

PROJECT LITERATURE CITED

1. Open Health Natural Language Processing (OHNLP) Consortium, www.ohnlp.org, Last Access Date: January 20, 2010
2. clinical Text Analysis and Knowledge Extraction System (cTAKES), www.ohnlp.org, Last Access Date: January 20, 2010
3. Clinical Element Model (CEM), www.clinicalelement.com, Last Access Date: January 20, 2010
4. Chute C, Beck S, Fisk T, *et al.*: The Enterprise Data Trust at Mayo Clinic: A semantically integrated warehouse of biomedical data. . *JAMIA* in press
5. Health Open Source Software Collaborative, <https://mi.regenstrief.org/wiki/display/hoss/Health+Open+Source+Software+Collaborative;jsessionId=C5FA8654DE95870C84B9925C66082FC8>, Last Access Date: January 20, 2010
6. LexBig and LexEVS, https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexBig_and_LexEVS, Last Access Date: January 20, 2010
7. National Center For Biomedical Ontology, <http://www.bioontology.org>, Last Access Date: January 20, 2010
8. Pathak J, Solbrig HR, Buntrock JD, *et al.*: LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime. *Journal of American Medical Informatics Association* 16:305-315, 2009
9. MirthConnect, www.mirthcorp.com/community/mirth-connect, Last Access Date: January 20, 2010
10. Institute (ANSI) of its Common Terminology Services (CTS): ANSI/HL7 CTS, V1-2005 Health Level Seven Standard: Common Terminology Services, Version 1, 2005
11. International Organization for Standardization: ISO International Standard (IS) 27951 Common Terminology Services Version 1, 2009

12. Noy NF, Shah NH, Whetzel PL, *et al.*: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 37:W170-173, 2009
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19483092
13. Savova G, Masanz J, Ogren P, *et al.*: Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, evaluation and applications. *JAMIA*, under review
14. Chapman W, Dowling J, Hripcsak G: Evaluation of Training with an Annotation Schema for Manual Annotation of Clinical Conditions from Emergency Department Reports. *Int J Med Inf* 77:107-113, 2008
15. Chapman W, Dowling J, Wagner M: Generating a reliable reference standard set for syndromic case classification. *J Am Med Inform Assoc* 12:618-629, 2005
16. Coden A, Savova G, Sominsky I, *et al.*: Automatically extracting cancer disease characteristics from pathology reports into a cancer disease knowledge model. *Journal of Biomedical Informatics* 42 (2009):937-949, 2009, doi:10.1016/j.jbi.2008.12.005
17. Ogren P, Savova G, Chute C: Constructing evaluation corpora for automated clinical named entity recognition, in *LREC*, Marakesh, Morrocco, 2008, pp 3143-3150, <http://www.lrec-conf.org/proceedings/lrec2008/>
18. Savova G, Bethard S, Styler W, *et al.*: Towards temporal relation discovery from the clinical narrative, in *AMIA*, San Francisco, CA, 2009
19. Uzuner Ö: Recognizing Obesity and Co-morbidities in Sparse Data. *Journal of the American Medical Informatics Association*. 16:561-570, 2009
20. Uzuner Ö, Goldstein I, Luo Y, *et al.*: Identifying Patient Smoking Status from Medical Discharge Records. *Journal of the American Medical Informatics Association* 15:14-24, 2008
21. Uzuner Ö, Luo T, Szolovits P: Evaluating the State-of-the-Art in Automatic De-identification. *Journal of the American Medical Informatics Association* 14:550-563, 2007

22. Chen J, Schein A, Ungar L, *et al.*: An Empirical Study of the Behavior of Active Learning for Word Sense Disambiguation, in *Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL)*, New York, NY, 2006
23. CLEAR-TK, <http://code.google.com/p/cleartk/> Last Access Date: January 20, 1999
24. Palmer M, Gildea D, Kingsbury P: The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics* 31, 2005
25. Pradhan S, Hacioglu K, Krugler V, *et al.*: Support vector learning for semantic argument classification. *Machine Learning* 60:11-39, 2005
26. Hacioglu K, Pradhan S, Ward W, *et al.*: Semantic Role Labeling by Tagging Syntactic Chunks, in *Proceedings of the Eighth Conference on Natural Language Learning (CONLL-2004)*, 2004
27. Bethard S, Lu Z, Martin J, *et al.*: Semantic Role Labeling for Protein Transport Predicates. *BMC Bioinformatics* Jun 11:9:277, 2008
28. Bethard S, Martin J, Klingenstein S: Finding Temporal Structure in Text: Machine Learning of Syntactic Temporal Relations. *International Journal of Semantic Computing (IJSC)* 1, 2007
29. Chapman W, Bridewell W, Hanbury P, *et al.*: A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics* 34:301-310, 2001
30. Harkema H, Thornblade T, Dowling J, *et al.*: ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *J Biomed Inform* 42:839-851, 2009
31. Chapman W, Chu D, Dowling J: ConText: An algorithm for identifying contextual features from clinical text., in *BioNLP Workshop of the Association for Computational Linguistics*, Prague, Czech Republic, 2007, pp 81-88
32. Christensen L, Harkema H, Irwin J, *et al.*: ONYX: A System for the Semantic Analysis of Clinical Text, in *Proceedings of the BioNLP2009 Workshop of the ACL Conference*, Denver, CO, 2009

33. Aronsky D, Fiszman M, Chapman WW, *et al.*: Combining decision support methodologies to diagnose pneumonia. *Proc AMIA Symp*:12-16, 2001
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11825148
34. He T: Coreference Resolution on Entities and Events for Hospital Discharge Summaries, in *EECS*, Cambridge, MA, MIT. M.Eng, 2007
35. Sibanda T: Was the Patient Cured? Understanding Semantic Categories and Their Relationships in Patient Records, in *EECS*, Cambridge, MA, MIT, 2006
36. Uzuner Ö, Mailoa J, Sibanda T: Semantic Relations for Problem-Oriented Medical Records, in *Fall Symposium of the American Medical Informatics Association (AMIA 2009)*, San Francisco, CA, 2009, p 661
37. Uzuner Ö, Zhang X, Sibanda T: Two Approaches to Assertion Classification, in *Fall Symposium of the American Medical Informatics Association (AMIA 2008)*, Washington, DC, 2008, p 752
38. Uzuner Ö, Zhang X, Sibanda T: Machine Learning and Rule-based Approaches to Assertion Classification. *Journal of the American Medical Informatics Association* 16:109-115, 2009, DOI 10.1197/jamia.M2950
39. Unstructured Information Management Architecture (UIMA), <http://incubator.apache.org/uima/>
Last Access Date:January 20, 2010
40. HL7, www.hl7.org/v3ballot/html/welcome/environment/index.htm, Last Access Date: January 20, 2010
41. Poesio M, Vieira R: A corpus-based investigation of definite description use. *Computational Linguistics* 24:183-216, 1998
42. Hripcsak G, Rothschild A: Agreement, the F-Measure, and Reliability in Information Retrieval. *J American Medical Informatics Association* 12:296-298, 2005
43. Marcus M, Santorini B, Marcinkiewicz M: Building a large annotated corpus of english: The penn treebank. *Computational Linguistics* 19:313-330, 1994

44. Kipper K, Korhonen A, Ryant N, *et al.*: Extensive Classifications of English verbs., in *Proceedings of the 12th EURALEX International Congress.*, Turin, Italy, 2006
45. Uzuner Ö, Sibanda T, Luo Y, *et al.*: A De-identifier for Medical Discharge Summaries. *International Journal Artificial Intelligence in Medicine* 42:13-35, 2008
46. Sowa J: Conceptual graphs for a database inference. *IBM Journal of Research and Development* 20:336-357, 1976
47. Sowa J: *Conceptual structures: information processing in mind and machine.* Reading, MA, 1984
48. The eMERGE Network,
https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page, Last Access Date: January 20, 2010
49. CDISC, <http://www.cdisc.org>, Last Access Date: January 20, 2010
50. Biomedical Research Integrated Domain Group (BRIDG), <http://www.bridgmodel.org>, Last Access Date: January 20, 2010
51. Liu B, Hsu W, Ma Y: Integrating classification and association rule mining, in *Intelligence AaFA*, New York, 1998
52. Thabtah F: A review of associative classification mining. *Knowledge Engineering Review* 22:37-65, 2007
53. Wei W, Chute C: Identification of Type 2 Diabetes Mellitus Patients by SNOMED CT Concept Frequency. . *AMIA Annual Symposium*, 2009
54. CDISC Share, <http://www.cdisc.org/cdisc-share>, Last Access Date: January 20, 2010
55. LexWiki, <https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexWiki>, Last Access Date: January 21, 2010
56. International Health Terminology Standards Development Organization (IHTSDO),
http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/About_IHTSDO/Publications/CompositionalGrammar_20081223.pdf, Last Access Date: January 20, 2010

57. Rector AL, Brandt S: Why do it the hard way? The case for an expressive description logic for SNOMED. *J Am Med Inform Assoc* 15:744-751, 2008,
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18755993
58. caBIG® Vocabulary Knowledge Center, https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/Main_Page. Last Access Date: January 21, 2010
59. Agrawal R, Imielinkski T, Swami A: Mining Associations between Sets of Items in Large Databases. *ACM SIGMOD Int'l Conf on Management of Data*: Washington, DC, 1993.
60. Witten I, Frank E: *Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations*, Morgan Kaufmann Publishers, 2000
61. Apache Open Source, <http://incubator.apache.org/>, Last Access Date: January 20, 2010
62. Apache UIMA, <http://incubator.apache.org/uima/>, Last Access Date: January 20, 2010
63. OASIS, <http://www.oasis-open.org/news/oasis-news-2009-03-19.php>, Last Access Date: January 20, 2010
64. Text Analytics Tools and Runtime for IBM LanguageWare,
<http://www.alphaworks.ibm.com/tech/lrw>, Last Access Date: January 20, 2010
65. Open Health Natural Language Processing (OHNLP) Consortium, <https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/OHNLP>, Last Access Date: January 20, 2010
66. U-Compare, <http://u-compare.org/>, Last Access Date: January 20, 2010
67. Getting Started: UIMA Asynchronous Scaleout, <http://incubator.apache.org/uima/doc-uimaas-what.html>, Last Access Date: January 20, 2010
68. “Question Answering” is technology's next grand challenge,
<http://www.research.ibm.com/deepqa/index.shtml>, Last Access Date: January 20, 2010
69. Rubin D: Inference and missing data. *Biometrika* 63:581-592, 1976,
70. rJAVA, <http://rosuda.org/rJava/> Last Access Date: January 21, 2010

71. Melton LJ, 3rd: History of the Rochester Epidemiology Project. *Mayo Clin Proc* 71:266-274, 1996,
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8594285
72. Kurland LT, Molgaard CA: The patient record in epidemiology. *Sci Am* 245:54-63, 1981,
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7027437
73. Project Management Institute: *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) (4th ed.)*. Newtown Square, PA Project Management Institute, Inc., 2008
74. Blue Gene, <http://www.research.ibm.com/bluegene>. Last Access Date: January 21, 2010