**Topic:**

Automatically building a reliable graph database for natural language querying

**Content:**

Acknowledgments

Abstract

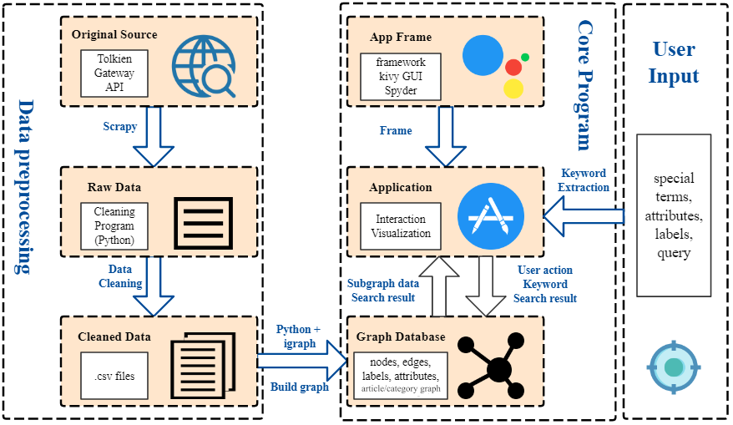
Dedication

1. Introduction and Motivation
2. Related Work

Mainly based on:

* 1. [Automatically creating datasets for measures of semantic relatedness](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.5257&rep=rep1&type=pdf)
  2. [An improvement of TFIDF weighting in text categorization](http://www.ipcsit.com/vol47/009-ICCTS2012-T049.pdf)
  3. [Analysis of the Wikipedia Category Graph for NLP Applications](http://www.aclweb.org/anthology/W07-0201)
  4. [Wikipedia-based Semantic Interpretationfor Natural Language Processing](https://www.jair.org/index.php/jair/article/view/10595/25347)

1. The method used in the repost/Problem Definition
   1. Build an article graph from the plain text of wiki original data, a graph with attributes from the semi-structure of wiki Infobox and a category graph from wiki categories. [from 2(c)(d)]
   2. Use TF-IDF, TF-IDF-CF, LTC, TFC weighing method separately on Naïve Bayes, Bayes Network, KNN, SVM to find the feature importance. Pick up the group with highest accuracy to do the ranking of searching result. [from 2(b)(d)]
   3. Find the linguistic distance between words by Semantic Relatedness for both query and article graph. [from 2(a)(d)]
2. Implementation
   1. Describe the architecture of the application
   2. Sketch the structure of the article and category graphs
   3. Expound the query processing
   4. Design the user enhancing function: when the user pick up a result from the list, it will add weight to the link between keywords in the query and the chosen result.



1. Experiment

Set 4 scenarios to test the function of the application (haven’t decided yet)

Thinking about using single keyword, multiple keywords, different sentences included those keywords to test the result of this application.

1. Conclusion and future work

Updating the report future on GitHub in the future (still a template yet):

<https://github.com/programeggsoup/2019SpringFinalExam/tree/master/Main/dissertation.pdf>