Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master thesis

Autolinks: Adaptive Hypergraph-based Information Management of Semantic Triples

Alvin Rindra Fazrie

4fazrie@informatik.uni-hamburg.de
Intelligent Adaptive Systems Master program
Matr.-Nr. 6641834

First Supervisor: Prof. Dr. Chris Biemann
Second Supervisor: Steffen Remus MSc.

Abstract

Autolinks 'automatic proactive researching' is a tool that provides a quick researching platform based on a text or a sentence by visualizing the results together with their semantic relations. In this internet era, people could get information easily with search engines. They will give us a ton of hyperlinks clustered by multiple pages by entering a single query to the input, then we could select a specific link we think the most relevant. The process of learning takes a time sometimes. After the chosen web page rendered, we need to read through a page to get a specific information related to the query given and sometimes we still have to deal with a number of hyperlinks to get further information. Even worse, most of the website nowadays exploit the curiosity gap of the reader, providing just enough information and not enough to satisfy the reader's curiosity, without clicking through another linked content. This clickbait phenomenon becomes so normal today and it makes our time to study longer.

Autolinks optimizes these concerns and is intended to make the learning process faster and more efficient. Instead of reading papers, websites, and other resources to understand a specific term, this machine will do it for us. From a text or a sentence given by the user, it will read and learn from multiple resources and digests the core related information by visualizing the information in the most convenient way. The information is visualized by a force-directed graph, a graph which contains nodes for the information and edges for the semantic relation so that it will ease the reader to understand how pieces of information correlate each other.

Autolinks is built with machine learning paradigm. Natural Language Processing (NLP) takes a responsibility to understand a given text and to comprehend which information from the sources have a relation to the given text and correlate each other. The reader could evaluate the results given and Autolinks will learn and correct the mistakes so that it could improve the precision and confidence in the next iteration. Bundled with this capability, Autolinks accelerates the process of researching and understanding during the study.

With respect to the background and the purpose of Autolinks, we address some research questions in this master thesis, including the following: how can a user interface be devised that is non-intrusive, i.e. helping users solve their information needs faster instead of impeding them?; which semantic services, realized with NLP technologies, are the most useful?; how can we measure success, i.e. showing that Autolinks really live up to its premise?

Acknowledgement

Here comes the acknowledgement...

Contents

1 Introduction

Interaction between Artificial Intelligence and humans has been growing to be a vital part in daily life in the past decade. Natural Language Processing has become one of the fundamental topics for scientists in the field of Data Science. In this thesis, we introduce the basic concepts of Language Technology and expand on the idea of visual interaction to support the learning / research process. To this end, we develop autolinks, cytoscape [4].

- Graph Theory, cytoscape.js [4]

- Hypergraph Papers:

An Algorithm for drawing compound graphs [1]

A Layout algorithm for undirected compund graphs [3]

Graph Visualization Techniques for Web Clustering Engines [5]

1.1 Motivation

Visual Analytics [10] Visual Analytics methode [7] Visual Analytics Process [8]

Taxonomy of Information Visualization [9]

An algorithm that learn what is in name [2]

1.2 Research question

Here comes the the Research Questions...

1. How can a user interface be devised that is non-intrusive, i.e. helping users solve their information needs faster instead of impeding them?
2. Which semantic services, realized with NLP technologies, are the most useful?
3. How can we measure success, i.e. showing that Autolinks really live up to its premise?,

1.3 Contributions

Here comes the contributions..

2 Related Work

Software for Social Network Analysis: [6]

2.1 Information Management tools

Here comes information management tools...

2.2 Text annotation tools

Here comes text annotation tools...

3 Background Study

Here comes Background Study...

3.1 Language Technology

Here comes contributions...

3.2 Machine learning

Here comes Machine Learning...

3.3 Data Visualization

Here comes Data Visualization...

3.4 Hypergraph

Here comes Hypergraph...

3.5 Web technologies

Here comes Web technologies...

4 System Overview

Here comes System Overview...

4.1 Autolinks Introduction

Here comes Autolinks Introduction...

4.2 Components in Autolinks

Here comes components in Autolinks...

5 Data Extraction

Here comes Data Extraction...

5.1 Broker

Here comes broker...

5.2 Wiki Service

Here comes Wikiservice...

6 Information Management Visualization

Here comes Information Management Viz...

6.1 Concept and Visualization

Here comes Concept and Visualization...

6.2 Compound Nodes / Parent

Here comes Compound Nodes / Parent...

6.3 System Overview

Here comes the System Overview...

6.4 Data Extraction

Here comes the Data Extraction...

Example of lists:

1. Fachbücher, Standards,
2. Wiss. Zeitschriftenartikel, Survey-Artikel,
3. Konferenzbeiträge,
4. Technical Reports, graue Literatur,
5. Online-Material, Arbeitspapiere, Firmenmaterial, Ausarbeitungen.

Im Internet können zur Feststellung der Qualität und Recherche von Publikationen

- Google Scholar (http://scholar.google.com),
- Microsoft Academic Search (http://academic.research.microsoft.com) → computer science → security & privacy,
- Computer Science Bibliography (http://dblp.uni-trier.de/) und die
- Scientific Literature Digital Library (http://citeseer.nj.nec.com/)

7 Evaluation

Here comes the evaluation...

7.1 Case Study

Here comes the case study...

7.2 User Experiment

Here comes the user experiment...

7.3 Evaluation details

Here comes the evaluation details...

```
1  int getGGTOf(int a, int b) {
2      // requires ((a > 0) && (b > 0)); ensures return > 0;
3      int h;
4      while (b != 0) {
5          h = b;
6          b = a % b; // % is the modulo operator. This line is long
               enough to show how line breaks in lstlisting are handled.
7          a = h;
8      }
9      return a;
10 }
```

Listing 7.1: Example of algorithm

8 Future Work

Here comes the future work...

9 Conclusion

Here comes the conclusion...

Bibliography

[1] François Bertault and Mirka Miller. An algorithm for drawing compound graphs. In Jan Kratochvíyl, editor, *Graph Drawing*, pages 197–204, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[2] Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel. An algorithm that learns what's in a name. *Machine Learning*, 34(1):211–231, Feb 1999.

[3] Ugur Dogrusoz, Erhan Giral, Ahmet Cetintas, Ali Civril, and Emek Demir. A layout algorithm for undirected compound graphs. *Information Sciences*, 179(7):980 – 994, 2009.

[4] Max Franz, Christian T. Lopes, Gerardo Huck, Yue Dong, Onur Sumer, and Gary D. Bader. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, 32(2):309–311, 2016.

[5] E. D. Giacomo, W. Didimo, L. Grilli, and G. Liotta. Graph visualization techniques for web clustering engines. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):294–304, March 2007.

[6] Mark Huisman and Marijtje Duijn. Software for social network analysis. In *Models and Methods in Social Network Analysis*, pages 270–316, 01 2005.

[7] D. A. Keim, C. Panse, M. Sips, and S. C. North. Visual data mining in large geospatial point sets. *IEEE Computer Graphics and Applications*, 24(5):36–44, Sept 2004.

[8] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann. Mastering the information age - solving problems with visual analytics, 2010.

[9] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, Sep 1996.

[10] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, Jan 2006.