# Group 1: Web Retrieval

**Project ID: W1**

**Project Title: Non-English to English Flickr Image Search**

**Proposed people: Daqing He (dah44@pitt.edu)  
  
Background:**

Flickr has large collection of images with many image descriptions in the form of English tags. But not all users know how to use English to express the query. Therefore this project aims to support the user issue a query in a language that is not English and find the images in the Flickr collection that satisfy the user’s query. The team can use a machine readable dictionary or a online translate service (such as Google Translate or Bing Translator) to perform the translation, and when the image is returned, the tags associated with the image should be translated back to the language of the query too.

**Outcome:**

A search engine with a graphic interface that takes users’ query in a non-English language (say Chinese or French), and return Flickr images in a format that is suitable for presenting images, and also display the corresponding tags in the same non-English language (Chinese or French).

**Hint:**

1. using an online translate service would be easier and maybe in better quality, but in that case, the focus of the project should be more on the interface design. Using a machine readable dictionary would know more about cross-language retrieval, but the performance would be less effective than using online translate service.
2. Flickr has API

**Project ID: W2**

**Project Title: Finding similar questions in Yahoo! Answers for my non-English question**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:**

Yahoo! Answers contain a huge collection of questions and answers. It is a great site for exploring answers to a question. This project will design and develop a search engine that take a question in a non-English language (say Chinese or French), and return a set of Yahoo! Answers questions that are similar to the question. To help the users read the returned questions, at least the returned questions should be translated to the language that the user can understand.

**Outcome:**

A search engine with a graphic interface that takes users’ question in a non-English language (say Chinese or French), and return a ranked list of Yahoo! Answers questions that match to the query. All returned questions should be translated back to the language of the query.

**Hint:**

* Use Google Translate or Bing translator to translate the query question and the returned questions from Yahoo! Answers.
* This project should focus on the careful design of accepting the user’s query, the display of returned questions in a user-friendly and meaningful way
* Yahoo Answers have API

**Project ID: W3**

**Project Title: Web QA**

**Contacts: Qikai Cheng (chengqikai0806@gmail.com)**

 **Background**

The common paradigm of searching and retrieving information on the Web is based on keyword-based search using one or more search engines, and then browsing through the large number of returned results. However, more and more people use WebQA for information retrieval. The goal of WebQA is to retrieve small snippets of text that contain the actual answer to a question. In this project, we want to utilize meta-search techniques and information extraction techniques to develop a WebQA engine. The approach consists of first using meta-search techniques in an open environment to gather candidate responses from search engines and other on-line databases, and then using information extraction techniques to find the answer to the specific question from these candidates.

In this project, you need to complete the following requirements:

* Builds a prototype system which
  + returns relevant sentences rather than documents to given questions
  + returns relevant entity for questions about common entity(time, number, people, location, institution). Sample questions: How long is Yangtze River in china? How old is Obama?

**Outcome:**

# A prototype system and a report

# Project ID: W4

**Project Title: User-specific Discussion Search in LinkedIn**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:**

Users in LinkedIn can subscribe to many groups, and each group can have many discussions that the user may be interested in. This project takes advantage of LinkedIn API to search within the discussions posted in the groups that the user subscribe, and return the latest discussions from a time stamp specified by the user, and the filter out those posts that do not contain any query term issued by the user, and then display the posts on the screen. The user then can click to like or unlike any returned posts, and such interaction is then recorded with the corresponding posts in LinkedIn.

**Outcome:**  
a search engine satisfies the above requirement

**Hint:**

Your system should use corresponding LinkedIn APIs for user login, for obtain users groups, for retrieving posts, for interacting with posts, and you also need to build code to perform the filtering function mentioned in the description.

**Project ID: W5**

**Project title: Federate search on Facebook and Google Plus**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:** This project builds a search engine that takes queries from users to search for posts related to a person, an event, a place, or other things on both Facebook and Google Plus, and return the results from the two sites with one coherent interface for displaying them.

**Outcome:**

an online search engine that takes queries and generate one set of results. The results should indicate which site the posts come from.

**Hint:**

Both Facebook and Google Plus provide API for retrieving information. Your tasks include the design of the interface, the querying of Facebook and Google Plus, and the result combination. You should try to rank the two sets of results into one ranked list based on their relevance to query. Besides this relevance-based ranking, you can decide other ways to combine the two sets of results.

**Project Id: W6**

**Project Title: Find the Right Answerers Twitter Questions**

**Contacts: Shuguang Han (hanshuguang@gmail.com)**

**Background**

People use social media websites in their daily basis. They posted different topics for different purposes. It is very common to see people ask questions on Twitter. One important follow-up support for question-liked tweets is to find the best answerers. In this project, we would like to use the real Twitter data (which we have already crawled) to explore proper methods for recommending the answers for Twitter questions. To be specific, you need to explore:

* The descriptive statistics for question-like tweets in Twitter for a specific domain (or a general domain)
* Proposes different algorithms for finding the best answerers
* Integrates one algorithm in a live system.

**Outcome:**

A report for the comparisons of different algorithms, and a live recommendation system

# Project ID: W7

**Project Title: Academic Search Engine Google Scholar ++**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:** Google Scholar does not have an official API. The Python code at [**https://github.com/ckreibich/scholar.py**](https://github.com/ckreibich/scholar.py)provides a set of simple API functions. Wrapping the Python codes into an implementation of an academic search engine so that the users can specify the queries as typing in Google Scholar textbox, but they can organize the returned results based on citation counts, based on grouping by journal articles, conference papers, or books. More ideas can also be added by the team.

**Outcome:**

A prototype academic search engine.

# Project ID: W8

**Project Title: Classical Music Information Representation with FRBR (Functional Requirement for Bibliographic Records) Entities  
Proposed People: Sung-Min Kim (**[**suk30@pitt.edu**](mailto:suk30@pitt.edu)**)**

**Background**: FRBR (Functional Requirement for Bibliographic Records) is a conceptual entity-relationship model which describes the bibliographic information of an entity and the relationship between entities in online library catalogues and bibliographic databases. Different from old catalog system, FRBR consists of four entities in Group one (Work, Expression, Manifestation, and Item), and two in Group 2 (Person and Corporate Body). Each entity has featured attributes and entities are connected by the relationships. Classical music domain is one of main beneficiary of FRBR model because a classical music (Work) reproduces a large number of performances (Expression) and sound recordings (Manifestation). For better relationship description between entities, additional relationship and attributes will be added.   
  
Currently, I borrow the system from Ontopia (<http://www.ontopia.net/>), and need to extend the features of search functions and enhance the display of search results and detailed page.

**Outcomes:** This project will establish a classical music information retrieval system based on FRBR model.

1. Search function based on each entity, attributes, and keywords. The search results should be displayed by entity based view.
2. Each page should be separated into two frames: 1) attributes display and 2) relationship display

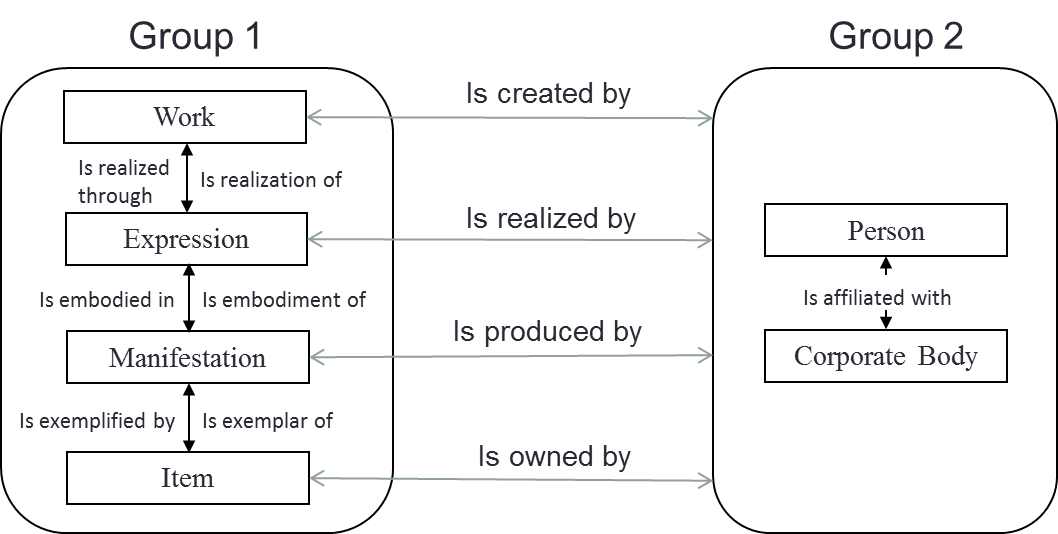
****

Figure 1 General FRBR Relationship Diagram

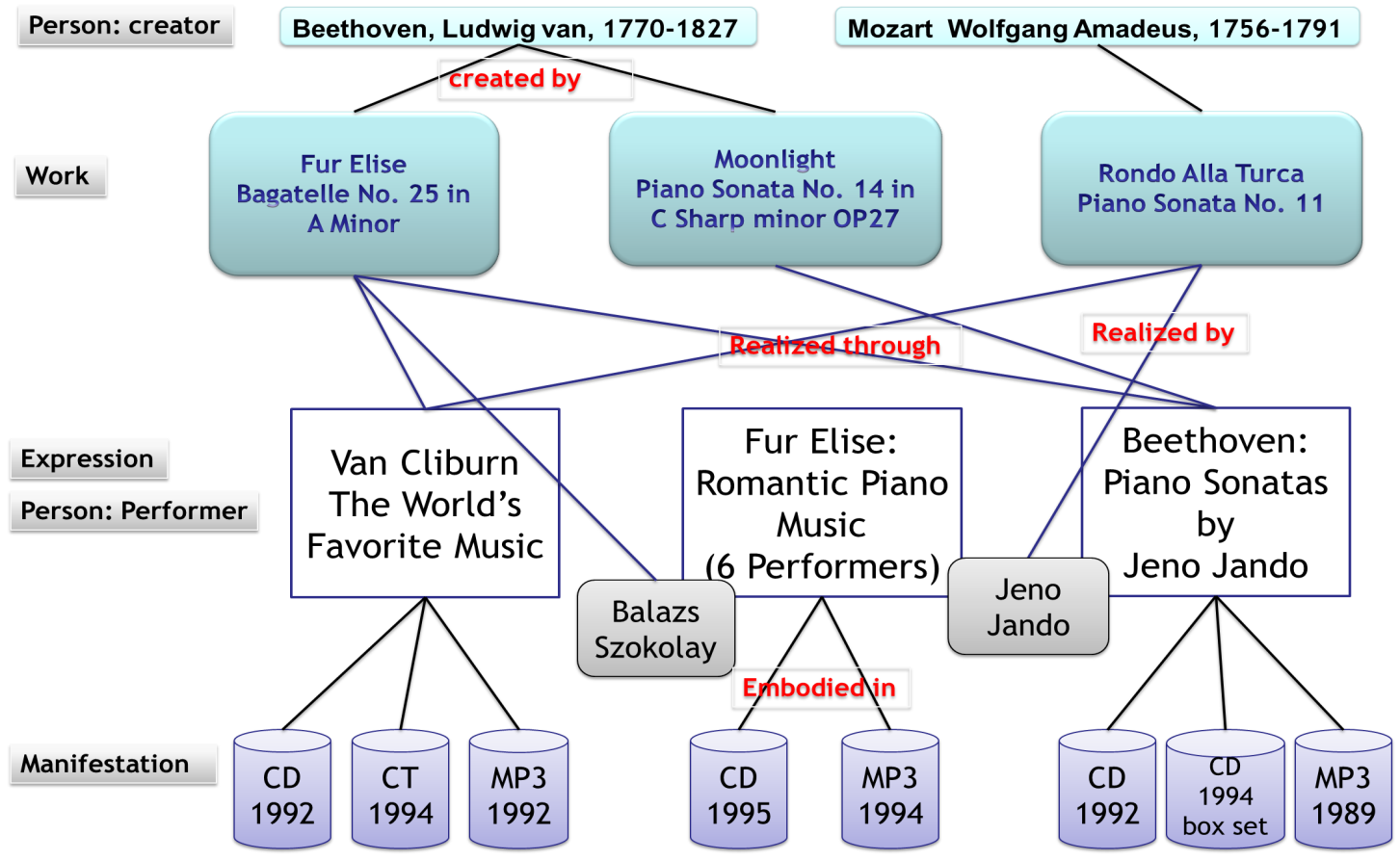
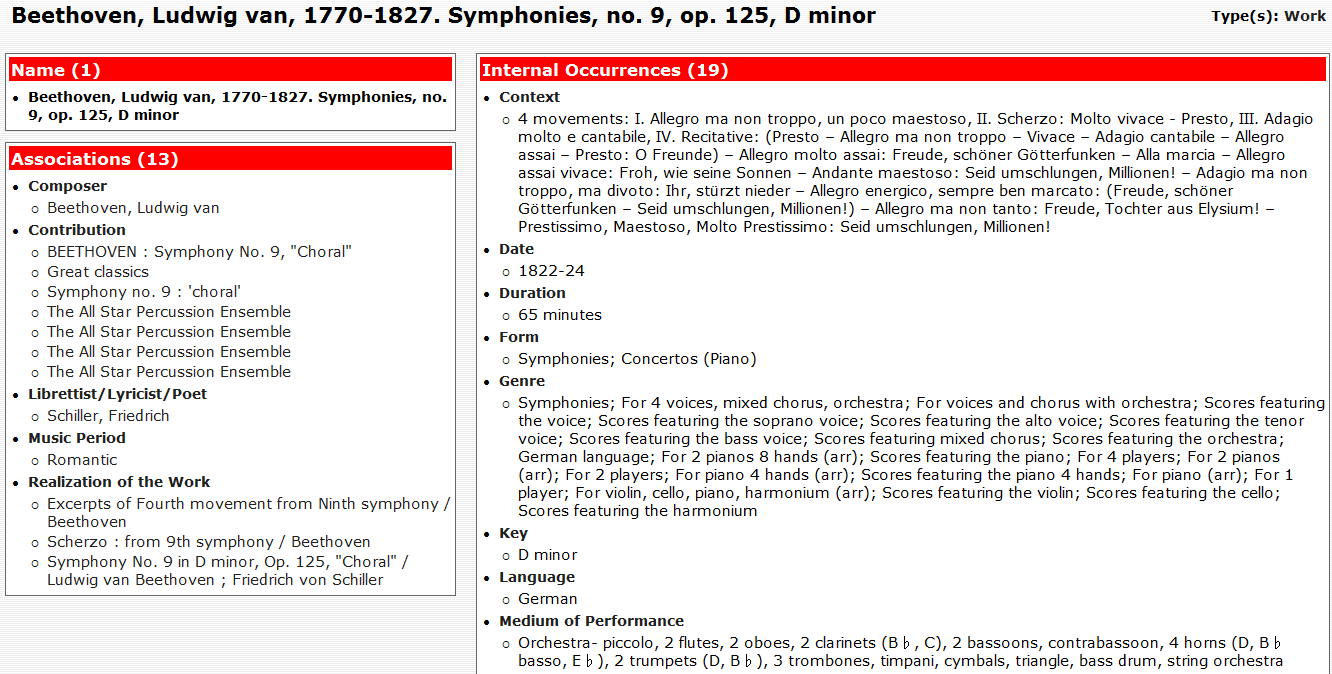


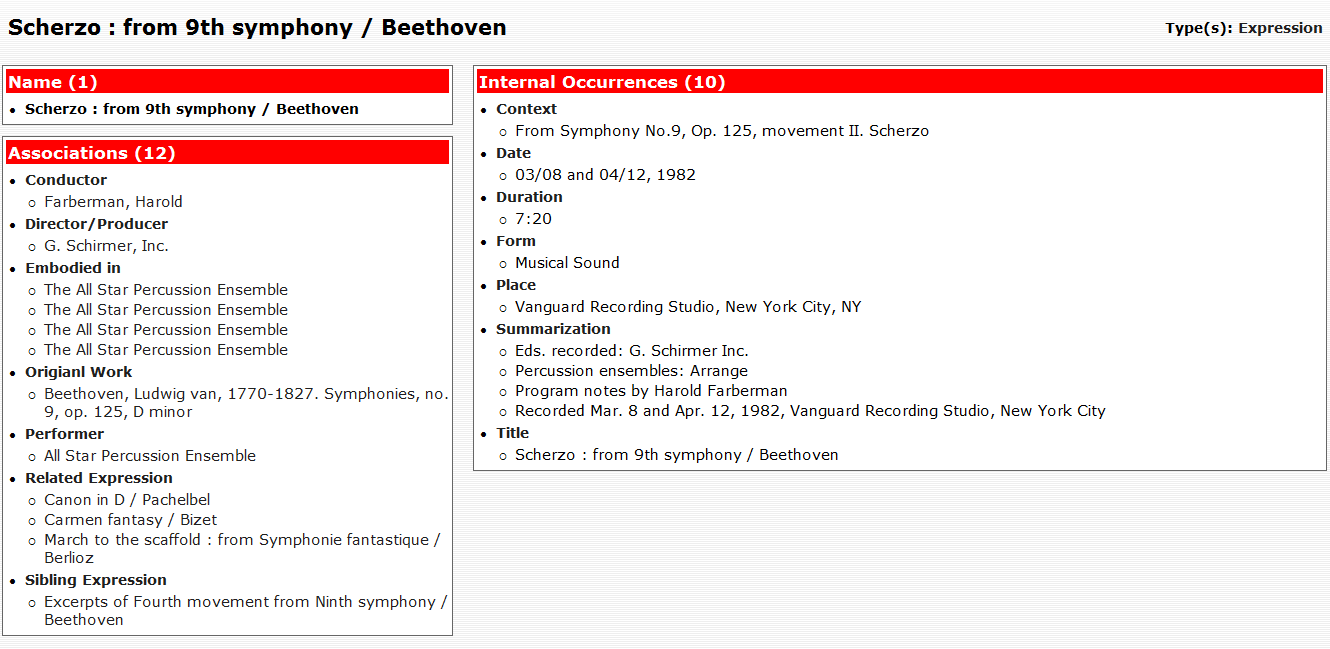
Figure 2 Example Diagram of Classical Music FRBR



**Relationship Description**

**Attributes Description**

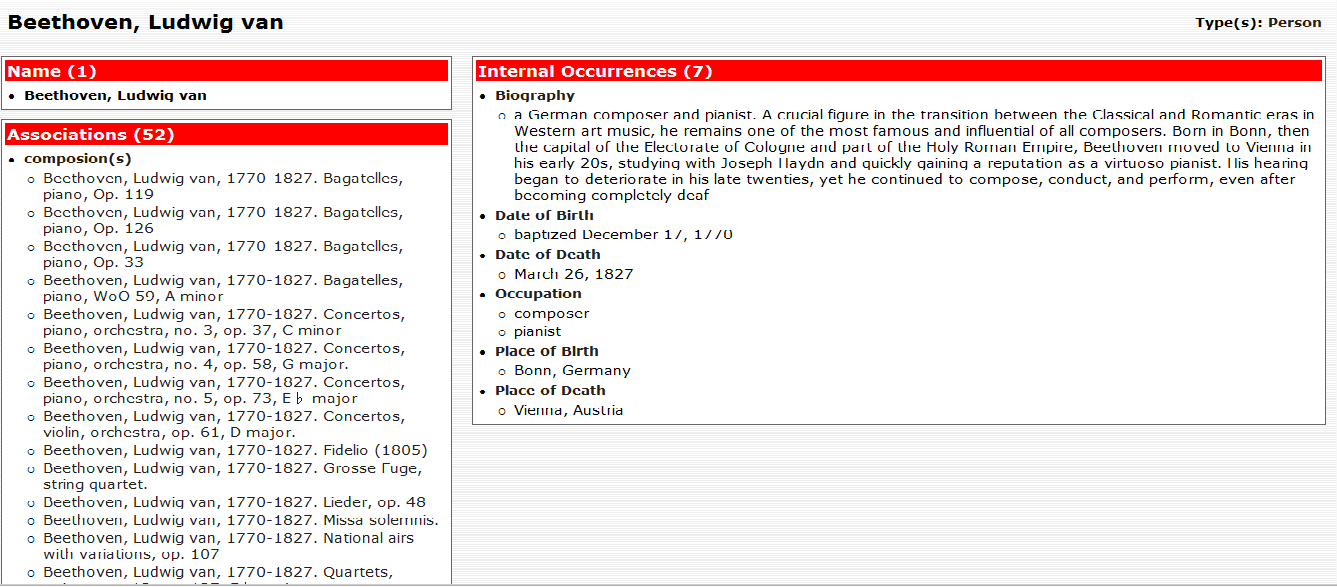
Figure 3 Screenshot of Work Page in Ontopoly



**Relationship Description**

**Attributes Description**

Figure 4 Screenshot of Expression Page in Ontopoly



**Relationship Description**

**Attributes Description**

Figure 5 Screenshot of Person Page in Ontopoly

# Project ID: W9

**Project Title: Entity and contents based search system in International Music Score Library Project (IMSLP)  
Proposed People: Sung-Min Kim (**[**suk30@pitt.edu**](mailto:suk30@pitt.edu)**)**

**Background**: International Music Score Library Project (http://imslp.org/) also known as the Petrucci Music Library, is an open online library project of storing scanned music scores, music media files, and work information mainly in classical music. IMSLP provides its users access to classical music scores from both the public domain and from composers who are willing to share their music without charge. Although IMSLP provide many contents in work pages, the major problem of this site is that it provides only keyword based search because it follows a wiki-like structure. The wiki-like structure is not designed for detailed searches; its users must know the name of the score or music work they are looking for or its composer or performers’ name. It would be helpful to provide better search functions based on its contents, for example, title, music period, composition year, etc.

**Outcomes:** This project will improve music information search functions in IMSLP.

1. A prototype of search system
2. Search function based on contents

**Project Id: W10**

**Project Title: Improving the interface design of Google Voice Search**

**Contacts:  Wei Jeng <cerises@gmail.com>**

**Background**

Voice search such as Siri or Google Voice Search recently became increasingly popular for both mobile and desktop devices. Compared to a conventional search interface when using the keyboard for query input, a user’s interaction with voice search systems can be much more complex. Researchers [1] found that the users’ voice input errors (IE), topic familiarity, and topic complexity are the major obstacles for the effective voice search. Therefore, an improving design of current interface is needed.

**Outcome**

With the understanding of current issues with user interactions in Voice search, students need to design and implement a effective interface that 1) supports voice query reformulation and 2) reduces users’ overall efforts in their IE correcting.

Reference:

[1] Jiang, J., Jeng, W., & He, D. (2013). How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search. In Proceedings of the SIGIR'13.

# Group 2: Mobile Retrieval

**Project ID: M1  
Project Title: Inferring Users’ Eye-gaze Attention from Mobile Touch Interaction  
Contacts: Shuguang Han (hanshuguang@gmail.com)**

Mobile devices, particularly the smartphones, are widely used in people’s daily lives. Comparing to the desktop computers, mobile devices have relatively small screens. On the one hand, this prevents users from fully exploring the content on web pages; on the other hand, it provides the opportunities for us to accurately infer user’s eye gaze attentions (Han, Hsiao, & Parra, 2014). The eye-gaze attentions can be used for better modeling of users’ interests (Buscher, Dengel, & van Elst, 2008). The estimation of eye-gaze attention based on Mobile Touch Interactions (MTIs) is not an easy task due to a lack of ground-truth data to evaluate the estimation algorithms – some researchers has tried the eye-tracking experiments (Biedert, Dengel, Buscher, & Vartan, 2012), but the experiments are too expensive and require too many resources for replications. In additional, seeing is not aware (Merikle, 2001) – it is probably that a user does not aware the information he/she reads.

In this project, we want to perform a free-recall experiment that follows the paradigm in several psychology experiments, that is, asking users to recall the information (or words) he/she just read – this information is then treated as what a user can aware. At the same time, we also log users’ MTIs so that we can figure out the best approach to infer the information that a user can capture through the interactions on mobile devices. The purposes of this project are:

* Builds a prototype system that provides two functions: 1) mobile touch interaction logging; and 2) receives users’ free recall information in the prototype system and writes it back into the database.
* Designs a user study and conducts preliminary analysis on inferring eye-gaze attention.

**Outcome:**

A prototype system, a preliminary user study and its result analysis

Biedert, R., Dengel, A., Buscher, G., & Vartan, A. (2012). Reading and estimating gaze on smart phones. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’12* (p. 385). New York, New York, USA: ACM Press. doi:10.1145/2168556.2168643

Buscher, G., Dengel, A., & van Elst, L. (2008). Query expansion using gaze-based feedback on the subdocument level. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’08* (p. 387). New York, New York, USA: ACM Press. doi:10.1145/1390334.1390401

Han, S., Hsiao, I., & Parra, D. (2014). A Study of Mobile Information Exploration with Multi-touch Interactions. *Social Computing, Behavioral-Cultural  …*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-05579-4\_33

Merikle, P. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition*, *79*(1-2), 115–134. doi:10.1016/S0010-0277(00)00126-8

**Project ID: M2**

**Project Title: Linkedin for android**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:** LinkedIn is the largest online professional network, which contains rich profiles of people’s experience, skills, and social networks. LinkedIn provides API for people search and group search, and various other functionalities.

This project is to explore the extension of LinkedIn on Android platform. You have the freedom to define what is the right LinkedIn search and presentation on a smartphone screen, and make it fashion, make it right.

**Outcome:**  
a nice interface for searching and presenting LinkedIn data on android phone.

**Project id: M3**

**Project title: Mendeley for android**

**Proposed people: Daqing He (**[**dah44@pitt.edu**](mailto:dah44@pitt.edu)**)**

**Background:** Mendeley is a well known online social reference management site, where people can upload articles to the site, and create their own personal library for the articles that they have read or they want to read. There are communities on it for sharing articles too.

This project is to explore the extension of Mendeley on Android platform. You have the freedom to define what is the right Mendeley search and presentation on a smartphone screen, and make it fashion, make it right.

**Outcome:**  
a nice interface for searching and presenting Mendeley data on android phone.

**Project id: M4**

**Project title: Yahoo! Answers Search for android**

**Proposed people: Daqing He (**[**dah44@pitt.edu**](mailto:dah44@pitt.edu)**)**

**Background:** Yahoo! Answers is a huge collection of online questions and their answers. There are many searches performed on Yahoo! Answers everyday to look for questions and answers satisfying the queries, either as keywords or a whole natural language question. This project is to explore a nice mobile search interface for retrieving similar questions for a given query on Android platform. You have the freedom to define what is the right Yahoo! Answers mobile search and presentation on a smartphone screen, and make it fashion, make it right.

**Outcome:**  
a nice interface for searching and presenting Yahoo! Answers data on android phone.

# Group 3: TREC Retrieval tasks

**Instructions**

1. visit <http://trec.nist.gov/tracks.html> to look for current and past retrieval tasks conducted by top information retrieval scholars,
2. work out a team first,
3. identify a potential track that the whole team is interested in, get familiar with the document collection and the search tasks,
4. work out a draft of the term project,
5. send to me a message with your team information and your project proposal draft

# Group 4: Self-Proposed Retrieval Projects

**Instructions**

1. read the proposed projects carefully. Only proceed with self-proposed retrieval projects when none of the above projects are suitable to you
2. work out a team first,
3. identify a potential retrieval task that the whole team is interested in
4. work out a draft of the term project in the same format as shown in this document,
5. send to me a message with your team information and your project proposal draft

**Project ID: W5**

**Project title: Federate search on Facebook and Twitter**

**Proposed people: Daqing He (dah44@pitt.edu)**

**Background:** This project builds a search engine that takes queries from users to search for posts related to a person, an event, a place, or other things on both Facebook and twitter, and return the results from the two sites with one coherent interface for displaying them.

**Outcome:**

an online search engine that takes queries and generate one set of results. The results should indicate which site the posts come from.

**Hint:**

Both Facebook and Google Plus provide API for retrieving information. Your tasks include the design of the interface, the querying of Facebook and Google Plus, and the result combination. You should try to rank the two sets of results into one ranked list based on their relevance to query. Besides this relevance-based ranking, you can decide other ways to combine the two sets of results.