

Delroy Cameron
Ohio Center for Excellence in Knowledge-enabled Computing (kno.e.sis Center)
College of Engineering and Computer Science
Wright State University
3640 Colonel Glenn Hwy., Dayton, OH 45435
Webpage: <http://knoesis.wright.edu/researchers/delroy>
Email: delroy@knoesis.org

RESEARCH SUMMARY

A critical direction for the progression of Semantic Web technologies is adoption and application by multi-disciplinary fields. Clinical, Biomedical and Sensor research are among pioneering disciplines in the Health Care and Life Sciences (HCLS) domain that have already exploited semantic web standards to facilitate data representation, interoperability and reasoning. In my PhD research, I have developed an interest in applying Semantic Web techniques to multi-disciplinary fields, to enhance information extraction and knowledge discovery. The ultimate goal is to automate knowledge discovery instead of the many cumbersome existing approaches.

PREDOSE: PREscription Drug abuse Online-Surveillance and Epidemiology Project

My current research project is a multi-disciplinary project between the Center for Interventions, Treatment and Addictions Research (CITAR) at the School of Medicine at Wright State University and the Kno.e.sis Center. PREDOSE is geared at identifying and understanding human behavior and attitudes related to pharmaceutical drug abuse, using web forum data. In preliminary work, we manually collected triples from qualitative interviews and infused them into a Semantic Browser, (called Scooner) to demonstrate the benefits of text exploration driven by semantics-based features. This preliminary work led to an NIH/R21 grant (400k, 2yrs, July 2011 - July 2013) awarded to Kno.e.sis and CITAR. The R21 was preceded by a Department of Community Health grant (which supports my research assistantship) also awarded jointly to Kno.e.sis and CITAR.

We have since built a comprehensive lexicon of drug references from both slang sources as well as standard terminologies such as the UMLS and Drugbank. Using this lexicon we have done some preliminary entity identification and are exploring ways of automating the entity disambiguation that is now necessary. Relationship extraction and triple extraction are subsequent steps. As we have already manually created an ontology schema using knowledge from existing structured sources and the specific project needs, we foresee the ontology playing a critical role in the ensuing phases of information extraction. On the second dimension of the project which centers around Trend Detection, we have recently made a somewhat significant discovery in which addicts have been found to abuse a common prescription drug. This information appears to be unknown by experts within the community. By leveraging a combination of sentiment analysis, entity identification and mood analysis we were able to slice data more effectively, thus leading to this outcome. This work is in the process of being submitted for publication.

Obvio: Literature-Based Knowledge Discovery (LBD) using Semantic Predications

My second research project involves analyzing RDF graphs to determine how assertions in a graph are connected across text documents that answer complex biomedical questions. The prevailing notion is that if assertions in a graph are connected across documents containing undiscovered knowledge, then knowledge discovery becomes a matter of restricting the search space to the relevant

subgraph covering the documents. Furthermore, this analysis ought to be completely graph-based and void of any syntactic document analysis, which is typical of classical information retrieval systems. To investigate this idea, we rely on a tool called SemRep¹ which extracts assertions (called semantic predications) from biomedical literature, maintaining the provenance information that binds each predication to its source document. The semantic predications together with facts from the UMLS Metathesaurus (a structured knowledge source maintained by NLM) make up the Biomedical Knowledge Repository (BKR). The BKR consists of more than 100 million triples, all with provenance information. In demonstrating connectivity among facts in the graph given a query and its relevant answer set, we first aim to establish high recall. To achieve this, we use the notion of *knowledge abstraction* to make logical leaps between concepts in the answer set, when direct connections between predications do not exist explicitly in the answer set. Second, we aim to obtain reasonably high precision in retrieving answer set documents. To do this, we eliminate irrelevant documents by computing a measure of *semantic relatedness* between a document, a predication and a given query. Semantic relatedness is measured by exploring the overlap in concept neighborhood from the graph. For this, we consider some domain specific attributes, such as Semantic Network Type, Semantic Network Group, Hierarchical and Associative concept proximity. This work was recently published in the 2011 Bioinformatics and Biomedicine (BIBM) Conference.

Scooner: Semantic Browser

Prior to my research on knowledge discovery using semantic predications at NLM, I was lead developer and subsequent team leader on the Trellis project (now Scooner²). Scooner is a user-driven tool for knowledge discovery using semantic predications, UMLS metathesaurus relations and previously, facts from the Linked Open Data (LoD). It is currently deployed as a web application for text exploration, that facilitates knowledge discovery, primarily from biomedical literature (i.e. Medline documents). Scooner superimposes background knowledge onto text by anchoring spotted entities in text with entities in the graph. Hence, as users browse documents in the search result page of the application, their navigation is guided by the predicates that connect entities in one document to entities in another. Scooner therefore facilitates entity-to-entity navigation through the semantics of the predicates between them. Additional features such as bookmarking, search result reranking, semantic trail aggregation and collaborative information sharing make the browser a useful tool for user-driven knowledge discovery if used by experts for document browsing. In fact, Scooner is used exactly for this purpose. An active prototype³ of the system has been deployed and is still in use by a group of more than 15 biomedical researchers at the Wright-Patterson Air Force Research Lab (AFRL) here in Dayton, Ohio.

¹SemRep - a tool created by Thomas Rindflesch at the National Library of Medicine for extrating triples from biomedical literature. Additional SemRep Literature - <http://skr.nlm.nih.gov/papers/index.shtml#SemRep>

²Scooner Demo - <http://knoesis.org/library/demos/scooner-demo/>

³Scooner prototype - <http://knoesis1.wright.edu/scooner-beta> (guest account is "knoesis")