DBpedia

October 2024

Table of Contents

Introduction

Purpose of DBpedia

- Organizes Wikipedia's knowledge in a structured way that is accessible by everyone
- Links people, places, events, and more in a way that other systems can understand and connect to.
- Enables smarter searches and more complex and efficent querying for information

History of DBpedia

- It was introduced in 2007 by people at Free University of Berlin and Leipzig University.
- Becomes a part of LOD Cloud in 2008 and is most interlinked part of it ever since.
- Wikidata was launched in 2012 and was integrated into DBpedia.
- In early versions, DBpedia faced delays in keeping the information up to date. However, now it is synced with Wikipedia in near realtime.

How DBpedia Works

Parsing Infoboxes

- DBpedia extracts data from infoboxes on Wikipedia pages, like "born in," "occupation," or "location," which are relatively standardized across similar types of articles.
- These help DBpedia identify key information that it can structure into a machine-readable format.

Figure 1: DBpedia Infobox Example
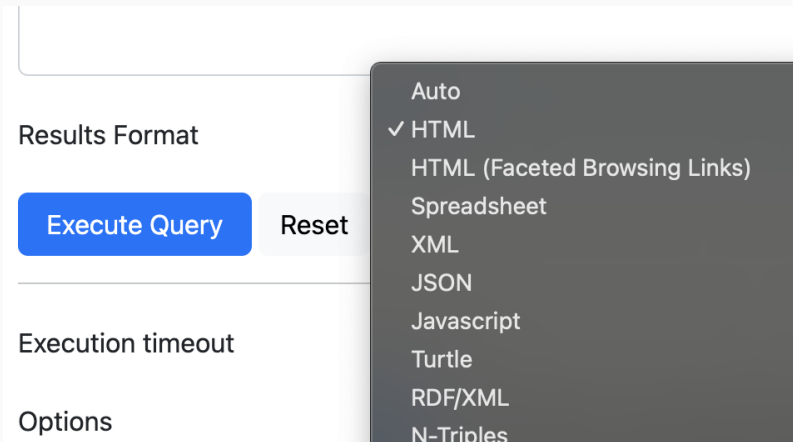
How DBpedia Works

DBpedia's Extraction Framework

- The DBpedia Extraction Framework is an open-source tool that scans Wikipedia's infoboxes, links, and categories, then extracts the data to create triples (subject-predicate-object format). For example:
- <Albert_Einstein> - <born_in> - <Ulm>
- This framework also extracts labels, abstracts (summaries), and links to other Wikipedia pages.
- Extracted information is converted into RDF format.
- A URI is assigned to each extracted resource which can be used to query and link it with other resources.

Representation

- RDF (Resource Description Framework) is the core format that DBpedia uses but there are other forms of expression in which data can be retrieved:

Results Format

Execute Query   Reset

Execution timeout

Options

Auto
✓ HTML
HTML (Faceted Browsing Links)
Spreadsheet
XML
JSON
Javascript
Turtle
RDF/XML
N-Triples

Comparing Wikipedia and DBpedia

The Lord of the Rings film series

Wikipedia : https://en.wikipedia.org/wiki/The_Lord_of_the_Rings_(film_series)
DBpedia: https://dbpedia.org/page/The_Lord_of_the_Rings_(film_series)

Challenges and Limitations

- Data Quality issues
- Can only caputre Information in the info boxes
- Linking with other databases to expand it's knowledge
- Huge database which makes querying less efficent

Sources

- https://www.youtube.com/watch?v=kxHTEZQSplA&t=389s
- https://www.dbpedia.org/about/

The End

Thank for your attention!