

Large Scale Semantic Annotation of Radiology Reports

Malika Mahoui, School of Informatics; Vinay Kashyap, Informatics Student; Patrick Jamieson M.D., Logical Semantics, Inc.; Josette Jones R.N., Ph.D., School of Informatics; Jeffrey Friedlin, D.O, Regenstrief Institute

The development and testing of automated information extraction (IE) systems depends on semantically annotated free text. This presentation reports on the results of a large scale annotation project of a radiology corpus, the Roentgen corpus, consisting of 594,000 de-identified radiology reports with 36 million words, and 4.3 million sentences supplied by Indiana University. The presentation highlights the (1) sentence-based approach in defining propositions annotating the corpus, (2) as well as the annotation framework that is incrementally built and refined in order to facilitate the process of annotation.