Methodology for Creating a Community Corpus using a Wikibase Knowledge Graph

Sara Assefa Alemayehu, Kushagra Singh Bisen, Pierre Maret, Alexandra Creighton, Rachel Gorman, Bushra Kundi, Thumeka Mgwgwi, Fabrice Muhlenbach, Serban Dinca-Panaitescu, Christo El Morr

Methodology for Creating a Community Corpus using a Wikibase Knowledge Graph

Sara Assefa Alemayehu[1], Kushagra Singh Bisen[1][0000−0003−0950−6043], Pierre Maret[1,4], Alexandra Creighton[2], Rachel Gorman[2], Bushra Kundi[2], Thumeka Mgwgwi[3], Fabrice Muhlenbach[1], Serban Dinca-Panaitescu[2], and Christo El Morr[2]

[1] Université Jean Monnet Saint Etienne, France
[2] York University, School of Health Policy and Management, Canada
[3] School of Gender, Sexuality and Women's Studies, York University, Canada
[4] The QA Company, France

Abstract. The Wikibase environment is provided by the Wikimedia foundation. It provides a suite of tools and a collaborative environment and implements Semantic Web techniques to model and store data. It has been shown that the Wikibase environment is well adapted to store and manage documents with their meta-data and that it can be associated with a search mechanism that facilitates the accessibility to this set of information. Several implementations of such a web platform have been proposed for various communities, but no methodology has been yet proposed. A methodology would help to identify tasks and easily reproduce them for the new implementation of community Wikibases, together with the search function that provides access to the information stored. In this paper, we propose a methodology for domain-specific Wikibase instantiation, for documents and their meta-data, and we identify some tools related to this methodology.

Keywords: Wikibase · Methodology · Semantic Web · Domain-Specific Corpus · Knowledge Graph

1 Introduction

With the rise of the Semantic Web [2], technologies assisting better representation of data and entities have emerged. Knowledge Graphs [11] aims to serve as an ever-evolving shared substrate of knowledge within an organisation or community [15]. Searching for domain-specific information on the web is rather tough. Communities employ domain-specific vocabulary and context while accessing information which is often overlooked while searching for information on the web. Knowledge Graph infrastructures such as Wikibase can be employed to build a community-specific knowledge graph to improve domain-specific representation and access. The Wikibase environment is a set of software provided by the Wikimedia Foundation and is the environment behind Wikidata[1]. Wikidata supports the Wikimedia family of projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others. Due to the success of Wikidata, the Wikibase environment has received increasing interest in storing textual data. It is well suited for setting up data archives that interoperate with the semantic web via open standards.

Large amounts of textual data are continuously communicated around the world. Textual data can be unstructured or structured. Unstructured textual data (also called free text) doesn't have a predefined data model, as opposed to structured data which is commonly modelled in a relational or graph data model [6]. A relational or graph data model is a robust data structure used for retrieving, organizing and managing data.

The Wikibase environment implements structured data, and more specifically a relational data model internally, and a graph data model for its interface. Several implementations of community-specific Wikibases have been created. Web sites built on top of these Wikibases facilitate the management and access to the information they contain [5] [7]. In [18] (manuscript in evaluation) it has been shown that the Wikibase environment is well adapted for the management of documents and their meta-data. In this implementation, the search function enables free text queries to uniformly access the document meta-data as well as the document contents.

Alternatives to Wikibase are Semantic MediaWiki (SMW) and Cargo [12]. SMW and Cargo are not developed for collaborative creation and maintenance of a knowledge base containing data that can be used by third parties systems. This is enabled in Wikimedia projects, and therefore, in the Wikibase environment. Thus, organizations can curate data for third-party consumption, for instance, GLAM (galleries, libraries, archives, and museums) institutions [3].

Therefore, the Wikibase environment seems well adapted for organizing data and serving as a data store to a web platform for efficiently accessing documents and their meta-data. Despite this, to the best of our knowledge, there is no methodology described to implement this kind of architecture. Thus, in this paper, we propose a methodology to organise the design and the implementation process of domain-specific Wikibases and a search function which substantially helps in the access to community information and documents.

The rest of this paper is organized as follows: In Section 2, we list some related work done on methods and tools for building and querying domain-specific corpora and for structuring and implementing community Wikibase for domain-specific documents. In Section 3, we describe the methodology we propose, we describe the tools that can be used in each task of the process, and we give an illustration of each task using examples from the Disabilitywiki project. This project aims at collecting information and improving access to disability and human-right related data and documents. It implements a Wikibase and a Natural Language Processing (NLP) enabled multilingual search engine [9]. We conclude and highlight future work in Section 4.

2 Related Work

Many studies have been recently conducted on the implementation of document repositories for accessing the information on domain-specific corpora. However unstructured data affect querying process performance and give the difficulty of the user to manage or retrieve it. Many attempts have been made to reorganize or directly process this data. Yafooz et al. discussed in [19] methods of managing unstructured data in the relational database management system and this study showed the significance of managing this data. Furthermore, he highlights the differences in managing such data with relational or NoSQL databases. He presents the methods that are often used to manage unstructured data in relational databases, however, how this structured data is queried by end-users, and what method and tools are used, are not specified.

Pampari et al. proposed in [16] a novel methodology to generate domain-specific large-scale question answering (QA) datasets by re-purposing existing annotations. This study demonstrates an instance of this methodology in generating a large-scale QA dataset for unstructured electronic medical records by leveraging existing expert annotations on clinical notes for various NLP tasks from the community-shared i2b2 datasets. The resulting corpus (emrQA) has 1 million question-logical forms and 400,000+ question-answer evidence pairs. The authors finally characterize the dataset and explore its learning potential by training baseline models for the question to logical form and question-to-answer mapping. Therefore, this paper addresses the lack of publicly available EMR (unstructured Electronic Medical Records) by creating a large-scale dataset emrQA. In this study, they show how to structure and access textual data but no methodology and tools are proposed.

The Enslaved project has developed a platform with linked open data (LOD) i.e. structured data available under an open license under the Wikibase environment, to interconnect individual projects and databases. A LOD-based approach facilitates federated searching and browsing across all linked project data on Enslaved.org [7]. They build a robust, open-source architecture to discover, connect, and visualize 600,000+ people's records and 5 million data points from archival fragments and spreadsheet entries that show the lives of the enslaved in richer detail. The enslaved project doesn't have any automatic data upload workflow and it always requires computer experts to feed information directly to the data store.

Zhou et al proposed in [20] a real-world dataset from the Enslaved project as a potential complex alignment benchmark. The benchmark consists of two resources, the Enslaved Ontology along with a Wikibase repository holding a large number of instance data from the Enslaved project, as well as a manually created reference alignment between them. The alignment was developed in consultation with domain experts in the digital humanities. The two knowledge graphs and the reference alignment were designed and created by ontologists and historians to support data representation, sharing, integration, and discovery. Additionally, they take advantage of Wikibase as a tool to represent the data, which is convenient for users with any level of expertise in its use. Zhou et al didn't introduce any methodology and tools suite for their process.

Diefenbach et al present in [5] how Wikibase is used as the infrastructure behind the "EU Knowledge Graph", which is deployed at the European Commission. This graph mainly integrates projects funded by the European Union and is used to make these projects visible to and easily accessible by citizens with no technical background. Moreover, the authors explain this deployment compared to a more classical approach to building RDF knowledge graphs, and they point to other projects that are using Wikibase as the underlying infrastructure. This paper presents how Wikibase can be used as an infrastructure for knowledge graphs and shows that while Wikibase is not as flexible as a traditional RDF deployment, it offers many out-of-the-box services that are either necessary or convenient for deploying a knowledge graph infrastructure. No methodology is described to build such a system.

Li et al in [13] propose AliMe KG, a domain knowledge graph in E-commerce that captures user problems, points of interest (POI), item information and relations thereof. It helps to understand user needs, answer pre-sales questions and generate explanation texts. They applied AliMe KG to several online business scenarios such as shopping guides, question answering over properties and selling point generation. In the paper, they systematically introduce how they construct a domain knowledge graph from free text, and demonstrate its business value with several applications.

3 Proposed Methodology

Our target to propose methodology for the design and implementation of systems that enable the implementation of Wikibases and facilitate access to community documents and their meta-data. These systems shall be composed of three main components: a Wikibase instance used to store the meta-data of the documents; File repositories used to store the documents; and a User interface (UI) used to upload, query, and access data and documents.

We propose the description of this methodology in three parts as illustrated in Figure 1. Each part is divided into several tasks. We use event-driven process chain (EPC) diagrams to describe the control-flow structure of each part as a chain of events and tasks. Part 1 concerns the *domain knowledge modelling and Wikibase setup*. Part 2 relates to the process of *uploading documents*. It is started when the knowledge environment is ready and when documents have to be loaded into the system. And part 3 relates to the implementation of *search function* over the documents and meta-data. It is started once the knowledge environment is ready.

3.1 Domain knowledge modeling

After the identification of the domain of interest for the community, this part aims at preparing the knowledge environment in which the targeted system will operate. The tasks of this part are presented in Figure 2.

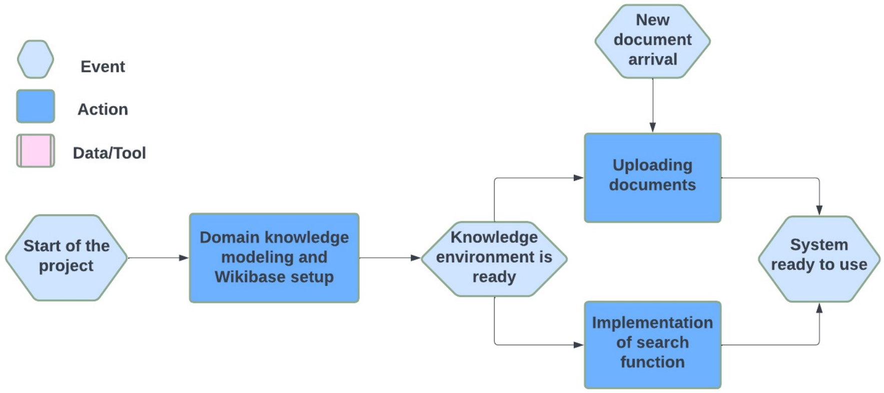

Event
Action
Data/Tool
New document arrival
Uploading documents
Start of the project
Domain knowledge modeling and Wikibase setup
Knowledge environment is ready
System ready to use
Implementation of search function

Fig. 1. General description of the methodology

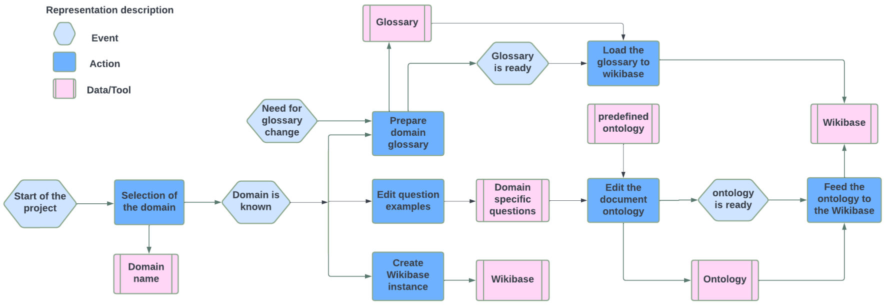


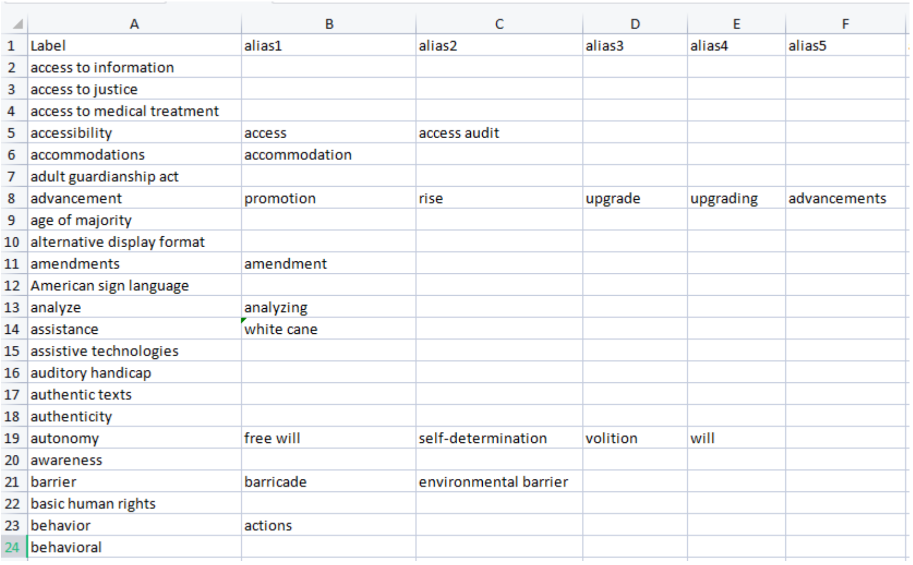
Fig. 2. Part 1 of the methodology: Domain knowledge modelling and Wikibase setup

Prepare domain glossary This task consists of the preparation of domain-specific vocabulary and the definition of terms. The output of this process consists of a list of terms, synonyms and definitions. The glossary content will be added to the Wikibase in a further task.

The glossary will serve as an input of the task *Annotate document* that will associate documents or sub-parts of documents (called Content box in the ontology) to glossary terms.

Preparing the glossary can be done manually by domain experts or with the help of tools reading automatically documents and extracting terms (i.e. MonkeyLearn)[8]. The automatic glossary term extraction shall be seen as a bootstrapping technique that generates an initial list of terms that domain experts will revise [17]. Some of examples of tools used to automatically read documents and extract terms are IBM Watson and MonkeyLearn . The output format of this task should be a CSV file that will serve as an input to the tasks *Load glossary to Wikibase* and *Annotate Document*.

In the Disability wiki project, domain experts have manually prepared in a table the glossary terms as shown in Figure 3. Each line represents a term and its synonyms. Definitions of terms are edited in a different file, each term is followed by its definition in the next column.



	A	B	C	D	E	F
1	Label	alias1	alias2	alias3	alias4	alias5
2	access to information					
3	access to justice					
4	access to medical treatment					
5	accessibility	access	access audit			
6	accommodations	accommodation				
7	adult guardianship act					
8	advancement	promotion	rise	upgrade	upgrading	advancements
9	age of majority					
10	alternative display format					
11	amendments	amendment				
12	American sign language					
13	analyze	analyzing				
14	assistance	white cane				
15	assistive technologies					
16	auditory handicap					
17	authentic texts					
18	authenticity					
19	autonomy	free will	self-determination	volition	will	
20	awareness					
21	barrier	barricade	environmental barrier			
22	basic human rights					
23	behavior	actions				
24	behavioral					

Fig. 3. Edited glossary for disabilityWiki

Edit question examples In this task, domain experts prepare prototype questions, which will be used to illustrate the types of knowledge expected by end-users and to test and train the search functionality. End-users can also be involved in this task. Experts can be guided by question types such as retrieval definitions, search for document meta-data, search for a unique answer from the documents, search for multiple answers from the documents, and search terms from the text.

As example, domain experts in the illustrative project prepared questions such as:

- What is the definition of health?
- Subject of crpd article 11
- When starts prison uprising? (grammatically wrong questions is supported by the search engine)

Edit the document ontology In this task, one has to create a formal description of the documents and their meta-data used for the targeted domain. The questions previously prepared by domain experts are used as illustrations of the need for knowledge to be represented in the documents. We propose a generic document ontology (Figure 4) that should be specifically adapted to the types and content of the documents used in the chosen domain.

Editing the ontology can be done manually or utilizing a graphical tool such as Protégé, which is the most used one [14] and supports many input and output formats. The output format of the ontology should be a file in CSV format (generated with a Protégé dedicated plugin). This file will be consumed later in the methodology by the automated task *Feed the ontology into the Wikibase.*

As an illustration from the Disabilitywiki project, we can mention that the generic ontology has been enriched with the concept *Article.* It is related to the concept *Paragraph* with a *Kind of* relation. The concept *Article* is necessary for this domain because the corpus will contain also legal texts which will be better represented by using this new concept.

3.2 Wikibase setup

Wikibase can be described as a tool that offers a connected infrastructure to implement and maintain a large-scale knowledge graph. Launching the Wikibase environment for collecting and accessing information for a specific domain requires two main processes:

Creating Wikibase instance The Wikibase proposes a Docker image for the Wikibase stack. After installing Docker, the user should clone locally this Wikibase Docker. image[5]. Wikibase requires also the installation of Docker images for the Blazegraph database, MySQL database and some other components. A Docker-compose file organizes these images. A bot account is necessary for

[5] https://medium.com/@thisismattmiller/wikibase-for-research-infrastructure-part-1-d3f640dfad34

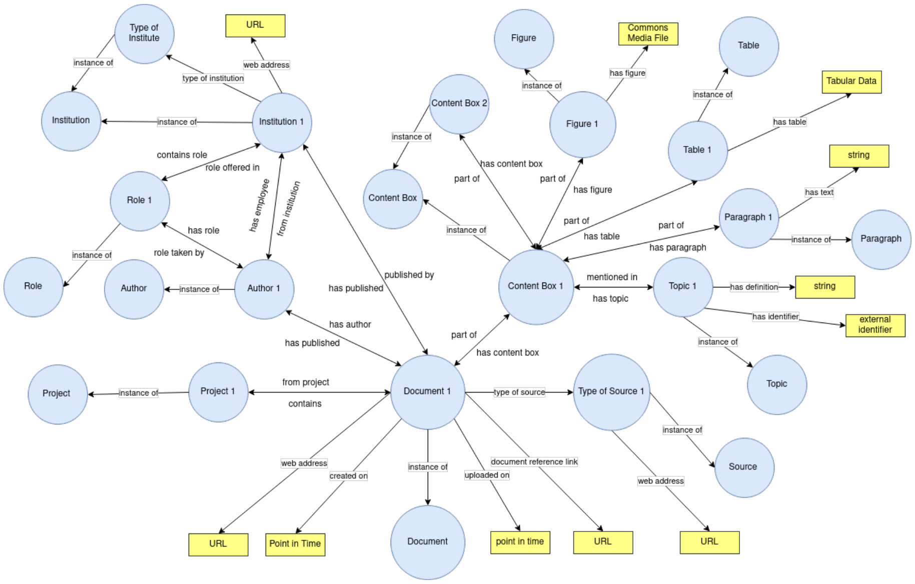


Fig. 4. Edited ontology for DisabilityWiki

launching bot programs for populating the wikibase with terms: items and properties. Terms can be also added by hand using special Wikibase pages.

Figure 5 shows the fresh look of the Disability wiki main page after installation of the Wikibase.

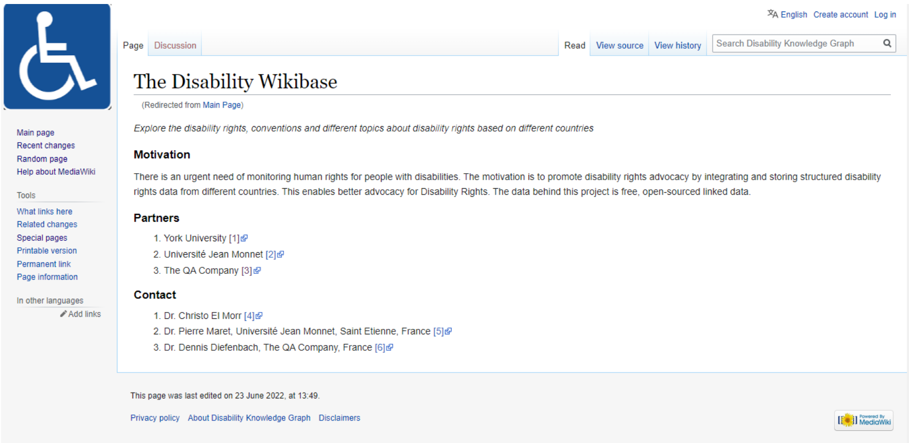
Load the glossary to the Wikibase Using the bot account, bulk edits can be launched using the pywikibot tool provided with the platform. Pywikibot is a Python library that interacts with the MediaWiki API. The terms from the domain glossary shall be loaded with this tool.

Load the ontology to the Wikibase As in the previous task, the prepared document ontology is loaded to the Wikibase by using the same bot program.

3.3 Upload Document Process

The uploading process contains several tasks to upload and manage documents and manage the meta-data. It is summarized in Figure 6. A repository is used to store the documents, and Wikibase is used to store the meta-data as well as the links to the documents.

Upload documents Figure 7 shows the user interface for the upload of documents and the edition of the first set of metadata related to them. Some of



The Disability Wikibase

(Redirected from Main Page)

Explore the disability rights, conventions and different topics about disability rights based on different countries

Motivation

There is an urgent need of monitoring human rights for people with disabilities. The motivation is to promote disability rights advocacy by integrating and storing structured disability rights data from different countries. This enables better advocacy for Disability Rights. The data behind this project is free, open-sourced linked data.

Partners

1. York University [1]
2. Université Jean Monnet [2]
3. The QA Company [3]

Contact

1. Dr. Christo El Morr [4]
2. Dr. Pierre Maret, Université Jean Monnet, Saint Etienne, France [5]
3. Dr. Dennis Diefenbach, The QA Company, France [6]

This page was last edited on 23 June 2022, at 13:49.

Privacy policy About Disability Knowledge Graph Disclaimers

Fig. 5. Disabilitywiki wikibase main page

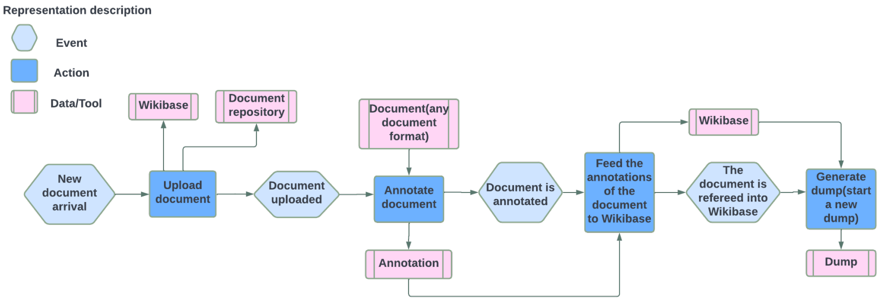
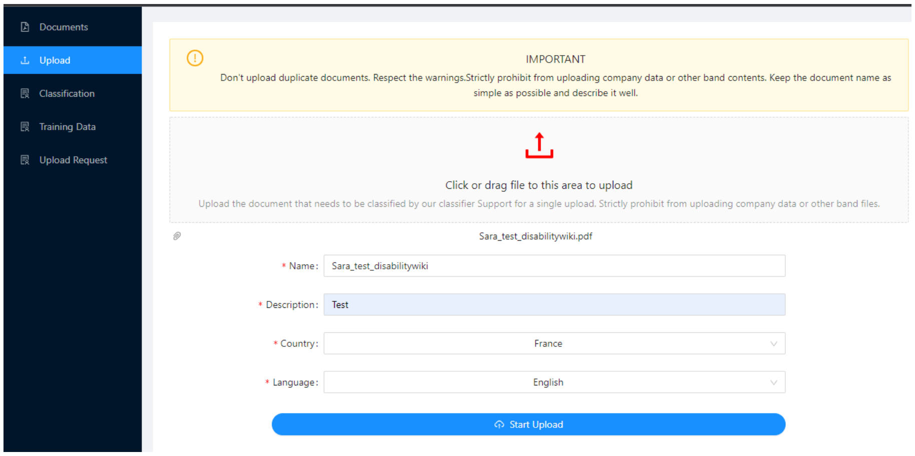

Representation description
Event
Action
Data/Tool
New document arrival
Upload document
Wikibase
Document repository
Document uploaded
Document(any document format)
Annotate document
Annotation
Document is annotated
Feed the annotations of the document to Wikibase
Wikibase
The document is refereed into Wikibase
Generate dump(start a new dump)
Dump

Fig. 6. Part 2 of the methodology: Document upload process

the meta-data defined in the document ontology (such as author, title, URL, country, institution, etc. in the Disabilitywiki project) is edited or automatically extracted by document analysis. Once checked (and possibly modified), this information is loaded to the Wikibase.



IMPORTANT

Don't upload duplicate documents. Respect the warnings.Strictly prohibit from uploading company data or other band contents. Keep the document name as simple as possible and describe it well.

Click or drag file to this area to upload

Upload the document that needs to be classified by our classifier Support for a single upload. Strictly prohibit from uploading company data or other band files.

Sara_test_disabilitywiki.pdf

* Name: Sara_test_disabilitywiki

* Description: Test

* Country: France

* Language: English

Fig. 7. Document uploading process in disabilitywiki

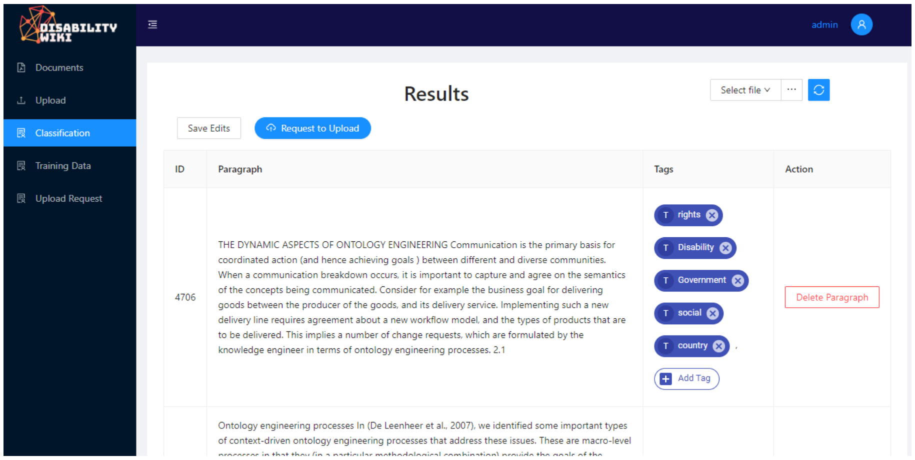
Annotation process The annotation process consists of labelling the content of the uploaded documents with glossary terms (meta-data). The document ontology is used to organize structured data from these annotations. The result of the annotation task is stored in a structured way(CSV, RDB).

In the Disabilitywiki project, each paragraph from uploaded documents is analyzed by an integrated machine learning algorithm to be tagged with glossary terms. The algorithm is using Latent Dirichlet Allocation method, which is an unsupervised learning that identifies hidden relationships in data. Figure 8 shows the user interface where a tagged paragraph has to be associated with glossary terms. The terms can be approved, deleted or extended by the user.

Feeding the annotations of the document to Wikibase Once the annotation process of a document is done, the Pywikibot tool can be used to load the CSV file resulting.

3.4 Implement Search Function

Once the knowledge environment is set up (part 1 described in 3.1), the implementation of a search function for end-users requires connecting a search engine



Results

ID	Paragraph	Tags	Action
4706	THE DYNAMIC ASPECTS OF ONTOLOGY ENGINEERING Communication is the primary basis for coordinated action (and hence achieving goals) between different and diverse communities. When a communication breakdown occurs, it is important to capture and agree on the semantics of the concepts being communicated. Consider for example the business goal for delivering goods between the producer of the goods, and its delivery service. Implementing such a new delivery line requires agreement about a new workflow model, and the types of products that are to be delivered. This implies a number of change requests, which are formulated by the knowledge engineer in terms of ontology engineering processes. 2.1	T rights ⊗ T Disability ⊗ T Government ⊗ T social ⊗ T country ⊗ , + Add Tag	Delete Paragraph
	Ontology engineering processes In (De Leenheer et al., 2007), we identified some important types of context-driven ontology engineering processes that address these issues. These are macro-level processes in that they (in a particular methodological combination) provide the goals of the		

Fig. 8. Tagged paragraph of documents with glossary terms

to a user interface and the domain wikibase and its related document repository. Also, the search mechanism has to be trained to this corpus of information. This 3rd part of the methodology is summarized in Figure 9 and is described hereafter.

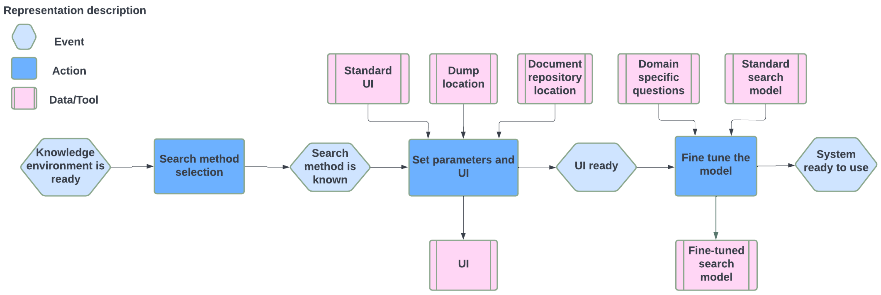


Fig. 9. Part 3 of the methodology: search process implementation

Search method selection The Wikibase user interface proposes a by-default search mechanism (based on Elasticsearch). However, this search function is not suitable for our methodology because the corpus that is created is composed of 2 parts: the domain documents and the Wikibase for the meta-data.

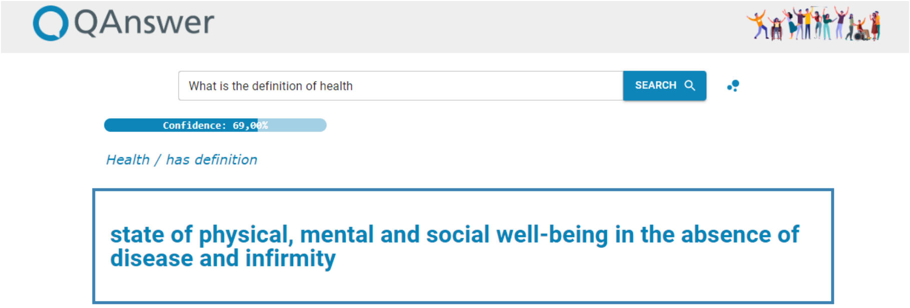
To query the heterogeneous corpus produced in our methodology, two customized search techniques are implemented: Elasticsearch and QAnswer. The

Elasticsearch technique is a classic full-text search. The advantages of this technique are that it is real-time, it can query any kind of textual content, and does not need a training step. However, it has the drawback that it does not consider semantic relations between data, natural language questions cannot be answered with precision, and it tends to present answers to user events if the corpus does not contain the information[18]. The QAnswer technique [4] proposes to analyse the semantic meaning of the question expressed with free-text or keywords and to search for the answers in the corpus. QAnswer can be executed on structured data (for example: knowledge graphs) or unstructured data (documents). The combination of the two techniques (heterogeneous corpus composed of structured and unstructured) has been proposed in [10]. The QAnswer search technique is therefore adapted to our methodology.

Parameters and user interface setup In this task, one has to set the parameters of the selected search method and implement a user interface (UI) for the domain-specific user searches. A standard user interface (if available) can be used as an input and can be customized, or a newly created UI shall be created. The document repository location and Wikibase dump location are examples of parameters that should be entered to parameterize the system.

Fine tuning of the search model This step is only valid in the case of the selection of a model that has to be trained. The questions previously prepared by domain experts in the *Edit question examples* task (See 3.1) are used as inputs for domain experts to train the system. A standard selected search model is used and trained with these questions to produce a fine-tuned search model. Additional domain questions can be added to possibly train the search model for better results.

An example of the search result on disabilitywiki UI is shown in Figure 10.



What is the definition of health

Confidence: 69,00%

Health / has definition

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

Fig. 10. Search result on Disabilitywiki UI

4 Conclusion and perspectives

In this paper, we proposed a methodology for the design of a community corpus using Wikibase. In our approach, the community corpus consists of domain knowledge and domain documents. The methodology consists of three parts: domain knowledge modelling, document upload, and implementation of a search function. We are not aware of the existence of such a methodology before, and we hope it could help some communities to implement Wikibases to share their corpora. The search function over the created corpus can take advantage of Elasticsearch or Question-Answering techniques. The QAnswer search engine has the advantage of accepting free text and keyword questions, and it can query uniformly the wikibase information as well as the document content.

The methodology could be integrated into a framework that would incorporate automated or semi-automated tools taking advantage of various techniques such as topic extraction, document summarization, etc.

References

1. Wikidata:wikibase. https://cutt.ly/aLF2R6Y, Mayr 2022. [Retrieved July 16, 2022].
2. Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001.
3. Jeroen De Dauw. Semantic mediawiki vs wikibase vs cargo. https://cutt.ly/5LnHLQP, May 2022. [Retrieved June 20, 2022].
4. Dennis Diefenbach, Andreas Both, Kamal Singh, and Pierre Maret. Towards a question answering system over the semantic web. *Semantic Web*, 11(3):421–439, 2020.
5. Dennis Diefenbach, Max De Wilde, and Samantha Alipio. Wikibase as an infrastructure for knowledge graphs: The eu knowledge graph. In *International Semantic Web Conference*, pages 631–647. Springer, 2021.
6. Adanma Cecilia Eberendu et al. Unstructured data: an overview of the data of big data. *International Journal of Computer Trends and Technology*, 38(1):46–50, 2016.
7. Enslave.org. People of the historical slave trade. https://enslaved.org/. [Retrieved June 20, 2022].
8. Jordán Pascual Espada, Jaime Solís Martínez, Irene Cid Rico, and Luis Emilio Velasco Sánchez. Extracting keywords of educational texts using a novel mechanism based on linguistic approaches and evolutive graphs. *Expert Systems with Applications*, page 118842, 2022.
9. Rachel Gorman, Pierre Maret, Alexandra Creighton, Bushra Kundi, Fabrice Muhlenbach, Alexis Buettgen, Enakshi Dua, Geoffrey Reaume, Thumeka Mgwigwi, Serban Dinca-Panaitescu, et al. The potential of an artificial intelligence for disability advocacy: The wikidisability project. In *Public Health and Informatics*, pages 1025–1026. IOS Press, 2021.
10. Kunpeng Guo, Dennis Diefenbach, and Christophe Gravier. Qanswer: Towards question answering search over websites. 2022.

11. Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. *Knowledge Graphs.* Number 22 in Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool, 2021.
12. Yaron Koren. Working with mediawiki. https://workingwithmediawiki.com/book/. [Retrieved July 19, 2022].
13. Feng-Lin Li, Hehong Chen, Guohai Xu, Tian Qiu, Feng Ji, Ji Zhang, and Haiqing Chen. Alimekg: Domain knowledge graph construction and application in e-commerce. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2581–2588, 2020.
14. MA Musen. The protégé project: A look back and a look forward. ai matters, 1 (4), 4-12, 2015.
15. Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale Knowledge Graphs: Lessons and Challenges. *ACM Queue*, 17(2), April 2019.
16. Anusri Pampari, Preethi Raghavan, Jennifer Liang, and Jian Peng. emrqa: A large corpus for question answering on electronic medical records. *arXiv preprint arXiv:1809.00732*, 2018.
17. Omar Qawasmeh, Maxime Lefranois, Antoine Zimmermann, and Pierre Maret. Computer-assisted ontology construction system: Focus on bootstrapping capabilities. In *European Semantic Web Conference*, pages 60–65. Springer, 2018.
18. Kushagra Singh Bisen, Sara Assefa Alemayehu, Pierre Maret, Alexandra Creighton, Rachel Gorman, Bushra Kundi, Thumeka Mgwgwi, Fabrice Muhlenbach, Serban Dinca-Panaitescu, Dennis Diefenbach, Kunpeng Guo, and Christo El Morr. Wikibase as an Infrastructure for Community Documents: The example of the Disability Wiki Platform. In *Semantics 2022 - 18th International Conference on Semantics Systems*, Vienna, Austria, September 2022.
19. Wael MS Yafooz, Siti ZZ Abidin, Nasiroh Omar, and Zanariah Idrus. Managing unstructured data in relational databases. In *2013 IEEE Conference on Systems, Process & Control (ICSPC)*, pages 198–203. IEEE, 2013.
20. Lu Zhou, Cogan Shimizu, Pascal Hitzler, Alicia M Sheill, Seila Gonzalez Estrecha, Catherine Foley, Duncan Tarr, and Dean Rehberger. The enslaved dataset: A real-world complex ontology alignment benchmark using wikibase. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 3197–3204, 2020.