From String to Thing
Wikidata based query expansion

Bernd Uttenweiler
IT Project Manager ETH Library
SWIB21 3. Dec 2021
Can we use Wikidata to improve our discovery system?

Context and preliminaries

- As part of the Swiss Wikidata Hackdays 2019, we were able to address several questions that we have encountered time and time again.
- In terms of resources, it was a smaller project. But it was all the more important to create concrete benefits for customers.

- Part of ETH team were Petra Maier, Michael Gasser and Bernd Uttenweiler.
- Further supporters were Annette Guignard, Noémie Ammann and Germano Giuliani.

- To say it in advance: We have dealt with the topics "person" and "place". I will focus on "person" here.
- Let's start with different questions that at first glance have nothing to do with each other!
A lot of data is organized by “person”

Here you can see "Hans Conrad Escher" on some pages of the partners of the Swiss project Metagrid.
Q 1: How to link persons?

• We have a lot of rare resources in our archives and collections. How can people find them if they do not use our library system?

• Solution: Linking of different systems/sites/databases/portals, even if they are structured by person.

• But where should the link in our Discovery System Primo VE lead to?

• We need something that represents a person and is linkable.
Q 2: Disambiguation: Which is the „Staudinger“ I am looking for?
Q 3: How to find all resources of a person?
Because of „other names“ / „Also Known as“ …

• … often different name variants are included in the metadata
• … and mostly not all name variants are included in the metadata
Q 4: How to use graphs in a discovery system?

• A library system like Primo VE offers only few options for flexibly linking records. A system like Wikidata is based on graphs. Can we use Wikidata graphs for our discovery system (Primo VE)?

• Could the user's search query be pre-processed to provide additional offers in Primo VE?

• What does the user search for when entering a search term (semantic interpretation instead of lexical search)?

• Or should there be some sort of advanced search whose fields and selections are then converted to SPARQL queries?

• Or can we identify some search patterns that we can implement in SPARQL queries and offer their results to the user as an option? E.g. Is the user looking for a person or a place?
Approach: offer a list of persons with the search results

When the user enters a search term, we offer him a list of persons, if possible.
List of persons video

ETH Library @ swisscovery

Search

ETH Library @ swisscovery

ETH Library & swisscovery is your gateway to over 30 million books, series, journals and non-book materials, as well as billions of electronic articles.

You can use ETH Library @ swisscovery to search the holdings of all ETH Zurich libraries, as well as more than 40 libraries across Switzerland.

Tips: If you want to search within the holdings of articles, select “swisscovery” instead of “swisscovery without CDI” in the search field.

More Information about swisscovery

Journal search for holdings at ETH Zurich

Using the journal search, you can search the journal holdings of the ETH Zurich libraries – both the e-journals and the printed journals. More about journals

Information and Support

info@library.ethz.ch
+41 44 632 21 23

Registration

You need to complete a one-time registration process to log in to swisscovery.

The basic instructions include:

- For members of ETH Zurich
- For private customers

Information about registration for companies

25.11.2021
List of persons - How is it done?

- Primo VE Customizing in AngularJS
- The search term (string) is wrapped in a SPARQL query and sent to the Wikidata Query Service (https://query.wikidata.org/sparql) via a proxy.
Where to put the person list?

Only a part of our customers use this feature, so it has to be visible, but not distracting or annoying.
Choose person page

- If the user selects a person, then a person page is rendered.
- A person page represents the person (“Thing”), is linkable in both directions, and collects resources belonging to that person.
- An URL represents this page.
Choose person page video
Person page links

Links from Metagrid
Links powered by Metagrid

Historisches Lexikon der Schweiz

Helvetica

Elites suisses au XIXe siècle

Links from beacon.findbuch
Links from the SeeAlso-Service pdx-aks, which is based on BEACON

Personenseite bei der Deutschen Digitalen Bibliothek (188)

Deutsche Biographie

Students or doctoral students (from Wikidata)

Leopold Ružička (Croatian scientist (1887-1976), Born: 1887-09-13)

Teachers (from Wikidata)

Daniel Vorländer (German chemist, Born: 1867-06-11)

Links in Archive
ETH Zurich University Archives (Inventarnummer: CH-001807-7:Hs 348)

Researcher profiles via IDs from Wikidata
ORCID record
Publons profile
Google Scholar profile
ResearchGate profile
Mendeley profile
Dimensions profile
Person page - How is it done?

- The collection of all information is offloaded to our express/nodejs system.

Primo VE

Nodejs route (cache): Multiple requests

WD: bibliographic data and researcher profiles

WD: archives at

WD: teacher and students

Metagrid beacon.findbuch Entityfacts
Precision vs Recall

- Here, we are in the area of tension between "important things are missing" and "there is too much wrong". This problem is also described by the term pair "Precision vs Recall".
- While "Precision" refers to the percentage of the actual result that is relevant, "Recall" refers to the percentage of the total relevant results that appear in the results list.

Wenjun Wu: More precise or more results?

You see the results of the search for a unique identification number (GND) of the person. Attention: You may not see all results because the identification number (GND) of the person is not assigned to all resources.

Select one of the search options:

9 results
- Search for the name, restricted by the identification number (GND) of the person OR the year of birth.
- You may not see all results for this person because the year of birth or identification number (GND) is not assigned to all resources.

24 results
- Search by name only
- You may also see wrong results if several people have the same name.

31 results
- Search for the name and name variants (from Wikidata)
- You may also see wrong results if several people have one of the name variants.
Precision vs Recall: Options

• Search for GND ID
• Search by name (from Wikdata, first name and last name) AND (GND ID OR date of birth)
• Search by name (from Wikdata, first name and last name)
• Search for name variants (from Wikidata)
Precision vs Recall - user survey

• This is not without complexity. Should we burden the users with this?
• This feature was welcomed because it makes transparent a situation that is there anyway.
• Which of the search alternatives is useful depends on the data and is different for different persons.
Person cards and links to person page video
Questions answered?

• Q1: The person page represents the person and is linkable in both directions (QID is part of URL).
• Q2: The user can choose between persons with the same name.
• Q3: With the person's metadata, controlled searches for the person's resources can be made in several ways.
• Q4: We test the search term to see if it could mean a person or a place and provide results for it. This could also be a possibility for other types, e.g. "buildings".
• Q4: We use wikidata graphs to create relationships between teacher and student in our discovery system.
Links

- Person page ETH Library @ swisscovery: https://eth.swisscovery.slsp.ch/discovery/search?search_scope=DiscoveryNetwork&tab=discovery_network&query=any,contains,%5Bwd%2Fperson%5DQ926125&vid=41SLSP_ETH:ETH&lang=en
- Person list in our test view: https://eth.swisscovery.slsp.ch/discovery/search?query=any,contains,staudinger&tab=discovery_network&search_scope=DiscoveryNetwork&vid=41SLSP_ETH:ETH_WD&lang=en&offset=0

Suggestions of people after the search input will come to the productive view in the first Q1 of 2022.
Thank you!
Questions?