

## Sujan Perera

**Phone:** 937-830-8087

**e-mail:** [sujan\[at\]knoesis.org](mailto:sujan[at]knoesis.org)

**Home Page:** <http://knoesis.org/researchers/sujan>

**Blog:** <http://sujan-udayanga.blogspot.com>

### RESEARCH INTERESTS

Knowledge Enhanced Natural Language Understanding and Text Mining, Knowledge Acquisition and Representation, Semantic Reasoning.

### RESEARCH EXPERIENCE

**Research Assistant** [Kno.e.sis Center](#), Wright State University Fall 2010 - Present

- Exploring the potentials of using domain knowledge for natural language understanding. This research is focused on extracting implicit information in the unstructured clinical notes.
- Developing techniques to, assess the completeness of the coded domain knowledge, identify the missing knowledge in the knowledge bases and acquire missing knowledge efficiently.
- Collaborating with an industrial engineering professor to enhance the current predictive models developed for proactive patient care. The input data for the current predictive models is proven to be inaccurate and incomplete. The main goal of this research is to enhance the input data being used for the current predictive models by mining the unstructured patient records.

**Summer Research Intern** IBM Watson Summer 2014, Summer 2015

- Worked with the IBM Watson Discovery Adviser for Healthcare and Life Sciences Team on finding complex, meaningful, and interesting relationships between biological entities.

**Summer Research Intern** [ezDI Inc](#) Summer 2012

- Worked on the problem of extracting meaningful information from the unstructured patient health records. This work built upon the cTAKES, a natural language processing engine to process clinical documents, and developed a semantic search engine for unstructured patient records by using semantic web technologies (RDF/RDFS/SPARQL).
- Developed a comprehensive knowledge base covering the healthcare domain to improve the machine understanding of the patient records. This knowledge base is used as the back bone of the search engine to answer complex queries by using semantic reasoning.
- The continuous collaboration with ezDI resulted in developing a computer assisted coding (CAC) product for 10<sup>th</sup> version of International Classification of Diseases (ICD 10). ezDI launched this product at annual convention of American Health Information Management Association (AHIMA) 2013.

### Undergraduate Research

- Worked on developing an [abnormal pattern detection on time series data by adapting the concepts from natural immune system](#). This project was selected for final 10 competitors for undergraduate category in National Best Quality Software Awards 2008 in Sri Lanka.

### PUBLICATIONS

- **Sujan Perera**, Pablo Mendes, Amit Sheth, Krishnaprasad Thirunarayan, Adarsh Alex, Christopher Heid, Greg Mott, *“Implicit Entity Recognition in Clinical Documents”*, Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (\*SEM) 2015, pp. 228-238. [\[Download\]](#)
- **Sujan Perera**, Cory Henson, Krishnaprasad Thirunarayan, Amit Sheth, *“Semantics Driven Approach for Knowledge Acquisition from EMRs”*, IEEE Journal of Biomedical and Health Informatics, vol.18, no.2, pp.515-524, March 2014 [\[Download\]](#)

- **Sujan Perera**, Cory Henson, Krishnaprasad Thirunarayan, Amit Sheth, “*Data Driven Knowledge Acquisition Method for Domain Knowledge Enrichment in the Healthcare*”, 6th International Conference on Bioinformatics and Biomedicine BIBM12, Philadelphia, 4-7 Oct, 2012, pp. 1-8 [Download]
- **Sujan Perera**, Amit Sheth, Krishnaprasad Thirunarayan, Suhas Nair, Neil Shah, ‘*Challenges in Understanding Clinical Notes: Why NLP Engines Fall Short and Where Background Knowledge Can Help*’, International Workshop on Data management & Analytics for healthcaRE at ACM Conference of Information and Knowledge Management (CIKM), pp. 21-26, Burlingame, USA, Nov 1, 2013.[Download]
- Daniulaityte R, Carlson R, Falck R, Cameron D, **Perera S**, Chen L and Sheth A. ‘*I just wanted to tell you that loperamide WILL WORK*’: A Web-Based Study of Extra-Medical Use of Loperamide. Journal of Drug and Alcohol Dependence, Nov 2012 [Download]

## PATENTS

- Pending patent application on ‘Mining Biological Networks to Explain and Rank Hypotheses.’

## RESEARCH PROJECTS

- **MIDAS**  
**MIDAS** is a collaborative project that aims to enhance the current capabilities of the Electronic Medical Record (EMR) System used by the Department of Defense. The broader goal includes enhancing the communication between the physicians and the patients who are currently limited to the regular visits, extracting and exploiting the knowledge buried in the medical records to identify more effective treatment plans, and empowering the patients by enabling them to monitor self health status.  
*My Contributions:*  
 Knowledge base development, extracting meaningful information from EMRs.
- **kHealth**  
**kHealth** is semantic platform which integrates patients data from passive and active sensors(including machine and human sensors) with domain ontologies. This integration allows to interpret data meaningfully, reasoning using domain ontologies and helps people to make decisions to improve health, wellness and fitness. This platform is being used to develop **mobile application** which aims to reduce preventable readmissions of patients with chronic heart failure.  
*My Contributions:*  
 Ontology Development, Develop Semantic Reasoning Capabilities.
- **PREDOSE**  
 The overall aim of **PREDOSE** is to develop automated techniques for web forum data analysis related to the illicit use of pharmaceutical opioids.  
*My Contributions:*  
 Ontology Development, Named Entity Identification, Implement Web Crawlers.

## EDUCATION

- **PhD student**  
 Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Computer Science and Engineering Wright State University, OH  
 Aug 2010 - Present  
 Advisor : Dr. Amit Sheth  
 GPA: 4.0
- **Bachelor of Science in Computer Science**  
 University of Colombo School of Computing (UCSC), Sri Lanka.  
 July 2004 - Aug 2008  
 GPA: 3.58

## PROFESSIONAL ACHIEVEMENTS AND AWARDS

- Won the George Thomas Post-Graduate fellowship for the academic year 2016
- One of the student contributors for the recently funded NIH R01 proposal on [monitoring cannabis and synthetic cannabis use in social media data](#).
- Won NSF travel awards to attend BIBM 2012 and ICHI 2015 conferences.
- External Reviewer : ISWC, ESWC, IJSWIS, AAAI, IEEE Intelligent Systems, Applied Ontology Journal, ODBASE.
- Program Committee Member : EKAW 2014, ISWC 2015, IJCAI 2016, International Workshop on Data management & Analytics for healthcaRE, co-located with ACM CIKM 2013
- Sun Certified Programmer for the *Java*<sup>TM</sup> 2 Platform 1.4

## PROFESSIONAL EXPERIENCE

- Software engineer (Aug 2008 to Aug 2010)  
[AePONA Inc](#) (acquired by Intel in 2013) - AePONA is a leading solution provider in the Network as a Service (NaaS) marketplace.
- Software Engineer(Intern) (Feb 2007 to Aug 2007)  
[epic Lanka technologies \(pvt\) Ltd](#) - epic is one of the leading security solution providers in Sri Lanka

## COMPUTER SKILLS

- Semantic Web Technologies: *RDFS, OWL, SPARQL, SPIN*
- Programming Languages: *Java, Python*
- Web Programming: *HTML, JavaScript, JSP, Java Servlet*
- Data Storages: *Oracle, MySql, Virtuoso*
- Other Technologies: *Shell Script, Ant, Ajax, MapReduce, git, NLTK, UIMA*

## SELECTED GRAD LEVEL COURSES

- Information Retrieval
- Data Mining
- Knowledge Representation
- Semantic Web
- Machine Learning

## REFERENCES

- Amit Sheth (Advisor)  
Director, Kno.e.sis Center  
Wright State University  
Dayton, OH  
Email: amit[at]knoesis.org
- Krishnaprasad Thirunarayan  
Professor, Kno.e.sis Center  
Wright State University  
Dayton, OH  
Email: tkprasad[at]knoesis.org
- Pablo Mendes  
Research Staff Member,  
IBM Research Almaden  
Email: pnmendes[at]us.ibm.com