

Master 2 Defense

Wikibase as an Infrastructure for Documents and Evaluation of Document Search

[Kushagra Singh BISEN](#)

Laboratoire Hubert Curien, Saint Etienne



Outline

1. Context and Problem
2. Contribution
3. Evaluation of Document Search
4. Conclusion

Context and Problem

Context and Problem

- Searching for domain specific information is tough
- The main focus in the work is building domain specific corpus, which can be searched and queried efficiently
- With Semantic Web, multiple tools such as Wikibase and QA systems over Knowledge Graphs have emerged

The issues are:

- The queries can be on the content or the metadata. Do we need 2 different answering systems?
- There are different existing search techniques. We need an evaluation for the search techniques to find what is the best solution?

Existing Solutions

There have been existing solutions to make a domain specific searchable and answerable system. In the paper, [\[1\]](#) specific information is extracted from the document with a neural network and generate triples from it. Although, the ontology and the triples generated are specific to that domain and cannot be extended. In the paper, [\[2\]](#) they proposed a neural network model for building a question answering system.

There have been various solutions to present data in structured fashion in Wikibase. The paper, [\[3\]](#) has proposed an ontology for wikibase but it is domain specific as well.

Research Questions

1. How can we have a uniform structured representation for documents of various types?
2. How to evaluate search for information?
3. Can or should we combine various search techniques for one unique solution method?

Contribution

Wikibase as Infrastructure for Documents

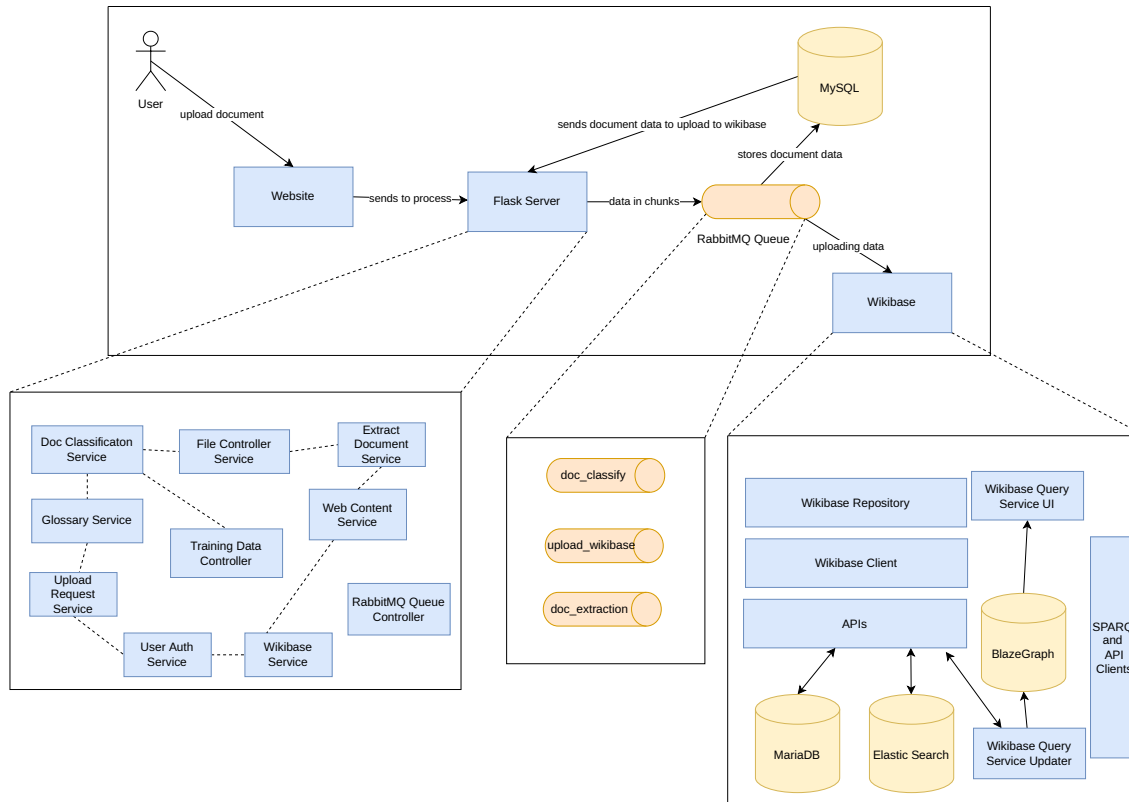
What is Wikibase?

- A free, open source knowledge graph
- It is the software behind Wikidata, one of the largest KG with 5 billion triples
- It is used to build open/enterprise knowledge graphs

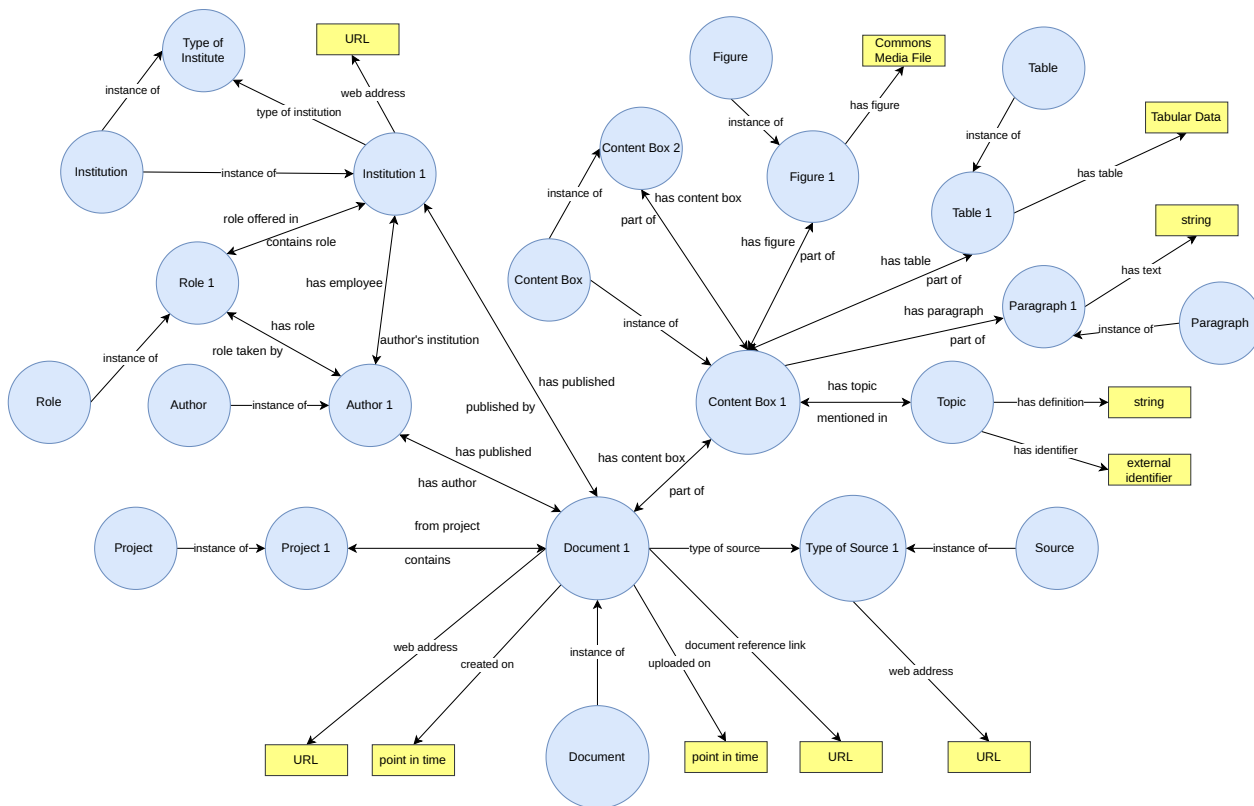
Why are we using Wikibase?

- Excellent for people out of scope of KG to interact with structured data
- Can add data by both humans and ingest heterogeneous data with bots
- Changes can be tracked and reversed, if needed
- Takes less time to setup with many features and scales up well

Wikibase as Infrastructure for Documents



Ontology for Wikibase as Infrastructure for Documents



Why do we need structured data about the document and its meta-data in RDF?

Example Use Case

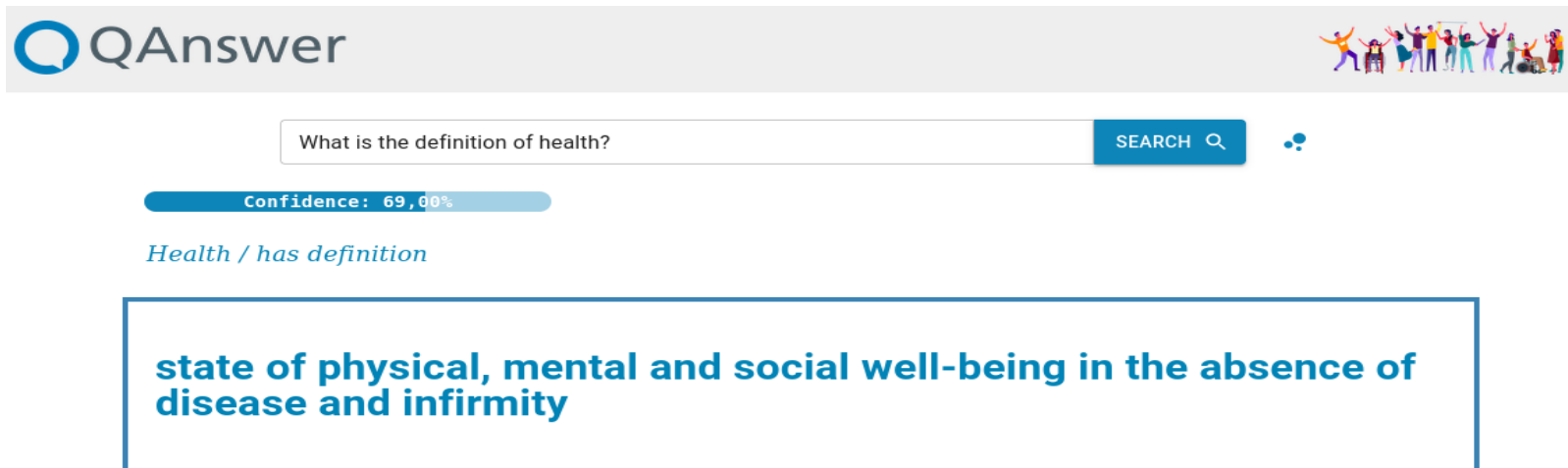
The documents we employed in the project for valorising them are in the domain of Disability Studies.

There are different types of questions that can arise, like,

- Single words and definitions
- Metadata of the document
- Unique answer from the documents
- Multiple answer from the documents
- Terms from the text

Combining Various Search Techniques

Response with Structured Data



The image shows a screenshot of the QAnswer search engine interface. At the top left is the QAnswer logo. On the right is an illustration of a diverse group of people. Below the logo is a search bar containing the text "What is the definition of health?". To the right of the search bar is a blue "SEARCH" button with a magnifying glass icon. Below the search bar is a blue progress bar indicating a confidence level of 69.00%. Below the progress bar is the text "Health / has definition". At the bottom, a blue-bordered box contains the definition: "state of physical, mental and social well-being in the absence of disease and infirmity".

QAnswer

What is the definition of health?

SEARCH


Confidence: 69,00%


Health / has definition

state of physical, mental and social well-being in the absence of disease and infirmity

Combining Various Search Techniques

Response with Free-Text

 QAnswer



meaning of ADAPT

SEARCH 🔍




Integrating Race Transforming Feminist Disability Studies by Sami Schalk and Jina B. Kim

Several histories of the disability rights movement present disability activism as a teleological narrative, with the 1990 passage of the ADA serving as that narrative's apex (Shapiro 1994; O'Brien 2004; Davis 2015). This momentous piece of legislation was the hard-won result of decades of protest by individuals and organizations like ADAPT (**American Disabled for Accessible Public Transit**), which continues disability rights activism in the present via protests against the so-called American Health Care Act in 2017, which aimed to partially repeal the Affordable Care Act, also known as Obamacare. However, histories of disability rights organizing render evident the white- ness of leadership in this wave of disability rights activism. In response to a white-dominated disability movement, an increasing number of voices have articulated the necessity of disability activism less centered on state legislation. As Mia Mingus writes, "I want us to tap into the transformative powers of disability, instead of only gaining access to the cur- rent system. . . . We don't simply want to join the ranks of the privileged, we want to challenge and dismantle those ranks and question why some people ...



[SHOW MORE](#)

Confidence: 89,69%

Open pdf 

Combining Various Search Techniques

Response with Keywords




children seperated from parents

SEARCH 🔍

☰

Autistic while black How autism amplifies stereotypes by Catina Burkett

Race is seldom reported in autism studies because the condition is often overlooked in minority **children** and adults. The statistics on autism maintained by the U.S. Centers for Disease Control and Prevention provide no information on race or ethnicity for autistic adults, although there has been an increase in diagnoses among minority **children**. Even so, one in four minority **children** — most of whom are black or Hispanic — still misses out on a diagnosis. People often blame low rates of diagnosis and intervention in minority **children** on the children's **parents** underreporting symptoms or lacking the resources to access expert care. But this is a stereotype as well: Plenty of minority **children** do not come **from** underprivileged families nor lack access to healthcare providers and teachers to screen them for autism. To dispel harmful stereotypes, researchers must include and track autistic black people. For adult black women with autism to get programs and services that address our needs, researchers first need to acknowledge that we exist.

[Open pdf](#) 

Evaluation of Document Search

Search Techniques Experiment

We undertook a search technique experiment for searching for information in a particular domain. The information was to be searched with 5 methods,

1. Searching with a PDF viewer
2. Searching over Wikibase
3. Searching over Structured Data
4. Elastic Search over Documents
5. Free-Text search over Documents

Search Techniques Experiment (Contd.)

Evaluation Techniques

We invited experimental subjects to search for information using different search techniques. We prepared a questionnaire for each search method which were divided into two parts,

- Search Instruction Questionnaire
- User Experience Questionnaire

There were 17 experimental subjects, 6 questions (5 true and 1 false) to search for and 1 UEQ questionnaire to be filled for each search method

Search Techniques Experiment (Contd.)

Search Instruction Questionnaire

The users had to search for information in 2 minutes and note the relevancy of the information retrieved on a scale of 1 to 7 (the higher the better), if they found an answer and the timestamp in seconds if they found the answer.

The questions answered from this experiment is,

- Did the user find an answer?
- What was the time taken through each method?
- Which method had the most relevant answers?

Search Techniques Experiment (Contd.)

Search Instruction Questionnaire (Contd.)

The user is expected to form his own keywords and questions to search for the answer by reading the instructions. We introduced a question with no answer present in the document corpus to count for false positives. The instructions were,

No.	Instruction to User	Answer available?
1	Find text about the racism faced by black feminists	Yes
2	Find text about elitism in american womens movement	Yes
3	Find text about human rights of minors	No
4	Find text about racism in United States	Yes
5	Find text about ableism in prison	Yes
6	Find text about police violence for disabled people	Yes

Search Techniques Experiment (Contd.)

User Experience Questionnaire

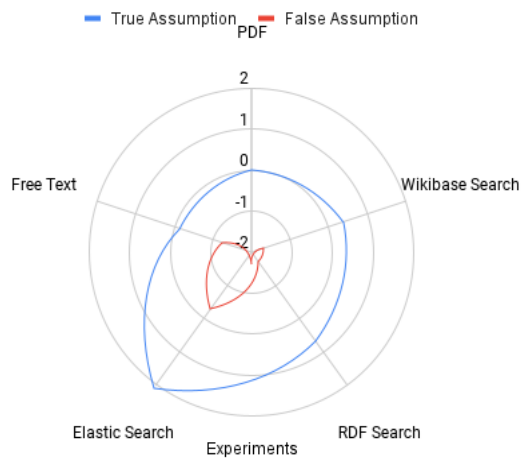
The users had to fill a questionnaire explaining their user experience after every search method. The questionnaire contains 26 individual items divided into 6 subscales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty)

The questions answered from this experiment are,

- What does the user feel about the usefulness of the method? i.e Pragmatic Value
- What does the user feel about the ease of use of the method? i.e Hedonic Value

Results

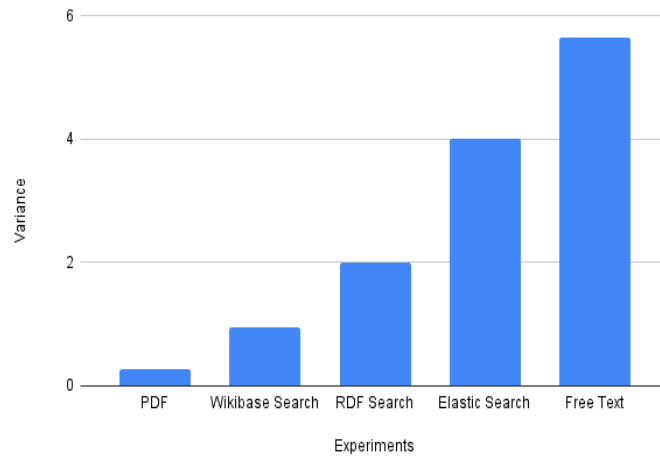
Search Methods



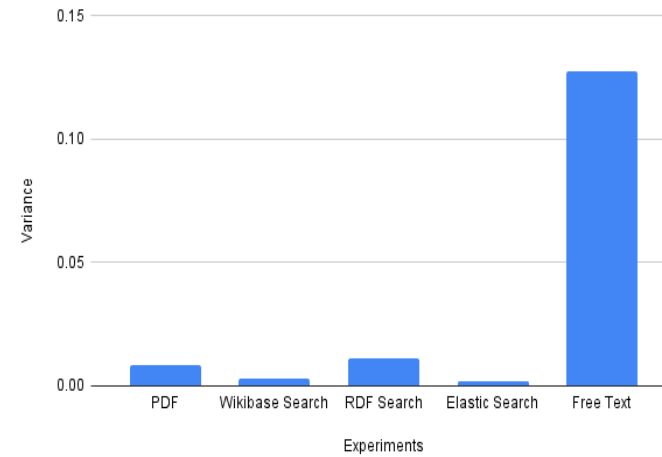
Values of relevant answer from a search method

Results

User Experience Questionnaire



UEQ Pragmatic



UEQ Hedonic

Results

Discussion - Search Methods

After analysing the results from search experiments. According to the users,

- Elastic search found the most relevant answers
- It was followed by QAnswer RDF Search and Wikibase Search
- With the question with no answer available, users found relevant answer with Elastic Search.

Results (Contd.)

Discussion - User Experience Questionnaire

After analysing the results from user experience questionnaire. According to the users,

- The users found free-text over the documents to be more efficient and useful
- It is followed by Elastic Search, and QAnswer Search over RDF in efficiency
- Users found searching over documents with free-text search with the most ease-of-use
- Users found other methods similar in ease of use

Conclusion

Conclusion

- We find that as elastic search had the most relevant answers. It also provided information for instruction with no actual answer in the document corpus
- Elastic Search is not a good method as it provides users with a false sense of information
- In questions we need an exact answer from, structured data is important
- In the subscales of UEQ, free-text performed the best which further supports the argument of having a definitive search system over document corpus.

Future Work

As I have 2 more weeks in my internship,

- The existing ontology is the one we adjusted to our needs for the project, I will work on aligning with existing ontologies such as Dublin Core, FOAF, LOINC, DoCO, UN Document Ontology
- I will work on better methods of analysing the data generated in the evaluation of document search
- We are preparing a demo paper on searching over free-text and RDF at the same time
- We are also preparing a paper demonstrating wikibase as an infrastructure for valorising documents