Integrating Publication signals for literature-based Knowledge discovery

Timothy Baldock

2385679

Proposal

Motivation

We have a vast database of biomedical research that is growing faster than anyone could read. Literature based discovery is needed to fully utilize the knowledge in this database, and to discover links between different research to help form new research ideas and explore treatments and discoveries. This project will explore using Pubmed's MeSH tags to discover these links, and what systems and information can generate the best link predictions using these tags.

Aims

Compare different link prediction systems and different metrics over a knowledge graph of mesh terms. This will use a time-split to create two separate knowledge graphs for evaluating the prediction systems.

Progress

- *Background reading and research.*
- *Explored dataset.*
- *Created subset of dataset.*
- *Filtered subset of dataset.*
- *Created knowledge graph.*
- *Implemented Swanson's abc.*
- *Implemented abc using different metrics.*
- *Evaluated abc using hits@k.*

Problems and risks

Problems

Filtering the dataset, how to do this meaningfully. This proved difficult. Finding other ways to make predictions using the knowledge graphs I have. Learning about literature based discovery. There's a lot of dense reading on the subject.

Risks

Learning how to create a machine learning system for making predictions using a knowledge graph. This will be started now, so that progress can be made well before second semester starts. Another risk is how long it will take to run the machine learning system on any significant dataset. To deal with this I will try to run it on my home PC, so that I can use the graphics card there for running it.

Plan

December – start learning and implementing machine learning system.

January – Finish implementation of machine learning system.

February – start writing dissertation.

March – Finish writing dissertation.