DIVA - CS526

Suggested Low Hanging Fruit Projects, Expected Project Outcomes, Project Design Methodology

By James Abello, www.mgvis.com

I. Low Hanging Fruit Projects

1. **CubeWall**: *An interface to Explore Multi-Faceted Data.*
   Reference: Pritish Sahu's MSCS Thesis, Computer Science Department, May 2017. Supervised by J. Abello.

2. **OEIS:** An Interface to explore the Online Encyclopedia of Integer Sequences-OEIS.
   See capstone flyer in MSCS web portal under the CAPSTONES TAB: mscs.rutgers.edu
   Reference: "Taming a Graph Hairball", Abello-Sun-Mackwither, 2018.

3. **Interactive Clustering**: Build on current work of Pranjal Awasti **on Interactive Machine Learning**.
   Reference: (Papers in Sakai)

4. **Data Stories**: Build on Current work on Culture Analytics.
   See capstone flyer in MSCS web portal under the CAPSTONES TAB: mscs.rutgers.edu
   Reference 1: *"Building in the Sea of the Great Unread",* Tim Tangherlini et al, to be posted on sakai
   Reference 2: "*Computational Folkloristics*", Abello, Broadwell, and Tangherlini, Communications of the *ACM, 2012.*
   Commentary: http://lab.softwarestudies.com/2012/07/computational-folkloristics.html

5. **MSCOIN**: Build an MSCS virtual currency and tools to navigate the Ledger (Ref: Nagamoto's paper). *See capstone flyer in MSCS web portal under the CAPSTONES TAB.*

6. **Graph Waves**: Build on current platform being worked on by Daniel Nakhimovih (CS PhD student).

Based on recent paper titled: *"***Atlas**: Local Graph Exploration in a Global Context. **James Abello***, Fred Hohman*, Varun Bezzam, Duen Horng (Polo) Chau.

7. **GPU Driven Processing**: Starting project using a new two GPU machine installed in Hill350.
8. For students interested in **Perception** devise a project to exemplify the essence of the two references below.
   a. *"Poetmotion 2", by Takahiro Kurashima, Lars Muller Publishers, 2014/2015.*
      Description. This is a book that contains only images. It contains only 12 lines of text. The author states:

"… If I am lucky, by playing with shapes of logic, I might understand a little bit more about the universe. And so, I will continue to make patterns."

b. "Semiology of Graphics", by Jacques Bertin (1967 in French), English Translation by W. Berg, ESRI Publisher, 1983.

9. For students interested in **Social Sciences**
   Select topic projects from the class reference book "Computational Social Sciences" **Edited by R.M. Alvarez, Cambridge University Press, 2016.**
10. **A variety of other types of projects can be found at glimpsed at ms.cs.rutgers.edu.**

II. Project Outcomes to be submitted on Sakai in Stages during the semester

a. **Latex Project Description** and associated pdf of the project following the template that will be made available on Canvas/sakai.
b. **One page summary description** to be posted on the MSCS/MSDS Portal if your project is successful.
c. **Power Point Presentation** according to the guidelines provided with the sample project presentation.
d. **Working Code plus documentation** including *installation procedures, *executables, *sample data
e. Short video (3'+) with or without audio.

Note: Best projects will be posted in the MSCS/MSDS Portal.

III. Project Design Methodology (by James Abello)

1. **DATA**
   a. Select Data Set –
      Identify Size, Format.
      Is the data at rest?
      Is it Streaming?
      Is the data "Time Variant data"?
      Is Time an essential component of the desired data processing?
   b. Identify central entities and relationships
   c. Identify basic Stats that that will be useful for processing: record size, distributions, etc. For Graphs (|V, |E|, cc, degree distributions, bicc, layers?, clusters?)
2. **QUESTIONS**
   a. What are the fundamental questions you want to answer about the data?
   b. What is the fundamental data representation that will allow the system and users to query extract information about the data?
   c. Create a name that faithfully describe the purpose of your Interactive System.

3. **MODE OF PROCESSING**
    a. How is the data going to be represented- manipulated internally?
    CSV file(s), SQL db?, NoSQL?, Graphs? HyperGraphs? Sets? Key-Value Pairs for MapReduce-Hadoop Processing?
    b. Data Store: Data Base + Network Representation?
    c. What questions can be answered via the type of Data Base Store being used?
    d. How is data input into the system after the system is set up initially? i.e. is the data Dynamic or is it Static?
4. **VISUAL REPRESENTATION**
    a. Plots: basic? Scatter? Labels? Colors? Adjustable size of basic elements?
    b. Matrix based?
    c. Node-Link based Representations
    d. Color, Texture
    e. How to link the different views
5. **INTERACTIVITY** (what is the level of interactivity desired? Definitely No more than one second per action: use look ahead to improve interactivity)
    a. **Interface Layout** : **Canvas,** Overview of the entire Data Space(Hierarchy Tree), Overall Sliders(one directional, bi-directional)
    b. **Interface Interaction mechanisms** and **answers representation** for further user interaction … (Textual, Graphic, Both?)
        a. Mouse Clicks, Mouse Hovering, Mouse Selection and their combinations
        b. Menu Driven Bottoms and Tabs
        c. Basic visual data elements controls : size, color, texture
        d. Fisheye Views
        e. Panning
        f. Zooming: Mouse control, Elastic Window
        g. Labeling: Zooming driven Hierarchical labeling
        h. Sliders(one sided, two sided)
        i. Linking of different Views
        j. User Annotations and Summary Constructors
6. **ANALYTICS and Evaluation**
    a. **Level of interactivity** achieved based on statistics collected by the system from **User Data Exploration Tasks Performance**.
    b. What have you **learned from the Data? i.e. summary construction based on targeted data samples for the** different User Data Exploration tasks.
    c. What are **the Data Characteristics** that will allow the **REUSE of your visual system interface. Give concrete examples.**
    d. What **type of Data is no suitable for the mode of exploration adopted by your system interface?**
    e. What are **possible enhancements that can be added to your system interface and/or processing?**

7. **DEVELOPMENT Documentation**

a. **Language(s)** used
b. Basic **Hardware - Software platform** required (Backend, Front End)
c. **Software Libraries** Used
d. Data Sets Sources
e. References from the literature**.**
f. Hosting and **Terms of usage**

8. Acknowledgements
a. Institutional
b. Funding Agencies and Companies
c. Personal