Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hepdyslip: Merged Ontologies For Hepatitis And Dyslipidaemia

A. Bashir[1], R. Nagpal[2], D. Mehrotra[3*], M. Bala[4]

[1,2]Amity University, Uttar Pradesh, India. *E-mails: zarabashir93@gmail.com, rnagpal1@amity.edu*
[3*]Jaypee Institute of Information and Technology, Noida, India, *mehdeepti@gmail.com*
[4]Indraprastha College for Women Delhi University, New Delhi, India.*manjugpm@gmail.com*.
***Corresponding author:** D. Mehrotra
*Email: mehdeepti@gmail.com

ARTICLE INFO

Received: 20 Dec 2024

Revised: 17 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

EHR systems has transformed healthcare data management but challenges related to interoperability, cost and security remain significant. Role of ontologies in EHR for representing the unstructured and semi-structured data found in different medical records. such as doctor's prescription, diagnostic report, and other treatment-related data and different biomedical knowledge bases are well established. The ontologies and knowledge graphs are used for representing and structuring the unstructured data so that it can be used for coding and retrieving the relevant data in knowledge bases. Knowledge Graphs are flexible and adaptable as it organizes data into interconnected structures that represent relationships between entities such as disease, treatments, and progressions. Implied relationship between many diseases is known however they involve complex interdependencies. Ontology merging plays an important role in showing the causal relationship among such diseases. The objective of ontology merging is to integrate pertinent features of ontologies, like axioms, persons, and annotations, into the resultant ontologies that can be used for better establishing the semantic relationship between two diseases. In this study two diseases considered are Hepatitis and Dyslipidemia and converging their ontologies a HypDysLip ontology is created. The created ontology is further enriched using the generative AI tool ChatGPT.

Keywords: Electronic health records, Knowledge base, Large Language models, Parsing, Resource Description Framework, Semantic web, Ontology, Semantic similarity.

1. Introduction.

Health care is a data-intensive industry that generates a lot of data. The information exchange among various healthcare systems and providers is challenging. Due to the unique methods every healthcare organization uses Electronic Health records (EHR) [1] for storing and exchanging information internally and externally ensuring data integrity and effective information exchange of EHRs, interoperability between the healthcare organizations. This underscores the need for semantic interoperability, which allows seamless sharing of data across all departments of the organization like the clinicians, nurses, laboratories, and the entire hospital as well as with the other healthcare centre that requires this information. Planning safe and effective treatment is made easier by the integration of patient health history made possible by electronic health records, or EHR. Aggregate-level EHR in conjunction with data analytics facilitate the research and development of successful chronic illness medications and treatments. [1]. One of the most extensively praised advantages of EHRs is quality improvement. Through a number of ways, EHRs can enhance clinical results and patient safety. [3] Given that non-technologists administer the system and that the majority of notes are written in natural language using terminology from the medical domain, semantic interoperability is more important in the healthcare setting. It is difficult to combine and extract meaning from the disparate standards that the current EHR solutions are adhering to. [4] EHRs also help healthcare providers communicate with one another, which improves care coordination and lowers the risk of medical errors. Furthermore, it has been demonstrated that EHRs enhance clinical decision-making and evidence-based practices, which in turn improve patient outcomes. EHRs can assist healthcare professionals follow best practices and recommendations by sending out notifications and reminders for screenings and preventive care [5]. This can be implemented by ensuring

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

interoperability among healthcare centres/clinicians in which the exchange of data is carried out in a semantic manner.

Semantic Interoperability refers to the capability of various systems to comprehend the context and meaning of data that is communicated during information exchange [2]. It involves understanding the meaning of the data and their inter-relationships. Healthcare interoperability has demonstrated the value of systems in accomplishing several objectives and providing solutions for various issues with EHRs. Healthcare interoperable systems will enhance data security, integrity, and cross-platform deployment with strong integration across many healthcare business ecosystems [2] Semantic interoperability seeks to facilitate data sharing across the hospital's lab, clinicians, nurses, and other departments. Therefore, to transmit information across organisational boundaries, eliminate data silos and maintain data regardless of vendors.[6] The semantic interoperability (SI) challenge in big data applications is not well-solved by the current methodologies due to a lack of well-defined standards and proven technologies. The collaborative strategy to addressing SI in big data and IoT for health care applications is suggested by [7]. Within the healthcare industry, doctors and patients can collaborate efficiently and conveniently with one another. To ensure semantic interoperability, Ontology, Semantic technologies, and Knowledge management systems are utilised. Gansel et al [3] explored the difficulties associated with Electronic Health Record maintenance, focusing on clarifying, identifying, and updating related concepts. In [4] the authors have explained the basic levels of interoperability which can enhance data and workflow within healthcare organizations by facilitating data interoperability. In healthcare domain, the interoperability refers to varied systems, applications and the devices that share, use and process information that is accessible from any place. In a traditional healthcare facility, the patients' records are stored manually or digitally but different hospitals have their own way of storing the data. Ontology designs a common shared platform which stores the unstructured and semi structured data of the patients and provides ease if access to the information. The data can be shared and accessed within and among healthcare facilities, thereby leading to interoperability of data. To bridge this gap in the existing literature, an Ontological framework is designed for the diseases Hepatitis Virus and Dyslipidaemia.

In past few years, ontology-based approaches have been widely adopted in many fields of research including the medical domain. Due to the disconnected development of individual ontologies, many ontologies in similar, identical, interrelated fields have been proposed and designed. Thereby not leading communication and interoperability among the applications or the information systems that rely on those ontologies. The issue can be resolved by integration of ontologies by merging the existing ontology into a new ontology that clumps the knowledge consisted by the individual ontologies that can be used by various heterogenous applications. Ontology mapping and merging are prominent fields of research in Artificial Intelligence. An Ontology is designed to be shared among various information systems. The mapping or merging of ontologies builds a meta-layer, thereby allowing various applications to access the merged ontology and sharing the information among various users/applications/information systems. Merging of ontologies could prove beneficial in the field of medicine as well. By merging the disease ontologies, we can gather the knowledge in a single system and by merging similar ontologies, the relationship among the diseases can also be predicted. The ontology could also aid the clinician in decision making based on the knowledge repository and association of the diseases in the individual ontologies.

Additionally, The most effective technology for processing and manipulating textual data is thought to be large language models, or LLMs. These models do perform better when adjusted to the medical terminology and causal linkages. Nonetheless, structured knowledge representation should be used to supplement the benefits of LLMs. In this case, Knowledge Graphs (KG) offer a means of exploring and utilising the retrieved material. There are several difficulties in determining causal relationships from medical abstracts. Abbreviations, specialised terminology, and terms with numerous meanings are common in medical language.

Furthermore, relationships between entities are frequently indicated rather than clearly stated, necessitating the ability to grasp context and draw conclusions. The availability and quality of data is another problem. Effectively fine-tuning a model may be more challenging because there aren't many annotated datasets made for relation extraction.

Example 1 (Cause). Given the text "The patient suffers from Hepatitis B virus disease. Patient has loss of appetite, fatigue, headache. Liver is dysmorphic. Patient has low cholesterol", the system extracts the Knowledge Graph. One component identified Hepatitis B as a Disease, loss-of-appetite, fatigue, headache as a Symptom, liver as a Body Part and cholesterol as a Risk factor. It identified two relations of type cause and one relation of type affect.

2. Related work

There has been a wide use of Semantic similarity in machine translation, retrieval of information, data mining and artificial intelligence over past few years. A comparative study by Sana Ben Abdallah et.al [5] is carried out based on the preprocessing phase, classification phase, deep learning models to extract the relevant terms and

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

semantic enrichment [6] of the ontology. the implementation and evaluation is done on various medical datasets.

A computation method which was based on concept tree for calculating world similarity was introduced [7] which formulated a method of construction of a concept tree, in combination with a method based on depth and path of the tree that covers all kinds of sememes' influences in the Decomposing word concept definition (DEF), thereby avoiding the complexity of the similarity computation processes that require measurement of the similarity among sememe groups in DEF of concept.

By combining large language models (LLMs), ontologies, and causality, new avenues for improving natural language understanding are opened up. Although LLMs show promise in identifying and comprehending causal relationships in textual data, a deeper domain knowledge and interdisciplinary approaches are needed once the task goes beyond language comprehension. After evaluating ChatGPT for causal reasoning, Gaoetal [6] discovered that it was prone to hallucinations. Furthermore, they noticed that ChatGPT is more appropriate at identifying explicit causal relationships than implicit ones. Their capabilities have been further enhanced by the integration of knowledge graphs with machine learning and artificial intelligence technologies, making them essential tools for intelligent systems and data-driven decision-making [8].

A case study was proposed by [8] who studied Alzheimer's disease and proposed a method to predict novel disease associated genes. Semantic similarity measures between the biomedical entities were calculated which then could be used to construct similarity between the network of diseases in a similar manner. Ontologies in medical field have been developed for supporting various areas in the field of medicine [9]. Ontologies and other terminological reserves have emerged for the retrieval of information to provide expansion of queries [10]. Yunzhi C et.al [11] proposed a methodology focusing on the construction of a Hepatitis Ontology and querying the same. It aims on providing a framework for Ontology based healthcare services. The work designed and proposed an algorithm for query expansion for Hepatitis Ontology. The query expansion included expansion of synonym, Hypernym and hyponym as well as the expansion of related words. The semantic similarity is being calculated to assess the similarity between the terms being retrieved. A Bayesian approach for construction of probabilistic models from knowledge graphs is discussed by Freedman et.al[12] in which an extension of SPARQL query language called as Orion DSL for the retrieval of computed probabilistic distributions against a base knowledge graph.

Computation of semantic similarity in a graph-based knowledge base is a challenging and an important issue. Majority of existing Machine-Human interaction systems with Natural language uses Ontologies for semantic clarification of the requests [13]. A new semantic similarity measure for Ontologies was proposed [14], which considers the object properties among the concepts. The method makes use of the patterns used by Hirst & St-Onge [15] for calculation of paths that are semantically correct. Runumi et al [16] proposed an Ontology design for patients suffering from dengue and used two approaches for generating RDF model from the unstructured records of the patients. The patterns were generated using parsing [17] of the sentences from the prepared case sheets of the patients. On comparing different approaches, the phrase structure based parsing model generation approach was found to be preferable considering the generation time of the model. Fareh M et al [18] presented an approach for merging OWL ontologies by enriching the initial ontologies semantically. The enrichment is done by a set of metadata which interprets their conceptions with homonyms and synonyms for each term with the usage of WordNet thereby generating a wordbook for each ontology to create a global one. The method is focussed upon computation of semantic relatedness between the concepts of ontologies, resulting in generation of a merged ontology using various techniques mentioned in their work. Taking all those research studies into consideration and adding furthermore to the work, two diseases have been taken into consideration which are to be modelled for Ontology as explained in the sections below.

3. Data Collection

The issue of structuring of unstructured and semi structured data in case of interoperability of data in healthcare domain as well as other fields of research is not clear in any of the research work. The aim of our work is to integrate unstructured and semi structured data from pathological reports and patients' summary of CT reports respectively and create a knowledge base of this data in a structured and well-formed manner. The structuring of data [19] is done using semantic techniques [20].

The data has been collected from Sher-I-Kashmir institute of Medical sciences, Jammu and Kashmir and government medical college Srinagar under the supervision of doctors who are treating patients suffering from Hepatitis virus disease and dyslipidaemia. Few reports have been taken from a local laboratory in Srinagar. The pathological reports of patients suffering from hepatitis virus disease and lipid profile reports of patients suffering from dyslipidaemia have been taken as unstructured format and the CT reports have been taken in textual format of patients suffering from hepatitis virus disease as well as Dyslipidaemia. The data is composed of 500 pathological reports, 240 lipid profile reports and 90 CT reports that have been procured under the supervision of a gastroenterologist and radiologist working in the above-mentioned hospitals. The data consists

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

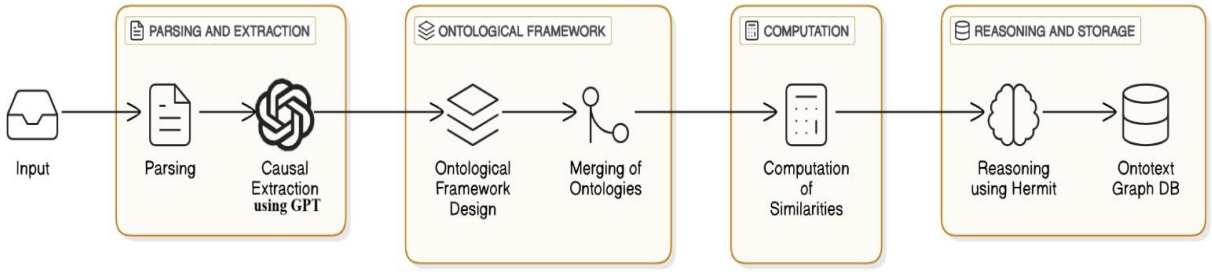
Research Article

of the reports of both patients suffering from hepatitis virus and dyslipidemia as well as patients having negative tests for the same diseases. The pathological report of a patient tested for Hepatitis virus disease is written and maintained by the gastroenterologist. This is a semi-structured text and is converted into tabular data using NLP [21] techniques. Patients are also tested for their lipid profiles Our approach also works upon the patient summary report prepared by the doctor after referring to the patient history, symptoms and the CT report given by the radiologist. A sample of the case sheet is shown in figure2 . It is an unstructured text which is parsed and converted into structured text.
This paper highlights the engineering methodologies and the reasoning rules for the treatment, design, organization, and interrogation of various Pathological and CT reports of patients suffering from hepatitis B, C virus and Dyslipidaemia.

4. Information Extraction

With the help of the data or the patient information extracted from the pathological reports, lipid profile reports and the Patient summary, a knowledge base is created. The approach is depicted clearly in figure 1 and is carried out in various stages as mentioned below:
a) Input is fetched.
b) Extraction using Parsing.
c) fined-tuned GPT-3.5 Turbo to extract causal relations by providing annotated abstracts
d) Ontological framework design
e) Merging of Ontologies
f) Computation of similarities
g) Reasoning using Hermit and export all inferred knowledge into a KG managed within the Ontotext Graph DB tool

Figure 1. System Architecture



However, knowledge bases are not easily integrated with the relational databases. To overcome this constraint, RDF[22] model is used instead, that can be easily modified and adapted according to the information needs. RDF
allows the declaration of classes, subclasses, defines the domains and ranges and thus integration of knowledge-based reasoning.

- **RDF model generation through Parsing**

The structural ambiguities from the unstructured documents can be resolved formally with the help of parsing [23]. As a case study, we consider the patients' summary report prepared and provided by a doctor that is shown in figure2. It represents a case sheet for a patient tested positive for hepatitis C virus disease, inputs of which are collected from the patient summary, pathological report and the CT report provided by the consultant doctor. The syntactic structure of a sentence is represented with the help of a phrase and dependency structure. The method for the extraction of triplets for generation of RDF model is based on rules and patterns that are applied to Constituent and dependency tree structure. Both of which are represented by a directed acyclic graph G=(V,E) of a set of vertices and edges,Where edge e E represents the dependency relation and Vertex v V represents a word.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

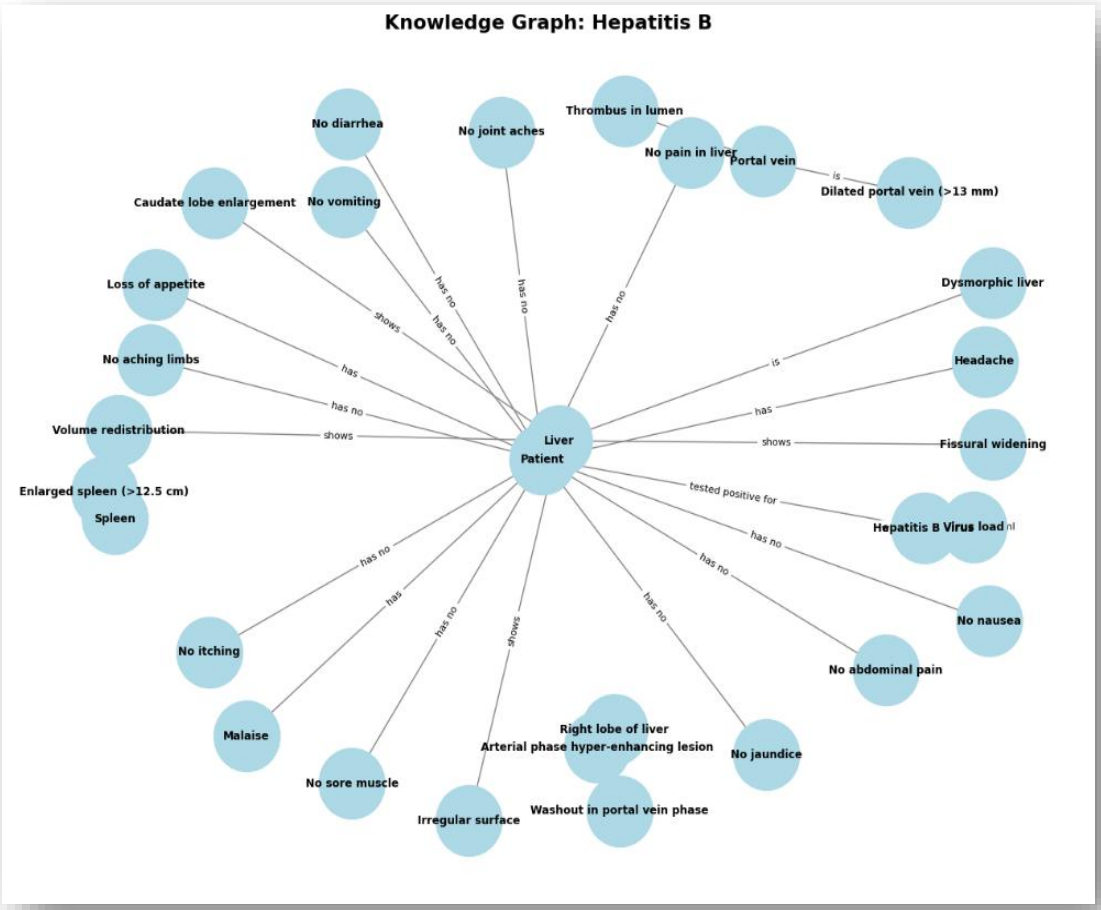
Figure 2. Sample case sheet from Pathology and CT report.

Sample case sheet for P16- *inputs from pathology and CT report*

Patient has fatigue. Patient has Headache and Malaise. Patient has loss of appetite. Patient has no pain in liver. Patient has no Diarrhoea. Patient has no Nausea. Patient has no sore muscle. Patient has no Vomiting. Patient has no Jaundice. Patient has no abdominal pain. Patient has no itching. Patient has no joint aches. Patient has no aching limbs. Patient was tested for Hepatitis Virus. Patient is tested positive for hepatitis B virus. Patient has load 1.1* 10'8 IU/mL.
Liver is dysmorphic. Liver shows irregular surface. Liver shows fissural widening. Liver shows volume redistribution. Liver shows caudate lobe enlargement. Liver shows small size. Liver shows heterogeneous density pattern. Arterial phase hyper enhancing lesion is seen in right lobe of liver. This lesion shows washout in portal vein phase. Portal vein is dilated (>13mm) with a thrombus in lumen. Spleen is enlarged (>12.5cm in size)

In a dependency tree [24], an extra dummy node is defined called as the ROOT node but is not a part of the sentence. The sentence from the above given case sheet is parsed and presented by a knowledge graph shown in figure 3.

Figure 3. Knowledge graph for the parsed case sheet.



Based on the structural representation of sentences, the documents are parsed in two ways: one is focussed upon the relationship between the words and the other on the identification of phrases along with their recursive structure. Each expression is comprised of three terms. The terms consist of an index that represents its position.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The traversal of trees is done in Breadth First order (BFS) and depends upon the syntax of the structure of the declarative sentences of the reports/documents. Each sentence written in the case sheet is fetch as input to the TypedDependency tree [24]. The algorithm to find the subject, object, and the predicate in the sentence of the case sheet is defined [24]. An RDF triplet is formed once the subject, object and the predicate are cached and are included as one instance of the RDF model. As defined in the algorithm, the patient-id is represented by the subject and the statement defines the fact of a patient.

An approach for RDF model generation

To build an RDF model by using facts that are stored in the case sheet documents as shown in the section above, a database of rules is to be created that consists of relation/phrase patterns that are extracted from the documents based on the case sheets. The unstructured documents of patients are transformed to RDF models by extraction of triplets that is performed in two phases. The first phase is focused upon the extraction of patterns using parsing of documents. The second phase is focused upon the generation of an RDF model for each document based on identification of a pattern while extraction during parsing of the document.. An RDF model is generated by extracting the triplets which consist of a subject, a predicate, and an object with the help of matching the structure of a sentence with the pattern that has been identified along with the triplets that are included in the RDF model. Once the sentences or the text is been subjected to parsing and RDF development, furthermore the summary reports are subjected to fine tuning using GPT 3.5 turbo.

5. Fine-tuning GPT-3.5 Turbo to extract causal relations by providing annotated abstracts:

utilising the fine-tuned GPT to extract causal relationships from medical abstracts and integrating these relationships into a knowledge graph. **There will also be examples of reasoning with Hermit[] and querying the enriched ontology in natural language based on Langchain.** Examine case sheet given above. When we refined the GPT model, this abstract was utilised in the validation dataset and annotated with five relationships.

Figure 4. Extracted relations

- **Output (Completion):**

```yaml
1. Cause: Hepatitis B virus (viral load 1.1 * 10^8 IU/mL) → Effect: Dysmorphic liver wit
2. Cause: Hepatitis B virus (viral load 1.1 * 10^8 IU/mL) → Effect: Enlarged spleen (>12
3. Cause: Arterial-phase hyper-enhancing lesion → Effect: Washout in portal vein phase.
4. Cause: Portal vein thrombus → Effect: Dilated portal vein (>13 mm).
5. Cause: Triglycerides at 120 mg/dL → Effect: Increased risk of dyslipidemia.
```

***Extracting relations*.**The model extracted the relations shown in figure 4. Extractions of the first, second, and third were done appropriately. We noticed that the mistakes are connected to the presence of abbreviations, even though the listed entities were correctly detected.The knowledge graph incorporates the same to address this. as connections among nodes. Therefore, the knowledge network fills up these gaps even when the model ignores these relationships. This suggests that employing LLMs may have the benefit of revealing more pertinent data that was overlooked during manual annotation. The extracted relations/entities are subjected to ontological development for interoperability of the system.
The patient case sheet is also subjected to various LLM models likeGPT, BART- Base, BART-Large, T5-Base and T-5 Large in order to compare the enhancement of the data shown in the table below:

Table 1. Generated examples for GPT, BART - Large, BART - Base, T5 - Large, and T5 - Base.

Model	Generated Explanation
Input Sentence-1	**patient has headache and malaise. patient has loss of appetite. patient has no pain in liver.**
GPT	"The presence of headache and malaise suggests a systemic condition, possibly due to an underlying infection, inflammation, or metabolic disturbance. Loss of appetite can be associated with a wide range of conditions, including infections, liver disease, metabolic disorders, or psychological factors. In the context of a Hepatitis B infection, this symptom

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	may indicate liver dysfunction, systemic inflammation, or an early sign of hepatic impairment. The absence of liver pain suggests that there may not be significant acute inflammation or stretching of the liver capsule. Imaging studies and liver function tests may still be necessary for a comprehensive assessment."
BART-Large	"The patient exhibits headache, malaise, and loss of appetite but does not report liver pain."
BART-Base	"A patient presents with headache, malaise, and loss of appetite but denies liver pain."
T5-Large	"A clinical case with symptoms of headache and malaise, along with reduced appetite, but without liver pain."
T5-Base	"Patient experiences headache and malaise but does not complain of liver pain."
Input Sentence-2	**patient has no aching limbs. patient has tested positive for hepatitis B virus with a load of 1.1 * 10'8 IU/ml.**
GPT	The patient does not report any limb pain, which suggests the absence of musculoskeletal or neuropathic conditions. This could indicate(explanation)…….. . "The patient has been diagnosed with an active Hepatitis B virus infection, as indicated by a high viral load of 1.1×10^8 IU/ml. This suggests significant viral replication, which may increase the risk of liver inflammation, fibrosis, or progression to chronic hepatitis. Close monitoring of liver function, antiviral therapy, and further diagnostic testing (such as liver biopsy or imaging) may be required to assess the disease stage and determine an appropriate treatment plan."
BART-Large	"The patient has tested positive for Hepatitis B virus with a high viral load of 1.1×10^8 IU/ml, indicating a significant infection that may require medical intervention."
BART-Base	"The patient is diagnosed with Hepatitis B, with a viral load of 1.1×10^8 IU/ml, suggesting an active infection."
T5-Large	"A patient has tested positive for Hepatitis B virus with a viral load of 1.1×10^8 IU/ml, which is indicative of an ongoing infection and potential liver damage."
T5-Base	"The patient has Hepatitis B with a viral load of 1.1×10^8 IU/ml, suggesting an active viral replication."
Input Sentence- 3	**Liver is dysmorphic. liver shows irregular surface**
GPT	"The liver exhibits structural abnormalities, characterized by dysmorphism and an irregular surface. These(explanation) ……. Further clinical evaluation, imaging studies, and liver function tests may be required to determine the underlying cause and severity of the condition."
BART-Large	"The liver appears structurally abnormal, showing signs of dysmorphism and an irregular surface, which may indicate underlying liver disease such as cirrhosis or fibrosis."
BART-Base	"The liver has an irregular structure, suggesting possible liver disease or chronic damage."
T5-Large	"The patient's liver is dysmorphic with an irregular surface, which could be a sign of progressive liver disease, including fibrosis or cirrhosis."
T5-Base	"The liver shows structural abnormalities and an irregular surface, which may indicate liver disease."

GPT model is superior to all models as it provides medical reasoning behind symptoms, suggests differential diagnoses and possible causes, recommends further tests and next steps and are more clinically relevant for decision-making**.** Compared to T5, BART-large can produce information that is more legible, useful, and concise. The performance of base models was lower than that of BART and T5 large. Bart-large explanations are generally more informative, readable, and well-explained. T5-large outputs, which are generally more medically informative and comprehensible, were equally good. [43].

6.Ontology development:

The voluminous mass of the pathological reports along with their unstructured data formats might lead to ambiguity and complexity in clinical data. To reduce the complexity, HepOnt Ontology is created. HepOnt Ontology consists of the main views of the pathological reports of the patient, their characteristics, and the relationships between them. HepOnt [25] is formed based on the Pathological reports. The ontology that has been proposed is defined according to the structure or format of the pathological reports. The material used for the reports is the blood sample of the patient. The protocol refers to the method or the type of test uses to detect

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the presence of absence of hepatitis B,C virus in the patient and the two tests carries out are Aggleutenation[26] and ELISA[27](Enzyme Linked Immuno Sorbent Assay. The subclass Hepatitis Virus refers to the type of Virus defined by a subclass Type. The subclass Type refers to the type of the Hepatitis Virus which can be either DNA based hepatitis virus[28] or RNA based hepatitis virus[29] defined by subclasses DNA_Based and RNA_based respectively. The serology is the outcome of the test i.e whether the patient is Hep B, C positive or negative.
In the similar manner an ontology for patient suffering or to be diagnosed of Dyslipidaemia is defined and developed. The Ontology is named as DysLipOnt Ontology.

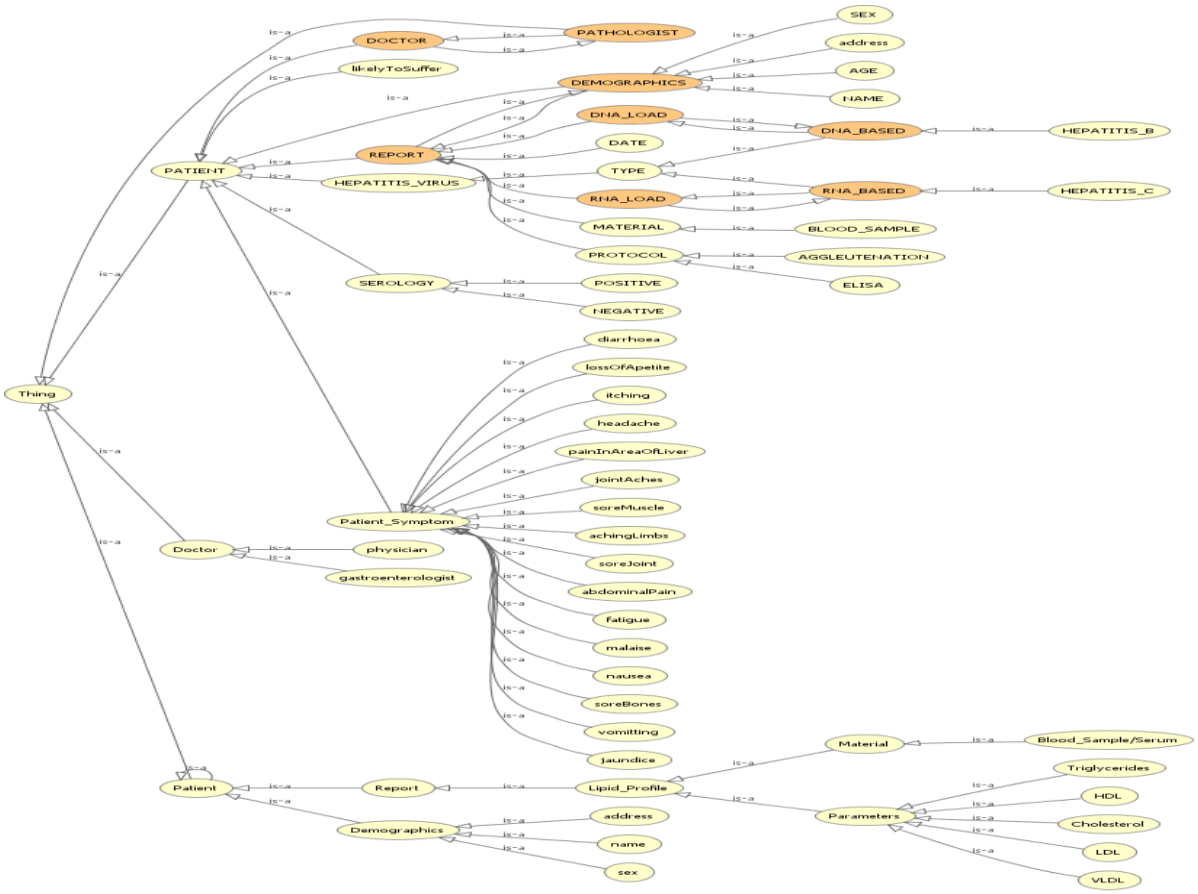
7.Merging Ontologies

The diseases that we have taken for the study are Hepatitis virus disease (Hepatitis B, Hepatitis C) and Dyslipidaemia. The criteria for selection of the diseases is the relationship [30][31] that exists between the two diseases. A patient suffering from Hepatitis B or Hepatitis C or both, is likely to have deranged lipid profile, hence Dyslipidaemia. The above-mentioned diseases have a cause-effect relationship and this causality became the basis of selection of those diseases. The data from the pathological reports is the semi-structured data. Unstructured data for the patients suffering from Hepatitis virus disease is also collected from the patient summary report and the CT report provided by the doctor which is also populated to the Ontological framework HepOnt. The knowledge base for the patients suffering from Dyslipidaemia is created by construction of an Ontological model DysLipOnt with the help of the patient data collected from the Lipid profile reports of the patients being tested for the same.
Since the two diseases have a causal relationship, the two ontologies constructed are merged [32] together to form a bigger ontology named as HepDysLipOnt that consists of both the knowledge bases i.e: records of patients suffering from Hepatitis virus disease as well as Dyslipidaemia. The ontologies are merged using Protégé 4.3. The application has a built-in plugin PROMPT[33] that enables the merger of the ontologies automatically once the plugin is started and commanded for the merger. The PROMPT has a built-in algorithm that requires two ontologies as input and returns a third ontology as an output which is the merged ontology. The merger algorithm is a semi- automated algorithm. It takes into consideration the structure of the ontology and not how the relations among the concepts are treated. The merged Ontology HepDysLipOnt is shown in figure 5 .

Figure 5. Merged Ontology- HepDysLip

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



Similarity measure

Quantification of similarities [34] between various concepts in an Ontology is measured as Semantic similarity. Prospective applications for those measures encompass search, mining of data and knowledge discovery in the knowledge base utilizing those ontologies. Semantic similarity reckons the alikeness between the words that are comprehended as the degree of taxonomical closeness. HepOnt taken as a case, Hepatitis B and Hepatitis C are similar because both are Hepatitis Virus diseases.

The Knowledge base is specifically modelled in a machine- readable format that is used for formalization of concepts with the use of a familiar nomenclature and is used to represent taxonomic and non-taxonomic associations through semantic links. In such a case, the resemblance or similarity between two concepts in measured by computing the inter-link distance [35]. The simplest manner in which the distance between two concepts c1 and c2 is measured is by computing the shortest path length (PL, which is the minimum number of links between those concepts c1, c2) connecting those concepts.

$dis_{PL}(c1, c2)$ = minimum number of edges connecting c1 and c2. (1)

Wu and Palmer [36] [37] (*W&P*) proposed a path-based measure shown in equation 2 that takes into account the depth of the concepts in the hierarchy.

$$Sim_{W\&P}(c1, c2) = \frac{2*N3}{N1+N2+2*N3} \quad (2)$$

where N_1 and N_2 are the number of *is-a* links from c_1 and c_2 respectively, to their least common subsume (LCS) N_3 is the number of *is-a* links from the LCS to the root of the ontology. It ranges from 1 (for identical concepts) to 0 (for non-identical concepts). Applying equation 2 to the subgraph given above in figure, the semantic similarity or the relatedness among Hepatitis B and Hepatitis C is computed and shown in figure as:

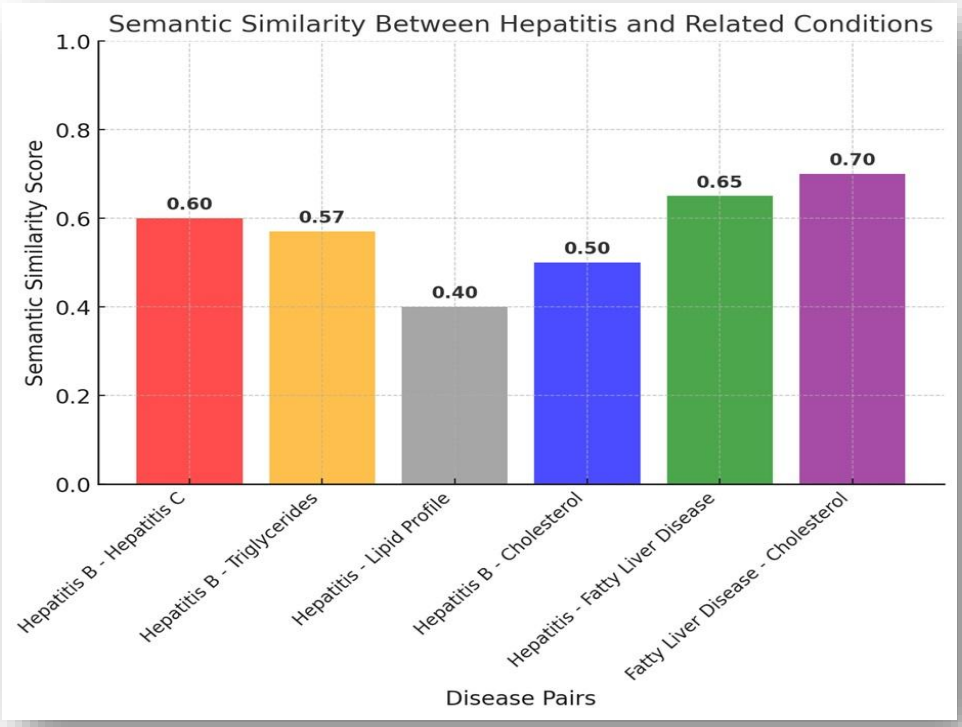
$Sim_{W\&P}$ (Hepatitis B, Hepatitis C) = 0.6

Semantic similarity of 0.6 indicates that the two viral diseases are 60% similar as they belong to a common disease Hepatitis. Similarly, $Sim_{W\&P}$ (Hepatitis B, Triglycerides) = 0.57 indicating 57% similarity between Hepatitis B and Triglycerides and $Sim_{W\&P}$(Hepatitis, lipid profile) = 0.4 indicating 40% similarity between Hepatitis virus and Lipid profile.

Figure 6. Semantic similarity between Hepatitis, Dyslipidaemia and other related conditions.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



7.Ontology enrichment

One of the most important procedures in Ontology integration [38] and ontology mining[39] is merging of Ontologies, which is useful for unification, extraction, or creation of semantic content from various ontologies that may or may not be defined on a similar domain. Merging ontologies based on alignment can lead to an enriched ontology with the knowledge bases of both the input ontologies. A strongly merged ontology comprises of all the concepts belonging to the individual ontologies along with their associated relationships.
The base metrics of the Hepatitis ontology and Dyslipidaemia Ontology were calculated [40]and then compared with those of the merged ontology HepDysLipOnt Ontology. the base metrics include the number of classes, individuals, data properties, object properties, logical axioms and total axioms as shown in table 2. Table 3 shows the base metrics for the ontology in comparison to the metrics after using Chatgpt for enriching the Ontology. it is clearly visible that there is an enrichment of the ontologies by 30% as the ontology after ChatGPT has a richer knowledge representation, better causal modeling, improved relationships, and integration of medical guidelines.
For example: usage of ChatGPT has enhanced the expansion of total axioms.
Before using Chat GPT: "Hepatitis B affects the liver."
After using ChatGPT: "Hepatitis B affects the liver, leading to fibrosis, cirrhosis, or hepatocellular carcinoma."

Table 2. Base metrics for Ontologies

Ontology	No. of classes	No. of logical axioms	No. of indivi Duals	Data prope rties	Object Prope rties	Total axioms
HepOnt	45	990	258	7	10	1197
DysLipOnt	19	126	120	9	9	179
HepDysLip Ont	63	1115	378	16	19	1375

Table 3. Base metrics for Ontologies after using ChatGPT

Ontology	No. of classes	No. of	No. of indivi	Data prope	Object Prope	Total axioms	Total

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		logical axioms	Duals	rties	rties		Axioms (after ChatGPT)
HepOnt	45	990	258	7	10	1197	1556
DysLipOnt	19	126	120	9	9	179	233
HepDysLipOnt	63	1115	378	16	19	1375	1788

Also, a number of graph metrics were used to evaluate and compare the quality of individual ontologies and the merged ontology. the graph metrics include the Horizontal and vertical measures [41] which are Absolute depth, average depth, maximal depth, Absolute breadth, average breadth, maximal breadth and tangledness as shown in table 4. The absolute depth indicates the depth of the relations in the ontology. absolute breadth indicates the average number of children in the ontology.

Table 4. Graph metrics for Ontology.

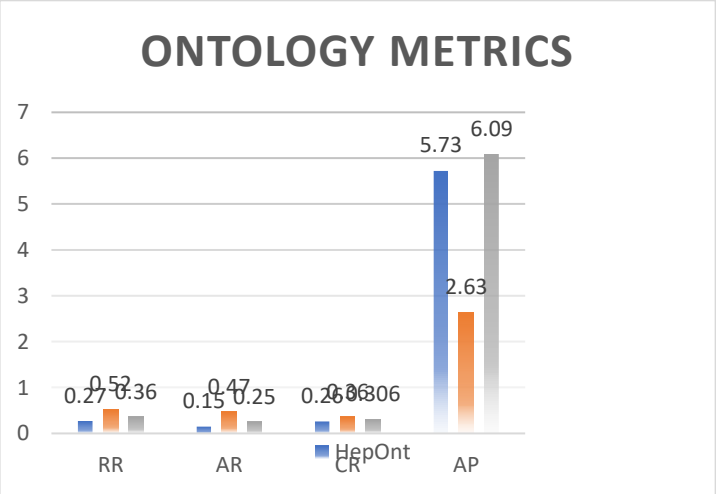
Graph metric	***HepOnt***	***DysLipOnt***	***HepDysLipOnt***
Absolute depth	99	42	110
Average depth	19.8	8.4	22
Maximal depth	5	5	5
Absolute breadth	44	18	62
Average breadth	2.75	3.6	3.875
Maximal breadth	16	5	16
Tangledness	1.24	1.28	1.125

fig.8 represents the Strong merged Ontology HepDysLip ontology.
The ontology is evaluated by computing various metrics [42] such as Relationship richness (RR), Attribute richness (AR), Class richness (CR) and Average Population (AP) for the individual ontologies and the merged ontology is shown in fig.10 given below. RR measures the number of relationships in the ontology and is computed as RR= Prop/ Sub-class + Prop; where Prop is the data properties and object properties. AR measures the average number of attributes in a class and is measured as AR= Attribute/ class; where attribute is the number of data properties. CR is defined as the knowledge metric as it signifies the real-world entities represented by the ontological model. It's the ratio of number of classes and total number of classes and is computed as CR= class $_{\text{with-instance}}$/ class. AR defines the number of population per class and is computed as AP= Individual/class.
It is clear that after merging the base ontology HepOnt and DysLipOnt Ontology, the merged ontology HepDysLipOnt Ontology has better and improved metric scores than that of the base ontology as the metrics have improved positively after merging.

Figure 7.Ontology metrics

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

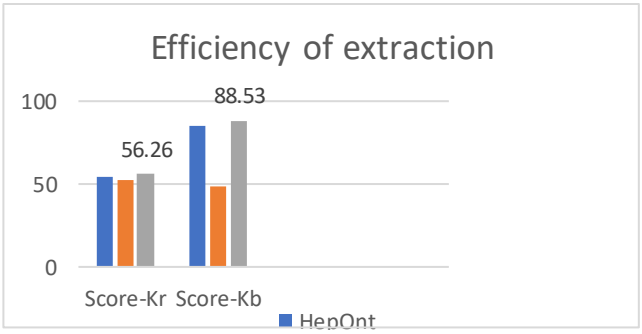


Furthermore, the work has also been evaluated based on the knowledge it represents, that reflects the relationships and attributes of the knowledge. It is represented by Score-Kr which is calculated as:

$$Score_{Kr} = \frac{((|Rel|*|Class|*100)+(Subclass|+|Rel|*|Prop|)}{(|Subclass|+|Rel|*|Class|)} \quad (3)$$

Also, the efficiency for extraction of base knowledge can be measured and defined as Score-Kb which is calculated as: $Score_{Kb} = \frac{((Class_{withInstance}*100)+Individual)}{|Class|}$ (4)

Figure 8. Efficiency of extraction.



Thus, the efficiency for extraction of base knowledge measured as Score-Kb has increased after merging the Hepatitis Ontology with score 85.73 and Dyslipidaemia Ontology leading to an enriched ontology with score 88.73.

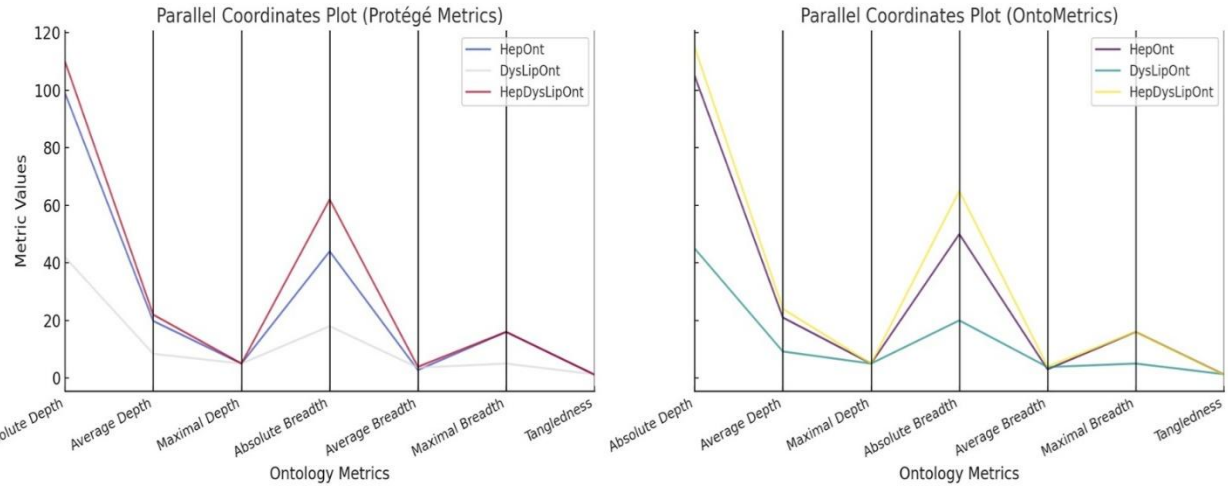
Comparison of Base Metrics Obtained with Existing Tools:

In addition to Protégé, another tool namely OntoMetrics, is used to verify the accuracy of the fundamental metrics of the output ontology developed. Following the merging procedure, the base metrics for the output ontologies were generated using both tools.These metrics are shown in the figure 9. It is discovered that there is a significant association between these parameters. This demonstrates the quality and accuracy of the suggested algorithm**.**

Figure 9. Base metrics of the output ontologies in OntoMetrics and Protégé.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



Conclusion

In this paper, a methodology for storing and accessing patient records suffering from Hepatitis virus and Dyslipidemia is designed by constructing Ontological frameworks for the same. The patient data is been collected from the pathological reports, CT reports and the patient summary report consisting of semi-structured and unstructured data. The data is collected and analyzed for parsing and structuring leading to formation of RDF framework. The data then is gathered on a single platform by constructing Ontologies named HepOnt and DysLipOnt for patients having Hepatitis virus disease and Dyslipidemia respectively. The Ontologies are further merged to combine all the patients' data in a single ontological framework HepDysLipOnt Ontology. the data is subjected to SWRL rules and can be queried by SPARQL queries for accessing the patient information. The model is validated by evaluation of semantic similarity measures followed by Fine-tuning GPT-3.5 Turbo to extract causal relations by providing annotated abstracts.
Reasoning is done using Hermit. Because of reasoning, the original Causal HepDysLipOnt ontology had 1217 logical axioms instead of 1115, 72 class assertions instead of 63, and 22 object property assertions instead of 19. Compared to the annotated relations (180), the GPT model extracted a greater number of relations (243).This suggests that a greater amount of relational data can be extracted from the abstractions using the GPT model. Additionally, the model's ability to identify different textual items is more extensive. We manually compared the extracted relations with the annotated relations in order to evaluate the accuracy of the GPT-extracted relations. The GPT model showed a high level of accuracy in determining the most prevalent relation categories, including "cause," and "affect". We compute the precision, which is defined as precision=Number of Correctly Identified Relations to the Total Number of Relations Extracted by GPT=0.782, in order to measure the precision. Comparing the GPT model's extracted relations to human annotations, this shows that about 78.2% of them were accurate over human annotations which deliver the accuracy of 70%. So, there is an improvement of 8.2% after using GPT 3.5 for extracting causal relations.
Furthermore, the ontology is enriched by computing the Ontology metrics and efficiency of extraction for the individual ontologies and the merged ontology. The results depict the merged ontology to be better in performance of metric evaluation and efficiency of information extraction. In future, the Ontology can be integrated with other domain ontologies. Also, the model is expected to define an association between the two diseases and explain the cause effect the relationship between the diseases.

Declaration:
Availability of data and material: The data is available with all authors.

Competing interests: The authors declare that they have no competing interests.

Funding: No funding was required for this study.

Authors' contributions: B.A, N.R, M.D, B.M conceptualized the study. B.A collected the data. B.A, N.R, M.D, B.M analysed the data. B.A wrote the manuscript which was edited by N.R, M.D, B.M. All authors have read and approved the manuscript and ensure that this is the case.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Acknowledgements: The authors are grateful to Department of Gastroenterology, Sher I Kashmir Institute of Medical Sciences Soura Srinagar and Government Medical College Srinagar for their valuable support

References

[1] de Mello, B. H., Rigo, S. J., da Costa, C. A., da Rosa Righi, R., Donida, B., Bez, M. R., & Schunke, L. C. (2022). Semantic interoperability in health records standards: a systematic literature review. *Health and technology*, *12*(2), 255-272.

[2] de Souza, P. L. D. L., de Souza, W. L. D. L., & Ciferri, R. R. (2022, May). Semantic interoperability in the Internet of things: A systematic literature review. In *ITNG 2022 19th International Conference on Information Technology-New Generations, S. Latifi, Ed. Cham: Springer International Publishing* (pp. 333-340).

[3] Gansel, X., Mary, M., & van Belkum, A. (2019). Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review. *European Journal of Clinical Microbiology & Infectious Diseases*, *38*, 1023-1034.

[4] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. (2011). Semantically enhanced information retrieval: An ontology-based approach. *Journal of Web Semantics*, *9*(4), 434-452.

[5] Ben Lamine, S. B. A., Dachraoui, M. A., & Baazaoui-Zghal, H. (2023). Deep learning-based extraction of concepts: A comparative study and application on medical data. *Journal of Information & Knowledge Management*, *22*(04), 2250072.

[6] Yuvaraj, D., Alnuaimi, S. S., Rasheed, B. H., Sivaram, M., & Porkodi, V. (2024). Ontology Based Semantic Enrichment for Improved Information Retrieval Model. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(15s), 70-7.

[7] Dai, L., Liu, B., Xia, Y., & Wu, S. (2008, August). Measuring semantic similarity between words using HowNet. In *2008 International Conference on Computer Science and Information Technology* (pp. 601-605). IEEE.

[8] Le, D. H., & Dang, V. T. (2016). Ontology-based disease similarity network for disease gene prediction. *Vietnam Journal of Computer Science*, *3*(3), 197-205.

[9] Ivanović, M., & Budimac, Z. (2014). An overview of ontologies and data resources in medical domains. *Expert Systems with Applications*, *41*(11), 5158-5166.

[10] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. (2011). Semantically enhanced information retrieval: An ontology-based approach. *Journal of Web Semantics*, *9*(4), 434-452.

[11] Yunzhi, C., Huijuan, L., Shapiro, L., Travillian, R. S., & Lanjuan, L. (2016). An approach to semantic query expansion system based on Hepatitis ontology. *Journal of Biological Research-Thessaloniki*, *23*, 11-22.

[12] Freedman, H., Metzger, J., Abolhassani, N., Tudor, A., Tomlinson, B., & Paul, S. (2024). A Bayesian Approach to Constructing Probabilistic Models from Knowledge Graphs. *International Journal of Semantic Computing*, *18*(1).

[13] Yahya, M. (2024). Building semantic models and knowledge graphs for intelligent smart manufacturing applications.

[14] Taieb, M. A. H., Aouicha, M. B., & Hamadou, A. B. (2014). Ontology-based approach for measuring semantic similarity. *Engineering Applications of Artificial Intelligence*, *36*, 238-261.

[15] Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert systems with applications*, *39*(9), 7718-7728.

[16] Devi, R., Mehrotra, D., Zghal, H. B., & Besbes, G. (2020). SWRL reasoning on ontology-based clinical dengue knowledge base. *International Journal of Metadata, Semantics and Ontologies*, *14*(1), 39-53.

[17] Devi, R., Mehrotra, D., Lamine, S. B. A. B., & Zghal, H. B. (2022). Constituent vs Dependency Parsing-Based RDF Model Generation from Dengue Patients' Case Sheets. *Journal of Information & Knowledge Management*, *21*(01), 2250013.

[18] Fareh, M., Boussaid, O., Chalal, R., Mezzi, M., & Nadji, K. (2013). Merging ontology by semantic enrichment and combining similarity measures. *International Journal of Metadata, Semantics and Ontologies*, *8*(1), 65-74.

[19] Hayman, J. A., Dekker, A., Feng, M., Keole, S. R., McNutt, T. R., Machtay, M., ... & James, B. Y. (2019). Minimum data elements for radiation oncology: An American Society for Radiation Oncology consensus paper. *Practical radiation oncology*, *9*(6), 395-401.

[20] Angula, N., & Dlodlo, N. (2018). Towards a framework to enable semantic interoperability of data in heterogeneous health information systems in Namibian public hospitals. In *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)* (pp. 835-845). Springer International Publishing.

[21] Maynard, D., Li, Y., & Peters, W. (2008). NLP Techniques for Term Extraction and Ontology Population.

[22] Buron, M., Goasdoué, F., Manolescu, I., & Mugnier, M. L. (2020, March). Ontology-based RDF integration of heterogeneous data. In *EDBT 2020-23rd International Conference on Extending Database Technology* (pp. 299-310).

[23] Lefrançois, M., Zimmermann, A., & Bakerally, N. (2017). A SPARQL extension for generating RDF from heterogeneous formats. In *The Semantic Web: 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28–June 1, 2017, Proceedings, Part I 14* (pp. 35-50). Springer International Publishing.

[24] De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006, May). Generating typed dependency parses from phrase structure parses. In *Lrec* (Vol. 6, pp. 449-454).

[25] Petrov, A., Popov, A., Chekardovsky, M., & Pushkarev, A. (2021). Methodology of application of open-source platform Protégé in the measurement and computing systems development for diagnostics of heat supply networks. In *CEUR Workshop Proceedings.-2021* (Vol. 2843, No. 20, p. 2021).

[26] Oseni Okolo, M. L., Omatola, C. A., Samson, S. O., & Idache, B. M. (2022). Evidence of hepatitis B infection and co-infection with enteric fever among febrile patients in a primary health facility in Kogi State, Nigeria. *Journal of Immunoassay and Immunochemistry*, *43*(5), 516-525.

[27] Wu, Y., Zeng, L., Xiong, Y., Leng, Y., Wang, H., & Xiong, Y. (2018). Fluorescence ELISA based on glucose oxidase-mediated fluorescence quenching of quantum dots for highly sensitive detection of Hepatitis B. *Talanta*, *181*, 258-264.

[28] Tsounis, E. P., Tourkochristou, E., Mouzaki, A., & Triantos, C. (2021). Toward a new era of hepatitis B virus therapeutics: The pursuit of a functional cure. *World journal of gastroenterology*, *27*(21), 2727.

[29] Zhang, Q., Huang, H., Sun, A., Liu, C., Wang, Z., Shi, F., ... & Zhang, Y. (2022). Change of cytokines in chronic hepatitis B patients and HBeAg are positively correlated with HBV RNA, based on real-world study. *Journal of Clinical and Translational Hepatology*, *10*(3), 390.

[30] Serfaty, L. (2017). Metabolic manifestations of hepatitis C virus: diabetes mellitus, dyslipidemia. *Clinics in Liver Disease*, *21*(3), 475-486.

[31] Joo, E. J., Chang, Y., Yeom, J. S., Cho, Y. K., & Ryu, S. (2019). Chronic hepatitis B virus infection and risk of dyslipidaemia: A cohort study. *Journal of Viral Hepatitis*, *26*(1), 162-169.

[32] Osman, I., Yahia, S. B., & Diallo, G. (2021). Ontology integration: approaches and challenging issues. *Information Fusion*, *71*, 38-63.

[33] Zailan Arabee Abdul Salam, Rabiah Abdul Kadir, & Azreen Azman. (2021). ONTOLOGY MERGING USING PROTÉGÉ – A CASE STUDY. *JOURNAL INFORMATION AND TECHNOLOGY MANAGEMENT (JISTM)*, *6*(22), 148–157

[34] Zhao, Y., Wang, J., Chen, J., Zhang, X., Guo, M., & Yu, G. (2020). A literature review of gene function prediction by modeling gene ontology. *Frontiers in genetics*, *11*, 400.

[35] Oussalah, M., & Mohamed, M. (2022). Knowledge-based sentence semantic similarity: algebraical properties. *Progress in Artificial Intelligence*, *11*(1), 43-63.

[36] Iana, A., Alam, M., & Paulheim, H. (2024). A survey on knowledge-aware news recommender systems. *Semantic Web*, (Preprint), 1-62.

[37] Kaddar, L. B., & Ben-Naoum, F. (2022). Novel approach for semantic similarity cross ontology. *Indonesian Journal of Electrical Engineering and Computer Science*, *26*(1), 493Ű504.

[38] Osman, I., Yahia, S. B., & Diallo, G. (2021). Ontology integration: approaches and challenging issues. *Information Fusion*, *71*, 38-63.

[39] Belhadi, H., Akli-Astouati, K., Djenouri, Y., & Lin, J. C. W. (2020). Data mining-based approach for ontology matching problem. *Applied Intelligence*, *50*(4), 1204-1221.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[40] Rudwan, M. S. M., & Fonou-Dombeu, J. V. (2024). A Novel Algorithm for Multi-Criteria Ontology Merging through Iterative Update of RDF Graph. *Big Data and Cognitive Computing*, *8*(3), 19.
[41] Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2005). Ontology evaluation and validation: an integrated formal model for the quality diagnostic task. *On-line: http://www. loa-cnr. it/Files/OntoEval4OntoDev_Final. pdf*.
[42] Ben Mahria, B., Chaker, I., & Zahi, A. (2021). A novel approach for learning ontology from relational database: from the construction to the evaluation. *Journal of Big Data*, *8*(1), 25.
[43] Latif A, Kim J. Evaluation and analysis of large language models for clinical text augmentation and generation. IEEE Access. 2024 Apr 3.