Scholar's Search Engine

Problem Definition:

The Scholar's Search Engine is an alternative to existing academic search engines, with better organization and visualization for similar connections based on various features such as field of study, papers, collaborations etc. It also has a powerful filtering mechanism based on academic affiliations, and will help people find scholars with the most impact in their field of study.

The system suggests scholars who have similar interests using a wide range of heuristics. It is targeted towards the research community to help them explore existing researchers in an easy and meaningful manner by expanding their network.

Heilmeier Questions:

- What are you trying to do?
    - Our objective is to help researchers connect efficiently. We are trying to build a search engine which will help people find researchers in their fields of interest, filtered by institution/locality/venue. Also, it allows them to explore the work of scholars working in similar fields or those having the greatest research fit.
- How is it done today, and what are the limits of current practice?
    - Currently researchers use text based academic search engines which are not particularly intuitive or easy to use. Also, these have very few filtering options. One of the limitations of current practice is that search results only show papers, without showing other researchers working on similar topics.
- What's new in your approach and why do you think it will be successful?
    - Our approach combines traditional searching with visualisation. Additionally, we would provide similarity based connections with powerful filtering based on fields, collaborations, affiliation etc. Such an approach will be successful because there is a dearth of such an intuitive search, which will help researchers find people working on similar fields, and potential collaborators.
- Who cares?
    - It can help researchers to connect with scholars working in their fields of interest and collaborate.
    - Upcoming graduate students can look for a potential guide efficiently.
- If you're successful, what difference will it make?

    - It will lead to a more open and collaborative research environment, where people can find researchers with similar interests easily.
- What are the risks and the payoffs?
    - Building a machine learning classifier and ranking heuristics with a good degree of accuracy will be a challenge. If we are able to achieve this, it will greatly reduce the effort that the researcher has to put in to find relevant information for his/her research.
- How much will it cost?
    - We will be using open source technology. The dataset is obtained from Microsoft Academic Graph[5] which is freely available. For dynamic visualization we will be using D3/Tableau. For database we will use MongoDB/SQLite
- How long will it take?
    - 10 weeks
- What are the metrics to check for success?
    - Fair Similarity results
    - Accurate filtering.
    - Reasonable Response time
    - Appealing Visualization

Survey / Literature review:

1 - This paper describes the microsoft academic service which is the basis for most of the data that we will be using for this project. We plan to learn from the recommendation model in this paper to implement our similarity algorithm.

2 - It provides an idea to build a framework for developing similarity based methods. This will help us in our project as we would need to run different kinds of clustering based on similarity, such as affiliation and field of study on our data.

3 - This paper provides different kinds of decision models which can be implemented based on similarity. We need to do such modelling as we will be providing similarity results in different domains.

4 -We are dealing with a big dataset and are trying to provide quick query responses, so our analytics need to work fast. This paper tells us how to speed it up by keeping the data in memory for most of the operations.

6 - Using the searches performed in the past to further improve results for future queries can be done using subspace clustering approach as given in this paper. The paper is about leveraging localized searches and recommending based on its analysis.

7, 17 - These papers provides an overview of various concepts and approaches applied to recommendation systems such as content based filtering and collaborative filtering, along with their advancements and shortcomings.

8, 15, 16 - These papers are dedicated to collaborative filtering based recommendations where citations among different papers are used to create a rating matrix. This will help in building up a ranking system for papers.

9, 14 - These papers propose ways to subspace clustering approach for recommender systems. We might use this for clustering the areas of interests into larger chunks and also for grouping researchers with similar profiles.

10 - This paper used geometry and shading approaches in order to visualise hierarchical data, and a cushion treemap method which could be used to manage large data sets. We don't intend to use new visualisation methods in our project, these ideas help us think about possible ways in which we can think about hierarchies of data.

11, 13 - These papers describe the implementation of a hybrid approach that uses content-based as well as collaborative recommendation. This might help us in finding emergent communities of interests among the scholars.

12- This paper proposed a personalized academic research recommender system by defining text similarity between two research papers. We will be applying a similar approach to find similarity between keywords in research papers to show most relevant search results.

Plan of activities and time estimates:

	Week 1	Week 2-5	Week 6-9	Final week
Pradeep	Brainstorm over topic, prepare project proposal, presentation	Backend Framework for frontend-backend communication	Work on visualization, Final integration of the backend and frontend	Prepare report, Give final touches to visualization
Ankita	Brainstorm over topic, prepare project proposal, presentation	Work on for field of study classifier.	Work on visualization, Final integration of the backend and frontend	Prepare report
Shubham	Brainstorm over topic, prepare project proposal, presentation	Work on for field of study classifier.	Machine learning technique to find similarity and rank the result	Prepare report
Ashish	Brainstorm over topic, prepare project proposal, presentation	Machine learning technique to find similarity and rank the result	Preprocessing grouping/techniques for backend	Give final touches to visualization, Prepare report
Ashita	Brainstorm over topic, prepare project proposal, presentation	Backend Framework, work on the UI	Preprocessing grouping/techniques for backend	Prepare report

Distribution of team member effort:

All team member will contribute similar amounts of effort. We will be switching the work between Visualization and Backend in order to balance the load and experience gained.

References:

[1] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, and Kuansan Wang, "An Overview of Microsoft Academic Service (MAS) and Applications, WWW – World Wide Web Consortium (W3C)", 18 May 2015.

[2] Włodzisław Duch, "Similarity-based methods: a general framework for classification, approximation and association", Available: http://www.fizyka.umk.pl/publications/kmk/00cc-kn.pdf

[3] "Similarity based Decision Models", Available: http://www.econ.core.hu/file/download/korosi/2010/kovacs10.pdf

[4] "Using In-Memory Analytics to Quickly Crunch Big Data", Available: http://www.computer.org/csdl/mags/co/2012/10/mco2012100016.pdf

[5] https://academicgraph.blob.core.windows.net/graph-2015-08-20/index.html

[6] Nitin Agarwal, Ehtesham Haque, Huan Liu, and Lance Parsons Arizona State University, Tempe AZ 85281, USA - "Research Paper Recommender Systems: A Subspace Clustering Approach"

[7]Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger, "Research Paper Recommender Systems: A Literature Survey" International Journal on Digital Libraries, Springer Berlin Heidelberg, 26 Jul 2015, Available: http://docear.org/papers/research_paper_recommender_system_evaluation--a_quantitative_literature_survey.pdf

[8] Available: http://files.grouplens.org/papers/techlens-cscw2002.pdf

[9] Megha Jain, "ALGORITHM FOR RESEARCH PAPER RECOMMENDATION SYSTEM",International Journal of Information Technology and Knowledge Management

[10] Jarke J. van Wijk, Frank van Ham, Huub van de Wetering , "Rendering Hierarchical Data", Available: http://www.win.tue.nl/~vanwijk/rhd.pdf

[11] Balabanovic, M., Shoham, "Fab: Content-based, Collaborative Recommendation. Communications of the ACM"

[12] Joonseok Lee, Kisung Lee, Jennifer G. Kim, "Personalized Academic Research Paper Recommendation System"

[13] Jie Tang, Duo Zhang, Limin Yao, "Social Network Extraction of Academic Researchers"

[14] "An index to quantify an individual's scientific research output", Available: http://arxiv.org/pdf/physics/0508025.pdf

[15] Jöran Beel Bela Gipp, "Google Scholar's Ranking Algorithm: An Introductory Overview." In Birger Larsen and Jacqueline Leta, editors, Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935, Available: http://www.sciplore.org/publications/2009-Google_Scholar%27s_Ranking_Algorithm_--_An_Introductory_Overview_--_preprint.pdf

[16] Bela Gipp, Jöran Beel, and Christian Hentschel - "Scienstein: A Research Paper Recommender System". In Proceedings of the International Conference on Emerging Trends in Computing (ICETiC'09), Available: http://www.gipp.com/wp-content/papercite-data/pdf/gipp09.pdf

[17] Karypis. "Evaluation of Item-Based Top-N Recommendation Algorithms" Available: http://glaros.dtc.umn.edu/gkhome/fetch/papers/itemrs.pdf