

# DBpedia

---

October 2024

# Table of Contents

1. Introduction
2. Purpose of DBpedia
3. How DBpedia Works
4. Comparing Wikipedia and DBpedia
5. Challenges and Limitations

## 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 efficient 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.

## 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.

# How DBpedia Works

## German Shepherd

84 languages

Article Talk

Read Edit View history Tools

From Wikipedia, the free encyclopedia

Checked

The **German Shepherd**,<sup>[a]</sup> also known in Britain as an **Alsatian**, is a German breed of **working dog** of medium to large size. The breed was developed by **Max von Stephanitz** using various **traditional German herding dogs** from 1899.

It was originally bred as a **herding dog**, for herding **sheep**. It has since been used in many other types of work, including **disability assistance**, **search-and-rescue**, **police work**, and **warfare**. It is commonly kept as a **companion dog**, and according to the **Fédération Cynologique Internationale** had the second-highest number of annual registrations in 2013.<sup>[4]</sup>

### History [edit]

During the 1890s, attempts were being made to standardise dog breeds.<sup>[5]</sup> Dogs were being bred to preserve traits that assisted in their job of herding sheep and protecting their flocks from predators.<sup>[6]</sup> In Germany this was practised within local communities, where shepherds selected and bred dogs. It was recognised that the breed had the necessary skills for herding sheep, such as intelligence, speed, strength and keen senses of smell.<sup>[6]</sup> The results were dogs that were able to do such things, but that differed significantly, both in appearance and ability, from one locality to another.<sup>[5]</sup>

### German Shepherd



Adult male

**Other names** German Shepherd Dog  
Alsatian  
Alsatian Wolf Dog  
*Deutscher Schäferhund*  
*Altdeutsche Schäferhunde*  
**Origin** Germany

		Traits	[hide]
<b>Height</b>	Males	60–65 cm (24–26 in) <sup>[1]</sup>	
	Females	55–60 cm (22–24 in) <sup>[1]</sup>	
<b>Weight</b>	Males	30–40 kg (66–88 lb) <sup>[1]</sup>	
	Females	22–32 kg (49–71 lb) <sup>[1]</sup>	
<b>Coat</b>	Double coat		

Figure 1: DBpedia Infobox Example

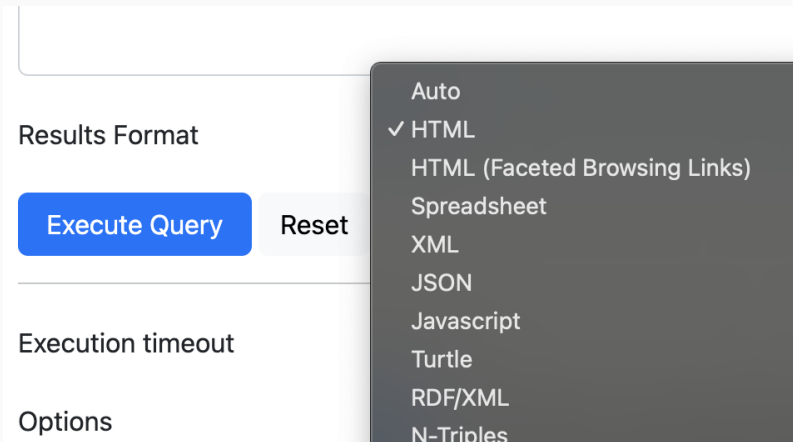
## DBpedia's Extraction Framework

- 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.

# How DBpedia Works

## 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:



The screenshot shows a web interface for querying DBpedia. It includes a text input field at the top, followed by the label "Results Format". Below this are two buttons: "Execute Query" (blue) and "Reset" (grey). Further down are labels for "Execution timeout" and "Options". A dropdown menu is open, displaying a list of available result formats: "Auto", "✓ HTML" (selected), "HTML (Faceted Browsing Links)", "Spreadsheet", "XML", "JSON", "Javascript", "Turtle", "RDF/XML", and "N-Triples".

Results Format

Execute Query Reset

Execution timeout

Options

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



## The Lord of the Rings film series

Wikipedia : [https://en.wikipedia.org/wiki/The\\_Lord\\_of\\_the\\_Rings\\_\(film\\_series\)](https://en.wikipedia.org/wiki/The_Lord_of_the_Rings_(film_series))

DBpedia: [https://dbpedia.org/page/The\\_Lord\\_of\\_the\\_Rings\\_\(film\\_series\)](https://dbpedia.org/page/The_Lord_of_the_Rings_(film_series))

## Challenges and Limitations

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

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

# The End

Thank for your attention!