Homework 12 (Due at 11:59pm on 2015/04/17)

Instructions:

In this homework, you will need to write a program to extract name entities from online news and load them into a triple store.

1. Choose 2 recent special art related news you are interested in (e.g., from artnews.com, Google news). "Special" means to avoid potential overlap with other students. You can simply copy the texts of the news and save them as local files.
2. Your program should use the REST API from OpenCalais: (http://www.opencalais.com) to get the name entities in RDF and load them into a triple store.
3. Use SPARQL queries to extract the "**Person**", "**Organization**", "**City**" and "**GenericRelations**" from the triple store. You can type the queries in the OpenRDF workbench or implement them in your code. Calais Viewer (http://viewer.opencalais.com) is a good for you to verify your results. All the extract entities are shown on the left side of the page.
4. To evaluate the recognition results, you need to manually label all the "**Person**", "**Organization**", "**City**" and "**GenericRelations**" in the two articles. You may use your manually labeled data as ground truth to calculate the precision, recall and F1 score (http://en.wikipedia.org/wiki/Precision_and_recall) of the article. You need to explain in which cases the recognition performs poorly and well, and why (the possible causes).

Submission guideline:

1. **hw12.pdf**: a report including:
   - The two Web articles (URLs) you chose (10 points).
   - A table with two columns. The URI of the extracted entity and the text corresponding to it. (20 points)
   - A table with your labeled entities and entities generated by OpenCalais in a table as follows. You also need to mark whether each recognized entity is correct, a false negatives or false positive. (20 points)

	Manually Labeled Results	Calais Results
Person	*List your results here*	*List your results here*
Organization	…	…
City	…	…
GenericRelations	…	…

   - A description of the recognition performance in terms of precision, recall and F1 score. Explanations about explain in which cases the recognition performs poorly and well, and why (the possible causes). (20 points)

2. **You code** with concise comments. (30 points)

Zip all your files into **hw12_[firstname]_[lastname].zip** and submit on Blackboard.