Project Proposal Write-up (1-2 pages)

1. Names of the group members and UFIDs
2. Title of the project
3. A list of keywords of the project
4. Proposal should include the following topics:
   a. What is the **dataset** you are planning to use? The data needs to be **ready**. If you still need to collect dataset at the time of project proposal, it means that you will not have enough time to explore it and develop the algorithm/system/evaluation. This means that you need to already have details and statistics on schema, content, data quality, data size, data integration plans if multiple data sources are used and give data example. The dataset should be large enough to justify parallel processing systems for the application.
   
   b. What **application/processing/analysis** you plan to do? What **model/algorithm** you plan to apply? Groups are encouraged to use various data mining algorithms discussed in class, and models/algorithms related to text/image retrieval and extraction, knowledge construction and querying, rule learning and graph/RDF database querying and analytics. Please provide pointers to the algorithm/model you plan to implementing – we encourage using state-of-the-art existing algorithms.
   
   c. What **system** you plan to use? Why do you need such a framework based on the data and analysis you plan to run? You have a choice among AWS EMR (Map-Reduce/Hive/Spark), parallel DB, GraphLab/GraphX, graph/RDF databases and other parallel/HCP systems. Please specify and highlight whether you are going to use AWS and how – credits will be distributed accordingly. Please provide pointers to the system/framework you plan to use – we encourage using state-of-the-art existing systems.
   
   d. What is the **outcome** of the data processing or analysis? You can generate one or more of the following data product: patterns, knowledge, prediction, ad-hoc query answering system, visualizations. What is the evaluation plan and metrics? What is measure of a successful project?

5. The proposal should also include
   
   a. Why is this problem important? What are the related work, papers, and state-of-the-art? What are the challenges? What are solutions to these challenges?
   
   b. What are the novelties of this project? Novel project will receive bonus marks.

6. The deadline of project proposal is due via Sakai @ 12pm noon EST on 29th Sept (Monday). Data Science course project from previous year can be found: http://dsr.cise.ufl.edu/?page_id=732.