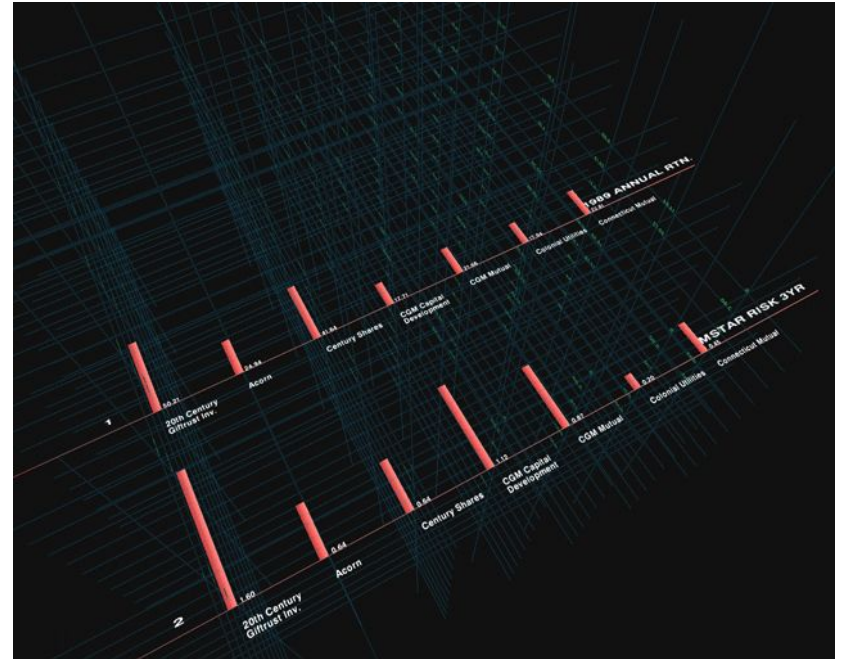


# Sorcero Data Visualization

**Forest Tennant, Michael Fahr, Bryce  
Mendenhall, Pao Yang, Rafael Carmen**

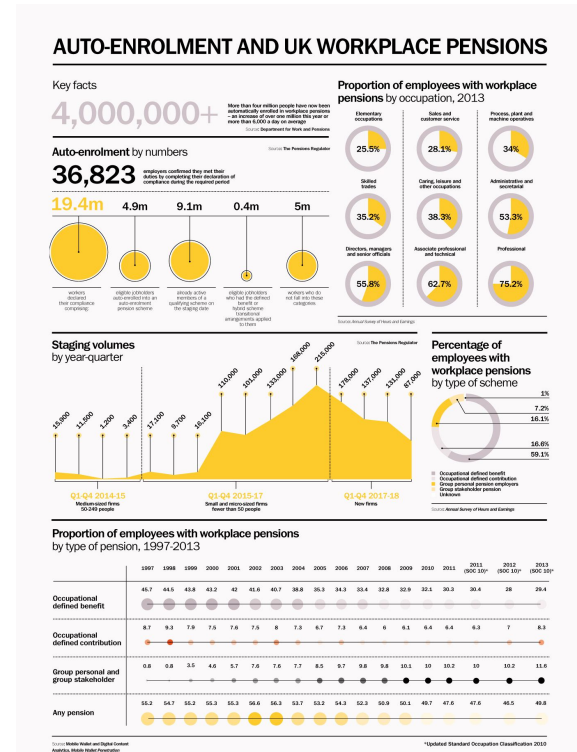
# Problem

- Corporations generate large volumes of data
  - Information about documents
  - How users interact with these documents
- As the amount of data grows, it becomes progressively more difficult to visualize core information
- It is important for companies to have access to key performance indicators (KPI)
  - KPI: Key indicators of progress towards an intended result
- Solutions are required to quickly analyze data



# Objective

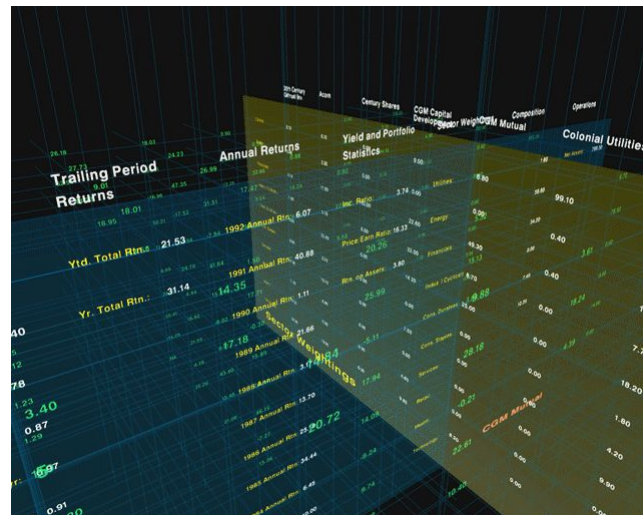
- Receive and analyze large volumes of data
- Use the data to provide
  - Meaningful visualization
  - Means of navigation
  - Insights into key performance indicators (KPIs)



# Background

- Key Concepts
  - Natural Processing Language (NLP)
    - Focuses on how computer interpret and analyze natural language data
    - Examples
      - Speech recognition
      - Text-to-speech
  - Apache Hadoop
    - A framework for processing large data sets in parallel across a Hadoop cluster.
    - MapReduce

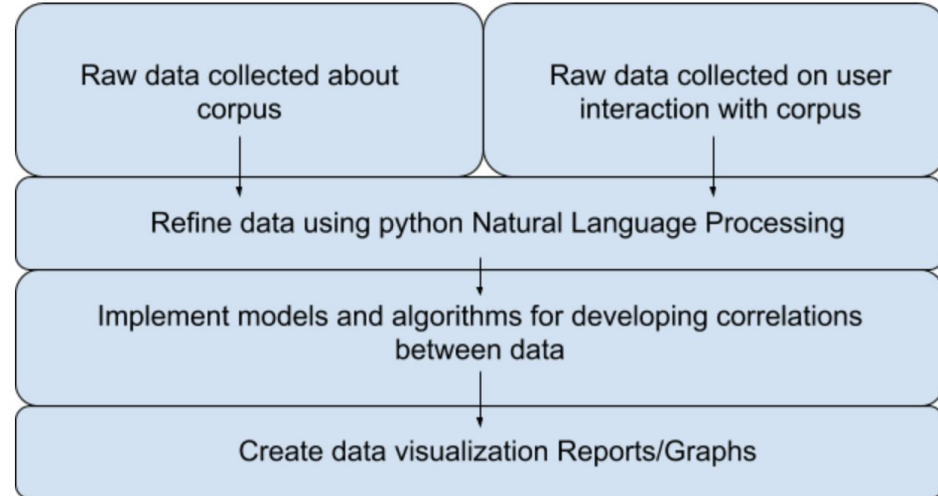
- Related Work
  - MATLAB
  - IBM Watson
  - Information Landscapes



<http://www.inventinginteractive.com/2010/02/01/information-landscapes/>

# Design

- Design Goals
  - Store the data
  - Group data into similar data sets
  - Provide additional means of navigation through data sets
  - Generate a communicable visual context based on the data provided
  - Show significant results with insights into KPIs
- High Level Architecture
  - Two sets of data
    - One set of data from corpus
    - One set of data from user interaction with the corpus
  - Use algorithms to sort the data
  - Run these two sets of data through a visualization tool to collect and process the data
  - Generate a straightforward visual based context on the data provided



# Tasks & Schedules

## Tasks

1. Explore and understand the background of data visualization and natural language processing.
2. Research other modern implementations to get an idea of other approaches.
3. Finalize architecture design and language of implementation.
4. Develop code to process the large volumes of data.
5. Create an algorithm for grouping/sorting the data into related fields.
6. Use the data to provide a meaningful visual context that suits the data.
7. Finalize the program by testing the application on multiple large sets of data
8. Document the final results.

Tasks	Dates
1. Do some research and understand the background of data visualization and natural language processing.	1/13-1/20
2. Research other modern implementations to receive an idea of other approaches.	1/21-1/27
3. Finalize architecture design and language of implementation.	1/28-2/10
4. Develop code to intake the large volumes of data.	2/11-2/24
5. Create an algorithm for grouping/sorting the data into related fields.	2/24-3/9
6. Use the data to provide a meaningful visual context that suits the data.	3/9-3/23
7. Finalize the program by testing the application on multiple large sets of data	3/23-4/6
8. Document the final results.	4/7-4/21

# Deliverables

- Design Document: Contains information on the programming languages that are used in the implementation process. This also contains the design process for implementing the NLP with the database. Furthermore, it includes different design features that were used to implement the final software and explains various aspects of the program such as the inputs and outputs.
- Database schema and initial data: Schema for storing the relations concluded from the NLP and the initial data these relations are created from.
- Python code: Code used for analyzing the data with NLP and creating the visuals.
- Website code: The PHP code for the web site split into three main subdirectories.
- Final Report: A report that summarizes the process and outcomes of the program that was implemented.

# References & Key Personnel

- [1] "Tableau: Business Intelligence and Analytics Software." Tableau Software, [www.tableau.com/](http://www.tableau.com/).
- [2] "MATLAB." *MathWorks*, [www.mathworks.com/products/matlab.html](http://www.mathworks.com/products/matlab.html).
- [3] "Smart Data Analysis and Visualization." *Watson Analytics*, [www.ibm.com/watson-analytics?lnk=hmhm](http://www.ibm.com/watson-analytics?lnk=hmhm).
- [4] "Muriel Cooper: Information Landscapes." *Inventing Interactive*, [www.inventinginteractive.com/2010/02/01/information-landscapes/](http://www.inventinginteractive.com/2010/02/01/information-landscapes/).

**Michael Fahr** – Fahr is a senior Computer Science major in the Computer Science and Computer Engineering department at the University of Arkansas. He has completed Software Engineering and Programming Paradigms.

**Forrest Tennant** - Tennant is a senior Computer Engineering major in the Computer Science and Computer Engineering department at the University of Arkansas Fayetteville. He has completed Software Engineering, Programming Paradigms and Programming Foundations I/II.

**Bryce Mendenhall** – Mendenhall is a senior Computer Engineering major in the Computer Science and Computer Engineering department at the University of Arkansas. He has completed Programming Foundations I/II, Software Engineering, and Programming Paradigms.

**Rafael Del Carmen** – Del Carmen is a senior Computer Engineering major in the Computer Science and Computer Engineering department at the University of Arkansas. He has completed. He has completed Cloud Computing, Programming Foundations I/II, Programming Paradigms, and Software Engineering.

**Pao Yang** – Yang is a senior Computer Science/Computer Engineering major in the Computer Science and Computer Engineering department at the University of Arkansas. He has completed Programming Foundations I/II, Software Engineering, and Programming Paradigms.