

A Domain-Specific Knowledge Graph for News Item Recommendations and Information Retrieval

Author: Christopher Eaves-Kohlbrener
Supervisor: Michael Zakharyashev

July 13, 2021

MSc Data Science project
Department of Computer Science and Information Systems
Birkbeck College, University of London
2021

This report is substantially the result of my own work, expressed in my own words, except where explicitly indicated in the text. I have read and understood the sections on plagiarism in the Programme Handbook and the College web site. I give my permission for it to be submitted to the JISC Plagiarism Detection Service.

The report may be freely copied and distributed provided the source is explicitly acknowledged.

Abstract

TBD...

Contents

1	Introduction	4
1.1	Problem	4
1.2	Project Objectives	4
1.3	Summary of Findings	4
1.4	Report Outline	4
1.5	Technical Specifications	4
2	Context and Background	4
2.1	Knowledge Graphs	4
2.2	Network stuff?	4
3	Data	4
3.1	Data Set Collection	5
3.2	Data Input and Preprocessing	5
4	Knowledge Graph Design and Implementation	5
4.1	Graph Database Creation	5
4.2	Semantic Enhancements	5
4.3	Natural Language Processing	5
4.4	Models, etc.???	5
5	Knowledge Graph performance evaluation	5
5.1	Interpretability	5
5.2	Others...confusion matrix, predictions, clustering, between- ness, etc.	5
6	Conclusions	5
6.1	Future Work	5
6.2	Conclusion	5
7	Appendix	5

1 Introduction

Introduction

1.1 Problem

Problem...

1.2 Project Objectives

Project Objectives...

1.3 Summary of Findings

Summary of Findings...

1.4 Report Outline

Report Outline...

1.5 Technical Specifications

Technical Specifications...

2 Context and Background

2.1 Knowledge Graphs

2.2 Network stuff?

similar to this? <https://www.dcs.bbk.ac.uk/r/studentprojects/exampleprojects/msc/>
"A graph neural network path to network centralities.pdf"

3 Data

What is Reuters data set, how did I get it, etc.

Jupyter notebook – data exploration and preprocessing, diagrams, histograms, etc.

Simple dimensions (properties like subjects, genres, word count) and complex NLP ones (XML body)

Challenges / edge cases: <https://github.com/heychrisek/msc-data-science-project/> - 5x-challenges

3.1 Data Set Collection

3.2 Data Input and Preprocessing

4 Knowledge Graph Design and Implementation

Continuation of previous section (Data Input and Preprocessing), load into Neo4j...

4.1 Graph Database Creation

Neo4j ingest

4.2 Semantic Enhancements

4.3 Natural Language Processing

4.4 Models, etc.???

5 Knowledge Graph performance evaluation

5.1 Interpretability

5.2 Others...confusion matrix, predictions, clustering, betweenness, etc.

6 Conclusions

6.1 Future Work

6.2 Conclusion

7 Appendix

References

- [1] A. Berven, O. A. Christensen, S. Moldeklev, A. L. Opdahl, and K. J. Villanger. A knowledge-graph platform for newsrooms. *Computers in Industry*, 123:103321, 2020.
- [2] P. Castells, F. Perdrix, E. Pulido, M. Rico, R. Benjamins, J. Contreras, and J. Lorés. Neptuno: Semantic web technologies for a digital newspaper archive. In *European Semantic Web Symposium*, pages 445–458. Springer, 2004.
- [3] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmman, S. Sun, and W. Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 601–610, 2014. Evgeniy Gabrilovich Wilko Horn Ni Lao Kevin Murphy Thomas Strohmman Shaohua Sun Wei Zhang Jeremy Heitz.
- [4] M. Färber, B. Ell, C. Menne, and A. Rettinger. A comparative survey of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web Journal*, 1(1):1–5, 2015.
- [5] N. Fernández, D. Fuentes, L. Sánchez, and J. A. Fisteus. The news ontology: Design and applications. *Expert Systems with Applications*, 37(12):8694–8704, 2010.
- [6] U. Germann, R. Liepins, D. Gosko, and G. Barzdins. Integrating multiple nlp technologies into an open-source platform for multilingual media monitoring. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 47–51, 2018.
- [7] G. Leban, B. Fortuna, J. Brank, and M. Grobelnik. Event registry: learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 107–110, 2014.
- [8] D. Liu, T. Bai, J. Lian, X. Zhao, G. Sun, J.-R. Wen, and X. Xie. News graph: An enhanced knowledge graph for news recommendation. In *KaRS@ CIKM*, pages 1–7, 2019.
- [9] X. Liu, A. Nourbakhsh, Q. Li, S. Shah, R. Martin, and J. Duprey. Reuters tracer: Toward automated news production using large scale

- social media data. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1483–1493. IEEE, 2017.
- [10] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2015.
 - [11] C. Rudnik, T. Ehrhart, O. Ferret, D. Teyssou, R. Troncy, and X. Tannier. Searching news articles using an event knowledge graph leveraged by wikidata. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1232–1239, 2019.
 - [12] P. Vossen, R. Agerri, I. Aldabe, A. Cybulska, M. van Erp, A. Fokkens, E. Laparra, A.-L. Minard, A. P. Aproso, G. Rigau, et al. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85, 2016.
 - [13] Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.
 - [14] Z. Zhao, S.-K. Han, and I.-M. So. Architecture of knowledge graph construction techniques. *International Journal of Pure and Applied Mathematics*, 118(19):1869–1883, 2018.