

DBpedia Metadata Datasets Project Proposal

Contact information

- Your Name: Mingzhe Du
- Email Address: mzdu99@gmail.com / dum@email.sc.edu
- Instant message: Google Hangout mzdu99@gmail.com
- Linkedin: <http://www.linkedin.com/in/mzdu99>
- Github: <https://github.com/mzdu>

Project

- Project Name: 5.10 DBpedia Metadata Datasets
- Project Description: DBpedia is a project to create a graph from Wikipedia data. Due to the limit of RDF notations, the extracted information is not structured natively at the start. The goal of this project is to create an extension of the DBpedia extraction framework which could extract the knowledge from Wikidata with the selected format(RDF,OWL or Wikidata model). This project could provide better quality, and well formatted data.
- Timeline:
 - Gathering information about different annotations models. 1 week
 - Clear the procedure, tools and methods for implementation. 1 week
 - Implement all functional requirements. 2 weeks
 - Prototype Testing and Gathering feedback. 1 week
 - Bug fixing. 1 week
 - Implement the project with more detailed functional requirements. 1-2 weeks
 - Testing and bug fixing 1-2 weeks.
 - Release application

Technical skills

- I am proficient in Python and Java. I have more than three years experience on Python and Python related web programming. I could well program with Javascript, jQuery, NoSQL database and relational database. I am a certified Java programmer since 2005.
- I have a good understanding of RDF, OWL and SPARQL. I gain my experience through my research and school courses like Natural Language Processing and Service Oriented Computing.

Open Source

- I have been proactively participating Open Source development and our local Palmetto Open Source Software Conference for years. I have seen a lot of smart people and stunning projects. It feels good that everybody could freely present and discuss their code and theories behind code. I really like this community, and I wish I could make more contributions to this community by working with DBpedia. Then I could utilize the skills and experiences to better develop Wikitheoria.com.

Background & education

University of South Carolina

Ph.D. student in Computer Science, 2011 - present

Columbia, SC ,US

University of South Carolina

Master of Science in Software Engineering, 2011

Columbia, SC, US

Tsinghua University (prior to transfer)

Master of Science in Software Engineering, 2009

Beijing, China

Harbin University of Science and Technology

Bachelor of Science in Software Engineering, 2007

Harbin, China

Research

- My current research involves two parts. The first part is to develop the website Wikitheoria.com with Python, Google App Engine with Jinja2 templates. The other part of my research is to experiment and implement text classification, ontology building with Natural Language ToolKit and Protégé on Wikitheoria.com. One book I like the most is Natural Language Processing with Python.
- Yes, I would like to co-author a conference paper with my mentors. That would be great.

Summer plans

- I will stay in Columbia, SC, United States during the whole summer. I expect to work 4 hours per day and 25-30 hours per week on this project. I will not take any class. Other than this GSoC project, I will spend the rest of my time on my own academy research. I may take a five days break at the end of this summer, but it is not decided yet.

GSoC Experience

- This is my first time to participate GSoC project. I decided to join this year's GSoC for several reasons. One important reason is that I really wish to see how RDF, OWL and those Natural Language Processing algorithms and methods are integrated collaborated with DBpedia. I wish I could proactively make contributions to the Linked Data community. I also believe the "DBpedia Metadata datasets" project will help me better understand the next step of my academy research project.

Why Me?

- I enjoy programming and learning new programming skills. I believe I could solve problems systematically and scientifically, then implement it with my solid programming skills.