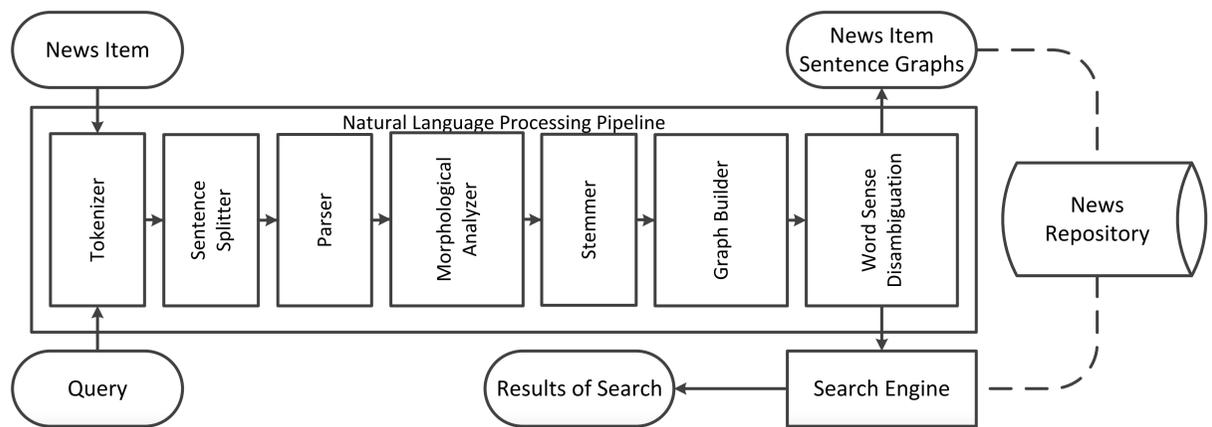




Dependency Graph Isomorphism for News Sentence Searching

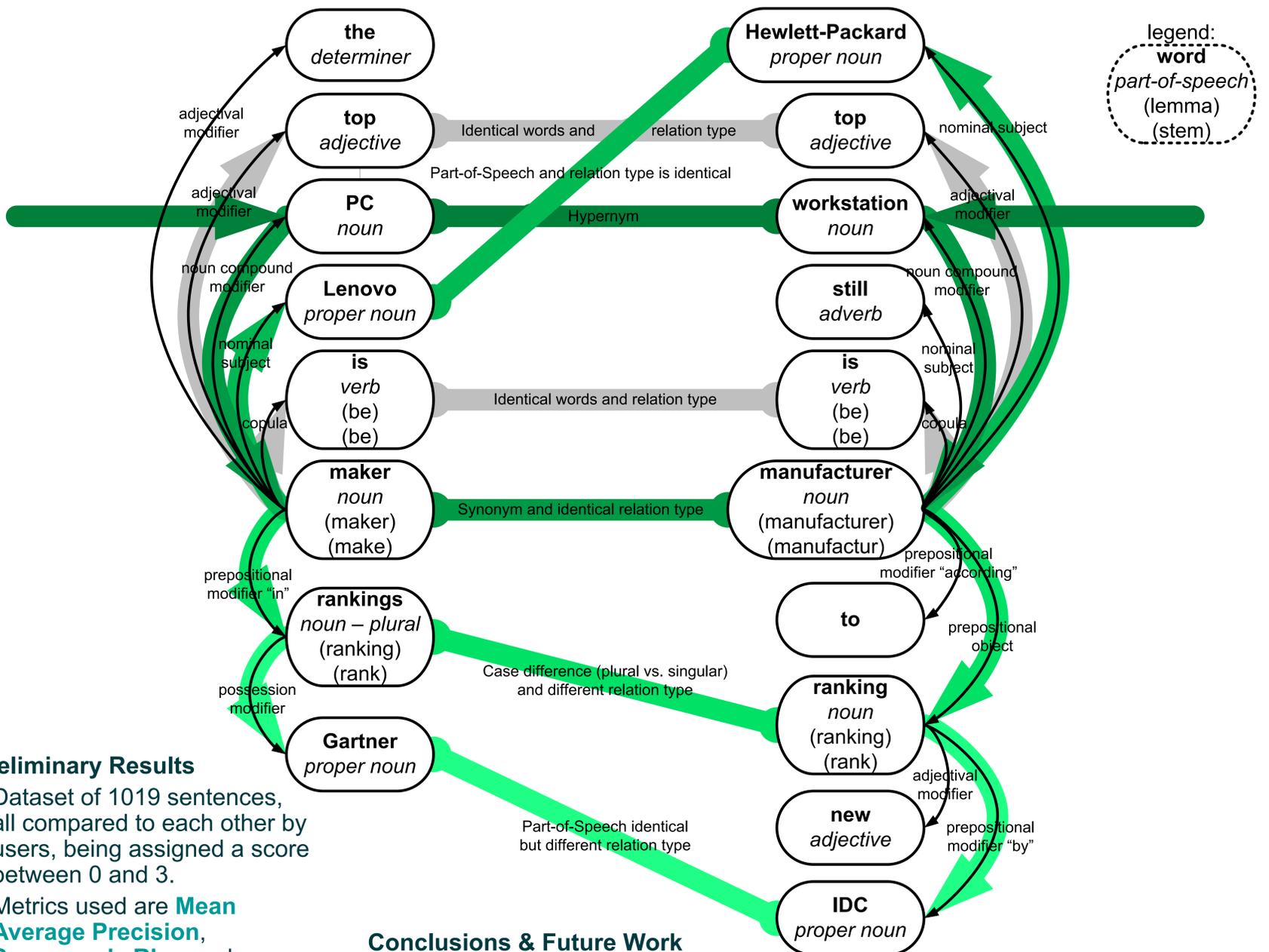
Intuition

- Transform text into graph structure based on **dependency parser** output.
- Search text by comparing these **graphs** instead of the words only.
- In this way, **grammatical context information** is incorporated into search.
- Additionally, **synonymy** and **hypernymy** are considered as well.



"In Gartner's rankings, Lenovo is the top PC maker."

"Hewlett-Packard is still top workstation manufacturer according to new ranking by IDC."



Preliminary Results

- Dataset of 1019 sentences, all compared to each other by users, being assigned a score between 0 and 3.
- Metrics used are **Mean Average Precision**, **Spearman's Rho**, and **normalized Discounted Cumulative Gain**.
- For all these metrics, a significantly higher score was achieved compared to the **TF-IDF baseline**.

Conclusions & Future Work

Useful additions to the framework might include **named entity recognition**, **co-reference resolution**, and a **knowledgebase** to compare entities (i.e., Gartner and IDC are both research firms, but the algorithm does not know this). Incorporating a **graph edit distance** metric into the algorithm could be beneficial as well.

Acknowledgments

The authors are partially supported by the Dutch national program COMMIT.