**Samet Ayhan**

**10/28/2014**

**Lab #5**

**Part #1**

Fields I used and the comparison function:

fields = [

{'field' : 'title', 'type': 'String'},

{'field' : 'price', 'type': 'Custom', 'has missing':True, 'comparator' : customComparator},

{'field' : 'manufacturer', 'type': 'Exact', 'has missing' : True}

]

The precision and recall the algorithm achieves:

found duplicate

1286

precision

0.206065318818

recall

0.171521035599

For the evaluation, please see **product\_evaluation.py**

**Part #2.1**

For the script, please see **Assignment 5 - Part 2.ipynb**

File: news1.html

Recognized Named Entities:

Reuters , ORGANIZATION

Ralph Nader , PERSON

Apple Inc , PERSON

Tim Cook , PERSON

Wall Street Journal , FACILITY

File: news2.html

Recognized Named Entities:

Reuters , ORGANIZATION

Google Inc , PERSON

Larry Page , PERSON

Sundar Pichai , PERSON

File: news3.html

Recognized Named Entities:

Reuters , ORGANIZATION

Hewlett-Packard Co , ORGANIZATION

Wall Street Journal , FACILITY

File: news4.html

Recognized Named Entities:

Reuters , ORGANIZATION

White House , FACILITY

West Africa , LOCATION

Ebola , PERSON

File: news5.html

Recognized Named Entities:

Reuters , ORGANIZATION

Sevilla , PERSON

Barcelona , PERSON

La Liga , PERSON

File: news6.html

Recognized Named Entities:

Reuters , ORGANIZATION

Gonzalo Higuain , PERSON

Rafael Benitez , PERSON

File: news7.html

Recognized Named Entities:

Reuters , ORGANIZATION

PSV Eindhoven , ORGANIZATION

File: news8.html

Recognized Named Entities:

Reuters , ORGANIZATION

Michael Fallon , PERSON

File: news9.html

Recognized Named Entities:

Reuters , ORGANIZATION

Kobani , PERSON

Islamic , ORGANIZATION

File: news10.html

Recognized Named Entities:

Reuters , ORGANIZATION

Gaza , LOCATION

Sinai , ORGANIZATION

**Part #2.2**

For the script and the results, please see **Assignment 5 - Part 2.ipynb**