Scientific Journal of Faculty of Education, Misurata University-Libya, No26, Mar 2025
Published online in March
المجلة العلمية لكلية التربية، جامعة مصراتة، ليبيا -العدد السادس والعشرون، مارس 2025
Issn: 2710-4141
تاريخ الاستلام:2024/12/31 **تاريخ النشر:2025/1/27**

Design and Implementation of a Proposed Model for a Semantic-Based Search Engine

Wejdan Mohamed Elhaj-Suliman
Misurata University, Faculty of Education, Information Technology department, Misurata, Libya
w.suliman@edu.misuratau.edu.ly

Abstract:

Conventional search engines that depend upon keyword-based retrieves have resulted in inaccuracies and non-relevance with respect to context. This research aims to overcome the drawbacks of traditional search methodologies by presenting a semantic model based on search engine that improves information retrieval using organized, ontology-oriented approach. This research aims to attain a model that can understand the queries and give accurate replies based on context. According to the company, this process includes building an ontology in OWL (Ontology Web Language) that allows and helps a semantic search engine understand nuanced academic relationships within university course data. This model has been implemented and tested in a well-controlled academic setting where the dataset was made to befit its exact performance.The results obtained show that the proposed model is indeed able to perform much better than regular search engines in terms of precision, relevance and finding time. This study provides, to the best of our knowledge, one of the first reported examples using domain-specific relatedness measures for rapid semantic search in specialized fields where accurate data retrieval is a must-have. Major contributions also include an efficient implementation framework that can be scaled up and adapted towards specific scenarios involving large medical databases which use more complex terminologies such as UMLS, ICD terms (e.g., BCM), etc. In an attempt to fill this research gap, we provide a baseline model that can be further developed and extended for other domains with contextual search problem.

Keywords Semantic web, search engine, semantic search, Ontology Web Language, OWL, Protégé, Jena, Course Finder

Scientific Journal of Faculty of Education, Misurata University-Libya, No26, Mar 2025
Published online in March
المجلة العلمية لكلية التربية، جامعة مصراتة، ليبيا -العدد السادس والعشرون، مارس 2025
Issn: 2710-4141
تاريخ الاستلام:2024/12/31 تاريخ النشر:2025/1/27

تصميم وتنفيذ نموذج مقترح لمحرك البحث دلالي

أ.وجدان محمد سليمان

جامعة مصراتة ،كلية التربية، قسم تقنية المعلومات ، مصراته ليبيا

الملخص:

تعتمد محركات البحث التقليدية القائمة على الكلمات المفتاحية على أساليب استرجاع غالبًا ما تؤدي إلى نتائج غير دقيقة وغير ذات صلة بالسياق. يهدف هذا البحث إلى التغلب على عيوب منهجيات البحث التقليدية من خلال تقديم نموذج دلالي يعتمد على محرك بحث يعزز استرجاع المعلومات باستخدام نهج منظم وموجه نحو الأنطولوجيا. يسعى البحث إلى تطوير نموذج يمكنه فهم الاستفسارات وتقديم إجابات دقيقة بناءً على السياق. يشمل هذا النهج إنشاء أنطولوجيا باستخدام لغة OWL (لغة الويب الدلالية)، مما يُمكن محرك البحث الدلالي من فهم العلاقات الأكاديمية الدقيقة داخل بيانات المقررات الجامعية.

تم تنفيذ النموذج واختباره في بيئة أكاديمية مُحكمة، حيث تم تصميم مجموعة البيانات بما يناسب تقييم أدائه بدقة. أظهرت النتائج أن النموذج المقترح يتمتع بأداء أفضل بكثير مقارنة بمحركات البحث التقليدية من حيث الدقة، الصلة، وسرعة الاسترجاع. وفقًا لما نعرفه، تُعد هذه الدراسة واحدة من أوائل الأمثلة الموثقة التي تستخدم مقاييس الصلة المتخصصة بالمجال للبحث الدلالي السريع في مجالات متخصصة تتطلب استرجاع بيانات دقيق.

تشمل المساهمات الرئيسية أيضًا إطار عمل تنفيذي كفء يمكن توسيعه وتكييفه مع سيناريوهات محددة تشمل قواعد بيانات طبية ضخمة تحتوي على مصطلحات معقدة مثل UMLS و ICDمثل BCMوغيرها. وفي محاولة لسد هذه الفجوة البحثية، يقدم هذا البحث نموذجًا أساسيًا يمكن تطويره وتوسيع نطاقه ليشمل مجالات أخرى تتطلب حلول بحث سياقية دقيقة.

الكلمات المفتاحية: الويب الدلالي، محرك البحث، البحث الدلالي، لغة الويب الأنطولوجية، OWL، بروتيجي، جينا، باحث المقررات.

1. Introduction

The rapid rise of online information has made search engines that provide accurate and relevant results essential. Standard keyword searches often fall short, as they match words without understanding the true meaning or context of user queries. This can result in irrelevant results and dissatisfied users (Pimentel, D. R,2024).

To fix these problems semantic web tech has come into play aiming to create a smarter search system that can grasp and work with the meaning hidden in data. By using structured knowledge maps, these semantic search engines show info in a way that lets them spot tricky links between ideas, which helps make searches more accurate and on-point . (Debellis, M, 2021)

This study introduces a groundbreaking method for semantic search by creating a new model that merges the Web Ontology Language (OWL) with Protégé and Jena frameworks to back an ontology-based search process. Unlike earlier models that often struggle with growth and flexibility, this model tackles these issues by using a adaptable structure that enables more intricate, field-specific queries in academic environments. By using OWL, the proposed model boosts semantic interpretation abilities offering a stronger solution than standard keyword-based search engines .(Geng, Y. et al. ,2021) Devlin et al. 2019

Beyond its use in schools, this model has a chance to be put into action in many areas that need spot-on data retrieval, like healthcare, law research, and science databases. By allowing for detailed context-based search, this model shows it can adapt well and might change how we find info across special fields. This makes it useful in areas where being right and on-point matters most. The study's findings show that using an ontology-driven approach, which this model does boosts the accuracy, relevance, and speed of search results in controlled school settings.

2. Literature Review

Semantic search engines are one step ahead of search systems which attempt to eradicate the flaws of old keyword based systems using ontologies and knowledge graphs. According to Ji et al. (2022), "knowledge graphs improve context for queries such that data representation is enhanced along with increased accuracy in information retrieval.".

Zouaoui and Rezeg (2021) developed a philosophy of the Quranic search engine based on ontology. This is semantically augmented semantic indexing to enhance and optimize and accurate retrieval of complex religious texts. Staying in that trend, Islam, Syed, and Shaikh (2023) posited that semantics web technologies had to be adapted in the formulation of tools designed to extract ontology and knowledge extraction, which, in turn would lead to more intelligent industrial applications.

In practical applications, researches of Azad et al. (2022) show that linked data search engines with the help of tools of data mining can enhance the effectiveness and accuracy of searches through enhancing latent semantic features of data. Similarly, Cao and Ngo (2018) developed a latent semantic-based model to enhance the process of semantic searches using ontological features.

3. The Semantic Web

The Semantic Web builds on the current web by giving information clear meanings making it easier for people and computers to work together. This setup helps to find, combine, automate, and reuse data more across different apps (Devlin et al. 2019). Instead of just linking web pages, the Semantic Web arranges data like a directed labeled graph. In this graph, each node stands for a resource, and each line shows a type of relationship, which is also treated as a resource. This model lets systems create complex links such as top-down or related connections between things, based on features like color, price, and where they are (Ji et al. 2022).

Figure 1 shows how the current web is built. It displays web pages linking to each other without meaningful connections. This points out the shortcomings of old-school keyword searches.

Figure 2 depicts how the Semantic Web is set up. It stresses the meaningful links between data. These links allow for more accurate and relevant information searches.

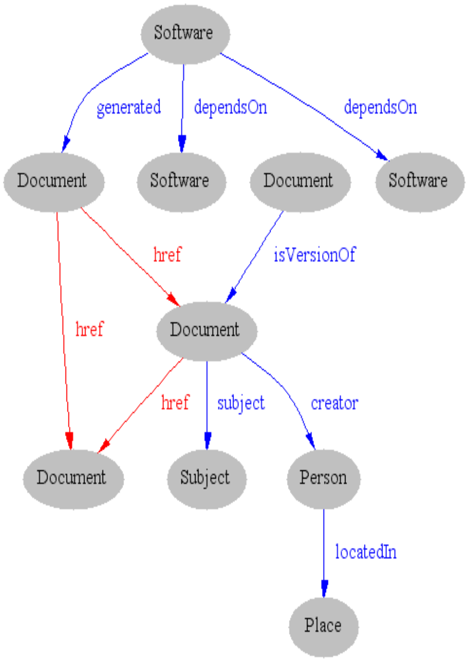

Figure 2 Semantic Web

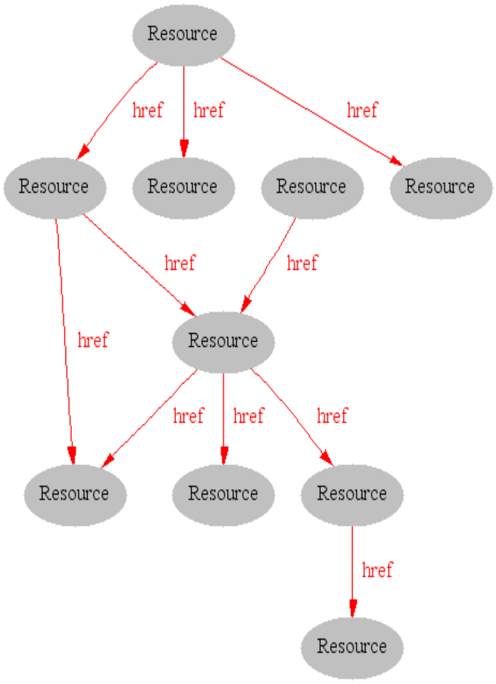

Figure 1 Current Web

4. Web ontology language (OWL)

The Semantic Web lets people work together to create ontologies and build shared vocabularies without central control. People use OWL when apps need to process information in documents, not just show it to humans.

OWL can express the meanings of terms in vocabularies and how those terms relate to each other. This description of terms and their connections is called an Ontology. OWL offers more ways to express meaning and semantics than RDF, RDF-S, and XML. This makes OWL the top choice for representing content on the Web that machines can understand. (Debellis, M. , 2021)

5. Search Engines

A search engine is a tool that looks for documents on the internet or in a specific database using keywords. It gives you a list of results based on what you type in (H Azad, A Deepak, & A Azad 2022). It has a big impact on helping people find websites. Better listings often bring lots of visitors to sites (Hogan, A. ,2020).). , search engines use web crawlers to gather info and indexers to make word indexes you can search. When you type in something like "buy PC online," the engine finds all pages with those words. Search engines often support Boolean searches by default. This lets you add or leave out terms without typing in Boolean operators yourself .(Hogan, A. ,2020).

5.1 The Semantic Search Engine

Latent Semantic Indexing is a specific mathematical process. It uses pure math to create semantic connections between document sets.

Semantic Search Engines collect word content in a way similar to current methods. They also weigh words like modern techniques do, but they can take extra steps beyond where current ones stop. (Beal, 2021)

The Semantic Search Engine has the ability to analyze word compilations in content assessing their relative importance, coherence, and semantic connections. The Search Engine then identifies other pages or page collections with matching semantic profiles or those that fall within an acceptable threshold of similarity. (Beal, 2021)

5.2 Semantic Search

The Semantic Web apps are expanding at the same rate as the WWW. Semantic search, a type of semantic Web search app, has an impact on

Web development as one of its most well-known and important advancements. This extra search has the potential to make searching better. Semantic Search aims to boost and upgrade the outcomes of traditional search methods. These conventional approaches rely on Information Retrieval tech that taps into Semantic Web data .(Cao, T. H., & Ngo, V. M, 2018).

Traditional Information Retrieval (IR) tech relies on word accuracy in documents. It adds to this in the Web setting with details about the Web's hyperlink structure. The Semantic Web makes available vast amounts of structured, machine-readable data about a wide range of objects. This offers many chances to improve on standard search methods. (H Azad, A Deepak, and A Azad 2022).

5.2 Free-Text vs Semantic Search (Sahu, Mahapatra, & Balabantaray, 2016).

Free-Text	**Semantic Search**
Lack of accuracy	Accurate results
Has to rely on the annotator to provide all possible keywords	Does not have to rely on keywords
Repetitive annotation effort	Provide for Inference (Implicit reasoning)
Lack of structure, concepts and relationships	Adds structure, Concepts-Relationship

6. University Courses Search Engine

In this paper we are proposing a model named The “Course Finder” as prove of concept, it is an online resource that allows people to search for particular course in a particular College/University in a particular country. Anyone can use the Course Finder to search for a course and compare these courses within diversity of criteria, including by specific provider, entry cut-offs, and course fees.

The Course Finder is basically designed for students who are looking for information on the Web that has been provided by the organization such as Colleges or Universities.

7. Methodology

We well be mentioning the technologies and the tools that are related to the main work, in addition to the description of the characters and the way of working for each one.

7.1 Protégé

Protégé considered as a free open-source platform which offers tools suite to make models of domain and knowledge-based applications within ontologies. (Protégé web site, 2020)

In this model, the ontology includes group of classes that are organized a subsumption hierarchy to signify the concepts of salient domain, a group of linked to classes describing their relations and properties, in addition a group of those classes – individual exemplars of the conceptions that give their properties specific values.

7.2 Jena - A Semantic Web Framework for Java

Jana is a toolkit that was developed by HP. It proposed to develop applications with the Semantic Web. (Jena web site, 2024) Basically, Jena is a framework of Java that used to build Semantic Web applications. It can provide a programmatic environment for RDF, RDFS, OWL, and SPARQL; in addition it contains a rule-based interface engine.

7.3 Integration of Jena in protégé-owl

The relationship between Protégé-OWL with Jena is constantly. it is possible to switch a Protégé *OWLModel* to Jena *OntModel*, to obtain a static snapshot of the model during the running time. However, the model would be built again following every model change. . (Protégé web site, 2020)

The main concept of this integration is that the operation for both systems occurs on a low level "*triple*" representation of the model.

Figure 3 shows the integration architecture for the Jena system within the Protégé-OWL environment, illustrating how triple-based data representation supports semantic search queries.

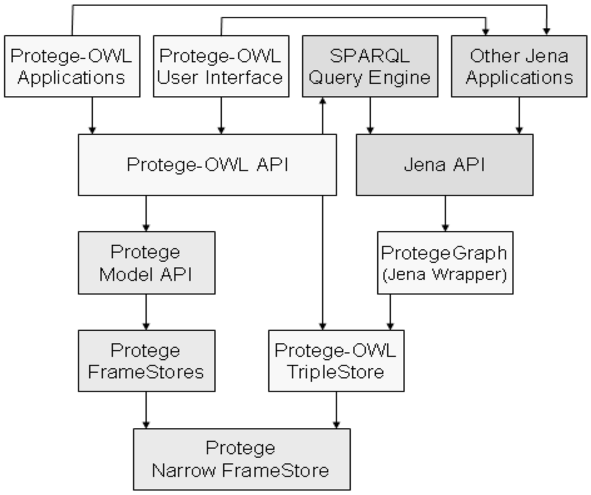

Figure 3: The architecture integration of Jena

The Protégé has its own native frame store mechanism *that TripleStore classes* wrapped it in *Protégé-OWL*. the corresponding interfaces in the Jena world, are known as Graph and Model. The process structure in this Model will be

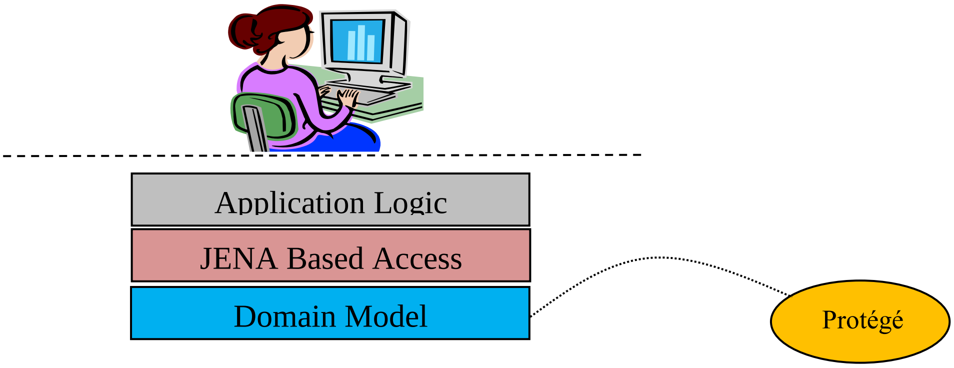

Figure 4 : The Architecture of the Model

presented in the figure 4 below. . (Protégé web site, 2024)

8 The Design

8.1 Domain Analysis: OWL

5.3 The first step in implementing this model involved conducting domain analysis using OWL to create an ontology representing the subject domain (Debellis, M. 2021). This ontology provides a structured conceptualization of the domain, detailing the key concepts and relationships necessary for the search engine to interpret queries accurately ((Sahu, Mahapatra, & Balabantaray,(2016)

. OWL-DL was chosen over OWL-Lite due to its additional features, such as unlimited cardinality and disjoint classes, as well as its full support from the reasoner, which is essential for comprehensive domain analysis. (Allemang, D., Hendler, J., & Gandon, F, 2020).

8.2 Classes Structure

There are five classes (*LevelOfStudy, Course, Module, University and CourseCost*) that have been created from the main class (*Thing*), Figure 5 illustrates the hierarchical structure of the primary ontology classes, including '*LevelOfStudy*,' '*Course*,' '*Module*,' '*University*,' and '*CourseCost*.' Each class represents a critical aspect of university courses, enabling more accurate categorization and relationship definition within the model. as bellow:

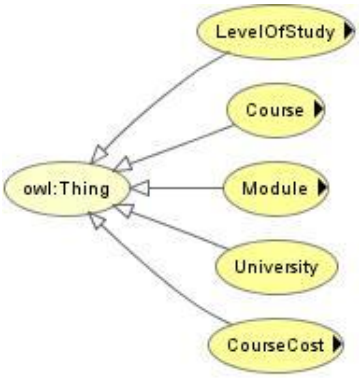

Figure 5 : Main five classes

- *LevelOfStudy:* There two classes belong this class, the first class called *Postgraduate* which describes the postgraduate courses, and the other course called *Undergraduate* which describes the undergraduate courses.
- *CourseCost:* This class is to describe the fees for two class. The first class called *FullTime-StudentFees* is describes the fees of full time student, the class *PartTimeStudentFees* is describes the fees of part time student.
- *University:* In this class, there will be a number of Individuals that are represent the names of some universities.
- *Model:* This class has a number of classes that represent the modules classification regarding to the subject title or name. For instance, *IT module* consists of number of *Individuals* representing that IT module, and so.
- *Course:* This class has the most important class which is the *SubjectArea* class. This class includes the subjects' classifications which are two in our course and can be increased easily to include all the available courses in a university. For instance, the class *ComputingAndTechnology* consists of four courses (classes) and each course will include number of *Individual* Member represent the related.

The following diagram (figure 6) provides a comprehensive view of the ontology structure, highlighting parent and child classes and the relationships between them within the university course domain.

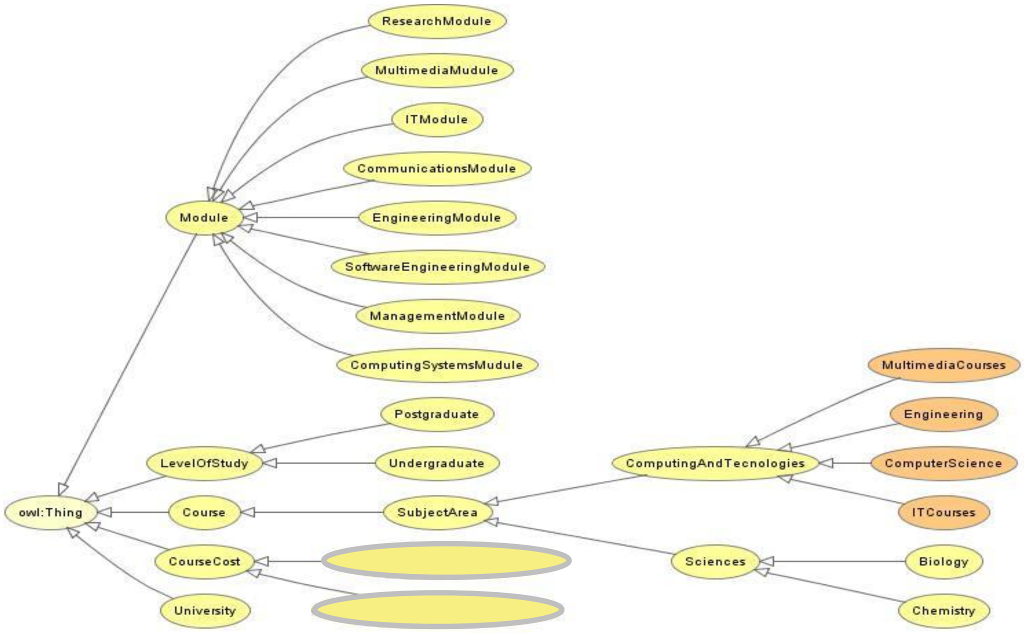

Figure 6: Ontology Structure for all Classes

In terms of they are Subclass, so for instance, all individuals that are members of the *ITCourses* class are members of the *ComputingAndTechnology* class and members of the *SubjectArea* class, as we have settled that *ITCourses* is a subclass of ComputingAndTechnology which is a subclass of *SubjectArea*. As depicted in Figure 7, the ‘Course’ subclass under ‘Computing and Technology’ further breaks down into specific categories, such as ‘IT Courses,’ illustrating a deeper level of subject classification.

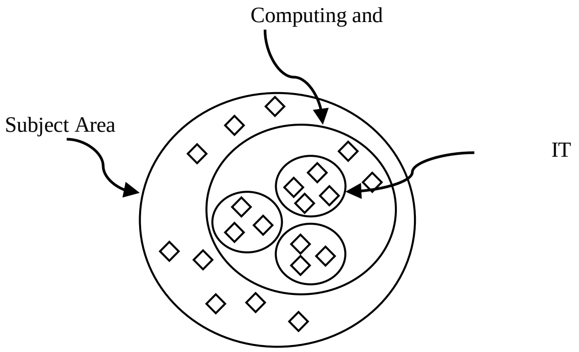

Figure 7 : Course subclass

8.3 Properties

The next step after creating the main structure for the university courses domain is to create some relationships between classes and individuals so the *Course* class has:

- Object properties: *hasModule*, *hasUniversity*, *hasCourseCost*, and *hasLevel*. The *Range* for these properties was: *Module*, *CourseCos*t and *LevelOfStudy* classese correspondingly.
- dataTypeProperties: *hasCourseName* and *hasStudyMode*

8.4 Individuals

The probable *Course Name* value will be an instance of *ComputingAndTechnology subclasses*. Which are *ITCorses, MultimediaCorses, EngineeringCourses and ComputingSinceCourses*.

To relate individuals to individuals we have used object properties, while for individuals to data types we have used data type properties. For instance, the property *hasITModule* is an *objectproperty* that links the individual information technology (IT courses) to the individual Enterprise Computing (IT module). While the property *hasStudyMode* is a *datatype* property that links individual (information technology) to *datatype* (xsd:string). Figures 8.a and 8.b display the object and data type properties associated with ‘Information Technology’ courses, including ‘hasITModule’ for object properties and ‘hasStudyMode’ for data type properties.

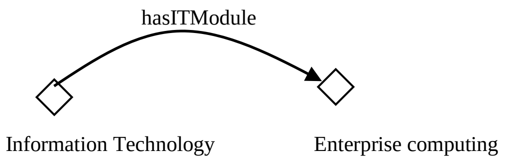

Figure 8.a: object property

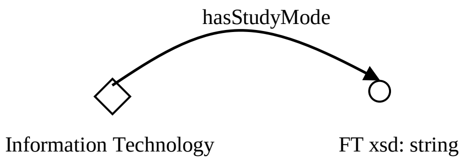

Figure 8.b: datatype property

To describe each course number of properties has been used. As can be seen in the figure below the course (Information Technology) has number of properties. Figure9 details the specific properties assigned to the 'Information Technology' course, such as links to modules, universities, course costs, and study levels.

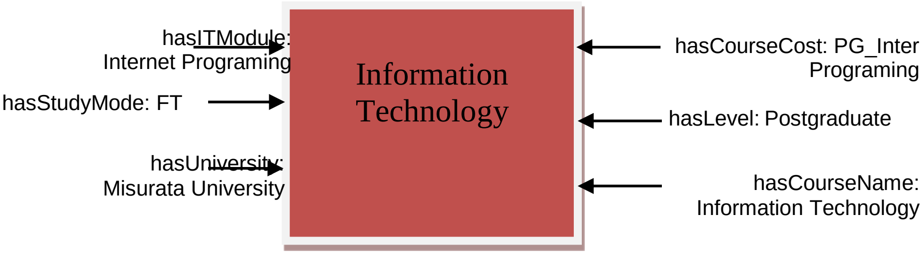

Figure 9: Individual properties

8.5 Disjoint Classes

The *Postgraduate* and the *Undergraduate* considered as *Disjoint Classes*. The class *LevelOfStudy* includes two subclasses which are *Postgraduate* and the *Undergraduate* classes, thus an individual/class cannot be an instance of more than one of these two classes. On the other hand, it's not assumed that the *Individual* is not a member of a particular class, basically because that Individual has not been declared to be a member of that class.

In sequence to 'separate' a group of classes, they have to be disjointed from each other. This will ensure that the individual who has been declared to be a class's member in the group cannot be a member of any other group classes. It might not be a sense to make *Individual* to be a Postgraduate and an Undergraduate; the same thing could be applied with classes *FullTimeStudentFees* and *PartTimeStudentFees*, where the student can only be either full time or part time.

9. Discussion

The design of the semantic search engine should trying to cover as much as possible the nearest meaning for each query. The full analysis for the domain should be considered when designing the search engine in terms of the data on the internet are not analysed semantically, thus the designed search engine in this Model is not a web-based.

Furthermore, it might be there an operational problem when the number of combination are not small in order to many keywords concerned, in addition to the multiple possible matches for each single keyword. Therefore, the design should chose the most specific class that could match among other class matches.

10 . Conclusion

Semantic search engines have the ability to intricate the search queries in a semantic way to logical relations between the documents and then get back the information sufficiently. The semantic search has more flexibility than other form-based query interface. Simply, Semantic search engines makes sense of the user queries by converting them into formal queries. It can find out the meaning of the entered key words and then querying the back-end semantic data repository. Moreover, the results will be ranked according to the matching or the nearest meaning.

There are many advantages the author has achieved while implementing this Model in terms of the analysing the information, designing, using the proper tools for the specific job.

References:

- Pimentel, D. R. (2024). *Learning to evaluate sources of science (mis)information on the internet: Assessing students' scientific online reasoning*. *Journal of Research in Science Teaching*.
- Geng, Y., Chen, J., Chen, Z., Pan, J. Z., Ye, Z., Yuan, Z., Jia, Y., & Chen, H. (2021). OntoZSL: Ontology-enhanced Zero-shot Learning. *Proceedings of the Web Conference 2021*, 3325–3336. https://doi.org/10.1145/3442381.3450042
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, 4171-4186
- Zouaoui, S., Rezeg, K. A Novel Quranic Search Engine Using an Ontology-Based Semantic Indexing. Arab J Sci Eng 46, 3653–3674 (2021). https://doi.org/10.1007/s13369-020-05082-5
- Islam, N., Syed, D., & Shaikh, Z. A. (2023). *Semantic Web: An overview and a .net-based tool for knowledge extraction and ontology development*. In *Semantic Technologies for Intelligent Industry 4.0 Applications* (1st ed., p. 29). River Publishers
- Cao, T. H., & Ngo, V. M. (2018). Semantic search by latent ontological features. *arXiv preprint arXiv:1807.05576*.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494-514
- H Azad, A Deepak, A Azad, , 2022, LOD search engine: A semantic search over linked data, Journal of Intelligent Information Systems,Volume 59, p 71–91.
- Sahu, S. K., Mahapatra, D. P., & Balabantaray, R. C. (2016). Comparative study of search engines in context of features and semantics. *Journal of Theoretical and Applied Information Technology, 88*(2), 212–218
- Protégé web site, 2020, Available at: http://protege.stanford.edu.
- Jena web site, 2024, Available at: http://jena.sourceforge.net

- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012). *OWL 2 Web Ontology Language Primer (Second Edition)*. W3C Recommendation. https://www.w3.org/TR/owl2-primer
- Debellis, M. (2021). A practical guide to building OWL ontologies using Protégé 5.5 and plugins. *arXiv preprint arXiv:2104.10123*
- Vangie Beal , 2021 ,Webopedia, search engine .Available at: https://www.webopedia.com/definitions/search-engine.
- Allemang, D., Hendler, J., & Gandon, F. (2020). *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL* (3rd ed.). ACM Books.
- Hogan, A. (2020). *Web Ontology Language*. In *The Web of Data* (pp. 185–322). Springer, Cham. https://doi.org/10.1007/978-3-030-51580-5_5