

Nobel Ontology

Group A3D

Andrea Bruttomesso, Alessandro Corrà, Davide Seghetto, Andrea Stocco

January 15, 2025

Overview

1. Domain of Interest

2. Ontology Design

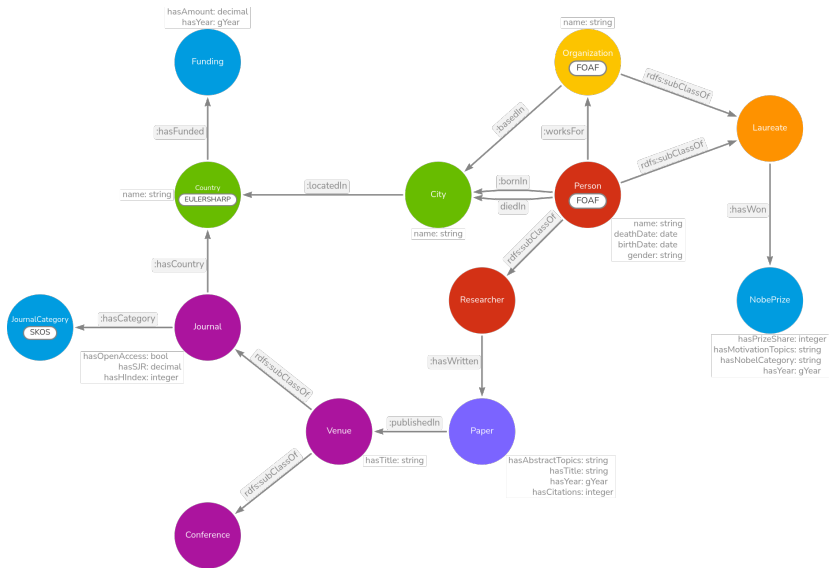
3. Challenges

4. Analytics

Domain of Interest



We have chosen the domain of scientific research. Specifically, we aim to analyze potential correlations among Nobel Prize winners, their publications, and the research funding invested by various countries. This domain was selected because it allows us to reveal potential historical and geographical patterns in scientific research.



Challenges

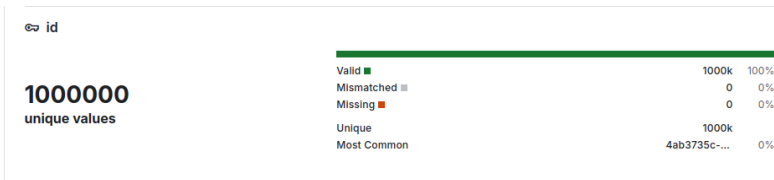
- Errors in Nobel Laureates dataset
- Topic Modeling
- Subset of papers dataset
- Researchers and Nobel Laureates matching names

Errors in Nobel Laureate dataset

Table: Example of dataset error

Year	Category	Prize Share	Full Name
1908	Medicine	1/2	Ilya Ilyich Mechnikov
1908	Medicine	1/2	Paul Ehrlich
1908	Medicine	1/2	Paul Ehrlich

Papers dataset was too big

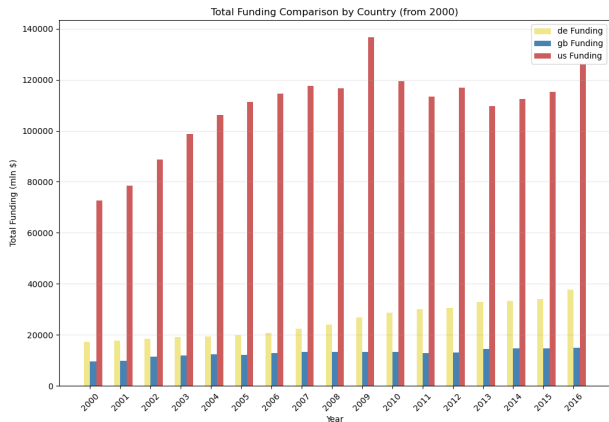


The Papers Dataset originally contained 1 million rows, which made it necessary to filter the data. After careful consideration, we decided to retain only the papers authored by a Laureate or published in a venue included in the Venues Dataset. Even with this filter applied, the dataset still contained approximately 300,000 rows so we further reduced it by selecting only the first 50,000 rows.

Researchers and Nobel Laureates matching names

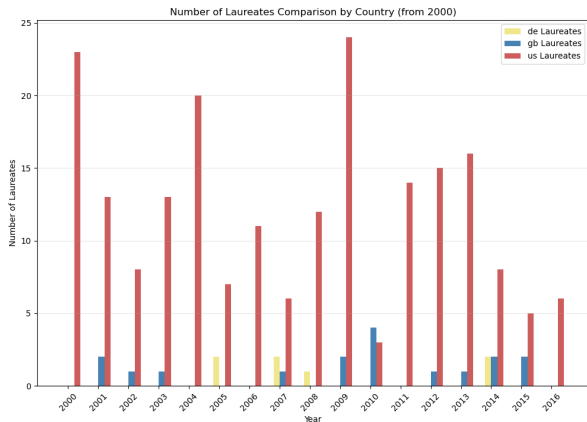
The Laureates and Papers Datasets often represent names differently, such as "Antoine Henri Becquerel" and "Antoine H. Becquerel." To effectively link these datasets, we utilized a library that implements a fuzzy matching algorithm with a similarity threshold of 90%. This approach allowed us to identify and match names that were not identical but sufficiently similar, ensuring a robust connection between the datasets.

Relationship between funding allocated for R&D and possibility of winning a Nobel Prize



Germany's R&D funding has been increasing over the years, whereas Great Britain's funding remains lower and more stable.

Relationship between funding allocated for R&D and possibility of winning a Nobel Prize (continued)



These two plots suggest that the link between funding and Nobel prizes is not always straightforward.

Age analysis of Nobel Prize winners

Table: Age statistics of Nobel Prize winners by category

Category	Min Age	Avg Age	Max Age
Chemistry	35	58	85
Economics	51	67	90
Literature	42	65	88
Medicine	32	58	87
Peace	17	61	87
Physics	25	55	88

- **Economics:** This category has the highest average age, with the eldest being Leonid Hurwicz.
- **Peace:** This category includes the youngest laureate, who is Malala Yousafzai at only 17 years old.
- **Chemistry, Medicine, and Physics:** These science-related categories show the lowest average and minimum age.
- **Literature:** This category has a high average age due also to the recognition of an author's entire career.

Questions?