Keeping Researchers Updated by Automatically Enriching an Ontology in the Medical Field

Emiliano Tramontana
Dipartimento di Matematica e Informatica, University of Catania, Italy

Gabriella Verga
Dipartimento di Matematica e Informatica, University of Catania, Italy

ABSTRACT

Ontologies provide a common and standard dictionary of terms in some domain for researchers to easily exchange data. Forming an ontology requires several years of work performed by human experts, and an ontology for a given domain is thought to be stable for many years. Nevertheless, as scientific articles are continuously published to gather knowledge on recent findings, existing ontologies risk becoming stale, or require further human effort. Moreover, searching new articles without referring to an ontology can be very time consuming and confusing, especially for novice researchers. We propose an approach for automatically relating newly available published articles to existing ontologies. By automatically selecting relevant scientific articles and making them appear besides other data in an ontology, we aim at supporting experienced and novice researchers. Therefore, as knowledge grows and articles are available, the ontology used by researchers will also be automatically connected, allowing them to readily discover new findings. To validate the effectiveness of the proposed approach, we have enriched OBIB, an ontology for biobanking, with a selection of articles extracted from PubMed. The approach is general enough and can be applied to other ontologies or publishers.

CCS CONCEPTS

• **Information systems**; • **Ontologies**;

KEYWORDS

Ontologies, e-learning, Knowledge management, PubMed, Ontology enrichment

ACM Reference Format:
Emiliano Tramontana and Gabriella Verga. 2022. Keeping Researchers Updated by Automatically Enriching an Ontology in the Medical Field. In *The 10th International Conference on Computer and Communications Management (ICCCM 2022), July 29–31, 2022, Okayama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3556223.3556262

1 INTRODUCTION

The internet and the World Wide Web have become predominant means for everyone to gain new knowledge quickly and easily in their research work or while studying. People have instant access to knowledge and information at the click of a mouse or a finger. Resorting to updated data online is made easy as most devices are always connected to the internet [1]. The global connection favors users in being always updated and for learning, giving a strong expansion to e-learning systems [2, 3]. E-learning refers to *online learning* and it entails having students to access materials and content from anywhere and at any time, proving useful in education and businesses [4]. Due to the spread of coronavirus disease, most educational institutions have adapted to remote learning, and, in the upcoming years, they will continue to adapt to e-learning. During the lockdown, e-learning proved to be a great support in education, and the students' interest in remote learning has grown [5].

As this studying approach is more widespread, students tend to carry out research by themselves, hence access several sources on the internet, however they need a little bit of backing. Indeed, the internet presents a lot of materials that without careful selection can easily confuse students [6]. For this reason, it is important to offer a clean guide that can facilitate learning. In the field of e-learning, an ontology can support teachers and students, helping them in the management of course material, e.g., allowing a more detailed organization and adaptation of the students' learning path [6, 7]. Ontologies represent a common and shareable view of a domain and specify the domain using concepts, attributes, and relationships. An *ontology* defines relevant concepts, *epistemology* deals with the theory of knowledge, and *methodology* is the work of the researcher when gathering knowledge.

This article proposes an innovative approach to e-learning, taking advantage of a new reinforced ontology that combines a known ontology with articles found in PubMed, the most popular search tool dedicated to biomedical and life sciences literature [8]. Our approach aims at providing a search tool that gives the benefits of an ontology, which guides users in discovering a domain, and the advantages of updated findings, in the form of articles. In addition, this approach keeps the ontology up to date, as while new articles are published, the terms of the ontology will be connected to pertinent ones. Hence, new findings will be related to the previously organized knowledge. The proposed search engine determines whether a searched keyword is pertinent with the ontology, then automatically extracts from PubMed a set of related articles, finally selects some of them to enrich the ontology. Therefore, overtime, as PubMed includes new articles, new relations are inserted into the ontology, providing more results to users.

The organization of the article is as follows. Section 2 describes concepts and gives the definitions for ontologies, e-learning, and PubMed. Section 3 reports the comparison with related works. Section 4 discusses the proposed system enriching and ontology and

ICCCM 2022, July 29–31, 2022, Okayama, Japan

ACM ISBN 978-1-4503-9634-9/22/07…$15.00
https://doi.org/10.1145/3556223.3556262

selecting the most suitable articles. Section 5 shows practical examples of using the approach for some searches. Finally, Section 6 reports our conclusions.

2 BACKGROUND

2.1 E-Learning

The term e-learning refers to education provided on a digital device that is intended to support learning [9]. It stands for the use of the internet and web-related technology in the context of education. A current definition of e-learning by the American Society for Training and Development (ASTD) states: *"e-learning is the use of electronic technologies to provide information and facilitate the development of skills and knowledge"* [10].

The growth in the adoption of e-learning is driven by the multiple advantages it offers to students, such as flexibility, internet accessibility, and cost-effectiveness [11]. According to [12], offering information to students at any time helps them assimilate better than traditional education.

2.2 Ontology

Ontology, a fundamental branch of philosophy, aims at providing a complete, definitive, and exhaustive classification of entities in all spheres of being, including properties and relationships [13]. An ontological model consists of *classes*, *attributes*, *individuals*, and *relations* between them. Each class defines a category for a set of object instances (or concepts or types of things) by describing their state and behavior; attributes are *properties*, characteristics, or parameters that objects can have and share; individuals represent domain objects (instances of classes); relations define the relationships between objects [14]. One representation of knowledge is via an RDF model[1]. An RDF model can be seen as an oriented graph whose nodes are resources, or primitive types, and whose edges represent properties.

Due to their power in semantic heterogeneity, ontologies have been extensively studied, especially in the medical and biomedical fields, and they allow a better interpretation of the data by offering a smoother and more efficient system [15–19]. In the field of biobanking, a reference ontology is the Ontology for Biobanking (OBIB[2]) [20].

2.3 Ontology for Biobanking

A biobank is a non-profit service unit, aimed at collecting, processing, storing, and distributing human biological samples and related data, for research and diagnosis. Biobanks are facilities that store specimens, such as bodily fluids and tissues, typically along with specimen annotation and clinical data. The Ontology for Biobanking (OBIB) is an ontology built for the annotation and modeling of biobank repository and biobanking administration [20]. OBIB is based on a subset of the Ontology for Biomedical Investigation (OBI), has the Basic Formal Ontology (BFO) as its upper ontology, and is developed following OBO Foundry principles[3]. OBIB resulted from the merging of two existing biobank-related ontologies, OMIABIS and biobank ontology. Thanks to the fact that OBIB derives from the merging of two ontologies, its coverage is wide. It covers information on samples and their processes, such as e.g., sample collection, handling, or storage; in addition, it contains organizational information such as the necessary maintenance, e.g., information on sample freezing. Table 1 shows relevant details of the latest OBIB version, i.e. November 19, 2021.

Table 2 shows some of the classes in OBIB. Each class contains data, as attributes, deriving from the represented concept. Among the main attributes are: *label*, identifying the class; *URI*, as Uniform Resource Identifier, which is a unique identifier for the class, and *Definition*, a brief description of the concept (or class itself).

2.4 PubMed

PubMed[4] is a free search engine of biomedical scientific literature, which has been operating from 1949 to the present. It was developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM), one of the National Institutes of Health (NIH). PubMed is the bibliographic component of the NCBI's Entrez retrieval system [21]. The main purpose of PubMed search engine is to provide access to biomedical journals, allowing the consultation of abstract and if available the full text. PubMed has a wide selection of scientific articles and consulting a relevant article can be difficult and time-consuming. Some effort has been carried out for improving the search mechanisms, hence the provided results [22], however, the abundance of articles makes it difficult to find relevant ones.

3 RELATED WORK

Ontologies have been widely studied to improve e-learning techniques, and offer help to teachers and students [23–26]. Given that one of the difficulties for students lies in choosing teaching materials, thanks to an ontology, an inexperienced person can fully understand a concept, allowing herself to be led by the knowledge acquired by experts.

The use of ontologies together with e-learning leads to multiple advantages, e.g., having material that is always updated and relevant. Greater benefits derive from the integration of ontologies and external sources [27–30]: corpus enrichment with external resources increases the accuracy of labels and the relation labeling, and it makes the data more flexible and manageable. Therefore, several approaches have addressed the problem of enriching ontological terms, many using unstructured text and others focusing on identifying new relationships of terms by using verbs from sentences [31–33].

Unlike previous approaches, this work offers a guide and an extension of the domain of knowledge, well-known and modelled by the ontology, with scientific articles, i.e., each ontology class (or term) will be used to refine the search and then the class will be given a relation with relevant scientific articles selected from PubMed, thus offering a broader and more easily searchable knowledge of the domain. This way of updating the ontology, however without changing its fundamental concepts and relations, is, to the best of our knowledge, a novel approach.

By using the ontology, we aim to guide the search performed on PubMed and then results are filtered. Several approaches have

[1]https://www.w3.org/TR/rdf11-concepts/
[2]https://bioportal.bioontology.org/ontologies/OBIB
[3]https://bioportal.bioontology.org/ontologies/OBIB
[4]https://pubmed.ncbi.nlm.nih.gov

Table 1: Relevant details for OBIB: number of named (non-anonymous) classes; number of individuals; number of properties or slots; maximum depth of the hierarchical tree; maximum number of children; average number of children at a level in the tree.

	Occurrences
Classes	1,854
Individuals	225
Properties	95
Maximum depth of the tree	15
Maximum number of children	129
Average number of children at a level of the tree	15

Table 2: A small selection of classes found in OBIB. Each class is identified by means of a URI (Uniform Resource Identifier) and includes a definition that explains the represented concept. Some classes in OBIB have been reused from other ontologies, hence for such classes the URI prefix indicates the origin ontology, firstly defining the class.

Label	URI	Description
blood specimen	OBI_0000655	A material entity derived from a portion of blood collected from an organism.
blood vessel	UBERON_0001981	A vessel through which blood circulates in the body.
brain	UBERON_0000955	The brain is the center of the nervous system in all vertebrate, and most invertebrate, animals. Some primitive animals such as jellyfish and starfish have a decentralized nervous system without a brain, while sponges lack any nervous system at all. In vertebrates, the brain is located in the head, protected by the skull and close to the primary sensory apparatus of vision, hearing, balance, taste, and smell.
brain specimen	OBI_0002516	A specimen that is derived from brain.
brain cerebellum specimen	OBI_0002538	A specimen that is derived from cerebellum.
specimen	OBI_0100051	A material entity that has the specimen role.

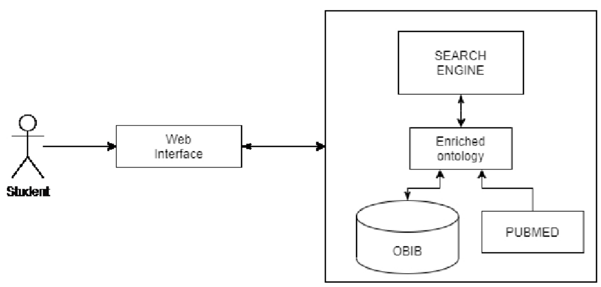

Figure 1: The proposed system having a front-end and a server-side consisting of components enriching the ontology

studied the efficacy and potential weaknesses of the search performed using PubMed [22, 34]. In our study, the articles found from PubMed searches are carefully selected to bypass weaknesses. For this we rank results according to the *Term Frequency-Inverse Document Frequency* (TF-IDF) score [35]. In the literature, this approach has been used to identify the most pertinent papers [36, 37]. Unlike the other approaches, however, we do not analyze the entire article, as we only work on specific sections considered the most relevant. In this way we offer a leaner and more focused approach.

4 PROPOSED SEARCH AND ENRICHMENT APPROACH

To offer an advanced and intelligent search engine, we propose to use both the data found within an ontology (organizing concepts in a domain) and the results coming from PubMed, hence giving life to a richer and strengthened ontology. Figure 1 shows the overall proposed system consisting of a front-end by means of which a user can search terms, and a back-end merging results of searches performed on both the ontology and PubMed. The main steps of the proposed approach are: (i) extraction of terms from OBIB ontology (the original version or an enriched version); (ii) selection of the most relevant articles from PubMed; (iii) enrichment of OBIB ontology by the addition of a new property.

Starting from the keyword inserted by a student performing a search, we firstly look up the keyword in OBIB, by comparing it with labels, descriptions, and synonyms. If there is a label (or a description, or a synonym) containing the keyword, then the set of words in such a label are used to perform a query in PubMed. The rationale here is that the ontology is used to suggest a *specialization* of the user search, i.e., a selection of a few among several meanings of the inserted keyword. Searching the set of words from the label in PubMed will usually provide many results in the form of a list of scientific articles, comprising titles, keywords, and abstracts. Such articles will be ranked and only a selection of them will be proposed to the user (see Subsection 4.2).

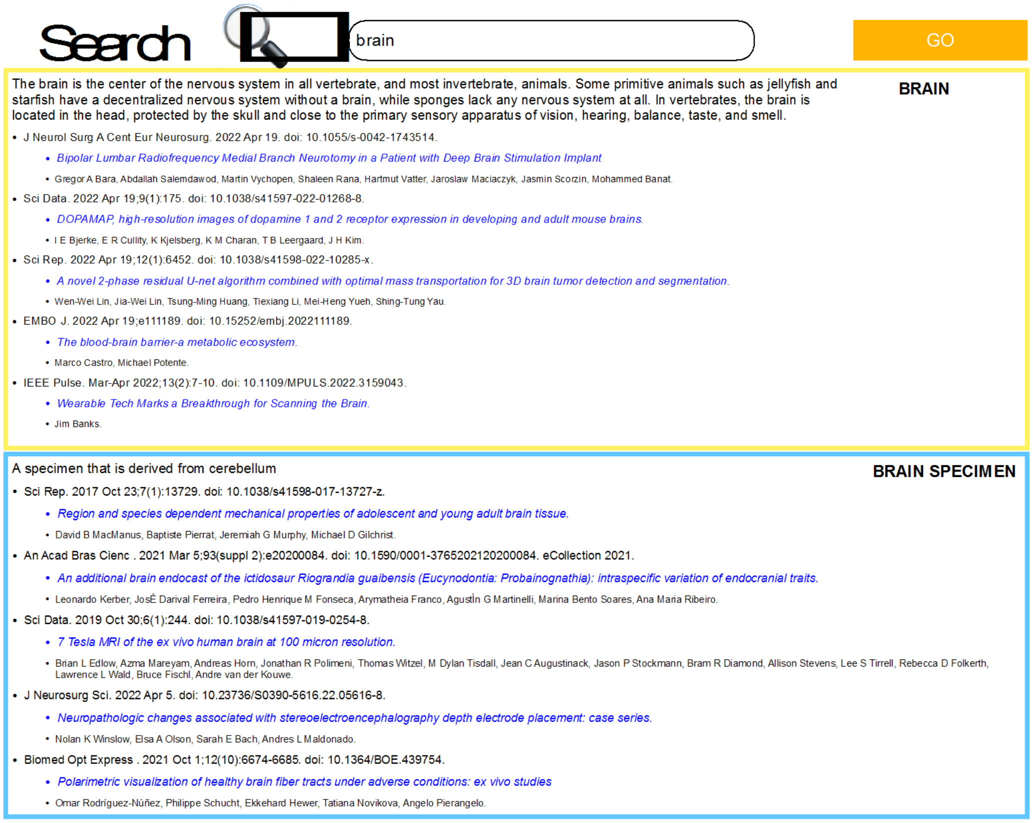

Figure 2: The user interface showing a portion of the results obtained when searching for the word *brain*.

Results will be given according to the following order. Firstly, the classes having a label that contains the searched keyword; secondly, the classes having a synonym containing the searched keyword, thirdly the classes having a description containing the searched keyword. Each part of the results will also have a list of articles that have been found in PubMed. Figure 2 gives an example of the results obtained when searching keyword *brain*. Each box is a match with an ontology class having a label that contains such a keyword.

The software system developed according to the above requirements makes use of Apache Jena framework [38] and two Python libraries. Apache Jena allows both reading terms within the ontology and creating the reinforced ontology (see subsection 4.1). Python libraries make it possible to connect to PubMed and have as a result the list of relevant articles, as well as select the most relevant results (see subsection 4.2).

4.1 Extracting terms from OBIB

Apache Jena is a Java framework for developing semantic web-oriented applications. It assists in creating and modifying models including ontologies. An ontology can be queried by SPARQL queries. Therefore, we implemented a query to extract all the class labels, as follows. Then, the query was executed using Apache Jena API *QueryExecutionFactory.create(query, model)*.

SELECT ?class **WHERE** { ?class rdf: type rdfs: Class. }

After results have been obtained from PubMed and have been selected (see subsection 4.2), the ontology is expanded by adding a new property, called *PubMedArt*, which lists the 5 most relevant articles. To add a new property, it was used *Resource.addProperty(Property p, RDFNode o)* API, where *p* is the new property, and *o* is the list of relevant articles.

4.2 Document Ranking

PubMed provides a list of articles when searching for a given keyword. Since results are usually numerous, checking them manually could be difficult. For this reason, in our approach each result will be automatically processed, and a score will be computed to represent its relevance to the search keyword. First, 5000 articles will be extracted from PubMed by using the library pymed. This library takes care of querying the PubMed database (using the standard PubMed query language) and parsing and cleaning the retrieved articles. For using it, we import the PubMed class, instantiate it and use it to perform a query. A relevant snippet of code is the following.

```
from pymed import PubMed
pubmed = PubMed(tool = "MyTool", email = "my@email.address")
results = pubmed.query("keyword", max_results = 5000)
```

For the above snippet, *tool* is the name of the application, as a string value; *email* is the e-mail address of the maintainer of the tool, and *keyword* is the searched keyword, that in our case is the label of an OBIB class, e.g., *brain*. For each article given as a result, to determine its relevance, a score was calculated using the function Term Frequency Inverse Document Frequency (TF-IDF). TF-IDF is a weight function used to measure the importance of a word within a document [39, 40]. This function increases proportionally to the number of times the term is contained in the document and increases inversely with the frequency of the term in the collection. The idea is to give more importance to the terms that appear in the document, but which in general are infrequent. The library used is tfidf_matcher. Here is the relevant code snippet implemented.

```
import tfidf_matcher as tm
results = tm.matcher([term_label], info_articles, len(info_article) -
1)
```

For the above code, *term_label* is the searched label, parameter *info_articles* is a list containing the corpus, i.e., the list of articles resulting from searching PubMed, each consisting of title, keywords and abstract. Then, *results* is an array holding the analyzed articles ranked by similarity.

Unlike other approaches, in our proposed system, each article consists of a list of words coming from title, keywords and abstract. The aim is to perform the search for a given word within the sections of the article conveying the objective of the article, rather than the details of the experiments. This should provide more relevant results.

5 RESULTS OF THE INNOVATIVE SEARCH APPROACH

Thanks to the innovative search engine described above, students can improve their learning process and outcomes, because they are given the knowledge of other professionals deriving from well-known ontologies, and a few carefully chosen scientific articles.

For our experiments, the chosen ontology is OBIB [6], an ontology designed for biobanks. Since PubMed is an engine designed for biomedical scientific literature, OBIB terms work well together with the scientific articles found in PubMed. In our tests, we searched all the 1854 OBIB classes labels, and 1453 had a satisfactory response, then each class was enriched with 5 pertinent articles. Therefore, the output of our test is an enriched ontology, available for the researcher to be used even off-line in her work. Moreover, the searching approach can be used online, and will benefit from articles that are included in PubMed overtime. Therefore, to perform the search, once one or more keywords (e.g., the words *sample* or *blood*) are provided, the system displays as output: the keywords and all terms found in the ontology (such as labels, synonyms, definitions); and the set of related scientific articles, as show in Figure 2. The graphic front-end has been implemented using specific languages and frameworks such as HTML and JQuery. The front-end communicates using appropriate APIs with the back end, which is built in Python, and returns a list of results, based on the keywords entered.

The following describes an example of using the proposed search engine. The searched keyword is *brain*. Of course, any word can be used, and then terms found in the ontology are used to assist an inexperienced user in his or her research (as described in Section 4). Once the word *brain* has been entered, our search engine will perform the following steps. Firstly, search the ontology for the word *brain*. The list of matching labels are: brain, brain specimen, brain cerebellum specimen, brain cortex specimen, cessation of brain activity death, brain death, brain bank, human brain biobank. Secondly, search PubMed using the above labels. Third, return the list of classes found in OBIB, together with the description of the class, and the list of articles selected from PubMed. Below we give a portion of the collected results. Each of the following results derives from having found the keyword *brain* in the label of a class.

- Label: brain (see the yellow box in Figure 2)
- Description: The brain is the center of the nervous system in all vertebrate, and most invertebrate, animals. Some primitive animals such as jellyfish and starfish have a decentralized nervous system without a brain, while sponges lack any nervous system at all. In vertebrates, the brain is located in the head, protected by the skull and close to the primary sensory apparatus of vision, hearing, balance, taste, and smell.
- List of Scientific Articles

(1) Bipolar Lumbar Radiofrequency Medial Branch Neurotomy in a Patient with Deep Brain Stimulation Implant. Pubmed_id: 35439825;

(2) DOPAMAP, high-resolution images of dopamine 1 and 2 receptor expression in developing and adult mouse brains. Pubmed_id: 35440585;

(3) A novel 2-phase residual U-net algorithm combined with optimal mass transportation for 3D brain tumor detection and segmentation. Pubmed_id: 35440793

(4) The blood-brain barrier-a metabolic ecosystem. Pubmed_id: 35437788

(5) Wearable Tech Marks a Breakthrough for Scanning the Brain. Pubmed_id: 35439116

- Label: brain specimen (see the blue box in Figure 2)
- Description: A specimen that is derived from cerebellum.
- List of Scientific Articles

(1) Region and species dependent mechanical properties of adolescent and young adult brain tissue. Pubmed_id: 29061984

(2) An additional brain endocast of the ictidosaur Riograndia guaibensis (Eucynodontia: Probainognathia): intraspecific variation of endocranial traits. Pubmed_id: 33681891

(3) 7 Tesla MRI of the ex vivo human brain at 100 micron resolution. Pubmed_id: 31666530

(4) Neuropathologic changes associated with stereoelectroencephalography depth electrode placement: case series. Pubmed_id: 35380201

(5) Polarimetric visualization of healthy brain fiber tracts under adverse conditions: ex vivo studies. Pubmed_id: 34745764

- Label: brain cerebellum specimen
- Description: A specimen that is derived from brain.
- List of Scientific Articles

(1) Mitochondrial Complex 1 Activity Measured by Spectrophotometry Is Reduced across All Brain Regions in Ageing and More Specifically in Neurodegeneration. Pubmed_id: 27333203

(2) Uncoupling of cerebellar blood flow and metabolism in paraneoplastic cerebellar degeneration: report of a case. Pubmed_id: 9366181

(3) Biopsy and post-mortem findings in a patient receiving cerebellar stimulation for epilepsy. Pubmed_id: 6405013

(4) Magnetic resonance imaging and three-dimensional reconstructions of the brain of a fetal common dolphin, Delphinus delphis. Pubmed_id: 11411314

(5) Brain size and brain organization of the whale shark, Rhincodon typus, using magnetic resonance imaging. Pubmed_id: 19729899

Note that by reading the description the student can select the word most pertinent to her need and analyze the selected scientific articles, as expected.

6 CONCLUSIONS

This work has proposed an innovative search engine that supports e-learning. The results given are complete and targeted thanks to the integration of specific terms provided by an ontology, and the selection of relevant scientific research articles. Thanks to our approach, the researcher is given novel published articles when searching for one or compound words. The proposed search engine can be used by any student or by a specialized user who wishes to have more scientific information on a topic. E.g., in the case of biobanks, a departmental user may want to research a type of sample to carry out surveys or simply to broaden their knowledge. Our proposed tool can be additionally used to enrich an ontology with selected scientific articles and then have the whole ontology updated with recent findings.

ACKNOWLEDGMENTS

The authors acknowledge project BIOSPEC3 funded by Regione Siciliana PO FESR 2014/20 Linea 1.1.5., and project TEAMS funded by University of Catania PIACERI 2020/22.

REFERENCES

[1] Ghulam Safdar and Abdul Wajid Khan. 2020. E-Learning: Current Scenario of Internet and Educational Learning among University Students of Punjab Pakistan. *Journal of Educational Research* 23, 1 (2020), 171.
[2] Stephen Downes. 2005. E-learning 2.0. *ELearn 2005*, 10 (2005), 1.
[3] Johan Ismail. 2001. The design of an e-learning system: Beyond the hype. *The internet and higher education* 4, 3-4 (2001), 329–336.
[4] Pankajdeep Kaur, Pallavi Sharma, and Nikhil Vohra. 2015. An ontology based E-learning system. *International Journal of Grid and Distributed Computing* 8, 5 (2015), 273–278.
[5] Rajapandian Radha, K Mahalakshmi, V Sathish Kumar, and AR Saravanakumar. 2020. E-Learning during lockdown of Covid-19 pandemic: A global perspective. International journal of control and automation 13, 4 (2020), 1088–1099.
[6] Francesco Colace and Massimo De Santo. 2009. Ontology for E-learning: A Bayesian approach. *IEEE Transactions on Education* 53, 2 (2009), 223–233.
[7] Igor Keleberda, Natalya Lesna, Sergiy Makovetskiy, and Vagan Terziyan. 2004. Personalized distance learning based on multiagent ontological system. In *Proceedings of IEEE International Conference on Advanced Learning Technologies*, 2004. 777–779.
[8] Ritu Khare, Robert Leaman, and Zhiyong Lu. 2014. Accessing biomedical literature in the current information landscape. *Biomedical Literature Mining* (2014), 11–31.
[9] Ruth C Clark and Richard E Mayer. 2016. E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. John Wiley & sons.
[10] Michael Green and Erin McGill. 2011. State of the industry, 2011: ASTD's annual review of workplace learning and development data. *Alexandria, VA: ASTD* (2011).
[11] Quadri Noorulhasan Naveed, A Muhammad, Sumaya Sanober, Mohamed Rafik N Qureshi, and Asadullah Shah. 2017. A mixed method study for investigating critical success factors (CSFs) of e-learning in Saudi Arabian universities. *Methods* 8, 5 (2017), 171–178.
[12] Shafqat Hameed, Atta Badii, and Andrea J Cullen. 2008. Effective e-learning integration with traditional learning in a blended learning environment. In *Proceedings of European and Mediterranean Conference on Information Systems*, Vol. 60. 14.
[13] Barry Smith. 2012. Ontology. In *The furniture of the world*. Brill, 47–68.
[14] Dragan Djurić, Dragan Gašević, and Vladan Devedžić. 2005. Ontology modeling and MDA. *Journal of Object technology* 4, 1 (2005), 109–128.
[15] Anita Bandrowski, Ryan Brinkman, Mathias Brochhausen, Matthew H Brush, Bill Bug, Marcus C Chibucos, Kevin Clancy, Mélanie Courtot, Dirk Derom, Michel Dumontier, *et al.* 2016. The ontology for biomedical investigations. *PloS one* 11, 4 (2016), e0154556.
[16] The Gene Ontology Consortium. 2020. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49, D1 (12 2020), D325–D334.
[17] Agnieszka Konys and Zygmunt Drążek. 2020. Ontology Learning Approaches to Provide Domain-Specific Knowledge Base. *Procedia Computer Science* 176 (2020), 3324–3334.
[18] Thomas Bittner, Maureen Donnelly, and Stephan Winter. 2005. Ontology and semantic interoperability. In *Large-scale 3D data integration*. CRC Press, 139–160.
[19] Luis Martín, Alberto Anguita, Víctor Maojo, Erwin Bonsma, Anca ID Bucur, Jeroen Vrijnsen, Mathias Brochhausen, Cristian Cocos, Holger Stenzhorn, Manolis Tsiknakis, *et al.* 2008. Ontology Based Integration of Distributed and Heterogeneous Data Sources in ACGT. In *HEALTHINF* (1). 301–306.
[20] Mathias Brochhausen, Jie Zheng, David Birtwell, Heather Williams, Anna Maria Masci, Helena Judge Ellis, and Christian J Stoeckert. 2016. OBIB-a novel ontology for biobanking. *Journal of biomedical semantics* 7, 1 (2016), 1–9.
[21] Kathi Canese and Sarah Weis. 2013. PubMed: the bibliographic database. *The NCBI Handbook 2* (2013), 1.
[22] Andreas Doms and Michael Schroeder. 2005. GoPubMed: exploring PubMed with the gene ontology. *Nucleic acids research* 33, suppl_2 (2005).
[23] Yas A Alsultanny. 2006. e-Learning System Overview based on Semantic Web. *Electronic Journal of e-Learning* 4, 2 (2006), pp111–118.
[24] Matteo Gaeta, Francesco Orciuoli, and Pierluigi Ritrovato. 2009. Advanced ontology management system for personalised e-Learning. *Knowledge-Based Systems* 22, 4 (2009), 292–301.
[25] James D Slotta and Michelene TH Chi. 2006. Helping students understand challenging topics in science through ontology training. *Cognition and instruction* 24, 2 (2006), 261–289.
[26] John K Tarus, Zhendong Niu, and Ghulam Mustafa. 2018. Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial intelligence review* 50, 1 (2018), 21–48.
[27] Aijuan Dong and Honglin Li. 2005. Ontology-based information integration in virtual learning environment. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence* (WI'05). IEEE, 762–765.
[28] Maryam Hazman, Samhaa R El-Beltagy, and Ahmed Rafea. 2009. Ontology learning from domain specific web documents. *International Journal of Metadata, Semantics and Ontologies* 4, 1-2 (2009), 24–33.
[29] David Sánchez and Antonio Moreno. 2008. Learning non-taxonomic relationships from web documents for domain ontology construction. *Data & Knowledge Engineering* 64, 3 (2008), 600–623.
[30] Emiliano Tramontana and Gabriella Verga. 2022. Ontology Enrichment with Text Extracted from Wikipedia. In *Proceedings of International Conference on Software Engineering and Information Management* (ICSIM'22). ACM. January 21-23, 2022: 113-117.
[31] Vivek Iyer, Lalit Mohan, Mehar Bhatia, and Y Raghu Reddy. 2019. A Survey on Ontology Enrichment from Text. In *Proceedings of International Conference on Natural Language Processing*. 95–104.
[32] Lalit Mohan Sanagavarapu, Vivek Iyer, and Y Raghu Reddy. 2021. OntoEnricher: a deep learning approach for ontology enrichment from unstructured text. *arXiv preprint arXiv:2102.04081* (2021).
[33] Paola Velardi, Paolo Fabriani, and Michele Missikoff. 2001. Using text processing techniques to automatically enrich a domain ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*. 270–284.
[34] Zuoshuang Xiang and Yongqun He. 2009. Improvement of PubMed literature searching using biomedical ontology. *Nature Precedings* (2009), 1–1.
[35] Akiko Aizawa. 2003. An information-theoretic perspective of tf–idf measures. *Information Processing & Management* 39, 1 (2003), 45–65.
[36] Jing Jiang and ChengXiang Zhai. 2007. An empirical study of tokenization strategies for biomedical information retrieval. *Information Retrieval* 10, 4-5 (2007), 341–363.
[37] Zhiyong Lu, Won Kim, and W John Wilbur. 2009. Evaluating relevance ranking strategies for MEDLINE retrieval. *Journal of the American Medical Informatics Association* 16, 1 (2009), 32–36.
[38] Stefan Siemer. 2019. Exploring the Apache Jena Framework. (2019).
[39] Prafulla Bafna, Dhanya Pramod, and Anagha Vaidya. 2016. Document clustering: TF-IDF approach. In *Proceedings of International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 61–66.
[40] Juan Ramos *et al.* 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242. 29–48.