Entity Resolution in the Presence of Constraints

Project Plan

Done By: Zeyu Shen (u4813857)
Supervisor: Dr. Qing Wang

Learning Objectives:
On the completion of the project, the following learning objectives are expected to achieve:
1. Have a good understanding for the literature of entity resolution techniques.
2. Implement a framework for entity resolution in the presence of constraints.

The project is divided to five phases:

1. Read relevant papers and learn necessary background knowledge, include entity resolution, constraints and dedupalog.

2. Set up working environment, typically eclipse and postgresQL.
   Set up right configuration to connect database. Understand the given package about similarity comparison.

3. Implement the algorithms that support the framework for entity resolution in the presence of constraints using Cora data.
   Incorporate weights for constraints into the framework.

4. Analyse the efficiency and effectiveness of the developed approach, and experimentally evaluate the approach using Cora data.

| Week 1 – Week 2 | Sign the study contract. Read the research thesis. | Phase 1 |
| Week 3 – Week 4 | Set up environment Conceptual understanding the algorithm Preliminary presentation. | Phase 2 |
| Week 5 – Week 7 | Implement the algorithm using Java | Phase 3 |
| Week 8 – Week 9 | Incorporate weights for constraints into the framework. Analyse the efficiency and effectiveness. Evaluate approach via experiment. | Phase 4 |
| Week 10 – Week 11 | Finalize the report. Prepare final presentation. | Phase 5 |
| Week 12 – Week 13 | Submit the report Final presentation. |  |

Report documenting is starting from Week 3. Efforts will be made to write report every week until submission.