Nobel Ontology

Group A3D

Andrea Bruttomesso, Alessandro Corrò, Davide Seghetto, Andrea Stocco

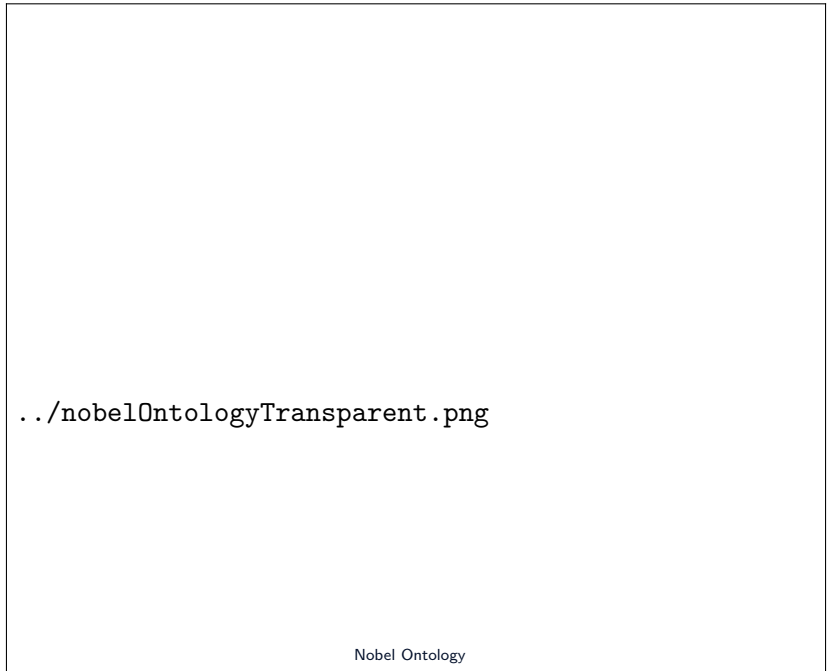
January 15, 2025

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.


../nobelOntologyTransparent.png

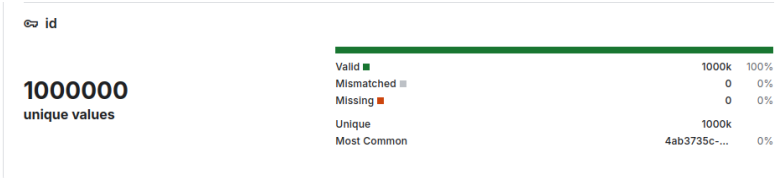
Challenges

- Finding Papers topics
- Restricting Papers dataset
- Matching Researchers and Nobel Laureates names

Finding Papers topics

The Papers dataset includes an "abstract" column, which is crucial for efficiently understanding and analyzing the content of scientific publications. To compare and explore this data, we applied topic modeling with the LDA algorithm, creating a bag of words for each paper. This allowed us to uncover deeper insights and establish meaningful connections across the dataset.

Restricting Papers dataset

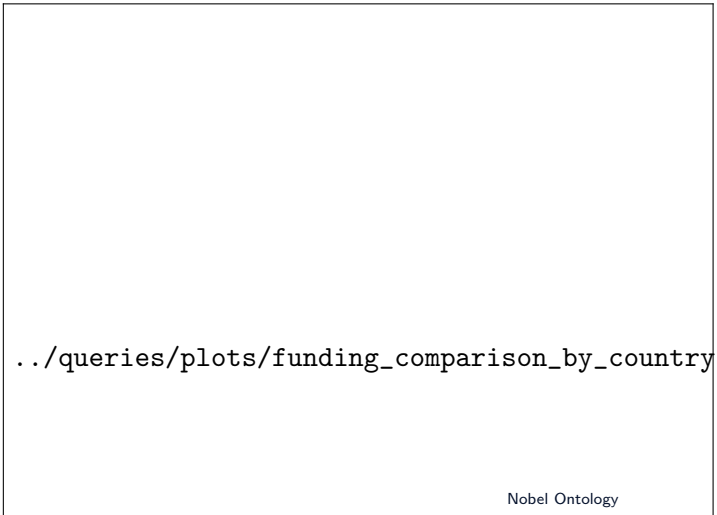


The Papers Dataset originally contained 1 million rows, which made it necessary to filter the data. After careful consideration, we decided to retain only the papers authored by a Laureate or published in a venue included in the Venues Dataset. Even with this filter applied, the dataset still contained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

Researchers and Nobel Laureates matching names

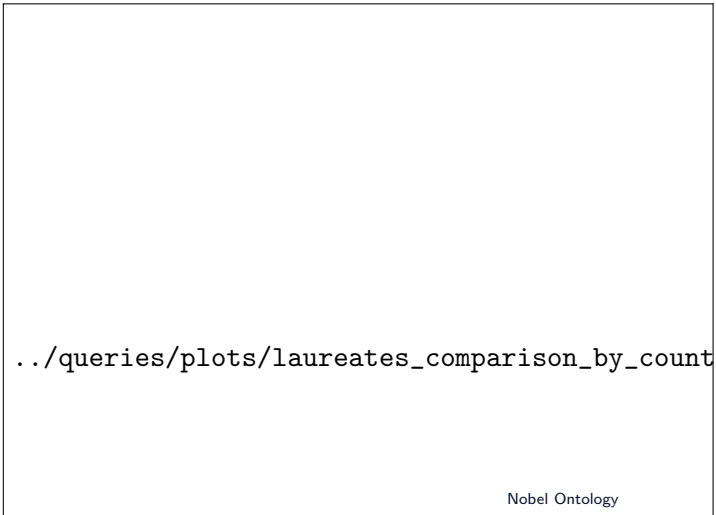
The Laureates and Papers Datasets often represent names differently, such as "Antoine Henri Becquerel" and "Antoine H. Becquerel." To effectively link these datasets, we utilized a library that implements a fuzzy matching algorithm with a similarity threshold of 90%. This approach allowed us to identify 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.

Nobel laureates: birthplace vs. research location by state

Table: States with lowest percentage of Laureates active in their home country

State	Laureates	Laureates active in their home country	Percentage
Canada	19	2	10,5%
China	11	2	18,2%
Italy	19	4	21%
Austria	17	4	23,5%
Australia	11	3	27,3%

Interesting insight: Japan, despite being the seventh country in terms of Nobel Prizes won, manages to retain nearly 67% of its laureates within its borders. On the other hand, Russia, despite being the sixth country in terms of Nobel Prizes won, has a retention rate of only 35%.

Nobel laureates: birthplace vs. research location by state

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?