Master thesis

Autolinks: Information Management on Hypergraph 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

Here comes the Introduction...

1.1 Motivation

Here comes the motivation...

1.2 Research question

Here comes the the Research Questions...

1.3 Contributions

Here comes the contributions..

2 Related Work

Here comes Related Work...

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

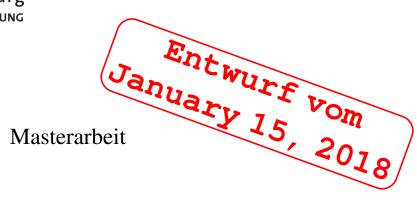Here comes the conclusion...

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Ggf. streichen: Ich bin damit einverstanden, dass meine Abschlussarbeit in den Bestand der Fachbereichsbibliothek eingestellt wird.

Hamburg, den January 15, 2018

_____

Alvin Rindra Fazrie

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Entwurf vom January 15, 2018

Masterarbeit

Master Thesis

vorgelegt von

Alvin Rindra Fazrie

geb. am 12. Juni 1991 in Jakarta

Matrikelnummer 6143847

Studiengang Informatik

eingereicht am January 15, 2018

Betreuer: Dipl.-Inf. Heinz Mustermann

Erstgutachter: Prof. Dr. Chris Biemann

Zweitgutachter: Steffen Remus M.Sc.

Todos im Text und Fragen an den Betreuer sind in dieser Form dargestellt

Thema: Privacy Enhancing Technologies zum Schutz von Kommunikationsbeziehungen

Bearbeiter: Eva Musterfrau, Heinz Mustermann

Datum: January 15, 2018

Literaturliste

David Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM 24/2 (1981) 84–88.

David Chaum: The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. Journal of Cryptology 1/1 (1988) 65–75.

David Goldschlag, Michael Reed, Paul Syverson: Onion Routing for Anonymous and Private Internet Connections. Communications of the ACM 42/2 (1999) 39–41.

Andreas Pfitzmann: Diensteintegrierende Kommunikationsnetze mit teilnehmerüberprüfbarem Datenschutz. IFB 234, Springer-Verlag, Berlin 1990.

Wei Wang, Mehul Motani, Vikram Srinivasan: Dependent link padding algorithms for low latency anonymity systems. Proc. 15th ACM conference on Computer and communications security. ACM, 2008, 323–332.

Todo list