<u>Original Paper</u>

A Virtual Community for Disability Advocacy: Development of a Searchable Artificial Intelligence–Supported Platform

Christo El Morr[1], PhD; Pierre Maret[2], PhD; Fabrice Muhlenbach[2], PhD; Dhayananth Dharmalingam[2], MSc; Rediet Tadesse[2], MSc; Alexandra Creighton[1], MA; Bushra Kundi[1], BHS; Alexis Buettgen[1], PhD; Thumeka Mgwigwi[3], MA; Serban Dinca-Panaitescu[1], PhD; Enakshi Dua[4], PhD; Rachel Gorman[1], PhD

[1]School of Health Policy and Management, Faculty of Health, York University, Toronto, ON, Canada

[2]CNRS, UMR 5516, Laboratoire Hubert Curien, Université Jean Monnet Saint Etienne, Saint Etienne, France

[3]Student Learning and Academic Success Department, York University Libraries, York University, Toronto, ON, Canada

[4]School of Gender, Sexuality and Women's Studies, York University, Toronto, ON, Canada

Corresponding Author:
Christo El Morr, PhD
School of Health Policy and Management
Faculty of Health
York University
Stong College Room 306
4700 Keele St
Toronto, ON, M3J 1P3
Canada
Phone: 1 4167362100
Email: elmorr@yorku.ca

Abstract

Background: The lack of availability of disability data has been identified as a major challenge hindering continuous disability equity monitoring. It is important to develop a platform that enables searching for disability data to expose systemic discrimination and social exclusion, which increase vulnerability to inequitable social conditions.

Objective: Our project aims to create an accessible and multilingual pilot disability website that structures and integrates data about people with disabilities and provides data for national and international disability advocacy communities. The platform will be endowed with a document upload function with hybrid (automated and manual) paragraph tagging, while the querying function will involve an intelligent natural language search in the supported languages.

Methods: We have designed and implemented a virtual community platform using Wikibase, Semantic Web, machine learning, and web programming tools to enable disability communities to upload and search for disability documents. The platform data model is based on an ontology we have designed following the United Nations Convention on the Rights of Persons with Disabilities (CRPD). The virtual community facilitates the uploading and sharing of validated information, and supports disability rights advocacy by enabling dissemination of knowledge.

Results: Using health informatics and artificial intelligence techniques (namely Semantic Web, machine learning, and natural language processing techniques), we were able to develop a pilot virtual community that supports disability rights advocacy by facilitating uploading, sharing, and accessing disability data. The system consists of a website on top of a Wikibase (a Semantic Web–based datastore). The virtual community accepts 4 types of users: information producers, information consumers, validators, and administrators. The virtual community enables the uploading of documents, semiautomatic tagging of their paragraphs with meaningful keywords, and validation of the process before uploading the data to the disability Wikibase. Once uploaded, public users (information consumers) can perform a semantic search using an intelligent and multilingual search engine (QAnswer). Further enhancements of the platform are planned.

Conclusions: The platform ontology is flexible and can accommodate advocacy reports and disability policy and legislation from specific jurisdictions, which can be accessed in relation to the CRPD articles. The platform ontology can be expanded to fit international contexts. The virtual community supports information upload and search. Semiautomatic tagging and intelligent multilingual semantic search using natural language are enabled using artificial intelligence techniques, namely Semantic Web, machine learning, and natural language processing.

Introduction

Human rights monitoring for people with disabilities is a pressing issue in Canada and France, as well as internationally. However, data and information about human rights policy and legal precedents, lived experiences, and media portrayals of people with disabilities are scarce. This limited data and information hinders the capacity for knowledge mobilization among researchers, nongovernmental organizations, and advocacy groups in their efforts to monitor the implementation and realization of human rights for people with disabilities. In addition, policy-based definitions of disability vary widely across local, national, and international jurisdictions, creating further challenges regarding the access of comparative data and information.

Internationally, a few sources provide access to some disability data (eg, the Disability Data Advocacy Working Group [1], Disability Data Portal [2], and Disability Rights Promotion International [DRPI] [3]). However, these databases have limited data while the need for such data is overwhelming. The incorporation of disability into human rights–focused legislations, constitutions, and international law (eg, the United Nations Convention on the Rights of Persons with Disabilities, or CRPD [4]) has highlighted an ongoing need for access to disability data for disability researchers, organizations of people with disabilities, and broader public policy and development sectors. The lack of availability of disability data has been identified as a major challenge for continuous disability rights monitoring (ie, collection, integration, analysis, interpretation, and mobilization of data and knowledge about the implementation and realization of human rights for people with disabilities) [5,6]. It is important to develop an inclusive approach to analytics and machine learning, one that can be scaled up to integrate disability data and enable the searching of such data to expose systemic discrimination and social exclusion, which increase vulnerability to inequitable social conditions [7].

Historically, there has been a lack of data on disability collected and shared on a global and national level [8]. The lack of aggregate and disaggregate disability data available is hampering researchers' ability to conduct critical health research [9]. It is also limiting our knowledge about the experiences of people with disabilities and the challenges and barriers they face [6]. Reed, Meeks, and Swenor [10] argue that "the disability data gap is more than just a surveillance oversight; social injustices exist that cannot be separated from this lack of information. The lack of data perpetuates the exclusion of disabled people from discussions of health equity and policies that are data driven."

Disability is a complex multidimensional experience that can be difficult to measure. Operational measures of disability vary according to the purpose and application of the data, the conception of disability, and the aspects of disability examined [10]. However, access to disability data is an issue of human rights. For example, Article 31 of the CRPD [11] notes the following obligation:

> *States Parties undertake to collect appropriate information, including statistical and research data, to enable them to formulate and implement policies to give effect to the present Convention… The information collected shall be disaggregated, as appropriate, and… States Parties shall assume responsibility for the dissemination of these statistics and ensure their accessibility to persons with disabilities and others.*

The disability Wikibase addresses this responsibility by making data accessible and available to a variety of stakeholders. Understanding, interpreting, and using disability data is essential to stakeholders with interests in disability issues, including advocacy groups, individuals, and organizations of people with disabilities, nongovernmental organizations, policy makers, and researchers [12]. Data contribute to decisions about poverty alleviation strategies, health and social service development, and humanitarian responses. Data collection is critical to the formulation of evidence-based policies and all aspects of the implementation of disability-inclusive policies and programs [13]. Reliable, accessible data are also crucial for monitoring progress and implementation of the CRPD.

Meanwhile, advancements are being made in the field of health informatics, which can assist in the organization and use of health data. Health informatics is an interdisciplinary field that uses computer science, decision-making, information processing, and management science to design and implement information systems in a user-centered manner to get, store, process, and provide information for users to assist them in achieving their objectives [14]. Some of the techniques used in health informatics are artificial intelligence (AI) techniques [15]. Machine learning techniques are trained algorithms that assist users in making sense of data to make organizational decisions [16] and health decisions [17,18]. Semantic Web techniques and natural language processing (NLP) provide computers with the ability to decipher the meaning of information and process phrases written or spoken in natural language [19]. The combination of these techniques can allow users to search for data using natural language sentences [20,21] in multiple languages, and they can also be used to find semantic similarities among disparate documents from different disciplines [22-26].

On the other hand, techniques to store, integrate, and search for data do exist; the publicly available Wikibase environment is an open-source collaborative tool for editing, integrating, and storing structured data. Wikidata is the most famous Wikibase example, and it targets public knowledge [27].

Our project aims to create a screen reader–accessible multilingual website on top of a Wikibase that structures and integrates data about people with disabilities and serves national and international disability advocacy communities. This pilot website will be endowed with a document upload function with an automated data-from-text extraction function for insertion into the Wikibase, and with an intelligent natural language search capability for querying the data.

Methods

Choice of Domain and Documents

To design and develop the underlying data model of the pilot platform, our domain of choice was the CRPD. The CRPD is an international human rights treaty intended to articulate and protect the rights and dignity of people with disabilities. The CRPD reflects a shift from viewing people with disabilities as objects of charity, medical treatment, and social protection toward viewing them as full and equal members of society with human rights. The CRPD contains 50 Articles covering a range of issues including health, accessibility, data collection, and monitoring mechanisms and is informed by the principles of dignity, autonomy, nondiscrimination, equality, and respect for difference. The CRPD has been signed by over 150 countries, including Canada and France.

This domain of choice allows for a large number of documents connected to the CRPD to be uploaded in the future, enabling the testing of the platform on a large scale. Our first set of documents to be shared included DRPI's reports of country- and region-specific rights monitoring based on the CRPD. DRPI is a collaborative project to establish a comprehensive, sustainable international system to monitor the human rights of people with disabilities in accordance with the CRPD. This project has been operating for over a decade and has collected data and information from North America, Europe, Latin America, Asia Pacific, Africa, the Middle East, and North Africa.

Platform Design

In early stages of the project, we collected the requirements of the platform and identified that the platform should act as a virtual community and incorporate the following characteristics:

1. Facilitate knowledge sharing by enabling the uploading of disability documents.
2. Provide validated information by allowing only authenticated external partners (eg, community organizations) to add documents and having a user approve them, acting as a gatekeeper.
3. Support disability rights advocacy by enabling dissemination of knowledge to the public by making the documents available to all potential users and using a natural language–enabled search engine.
4. Administer the platform to create users, continuously refine the machine learning models, provide reports, etc.

In summary, the aim of this project was to create a virtual knowledge network using the Semantic Web and other AI techniques. The virtual knowledge network includes 4 types of users:

1. Authenticated information producers that upload information of interest (eg, reports, blogs, articles) and assign tags (ie, keywords) to the new data.
2. Domain expert users that play the role of gatekeepers and validate the uploaded information (ie, documents and tags) by issuing an upload request to the platform administrator.
3. Platform administrators that manage the platform. Their tasks include approving/rejecting upload requests, continuously maintaining the platform, and finetuning the AI models.
4. Information consumers that search for information stored on the platform using natural language questions.

Platform Implementation

The virtual knowledge network front end includes a simple website connected to a search engine, while the back end uses the following:

1. A machine learning model to tag each paragraph of the uploaded documents with corresponding keywords that reflect the meanings embedded in the text.
2. A Wikibase structure that holds the disability rights knowledge base and the corresponding data (ie, tags, annotated paragraphs, links to the documents).
3. An intelligent search engine (QAnswer) that uses NLP to search the Wikibase for answers to users' questions.

To implement the different parts of the platform, the following major programming stacks were used:

1. React JS, a simple and flexible JavaScript-based library, was used to develop the front end of the platform [28].
2. MySQL was our choice for the database used to store temporary document information before its semantification and storage in the Wikibase [29].
3. Wikibase was used to store the disability rights knowledge base, which was represented in a knowledge graph and used the Resource Description Framework (RDF) format [30]. A knowledge graph is a network of entities, their semantic types and properties, and the relationships between the entities [31]. It is used to represent information and reasoning.
4. QAnswer, a platform that makes RDF data accessible via natural language, has been used to enable intelligent search using intelligent natural language questioning [32-34]. QAnswer is the first AI-driven platform to query RDF datastores in natural language. It uses semantic technologies (NLP, Semantic Web) as well as machine learning techniques [32].
5. Python was used to develop cross-platform and server-side applications [35].
6. FastText was used for NLP to create a machine learning model to tag paragraphs. FastText is often on par with deep learning classifiers in terms of accuracy, and is many orders of magnitude faster for training and evaluation [36,37].
7. PyMuPDF, a Python library, was used to access and process uploaded documents [38].
8. GitHub was used to maintain source code versioning [39].

Results

Disability Data Modeling

CRPD Ontology

Data modeling includes identifying the correct ontology for storing document data and finding appropriate relations between different data entities to map the data correctly within the Wikibase. Our team developed an ontology schema (Figure 1) for the CRPD that is flexible and can accommodate the CRPD Articles as well as incorporate jurisdiction-specific legislation so that users will be able to tag and search disability-related information in relation to internationally and locally recognized rights domains.

Figure 1. CRPD ontology. CRPD: Convention on the Rights of Persons with Disabilities; UN: United Nations.
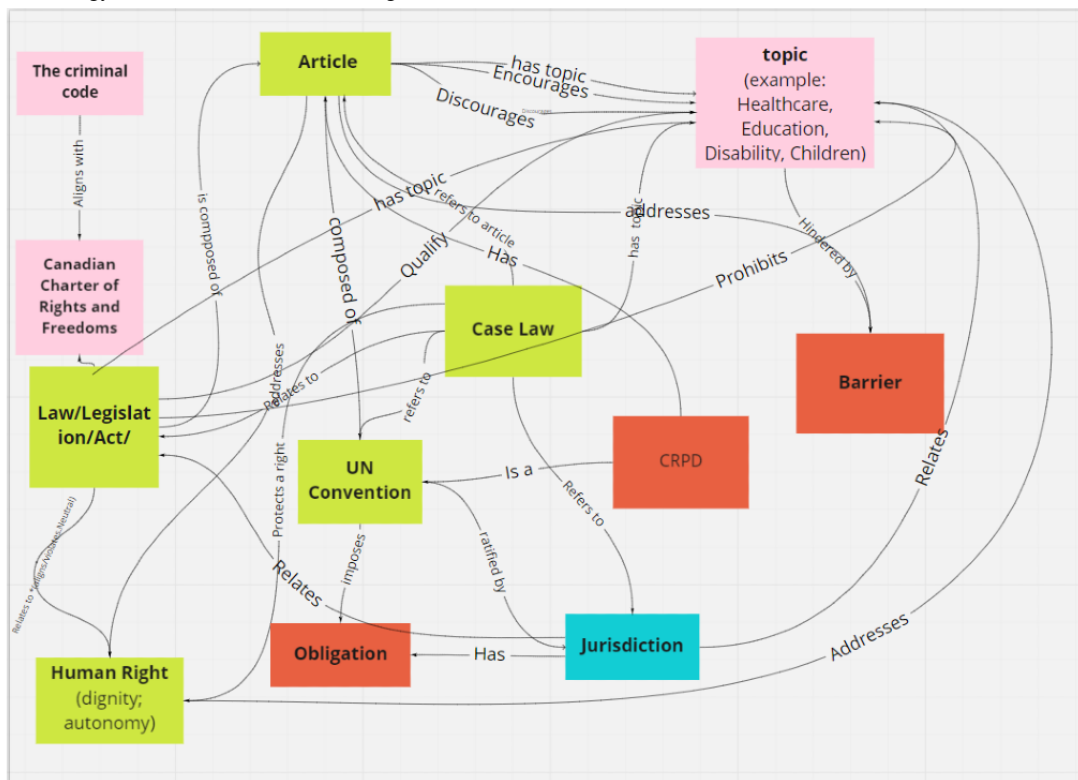
CRPD Knowledge Graph

Based on the ontology, domain experts developed structured data that describes the CRPD document. The structured data (Table 1) was used to create the CPRD knowledge graph.

The structured data was used to develop a detailed knowledge graph that describes the CRPD; for instance, Table 1 shows that Article 25 of the CRPD relates to the topic "Health." The knowledge graph (Figure 2) shows the same information as a graph. The knowledge graph is practically a set of triples (ie, subject-predicate-object) and was implemented in the disability Wikibase, which is a datastore that stores triples. Once incorporated in the disability Wikibase, the knowledge graph allowed semantification of the DRPI documents into triples (eg, document - has paragraph - the paragraph text; paragraph - has topic - discrimination). Semantification allows for a semantic search to be performed on the documents and their paragraphs.

Table 1. Structured data prepared by domain experts to develop the Convention on the Rights of Persons with Disabilities knowledge graph.

Article 25 (Health) topics	Has-topic	Text
Health	has-CRDP-Text	*Provide persons with disabilities with the same range, quality, and standard of free or affordable health care and programs as provided to other persons, including in the area of sexual and reproductive health and population-based public health programmes*
Discrimination	Mentioned-In-CRPD-Article	*Prohibit discrimination against persons with disabilities in the provision of health insurance, and life insurance where such insurance is permitted by national law, which shall be provided in a fair and reasonable manner*
Preventative measures	has-CRDP-Text	*Provide those health services needed by persons with disabilities specifically because of their disabilities, including early identification and intervention as appropriate, and services designed to minimize and prevent further disabilities, including among children and older persons*

Figure 2. A partial knowledge graph that depicts entities and relationships present in CRPD Article 25. CRPD: Convention on the Rights of Persons with Disabilities; DRPI: Disability Rights Promotion International.
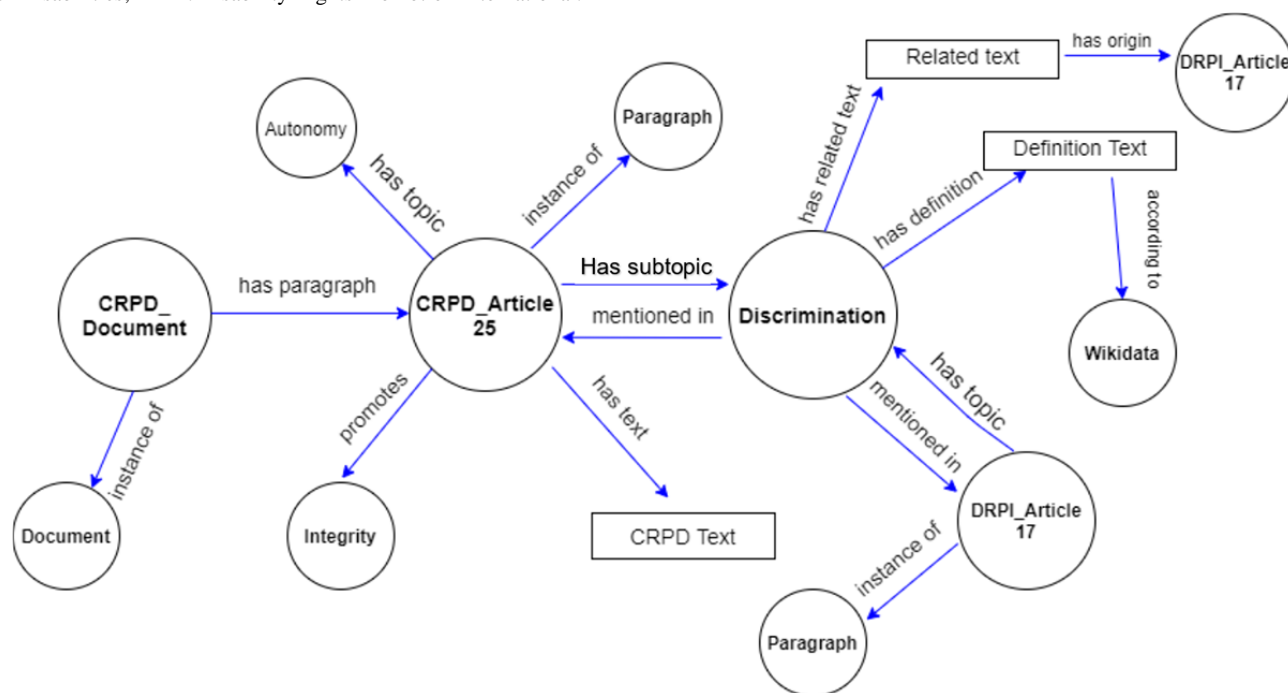

Figure 2 shows the relationships among the entities of articles 25 and 17 of the CRPD using a knowledge graph. Documents, topics, and sources are concepts connected by relations (eg, "has topic," "has paragraph," "according to," "mentioned in," "has text," "has definition," "has origin," "instance of"). For instance, paragraphs are connected to the document using the "has paragraph" relation since each document is formed of many paragraphs. In turn, each paragraph has one or more topics and subtopics, and so on.

Topic Modeling Using Latent Dirichlet Allocation

We used a latent Dirichlet allocation (LDA) [40] machine learning algorithm to model the topics in the available documents. To train the model, a set of 30 unstructured and unlabeled documents were provided in an unsupervised manner. To determine the number of topics to include in this unsupervised learning method, we conducted tests to select the best coherence measure score; this score is based on the degree of semantic similarity between high-scoring words in the topic.

The LDA result provided disability experts with an initial list of keywords that could be used to tag the documents' content. The disability experts added to the list of keywords and created a glossary of terms related to the disability field. The glossary terms were used as labels for tagging the paragraphs of a training set of the document corpus. Afterward, the labeled documents were used by the supervised machine learning algorithm proposed by the FastText library to design a text classification model that allows users to tag uploaded documents.

Initial Glossary Development and Enrichment

Disability domain experts developed a glossary of keywords that describe the main topics and subtopics embedded in the DRPI and CRPD documents. These keywords were then used to tag paragraphs of the uploaded documents with appropriate topics.

Following the development of the glossary, the Merriam-Webster Dictionary application programming interface (API) was used to get synonyms of the glossary terms. Thus, the glossary terms were enriched with synonym words.

Document Tagging

Overview

Once a user uploads a document, the platform splits the document into paragraphs and uses the glossary terms to tag them with appropriate topics. In the first phase, the extracted paragraphs and corresponding tags were verified by domain experts and the result was used as a training data set for a machine learning algorithm (Figure 3). The algorithm was used on the platform to automatically tag newly uploaded documents.

XSL•FO
RenderX

Figure 3. Document tagging flowchart. API: application programming interface.
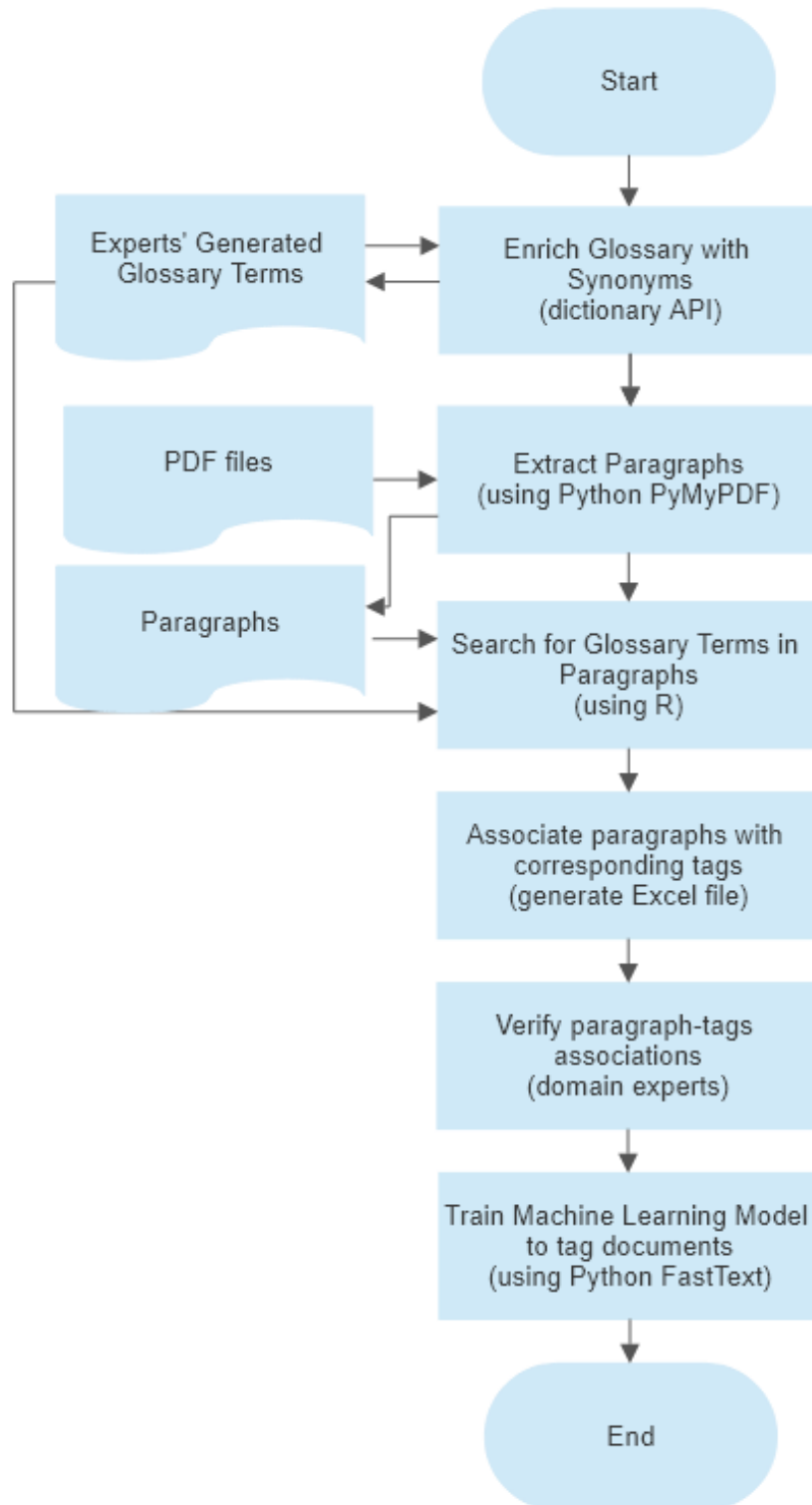
Paragraph Extraction and Tagging

Using the PyMuPDF library (a Python binding for MuPDF), the PDF files were parsed, and paragraphs were extracted. Afterward, R (version 4.0.5; R Foundation for Statistical Computing) was used to tag the paragraphs with appropriate glossary terms. The tags were reviewed by domain experts who added or deleted tags as needed. The result was a set of validated instances of paragraphs and their corresponding tags, which were used as the training data set.

Machine Learning Model Generation to Automate Tagging

The validated paragraph-tag instances were used as a training set for the FastText algorithm. FastText is a library for efficient learning of word representations and sentence classification. This library can be used to build a word representation model

with unsupervised training (using SkipGram or CBOW), or to design a text classification model with supervised learning. We used the latter approach and used the FastText library defaults for tagging the document paragraphs.

The FastText algorithm proposes tags from the glossary for the paragraphs and the result is displayed (Figure 4). Users can reject or add to the proposed tags, allowing for continuous enrichment of the glossary and training of the model, which make it more accurate. A request to upload the tagged paragraphs and other related information to the Wikibase is sent to the administrator.

Figure 4. Paragraphs and their tags are displayed to the user.
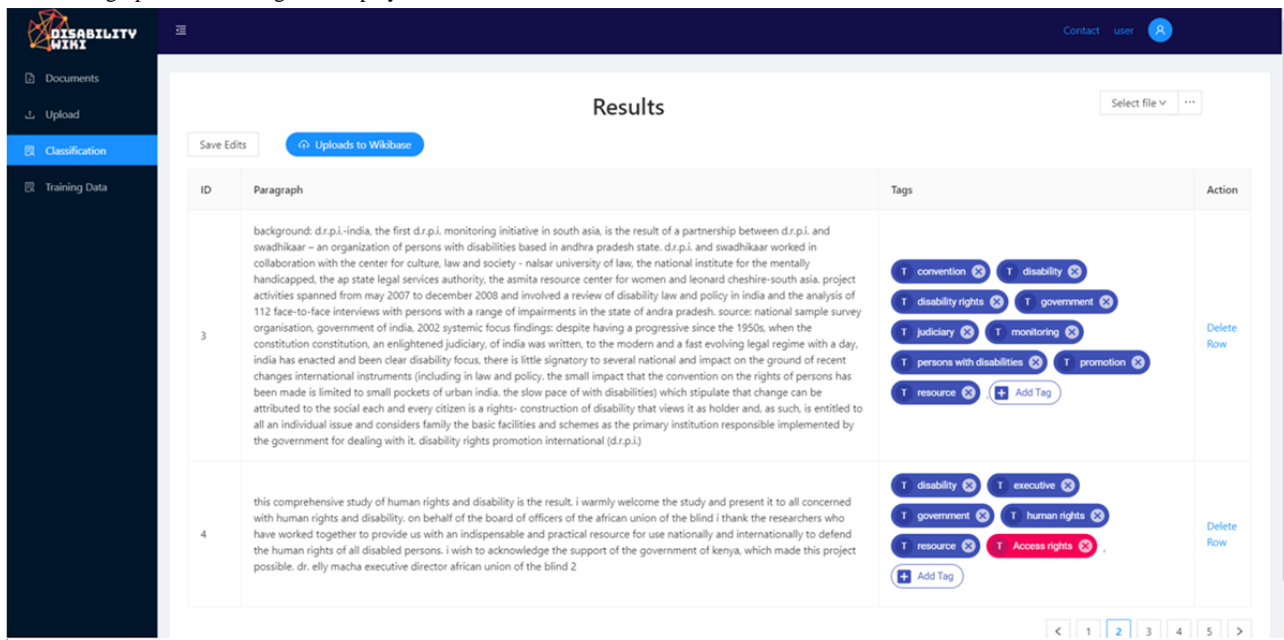
Document Upload to Wikidata

The administrator can review the upload requests and accept or reject them. Upon acceptance, the tagged paragraphs are semantified based on the ontology and uploaded to the disability Wikibase, where they are stored as triples that can be subject to semantic search (Figure 5).

Figure 5. Architecture diagram for the document upload process flow. ML: machine learning.
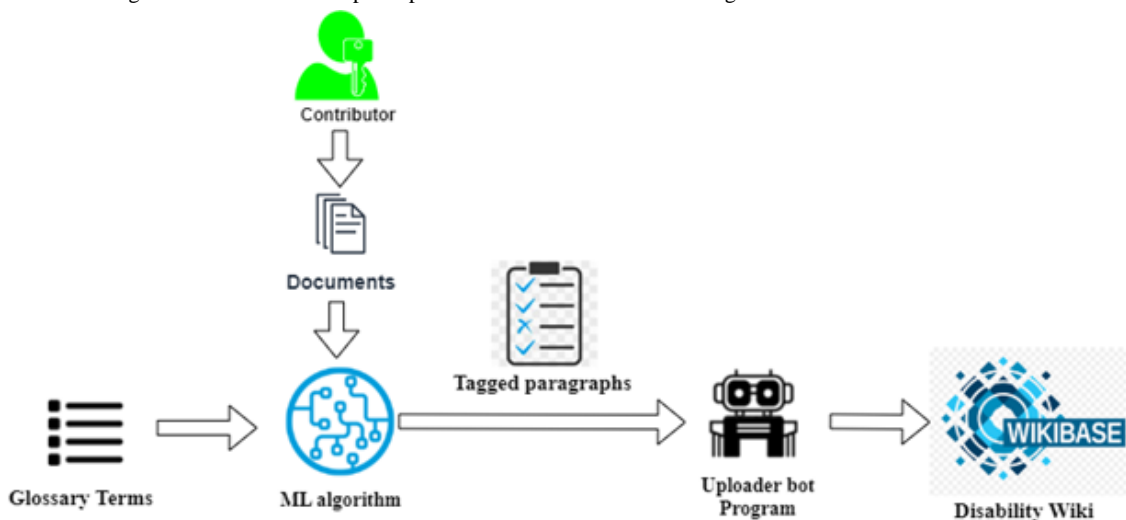

Natural Language Search Using QAnswer

The disability Wikibase is connected to the QAnswer search engine, which performs a natural language search (Figure 6). A simple search text field on the website is provided for users to enter their queries.

Although only authorized users can be producers and upload documents, any user can be a consumer and search for information. The disability Wikibase uses QAnswer API endpoints to send questions to, and retrieve answers from, QAnswer. We used the autocomplete feature in QAnswer to provide several precompleted questions while users are typing their questions. Thus, users can select from these autocompleted options instead of continuing to type their question.

Figure 6. Search function flow of the disability wiki website.
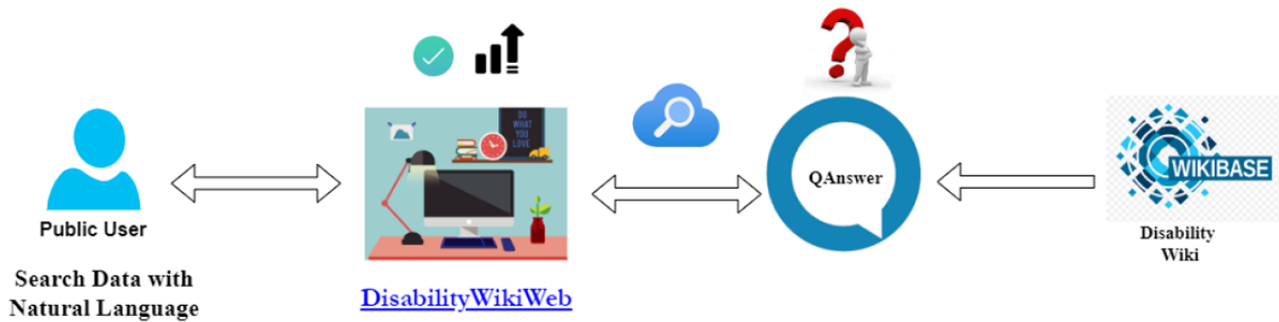
Discussion

Principal Results

There are currently no tools available to collect and share data related to people with disabilities. We have developed a platform that can fill this gap and respond to this need.

Using health informatics and AI techniques, we were able to develop a pilot virtual community that supports disability rights advocacy by facilitating the sharing and accessing of disability data. The virtual community accepts 4 types of users: information producers, information consumers, validators, and administrators. The virtual community enables the uploading of documents, semiautomatic tagging of document paragraphs with meaningful keywords, and validation of the process before the document and its associated tags are uploaded to the disability Wikibase. Once uploaded, public users can perform a semantic search using the QAnswer intelligent search engine. Currently, the platform allows information producers to upload documents and label their paragraphs with appropriate tags, and also allows information consumers to search for information using natural language. Hence, the platform answers a serious need in the disability community to store disability-related data in one database that includes an intelligent search function.

Virtual communities for health purposes (including disabilities) already exist; however, their purpose is to disseminate information only, with no means for members to contribute additional data and information. Wikipedia is wiki-based but it is not specialized and includes nonvalidated, general-purpose information, with no means to upload documents or perform a semantic search. This is the first time an AI approach was used to bolster disability advocacy. The proposed knowledge graph's content has been validated by disability domain experts, and it is also technically sound—Wikidata's best practices were followed. Future evolutions of the knowledge graph will have to maintain its accuracy, consistency, and conciseness, as defined by Zaveri et al [41].

This project advances our previous informatics work on a virtual knowledge network using documents from the DRPI [42]; in our previous work, we developed a virtual knowledge network to share documents [42], while on this pilot site we enabled semantic search and allowed accurate tagging of documents' content using machine learning approaches. The current platform simplifies document sharing and offers the public, disability organizations, researchers, and policy makers a tool to allow for more transparency in disability data.

Our proposed generic novel approach is to provide a platform for semantic free-text search on freely collected documents on a given topic. Our future research aims to answer the following question: what kind of information is searched for and by what types of users? To answer this question, we are planning to collect information regarding the users of the platform and the search terms they use; in addition, a thematic analysis will also be conducted to understand the users' needs.

Limitations and Future Perspectives

This work is still in its pilot stage; there is a limited number of documents available for machine learning training, while accuracy measurement and evaluation of the machine learning algorithms are underway. However, the more users upload documents and validate the automatic tagging, the more accurate the automatic tagging becomes due to continuous retraining.

There are other features and project phases planned for the future. Virtual communities have been used in many domains, including health care [43], to address chronic disease management, education [44], and mental health. Our current platform is expanding the use of virtual communities for social purposes in the domain of disability rights [45]. The next phases of the project will involve an evaluation of the adoption of this virtual community using evaluation frameworks [46], as well the use of background analytics to analyze usage patterns and enhance the search capacity. Moreover, users' experience will be formally evaluated to enhance acceptance and adherence. Universal accessibility allows equitable access to information; accessibility testing is underway, and improvements are being implemented. Additionally, while QAnswer can answer questions formulated in English and French, the current platform is written in English; translating the website into French and uploading documents written in French are both on our project agenda.

NLP has been used for many health care applications for information extraction, sentiment analysis, search extraction, and information synthesis [47], using the state-of-the-art NLP model Bidirectional Encoder Representations from Transformers (BERT) [48]. Our next steps include tagging optimization using newer approaches to NLP such as BERT. This platform may encourage the creation of a domain-specific DisabilityBERT, similar to the BioBERT adapted for biomedical text and SciBERT adapted for science text [49].

In our future work, we will continue to elucidate biases within the disability Wikibase search. Although data analytics is

important for collecting evidence and making decisions [50,51], neither machine learning nor data are value-free [52]. Deciding which data items will be collected is a process that embeds a worldview (Weltanschauung). For example, is the information that is collected and searched about "abnormality" [53] or "participation restriction" [54]? In addition, the production of information is done for a certain purpose and hence is not value-free; for example, were certain data collected to understand the efficiency of people with disabilities in doing a task, or to determine if people with disabilities face discrimination in schools, universities, or the workplace? In this way, our virtual community is informed and designed in accordance with the human rights principles of the CRPD. Finally, the computer algorithms are value-loaded; indeed, predictive machine learning algorithms "learn" (ie, create models) from past data to predict events in the future based on current data. However, past data embeds values and biases from the past; for example, in 2018, Amazon's AI-based recruiting tool was found to be sexist as it consistently suggested that male applicants be recruited [55], as it learned from resumes previously submitted to the company, which is in the male-dominated tech industry. Our choice to use the CRPD to build our ontology represents a first step in creating a search algorithm for disability data to be used by and for disability advocates and communities.

The current project requires many improvements and feature enhancements. First, the document data is semantified using a few relations, such as "has paragraph," "has topic," and "mentioned in." Future work will include other relations, such as "comply with," "supports," and "has legislation" and update the machine learning classifier to extract such semantic information from the uploaded document. Uploaded documents could be linked to their corresponding country as some documents are country- and region-specific. The current version of the application contains such properties in the interface; however, it is not embedded in the Wikibase data model. In the next phase, it is important to semantically link documents with countries and regions by importing countries and regions from Wikidata to the disability Wikibase.

Finally, the web application should be updated to include message broker architecture that allows server-side load balancing to handle large volumes of search requests and optimize the server's performance.

Conclusions

The platform ontology was based on the CRPD to create a platform for equity-focused disability data that can be used by and for disability advocates and communities; the data includes advocacy reports and disability policy and legislation from specific jurisdictions or international contexts. The virtual community accepts 4 types of users: information producers, information consumers, validators, and administrators. Semiautomatic tagging and intelligent semantic search using natural language are enabled using AI techniques. The search engine supports bilingual (English and French) search and further enhancements are planned.

Multidisciplinary research in the domain of disabilities is challenging. An iterative process completed in a collaborative atmosphere allows team members to elucidate ambiguities and helps the team face challenges and provide solutions.

Authors' Contributions

CEM is the primary investigator who received the funds for this project. CEM and RG co-designed the study. CEM, RG, PM, and FM contributed equally to this phase of the project. AC and BK worked on the model triples and tested the tagging process; DD and RT developed the platform and machine learning algorithms. All approved the final version and agreed to be accountable for all aspects of the submitted paper. It is the understanding of the university and researchers that the Project Intellectual Property belongs to CEM, RG, PM, and FM.

Conflicts of Interest

PM is co-owner of The QA Company, which develops the QAnswer technology. All other authors declare no conflicts of interest.

References

1. Disability Data Advocacy Working Group. International Disability Alliance. 2019. URL: http://www.internationaldisabilityalliance.org/content/disability-data-advocacy-working-group [accessed 2021-07-15]
2. Disability Data Portal. Leonard Cheshire. 2020. URL: https://www.disabilitydataportal.com/ [accessed 2021-07-15]
3. Rights Library - DRPI TRAINING. Disability Rights Promotion International (DRPI). 2016 Jul 06. URL: https://drpitraining.research.yorku.ca/rights-library/ [accessed 2021-07-15]
4. UN Convention on the Rights of Persons with Disabilities. New York: Department of Economic and Social Affairs Disability, United Nations; 2006 Dec 13. URL: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html [accessed 2021-07-15]
5. Loeb M. Disability statistics: an integral but missing (and misunderstood) component of development work. Nord J Hum Rights 2013;31(3):306-324 [FREE Full text] [Medline: 26925181]

XSL·FO
RenderX

6. Abualghaib O, Groce N, Simeu N, Carew MT, Mont D. Making Visible the Invisible: Why Disability-Disaggregated Data is Vital to "Leave No-One Behind". Sustainability 2019 May 31;11(11):3091. [doi: 10.3390/su11113091]

7. Rioux M, Parekh G. Disability, Rights Monitoring, and Social Change. Toronto: Canadian Scholars' Press; 2015.

8. World Report on Disability. Geneva, Switzerland: World Health Organization; 2011. URL: https://www.who.int/disabilities/world_report/2011/report.pdf [accessed 2021-10-28]

9. Casebolt MT. Availability and quality of global disability data: A commentary on the Demographic and Health Surveys. Disabil Health J 2021 Jan;14(1):100972. [doi: 10.1016/j.dhjo.2020.100972] [Medline: 32792331]

10. Reed NS, Meeks LM, Swenor BK. Disability and COVID-19: who counts depends on who is counted. The Lancet Public Health 2020 Aug;5(8):e423. [doi: 10.1016/s2468-2667(20)30161-4]

11. UN General Assembly. Convention on the rights of persons with disabilities: Article 31. New York, NY: The United Nations; 2007. URL: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-31-statistics-and-data-collection.html [accessed 2021-07-15]

12. Resources for Data Users. The Washington Group. 2021. URL: https://www.washingtongroup-disability.com/resources/resources-for-data-users/ [accessed 2021-07-15]

13. Strengthening the collection of data on disability. World Health Organization. 2021. URL: https://www.who.int/activities/collection-of-data-on-disability [accessed 2021-10-13]

14. El Morr C. Introduction to Health Informatics: A Canadian Perspective. Toronto, Canada: Canadian Scholars' Press; 2018.

15. El Morr C, Ali-Hassan H. Analytics in Healthcare: A Practical Introduction. Cham, Switzerland: Springer International Publishing; 2019:105.

16. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLoS One 2019;14(2):e0212356 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]

17. Lake IR, Colón-González FJ, Barker GC, Morbey RA, Smith GE, Elliot AJ. Machine learning to refine decision making within a syndromic surveillance service. BMC Public Health 2019 May 14;19(1):559 [FREE Full text] [doi: 10.1186/s12889-019-6916-9] [Medline: 31088446]

18. Sahoo A, Pradhan C, Das H. Performance Evaluation of Different Machine Learning Methods and Deep-Learning Based Convolutional Neural Network for Health Decision Making. In: Rout M, Rout J, Das H, editors. Nature Inspired Computing for Data Science. Cham, Switzerland: Springer International Publishing; 2020:201-212.

19. Chowdhary K. Natural Language Processing. In: Fundamentals of Artificial Intelligence. New Delhi, India: Springer India; 2020:603-649.

20. Edgcomb JB, Zima B. Machine Learning, Natural Language Processing, and the Electronic Health Record: Innovations in Mental Health Services Research. Psychiatr Serv 2019 Apr 01;70(4):346-349. [doi: 10.1176/appi.ps.201800401] [Medline: 30784377]

21. Velupillai S, Epstein S, Bittar A, Stephenson T, Dutta R, Downs J. Identifying Suicidal Adolescents from Mental Health Records Using Natural Language Processing. Stud Health Technol Inform 2019 Aug 21;264:413-417. [doi: 10.3233/SHTI190254] [Medline: 31437956]

22. Al-Natsheh HT, Martinet L, Muhlenbach F, Rico F, Zighed DA. Semantic Search-by-Examples for Scientific Topic Corpus Expansion in Digital Libraries. : IEEE; 2017 Presented at: IEEE International Conference on Data Mining Workshops (ICDMW); November 2017; New Orleans, LA, USA p. 747-756. [doi: 10.1109/icdmw.2017.103]

23. Al-Natsheh H, Martinet L, Muhlenbach F, Zighed D. UdL at SemEval-2017 Task 1: Semantic Textual Similarity Estimation of English Sentence Pairs Using Regression Model over Pairwise Features. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).: The Association for Computational Linguistics; 2017 Presented at: The 11th International Workshop on Semantic Evaluation (SemEval-2017); August 3-4, 2017; Vancouver, Canada p. 115-119. [doi: 10.18653/v1/s17-2013]

24. Bast H, Buchhold B, Haussmann E. Semantic Search on Text and Knowledge Bases. FNT in Information Retrieval 2016;10(1):119-271. [doi: 10.1561/1500000032]

25. Gromann D. Neural language models for the multilingual, transcultural, and multimodal Semantic Web. SW 2020 Jan 31;11(1):29-39. [doi: 10.3233/sw-190373]

26. Al-Natsheh HT. Text Mining Approaches for Semantic Similarity Exploration and Metadata Enrichment of Scientific Digital Libraries. Lyon, France: L'Université de Lyon; 2019.

27. Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase. Commun ACM 2014 Sep 23;57(10):78-85. [doi: 10.1145/2629489]

28. React. React: A JavaScript library for building user interfaces. 2021. URL: https://reactjs.org/docs/getting-started.html [accessed 2021-07-15]

29. MySQL Workbench: Enhanced Data Integration.: Oracle; 2021. URL: https://www.mysql.com/products/workbench/ [accessed 2021-07-15]

30. MediaWiki.: Wikipedia Inc; 2020 May 26. URL: https://www.mediawiki.org/wiki/MediaWiki [accessed 2021-10-27]

31. Ehrlinger L, Wöß W. Towards a Definition of Knowledge Graphs. : CEUR-WS.org; 2016 Presented at: The 12th International Conference on Semantic Systems - SEMANTICS 2016; September 12-15; Leipzig, Germany. [doi: 10.1007/978-1-349-19066-9_2]

XSL·FO

RenderX

32.   Accessing your Knowledge via Natural Language. QAnswer. 2021. URL: https://www.qanswer.eu/ [accessed 2021-07-15]

33.   Diefenbach D, Giménez-García J, Both A, Singh K, Maretditors. QAnswer KG: Designing a Portable Question Answering System over RDF Data. In: The Semantic Web. Cham, Switzerland: Springer International Publishing; 2020 Presented at: The Semantic Web, The 17th International Conference, ESWC 2020; May 31–June 4, 2020; Heraklion, Crete, Greece. [doi: 10.1007/978-3-030-49461-2_25]

34.   Diefenbach D, Migliatti P, Qawasmeh O, Lully V, Singh K, Maret P. QAnswer: A Question Answering prototype bridging the gap between a considerable part of the LOD cloud and end-users. : Association for Computing Machinery; 2019 Presented at: The World Wide Web Conference; 13-17 May 2019; San Francisco, CA, USA. [doi: 10.1145/3308558.3314124]

35.   Python. 2021. URL: https://www.python.org/ [accessed 2021-07-15]

36.   Alshdaifat E. Hierarchical Ensemble Classification: Towards the Classification of Data Collections that Feature Large Numbers of Class Labels. Liverpool, UK: The University of Liverpool; 2015.

37.   Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification. : Association for Computational Linguistics; 2017 Presented at: The 15th Conference of the European Chapter of the Association for Computational Linguistics; April 3-7, 2017; Valencia, Spain. [doi: 10.18653/v1/e17-2068]

38.   McKie JX. PyMuPDF (Version 1.18.19). 2021. URL: https://github.com/pymupdf/PyMuPDF [accessed 2021-10-15]

39.   GitHub. 2021. URL: https://github.com/ [accessed 2021-07-15]

40.   Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.

41.   Zaveri A, Rula A, Maurino A, Pietrobon R, Lehmann J, Auer S. Quality assessment for Linked Data: A Survey. SW 2015 Mar 17;7(1):63-93. [doi: 10.3233/sw-150175]

42.   El Morr C, Maret P, Dinca-Panaitescu M, Rioux M, Subercaze J. Virtual community life cycle: A model to develop systems with fluid requirements. In: Virtual Community Building and the Information Society: Current and Future Directions. Hershey, PA, USA: IGI Global; 2012:1-14.

43.   El Morr C, Maule C, Ashfaq I, Ritvo P, Ahmad F. A Student-Centered Mental Health Virtual Community Needs and Features: A Focus Group Study. Stud Health Technol Inform 2017;234:104-108. [Medline: 28186024]

44.   El Morr C, AlHamzah M, Ng P, Purewal A, Al-Omran M. Knowledge of peripheral arterial disease: Results of an intervention to measure and improve PAD knowledge in Toronto. Vascular 2017 Oct 25;25(5):479-487. [doi: 10.1177/1708538116689355] [Medline: 28121280]

45.   Gorman R, Maret P, Creighton A, Kundi B, Muhlenbach F, Buettgen A, et al. The Potential of an Artificial Intelligence for Disability Advocacy: The WikiDisability Project. Stud Health Technol Inform 2021 May 27;281:1025-1026. [doi: 10.3233/SHTI210338] [Medline: 34042833]

46.   El Morr C, Eftychiou L. Evaluation Frameworks for Health Virtual Communities. In: The Digitization of Healthcare. London, UK: Palgrave Macmillan; 2017:99-118.

47.   Elbattah M, Arnaud E, Gignon M, Dequen G. The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications. 2021 Presented at: The 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC); February 11-13, 2021; Vienna, Austria. [doi: 10.5220/0010414508250832]

48.   Devlin J, Chang M, Lee K, Toutanova K. Pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT); June 3-5, 2019; Minneapolis, MN, USA.

49.   Beltagy I, Cohan A, Lo K. Pretrained contextualized embeddings for scientific text. Stroudsburg, PA, USA: The Association for Computational Linguistics; 2019 Presented at: The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November 3-7, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-1371]

50.   Laudon K, Laudon J. Management Information Systems: Managing the Digital Firm. New York, NY: Pearson; 2017.

51.   Pearlson K, Saunders C, Galletta D. Managing and Using Information Systems A Strategic Approach. Hoboken, NJ, USA: John Wiley & Sons; 2016.

52.   Kelleher J, Namee B, D'Arcy A. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, Massachusetts, USA: MIT Press; 2015.

53.   The International Classification of Impairments, Disabilities, and Handicaps. Geneva, Switzerland: World Health Organization; 1976. URL: https://apps.who.int/iris/handle/10665/41003 [accessed 2021-10-28]

54.   Health topics - Disabilities. Geneva, Switzerland: World Health Organization URL: https://www.who.int/topics/disabilities/en/ [accessed 2021-07-15]

55.   Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. 2018. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [accessed 2020-01-16]

Abbreviations

AI: artificial intelligence
API: application programming interface

XSL•FO

RenderX

BERT: Bidirectional Encoder Representations from Transformers
CRPD: Convention on the Rights of Persons with Disabilities
DRPI: Disability Rights Promotion International
LDA: latent Dirichlet allocation
NLP: natural language processing
RDF: Resource Description Framework

XSL•FO
RenderX