanser S, Hahn U, Rogers J: The Semantics of Procedures and Diseases in SNOMED® CT. *Method Inform Med*, 2006, in press.

Schulz S, Kumar A, Bittner T: Biomedical ontologies: What part-of is and isn't. *J Biomed Inform*, 2006; 39 (3) : 350-361

Zaiss A, Hanser S, Schulz S: Mapping of ICHI to CCAM Basic Coding System. In: Reichert A, Mihalas G, Stoicu-Tivadar L, Schulz S, Engelbrecht R (Hrsg): *Integrating Biomedical Information: From e-Cell to e-Patient. Proc European Federation for Medical Informatics Special Topic Conference, April 6-8, 2006, Timișoara, Romania*. Berlin: Akademische Verlagsgesellschaft 2006; 281-292

Rector, A, Rogers, J, and Bittner, T, Granularity, scale & collectivity: When size does and does not matter. *Journal of Biomedical Informatics*, 2006. 39(3): p. 333-349

Fellbaum Christiane, Hahn Udo, Smith Barry: Towards New Information Resources for Public Health – From WordNet to Medical WordNet (<http://ontology.buffalo.edu/MFN/MWN-JBI.pdf>), *Journal of Biomedical Informatics*, 39 (2006) 3, 321-332.

Whetzel Patricia, Brinkman Ryan, Causton Helen, Fan Liju, Fostel Jennifer, Fragoso Gilberto, Heiskanen Mervi, Hernandez-Boussard Tina, Morrison Norman, Parkinson Helen, Rocca-Serra Philippe, Sansone Susanna-Assunta, Schober Daniel, Smith Barry, Stevens Robert, Stoeckert Chris, Taylor Chris, White Joe: The Development of FuGO – An Ontology for Functional Genomics Experiments, in: *Omics: A Journal of Integrative Biology*, 10(2) 2006, 199-204.

Erik Sundvall, Rahil Qamar, Mikael Nyström, Mattias Forss, Håkan Petersson, Hans Åhlfeldt, Alan Rector. Integration of Tools for Binding Archetypes to SNOMED CT. *Proceedings of SMCS2006; 1-3 Oct, Copenhagen, Denmark*, p 64-68; <http://www.hiww.org/smcs2006/proceedings/12SundvallSMCS2006final.pdf>

Erik Sundvall, Mikael Nyström, Mattias Forss, Rong Chen, Håkan Petersson, Hans Åhlfeldt. Graphical Overview and Navigation of Electronic Health Records in a prototyping environment using Google Earth and openEHR Archetypes. *Proceedings of MIE'06 - Studies in Health Technology and Informatics*.

Mikael Nyström, Magnus Merkel, Håkan Petersson, Hans Åhlfeldt. Evaluating bilingual medical terminologies with word alignment methods. Submitted to MEDINFO 2007.

Dietrich Rebholz-Schuhmann, Graham Cameron, Dominic Clark, Francesco Beltrame, Jean-Louis Coatrieux, Eva Del Hoyo Barbolla, Fernando Martin-Sanchez, Luciano Milanesi, Ioannis Tollis, Erik van Mullighan, Johan van der Lei. SYMBiomatics: Synergies in Medical Informatics and Bioinformatics – exploring current scientific literature for emerging topics. BITS 2006, Bologna. *BMC Bioinformatics* (Accepted for publication)



Rebholz-Schuhmann,D., Kirsch,H., Arregui,M., Gaudan,S., Riethoven,M., and Stoehr,P.
Protein Annotation by EBIMed. *Nat Biotechnol* 2006; 24(8):902-3.

M. Poprat, K. Markó & U. Hahn. A Language Classifier that Automatically Divides Medical Documents for Experts and Health Care Consumers In: *MIE 2006 – Proceedings of the 20th International Congress of the European Federation for Medical Informatics*. Ubiquity: Technologies for Better Health in Aging Societies. Ed. by A. Hasman, R. Haux, J. van der Lei, E. De Clercq, F. H. Roger France. Maastricht, The Netherlands, 27-30 August 2006. Amsterdam etc.: IOS Press, 2006, pp.503-508 (Studies in Health Technology and Informatics, 124)

K. Markó, P. Daumke & U. Hahn Cross-Lingual Alignment of Biomedical Acronyms and Their Expansions In: *MIE 2006 – Proceedings of the 20th International Congress of the European Federation for Medical Informatics*. Ubiquity: Technologies for Better Health in Aging Societies. Ed. by A. Hasman, R. Haux, J. van der Lei, E. De Clercq, F. H. Roger France. Maastricht, The Netherlands, 27-30 August 2006. Amsterdam etc.: IOS Press, 2006, pp.857-862 (Studies in Health Technology and Informatics, 124)

K. Markó, S. Schulz, U. Hahn: Automatic lexeme acquisition for a multilingual medical subword thesaurus. *International Journal of Medical Informatics (IJMI)*. 2006 (ePublication)

K. Markó, P. Daumke, U. Hahn: Cross-Lingual Alignment of Biomedical Acronyms and their Expansions. *Proceedings of the XX International Congress of the European Federation for Medical Informatics (MIE '06)*, Maastricht, Netherlands. 2006: 857-862.

M. Poprat, K. Markó, U. Hahn: A Language Classifier that Automatically Divides Medical Documents for Experts and Health Care Consumers. *Proceedings of the XX International Congress of the European Federation for Medical Informatics (MIE '06)*, Maastricht, Netherlands. 2006: 503-508.

P. Zweigenbaum, S. Schulz, P. Ruch, editors. *LREC Workshop Acquiring and representing multilingual, specialized lexicons: the case of biomedicine*, Genova, Italy, 2006. ELDA. Program available at: <http://estime.spim.jussieu.fr/~pz/lrec2006/programme-specialized-lexicons.html>

L. Deléger, M. Merkel, & P. Zweigenbaum. Using word alignment to extend multilingual medical terminologies. *Zweigenbaum et al. 5th International Conference on Language Resources and Evaluation (LREC '06): Workshop on Acquiring and Representing Multilingual, Specialized Lexicons*, Genoa, Italy. 2006.

P. Daumke, S. Schulz, K. Markó: Subword Approach For Acquiring and Cross-Linking Multilingual Specialized Lexicons. *Zweigenbaum et al. 5th International Conference on Language Resources and Evaluation (LREC '06): Workshop on Acquiring and Representing Multilingual, Specialized Lexicons*, Genoa, Italy. 2006.

K. Markó, R. Baud, P. Zweigenbaum, M. Merkel, M. Toporowska-Gronostaj, D. Kokkinakis, S. Schulz: Cross-Lingual Alignment of Medical Lexicons. *Zweigenbaum et al. 5th International Conference on Language Resources and Evaluation (LREC '06): Workshop on Acquiring and Representing Multilingual, Specialized Lexicons*, Genoa, Italy. 2006.

S. Schulz, K. Markó, P. Daumke, U. Hahn, S. Hanser, P. Nohama, R. Andrade, E. Pacheco, M. Romacker: Semantic Atomicity and Multilinguality in the Medical Domain: Design Considerations for the MorphoSaurus Subword Lexicon. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC '06)*, Genoa, Italy. 2006: 1684-1687.



-
- O. Medelyan, S. Schulz, J. Paetzold, M. Poprat, K. Markó: Language Specific and Topic Focused Web Crawling. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC '06), Genoa, Italy. 2006: 865-868.
- K. Markó, R. Baud, P. Zweigenbaum, M. Merkel, L. Borin, S. Schulz: Towards a Multilingual Medical Lexicon. In Proc AMIA Annual Fall Symposium 2006, Washington, DC, November 2006, 534 - 538.
- L. Deléger, M. Merkel, P. Zweigenbaum. Contribution to terminology internationalization by word alignment in parallel corpora. In Proc AMIA Annual Fall Symposium 2006, Washington, DC, November 2006, 185-189.
- L. Deléger, M. Merkel, P. Zweigenbaum. Enriching medical terminologies: an approach based on aligned corpora. In Medical Informatics Europe, 2006. Stud Health Technol Inform. 2006;124:747-52.
- M. Nyström, M. Merkel, L. Ahrenberg, P. Zweigenbaum, H. Petersson, H. Åhlfeldt. Creating a medical English-Swedish dictionary using interactive word alignment. BMC Medical Informatics and Decision Making, 2006. 6:35, doi:10.1186/1472-6947-6-35
- J. Wermter, K. Tomanek, F. Balzer. Automatische Erkennung und effiziente Annotation von anonymisierungsrelevanten Begriffen in klinischen Freitexten. GMDS 2006, Leipzig. Düsseldorf, Köln: German Medical Science; 2006. Doc 06gmds128