Web-Intelligence WS 25/26

Name: Irmak Damla Özdemir **Matrikel-Nr.:** 5123127

Identifizieren Sie Text-Daten und Graph-Daten im Web für Ihren Anwendungsfall. Demonstrieren Sie ihren Nutzen für Ihren Anwendungsfall. Vergleichen Sie diese beiden Arten von Daten.

Text Data in the Application

Text data is designed to convey meaning to human readers and often requires preprocessing before it can be analyzed automatically. On the web, text data appears in articles, comments, descriptions, summaries, and metadata fields such as titles.

In the Trending Health News system, text data is the core input of the analysis. Each collected news article contains multiple text-based attributes, including:

- Article title
- Article summary
- Source name
- Publication date as textual metadata
- Article URL

Source	**Title**	**Topic**
WHO	WHO issues global guideline on the use of GLP-1 medicines in treating obesity	Obesity
Tagesschau	Warken gibt Krankenkassen Mitschuld bei Beitragserhöhungen	Policy & system
Spiegel Gesundheit	Grippe: Die Zahlen steigen früher als in den Vorjahren	Influenza
Deutsches Ärzteblatt	Vorsicht vor hormonaktiven Substanzen in Kinderspielzeug	Other

Textual elements are fetched via RSS feeds and stored in a standardized format. Especially the title and summary fields play a crucial role, as they contain the actual content describing health-related topics.

Use of Text Data for Topic Detection

The main analytical task of the application is identifying trending health topics. This is achieved through keyword-based text analysis. Titles and summaries are converted to lowercase and scanned for predefined keywords related to topics such as Covid, vaccination, depression, nutrition, or healthcare policy.

For example, if a summary contains keywords like “vaccination”, “Impfstoff”, or “immunization”, the article is classified under the topic Vaccination. This demonstrates how raw text data is transformed into structured analytical results through preprocessing and categorization.

Benefits of Text Data in the Use Case

Text data provides important benefits for my application:

- It contains the semantic meaning of health news.
- It allows detection of emerging trends through keyword frequency.
- It supports further aggregation and visualization in tools such as AppSheet, KNIME, or Tableau.

Without text data, it would not be possible to understand what topics are being discussed. Therefore, text data forms the informational foundation of the entire system.

Graph Data in the Application

Instead of focusing on content, graph data models connections between entities. This structure is particularly useful for analyzing networks, dependencies, and relationships. Although the Trending Health News system primarily processes textual content, it can also be naturally modeled as a graph-based structure. As a high-level example:

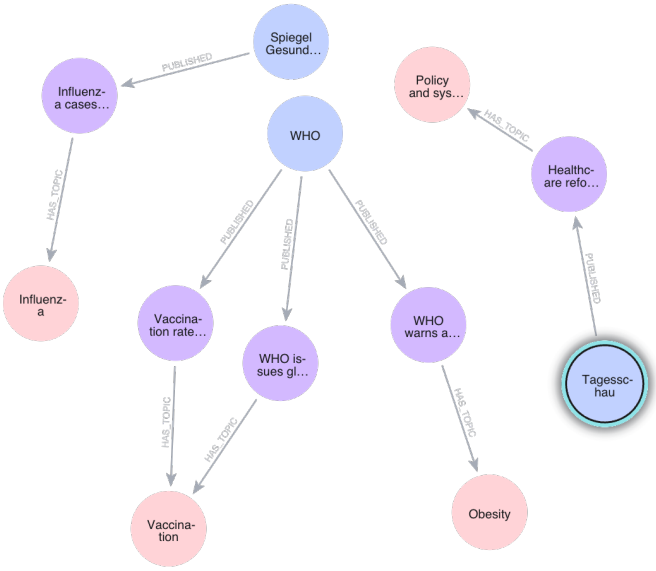

Abbildung 1: Example of a graph-based representation of health news data in Neo4j

In this context, nodes represent the main entities used in the graph-based prototype:

- News articles
- Topics such as Vaccination, Influenza, Policy and system, and Obesity
- News sources including WHO, Tagesschau, and Spiegel Gesundheit

Edges represent the relationships between these entities:

- A source publishes an article
- An article is categorized under a specific topic

For example, an article node can be connected to a Vaccination topic node and to a WHO source node. This creates a network that reflects how topics, sources, and articles are related.

Benefits of Graph Data

Graph data enables analysis beyond simple keyword counts. Using a graph-based view, the system can answer questions such as:

- Which sources contribute most to a specific topic?
- Which topics are shared across multiple sources?
- How strongly is a topic connected to a particular publisher?
- Which sources dominate the overall information network?

This structure is, where importance is derived from connections rather than isolated values. A topic discussed by many influential sources becomes more significant within the network.

Comparing Graph and Text Data

Text data and graph data serve different purposes in the Trending Health News application and are used at different stages of the analysis process. Text data is directly required for the core functionality of the system, since health topics are detected by analyzing article titles and summaries. Without textual content, it would not be possible to identify which health issues are currently discussed in the media.

Graph data, on the other hand, is mainly relevant when focusing on relationships rather than content. In this project, graph data is not stored explicitly in a graph database. Instead, relationships between sources, articles, and topics are represented implicitly through aggregation and visualization. Tools such as KNIME and Tableau show how often specific sources are connected to certain topics, which reflects graph-like structures without requiring a dedicated graph model.

Overall, text data plays a more central role in the application, as it forms the basis for topic detection and data processing. Graph data concepts support the interpretation of results by highlighting connections and patterns between entities, but they are not essential for the system's basic functionality.

Conclusion

In conclusion, text data represents the essential foundation of the Trending Health News application, as the core analytical task, which is topic detection, is entirely based on processing article titles and summaries.

Graph data, in contrast, is not strictly essential as a standalone data model or database in this application. The system does not require a dedicated graph database, since explicit graph operations such as path traversal or node ranking are not part of the core functionality. However, graph data concepts are still effectively applied through visualization tools such as KNIME and Tableau.

By visualizing relationships between sources, articles, and topics, these tools implicitly represent graph structures and enable relationship-oriented analysis. As a result, the application benefits from graph-based thinking without the need for explicit graph data storage. This demonstrates that, for the given use case, text data is fundamental, while graph data concepts enhance interpretation and insight through visualization rather than being an essential structural component.