DBpedia Metadata Datasets Project Proposal

Contact Information

- Name: Mingzhe Du
- Email Address: mzdu99@gmail.com
- Instant Messenge: (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 which extracts Wikipedia's structured data in the form of Linked Data ontologies to construct interconnected RDF standard graphs. 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: (My developing process follows Agile development.)
    - Gather information about different annotation models. 1 week
    - Analyze the procedure, tools and methods for implementation, prioritized the tasks. 1 week
    - Implement high-priority functional requirements. 2 weeks
    - Test the first prototype and gather feedback. 1 week
    - Bug fixing. 1 week
    - Implement the project with more detailed functional requirements. 1-2 weeks
    - Release beta version. 1 week
    - Test and fix bugs. 1 week
    - Release final application.

Technical Skills

- I am proficient in Python and Java. I could program well with Javascript, jQuery, NoSQL database and relational database. I have been a certified Java programmer since 2005. I am the principle developer of Wikitheoria.com, an NSF sponsored web-based crowd-sourcing tool to share and collaborate on sociological researchable ideas. The ultimate goal of this project is to contribute the well-structured sociology information and knowledge to our Linked Data community. I have also been using NodeJS, AngularJS and MongoDB during my development at HelpMonger.com.
- 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. Afterwards, I could utilize the skills and experiences I gained to better develop and extend Wikitheoria.com.

Background & Education

University of South Carolina Columbia, SC ,US
Ph.D. student in Computer Science, 2011 - present

University of South Carolina Columbia, SC, US
Tsinghua University (transferred to University of SC in 2009) Beijing, China
Master of Science in Software Engineering, 2011

Harbin University of Science and Technology Harbin, China
Bachelor of Science in Software Engineering, 2007

Research

- My current research involves two parts. The first part is to develop the website Wikitheoria.com with Python, Google App Engine plus 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 4-5 hours per day and 20-25 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 day break at the end of this summer, but it is not decided yet.

GSoC Experience

- This is my first time to participate in GSoC. I decided to join this year's GSoC for several reasons. One important reason is that I really wish to learn how RDF, OWL and Natural Language Processing algorithms and methods are cooperated and collaborated with DBpedia. I wish I could proactively make contributions to the Linked Data community. I also believe the "DBpedia Metadata Datasets" project will give me a deeper insight into the next stage of my academy research.

Why Me?

- During my years of study at school, I have learnt how to scientifically and systematically analyze problems. I plan to utilize my programming skills in developing solutions and contributing to linked data community.