Clean Coders

UNIVERSITY OF TEXAS AT DALLAS | 800 W CaMPBELL RD RICHARDSON TX 75080

Team Project Check Point

Semantic web



Table of Contents

[Group Information 2](#_Toc434605885)

[Details of team members 2](#_Toc434605886)

[Progress vs. Planned milestones 3](#_Toc434605887)

[Resolving Open Design Issues 3](#_Toc434605888)

[Results so far 3](#_Toc434605889)

# Group Information

Group Name: Clean Coders

Project Name: Relation extraction using semantic web technologies

Team Members: 4

## Details of team members

1. Ekal Golas – exg140230
2. Navaneeth Rao – nbv140130 (In a different section - 502)
3. Tejaswini Sirlapu – txs142130
4. Prashanth Govindaraj – pxg142030

# Progress vs. Planned milestones

Following are the milestones in the project and progress with respect to the planned milestones:-

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Planned Milestone** | **Progress** | **Status** |
| 25-Oct-15 | Coding the extractors and testing on some limited sample dataset | Used Stanford NLP and Apache Solr to extract the location-friends-organizations relationships | Done. |
| 1-Nov-15 | Describing the ontologies and structure of data by using Jena. | Used Dublin Core and FOAF ontologies. Still researching on whether we need a custom ontology. | Done. Needs revisiting |
| 8-Nov-15 | Use of SPARQL to derive results from data and present them in a user-friendly manner | Wrote few SPARQL queries and tested the results. To present these results in a user-friendly manner, the work on the UI has been started | In Progress |

## Resolving Open Design Issues

Following are the problem encountered and open design issues and their solutions:-

|  |  |
| --- | --- |
| **Issue** | **Solution** |
| Occurrence of pronouns in sentences | Co-reference resolution using Stanford Decoref API |
| Need of a custom ontology | Still researching. If required, we will have to define custom class hierarchies |

## Results so far

Following are the results obtained till now:-

1. Gave a sample set of sentences - "Satya Nadella is the CEO of Microsoft. Its headquarters is in Redmond, Washington. He was born in India."
2. Following output was observed for co-reference resolution

{1=CHAIN1-["Satya Nadella" in sentence 1, "the CEO of Microsoft" in sentence 1, "He" in sentence 3], 2=CHAIN2-["Microsoft" in sentence 1, "Its" in sentence 2], 4=CHAIN4-["Redmond" in sentence 2], 5=CHAIN5-["Its headquarters" in sentence 2], 7=CHAIN7-["Redmond , Washington" in sentence 2], 8=CHAIN8-["India" in sentence 3]}

1. Following output was observed for co-occurrence of named entities

Satya Nadella is the CEO of Microsoft.

Satya Nadella - PERSON

Microsoft - ORGANIZATION

Its headquarters is in Redmond, Washington.

Redmond - LOCATION

Washington - LOCATION

He was born in India.

India – LOCATION

1. Following is a screenshot for the home page of the UI designed:-

