Text-to-OWL: Automated Ontology Construction for Tuberculosis Treatment Recommendation using Generative AI

Zonabo Ouédraogo[1,2], Lydie Simone Tapsoba[2], Aminata Sabane[1,2],
Rodrique Kafando[1,3], Abdoul Kader Kabore[4], Tegawendé F. Bissyande[1,4]

Interdisciplinary Center of Excellence in AI for Development (CITADEL)[1],
University Joseph KI-ZERBO[2], Virtual University of Burkina Faso[3],
University of Luxembourg[4]
ozenabo3@gmail.com, tapslyd@gmail.com, rodrique.kafando@citadel.bf,
abdoulkader.kabore@citadel.bf, aminata.sabane@ujkz.bf,
tegawende.bissyande@citadel.bf

Abstract. This paper presents an automated approach for building ontologies to improve treatment recommendations for tuberculosis (TB), in particular multidrug-resistant tuberculosis (MDR-TB) cases in Burkina Faso, using generative language models such as GPT-3. The aim is to facilitate the personalization of treatments according to the patient profile and drug resistance. Two approaches were explored: a automated approach based on the DaVinci GPT-3 model to generate OWL axioms from natural language sentences and a semi-automated approach using text extraction and natural language processing (NLP) techniques. The automated approach was fine-tuned with a dataset consisting of technical guidelines on TB management. The automated approach created an ontology composed of 158 classes, 55 object properties and 57 data properties, outperforming the semi-automated approach in terms of efficiency and accuracy. The axioms generated were validated using Protégé and integrated into a formal knowledge base. The study demonstrates that the use of language models such as GPT-3 can efficiently automate ontology generation, reducing human intervention.This approach is particularly well-suited to the management of complex MDR-TB cases and paves the way for standardization of treatment recommendations, while remaining adaptable to local specificities.

Keywords: Tuberculosis · Generative AI · Ontologies · Medical Decision Support Systems · Text-to-Ontology.

1 Introduction

Tuberculosis (TB) remains a major public health issue in many developing countries, particularly in Burkina Faso, where health systems face increased challenges in managing cases of multidrug-resistant tuberculosis (MDR-TB) [1]. MDR-TB significantly complicates standard treatment regimens as it evades conventional treatments, thereby increasing mortality and prolonging treatment periods [2]. In this context, the integration of advanced technologies such as artificial intelligence (AI) and

[1] Rapportmondialsurlatuberculose(OMS)
[2] http://cns.bf/IMG/pdf/mshp_profil_du_burkina_sur_la_tuberculose.pdf

ontology-based systems becomes crucial to improve treatment recommendations and tailor medical interventions to the specific needs of patients [3].

Current therapeutic recommendation systems for tuberculosis generally rely on clinical guidelines and standardized protocols, often based on recommendations from the World Health Organization (WHO). While these systems are effective for common cases of tuberculosis, they show their limitations when faced with complex cases, such as those of multidrug-resistant tuberculosis (MDR-TB), where resistance to standard drugs complicates therapeutic decisions [1].

To address this gap, this study proposes an innovative model for the automatic generation of ontologies aimed at improving treatment recommendations for TB, particularly for MDR-TB cases [2]. The hypothesis is that the use of generative models based on Large Language Models (LLMs), such as GPT-3, can transform technical guides and local data into a formal ontology [3]. This ontology facilitates the personalization of treatment recommendations based on the patient's profile and drug resistances.

In the following sections of this article, we present a review of the literature on ontology construction approaches and recommendation systems for tuberculosis treatment. Next, we detail the methodology used to generate ontologies from management guidelines. Finally, we present the results obtained and discuss the implications for case management of multidrug-resistant tuberculosis (MDR-TB) in Burkina Faso.

2 Related Work

There are several relevant works related to the use of ontologies for the treatment of tuberculosis (TB). We have grouped the articles into two distinct groups. The first group consists of articles using ontologies in the treatment or monitoring of tuberculosis. The second includes articles that use tools to generate ontologies or make predictions to understand the construction approaches that can be considered for our work.

2.1 Ontologies in Tuberculosis Treatment and Monitoring

Several studies explore the use of ontologies in tuberculosis (TB) treatment and monitoring.
Surveillance Systems: Gadicherla *and al.* [4] propose a TB surveillance system using a manual ontology for real-time data collection and analysis. While effective for identifying high-risk areas, limitations include narrow scope (excluding financial aspects) and a lack of focus on resource management.
TB Adherence Ontology: Olukunle Ayodeji Ogundele et al. [5] develop an ontology to capture factors influencing TB treatment adherence behavior in sub-Saharan Africa. This ontology offers a new perspective on adherence factors and their interrelationships.
Epidemiological Surveillance in Gabon: Raymond Ondzigue Mbenga [6] developed a decision support system for the epidemiological surveillance of tuberculosis in Gabon. The system combines a

[3] `https://www.scidev.net/afrique-sub-saharienne/news/l-intelligence-artificielle-\une-solution-aux-problemes-de-sante-de-l-afrique/`

spatio-temporal decision-making tool (SOLAP-TB) and a multi-agent simulation model (SMA-TB). This approach is distinguished by the use of complex modeling to monitor the evolution of TB in real-time, although the complexity of the model may limit its transferability to other contexts.

Tuberculosis Management: Kumar Abhishek[1] presents an ontology-based formalization for the Revised National Tuberculosis Control Program (RNTCP) in India. He utilizes a decision support system (DSS) and semantic web technologies to improve tuberculosis management and control. The proposed system integrates ontologies and SWRL rules for diagnosis and treatment, following WHO recommendations. The ontologies facilitate knowledge representation and decision-making based on logical rules. The article concludes with the potential extension of the system to track patients with multi-drug resistant tuberculosis and those co-infected with HIV.

Ontology Enrichment: Desi Ramayanti [7] leverages the existing Epidemiology Ontology (EPO) to create a TB-specific ontology through a semi-automatic approach. This approach emphasizes reducing time and costs associated with manual enrichment. These studies showcase the potential of ontologies for TB management but also highlight limitations in scope and automation.

2.2 Ontology Generation and Prediction Tools

Research efforts explore tools for ontology generation and prediction:

Automatic Ontology Construction: Fatima N. Al-Aswadi [8] discusses challenges and approaches in automatic ontology construction from text. The focus is on reducing human intervention and improving relation/axiom discovery. Deep learning approaches are seen as promising avenues for future exploration.

Ontology Generation Frameworks: Samaa Elnagar and al. [9] propose a framework for automating ontology generation from unstructured knowledge sources. This framework promotes efficiency and reduces costs compared to manual methods, but limitations exist in addressing complex OWL structures.

Term Extraction and Taxonomy Generation: Gerhard W. and al. [10] (2016) evaluate the effectiveness of word2vec for ontology creation by extracting terms and generating taxonomies from text. While efficient, word2vec's results can be influenced by input terms, impacting relevance.

Large Language Models for Ontology Enrichment: Patricia M. and Adrian G. [11] present a Protégé plugin that leverages large language models (like GPT-3) to automatically convert natural language sentences into OWL axioms. This approach aims to accelerate ontology development but necessitates evaluation of its impact on overall quality and reliability. These studies explore diverse tools for ontology generation and enrichment, highlighting a trend towards automation and the use of advanced language models.

2.3 Comparison of Approaches

Approaches	Advantages	Disadvantages
Using word2vec for ontology learning	Simple implementation, strong ability to extract relevant terms	Results sometimes too closely related to seed terms, difficulty in generating accurate taxonomic relations (about 50% accuracy)
Methods for automatic extraction and generation of ontologies from unstructured corpus	Cost and time reduction through automation, ability to use Knowledge Graphs (KGs) for dynamic representation	Often domain-specific approaches, still require manual interventions, KG quality issues (uncertain reliability and completeness)
Semi-automatic ontologies (Desi Ramayanti)	Cost and time reduction in ontology development	Still partially manual, not fully automated
Large language models (Patricia M.)	Automation of OWL axioms from natural language sentences	Requires evaluation of the quality and reliability of the axioms

Table 1. Table summarizing the advantages and disadvantages of different ontology-building approaches.

Among the various articles, the works of Kumar Abhishek have addressed the complexity of managing cases of multidrug-resistant tuberculosis (MDR-TB) and comorbidities, highlighting the persistent challenges in this field. The main objective of this thesis is to create an ontology that proposes suitable treatments for these specific cases of tuberculosis. For the implementation of the knowledge base, the approach adopted relied on the works of Patricia M. and Adrian G., who utilize large-scale language models, such as GPT-3, to translate sentences written in natural language into OWL axioms. These works show progress towards the automation of ontologies in TB management, but they also reveal gaps in adapting to local specificities, such as managing MDR-TB in Burkina Faso. Our research proposes to explore these approaches while leveraging the use of large language models to generate ontologies adapted to local contexts.

3 Material and methodology

3.1 Material

This section describes the tools and methods used to construct an ontology from technical guidelines for tuberculosis management, particularly for managing multidrug-resistant tuberculosis (MDR-TB). Two approaches are presented: an automatic approach and a semi-automatic approach.

Protégé: protégé is a widely used open-source ontology editor for creating and validating ontologies using the OWL language[1]. It allows for the syntactic validity of generated axioms to be checked and integrated into a formal knowledge base.

[1] `https://www.imgt.org/textes/IMGTeducation/Enseignements/_FR/Giudicelli-MEDBioinformatique_ontologies_23_Mai_2013.pdf`

Apache Jena Fuseki: apache Jena Fuseki is a SPARQL server used to query, manipulate, and store RDF triples [12]. This server allowed us to execute queries on semantic data and integrate the results into the ontology.

GPT-3: designed by OpenAI in 2018, GPT-3 is a deep learning model belonging to the GPT (generative pre-trained transformer) family developed by the same company. It is a pre-trained generative model for automatic natural language processing (ANLP) or natural language processing (NLP) [3]. With its 175 billion parameters, GPT-3 is one of the large language models (LLMs) and is used for tasks such as question-answering, translation, text writing, problem solving and application code generation.

DaVinci: the DaVinci model was used for the automatic translation of natural language sentences into OWL axioms [4]. This natural language processing model, developed by OpenAI, was fine-tuned for this specific task of axiom generation. Its ability to understand and generate complex texts makes it an ideal tool for creating ontologies in specialized fields like tuberculosis.

PyPDF2: PyPDF2 is a Python library used to extract raw text from PDF files[5]. We used PyPDF2 to read the PDF files of the technical guidelines for tuberculosis management, thus enabling automated extraction of textual information.

SpaCy: SpaCy is an open-source natural language processing (NLP) library[6]. It was used to segment the extracted text into sentences, recognize named entities (NER), and perform syntactic and morphological analyses.

3.2 Proposed approaches

We explored two approaches to construct the ontology from textual documents: an automatic approach and a semi-automatic approach.

Automatic Approach: In this approach, we use the fine-tuning technique on the Davinci model [11]. Fine-tuning is a method that allows adapting a pre-trained Machine Learning model to a specific task.

Figure 1 illustrates this approach in detail, highlighting the various steps of the process.

[3] `https://www.ibm.com/topics/machine-learninghttps://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1516205-gpt-3-le-modele-de-langue-geant-d-openai/`

[4] `https://www.actuia.com/actualite/openai-officialise-une-nouvelle-version-pour-\-son-modele-de-langage-gtp-3/`

[5] `https://nanonets.com/blog/pypdf2-library-working-with-pdf-files-in-python/`

[6] `https://spacy.io/`

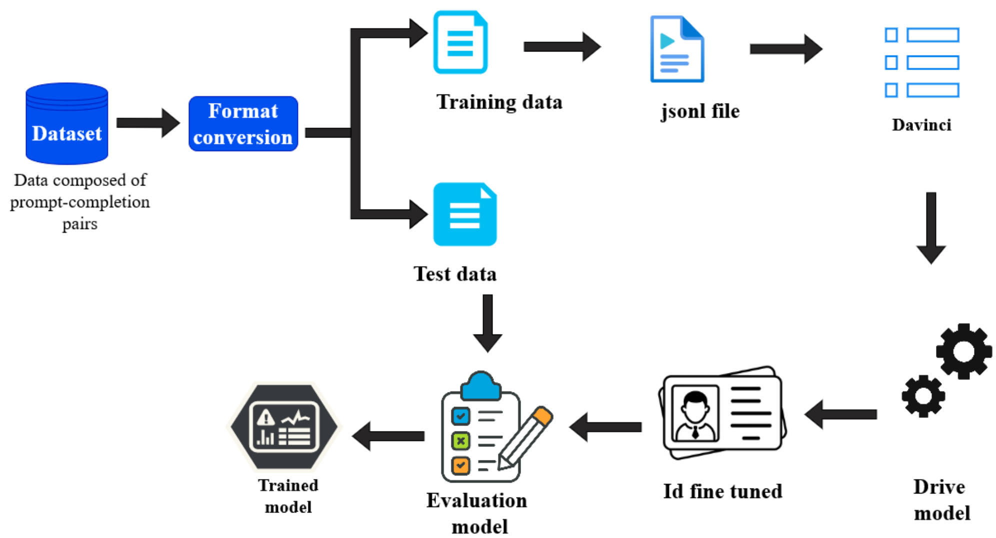

Fig. 1. Diagram of the automatic approach

According to Figure 1, the following steps are involved:

- Data Collection and Preparation: we built a dataset consisting of 182 auto-completion pairs from two technical guidelines for tuberculosis management in Burkina Faso;
- Conversion to JSONL Format: the sentences were converted to JSONL (JavaScript Object Notation Lines) format, a format suitable for training natural language processing models. Each line of the file represents a pair consisting of a sentence written in natural language and its translation into an OWL axiom;
- Training and Fine-Tuning the GPT-3 Model: the DaVinci model of GPT-3 was fine-tuned using these data pairs. This fine-tuning process adjusts the model's parameters to generate precise and coherent OWL axioms from sentences written in natural language. A test dataset was used to evaluate the model's performance after training;
- Generation and Validation of OWL Axioms: Once the model was trained, it was used to automatically generate OWL axioms. The generated axioms were then validated using Protégé to check their syntactic accuracy before integration into the ontology.

Semi-Automatic Approach The semi-automatic approach uses a combination of text extraction from PDF files and NLP (Natural Language Processing) analysis to extract triplets. Here are the main steps of this approach:

Figure 2 the steps involved in extracting triplets from PDF files.

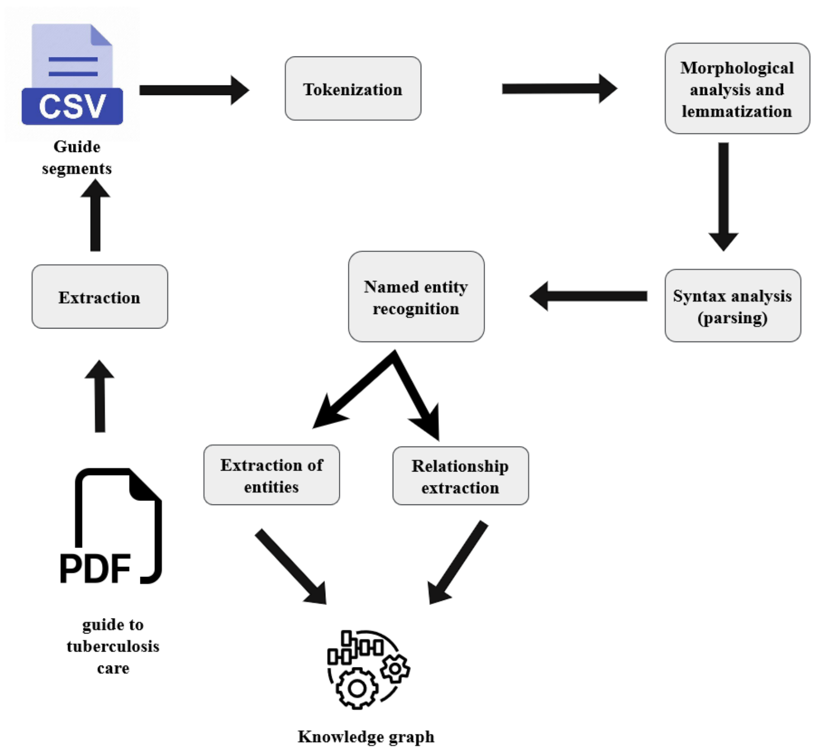

Fig. 2. Diagram of PDF text extraction.

According to Figure 2, the following steps are involved:

- Text Extraction from PDF Files : the text from the technical guides was extracted using a PyPDF2 library. This script automated the extraction of relevant information from large PDF files while preserving their structure. The raw text was then cleaned to remove unwanted characters and extra spaces. ;
- Sentence Segmentation and NLP Analysis: after extraction, the text was segmented into sentences using SpaCy. SpaCy identified named entities and performed syntactic and morphological analysis of the sentences, facilitating the extraction of RDF triples;

- Data Organization in CSV: the triples were organized into a CSV file for more structured data management;
- Knowledge Graph Generation: we built a knowledge graph to visualize the relationships between the triples.

Using the DaVinci model allows for fully automated generation of OWL axioms with increased accuracy, especially for complex sentences. Additionally, the fine-tuning process enables the model to adapt specifically to the domain of tuberculosis, thereby reducing the need for human intervention. In contrast, the semi-automatic approach, while functional, requires more manual intervention, particularly for text extraction and segmentation, as well as RDF triple generation. This approach is also less efficient in handling complex concepts, as it relies on additional steps of data cleaning and structuring. In summary, the automatic approach proved to be faster and more accurate for generating OWL axioms, requiring minimal human intervention. It is therefore more suitable for constructing a formal ontology in complex domains such as the management of multidrug-resistant tuberculosis.

4 Experiments

4.1 Results of the semi-automatic approach

The two approaches we have mentioned are two distinct methods for extracting information from texts and converting it into a structured form. Table 2 is an excerpt of the obtained triplets.

Table 2. Results of the Semi-Automatic Approach

Subjects	Predicate	Object
Patient	is	person
Patient	assumes	PresumedCase
PresumedCase	follows	PreventiveTreatment
Patient	isInfectedBy	TuberculosisCase
Patient	isUnder	Surveillance
Patient	takesA	XpertTest
Patient	interrupts	Treatment
TuberculosisCase	hasA	TherapeuticRegimen
Antitubercular	_hasFor_	MajorEffect
Antitubercular	hasFor__	MinorEffect
MajorEffect	__hasFor	RecommendedAction
MinorEffect	__hasFor_	RecommendedAction

The Semi-Automatic Approach focuses on extracting subject-predicate-object triplets from texts. These triplets can help identify different classes in PDF files. For example, in the context of tuberculosis treatment, a triplet could be “Patient assumes PresumedCase,” where “Patient” is the subject, “assumes” is the predicate, and “PresumedCase” is the object.

4.2 Results of the automatic approach

We used an automatic approach to generate ontologies from natural language sentences. The model, based on GPT-3 (DaVinci), was fine-tuned with 182 pairs of natural language sentences and their translations into OWL axioms.

- **Model performance:** After several training steps, the model showed high accuracy in generating axioms. For example, the sentence *'A patient is a person'* was correctly translated into an OWL axiom by declaring the `Patient` class and establishing a subclass relationship with `Person`.
- **Obtained metrics:** During training, the training loss (a measure of the model's error) gradually decreased, reaching a value as low as 0.0002, indicating that the model learned well from the training data and is capable of generating accurate results.

Table 3 The table below shows the different metrics.

Table 3. Training and Validation Results

Step	Train Loss	Train Accuracy	Valid Loss	Valid Mean Token Accuracy
1	4.0779	0.31818	-	-
2	2.69302	0.59259	-	-
3	3.85006	0.40816	-	-
4	3.23567	0.50980	-	-
5	2.58106	0.55556	-	-

Table 4 illustrates the results obtained using the automatic approach.

Table 4. Results of the Automatic Approach

Prompt	Results
A patient is a person.	Declaration(Class(:Patient)) Declaration(Class(:Person)) SubClassOf(Patient :Person)
A patient is a class, but disjoint from Doctor.	Declaration(Class(:Patient)) DisjointClasses(Patient :Doctor)
A person is a class. A patient is a subclass of person. A patient is a person.	Declaration(Class(:Patient)) SubClassOf(Patient :Person), Declaration(Individual(:Florence)) ClassAssertion(Patient :Person), Declaration(Individual(:Paul)) ClassAssertion(Patient :Person)
A tuberculosis case is any patient for whom tuberculosis has been confirmed bacteriologically or diagnosed clinically.	Declaration(Class(:Tuberculosis_Case)), Declaration(Class(:Patient)), Declaration(ObjectProperty(:confirmed_by_Bacteria)), Declaration(ObjectProperty(:diagnosed_Clinically))

Here are some examples of axioms generated by the automatic approach:

– **Natural language sentence:** "A patient is a person"

  • **OWL Axiom:** `Declaration(Class(:Patient)), SubClassOf(:Patient :Person)`

These results show that the model is capable of efficiently converting sentences into formal axioms, thereby facilitating the creation of a comprehensive ontology for tuberculosis care.

For the continuation of our work, we have opted for the automatic approach. Several reasons justify this choice. On the one hand, this approach does not require constant expert intervention, saving time. On the other hand, the axioms generated by this method greatly facilitate the creation of the ontology in the Protégé software. After generating these axioms, we obtained 158 classes, 55 object properties, and 57 data properties. Figure 3 illustrates the graph of the proposed ontology.

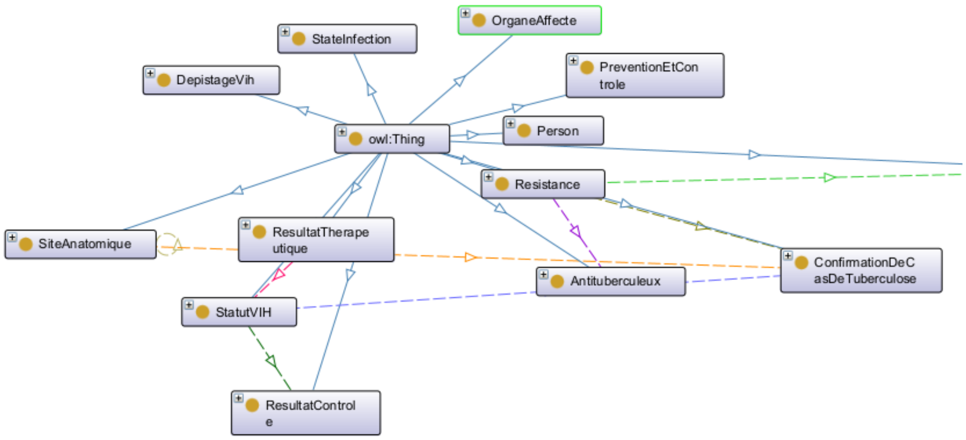

Fig. 3. *A sample of the proposed Ontology Graph*

Figure 3 shows a sample of the generated ontology. It presents the relationships between medical concepts related to tuberculosis, such as Patient, Tuberculosis Case, and Treatment. This graph highlights how these concepts are connected, thus allowing a better understanding of the relationships between the different entities.

We used an Apache Jena Fuseki to execute SPARQL queries. These queries allowed us to explore and query the data represented in the knowledge base. In Figure 4, a query was performed to select information on the phases of tuberculosis (TB) treatment and the details associated with these phases, such as treatment duration, antibiotics used, TB cases, and medication dosages.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX a: <http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#>

SELECT
  ?casTb
  ?Phase
  ?dureeTraitement
  (GROUP_CONCAT(CONCAT( STRAFTER(STR(?Antibiotiques), "#")); separator="; ") AS ?NomAntibiotiques)
  ?posologieTbResistant
WHERE {
  {
    ?Variable rdf:type a:PhaseIntensif.
    ?Variable a:nomPhase ?Phase.
  } UNION {
```

Fig. 4. *Query on the phases of tuberculosis (TB) treatment*

Table Response 10 results in 0.208 seconds Simple view Ellipse Filter query results Page size: 50

	casTb	Phase	dureeTraitement	NomAntibiotiques	posologieTbResistant
1	TB-MR/RR	Début de la deuxième phase ou...	Semaine 17 à 24	Bédaquiline; Moxifloxacine; Ethambutol; Pyrazinamide; Clofazimine	Bédaquiline 2 cp 3 fois par semaine (lundi, mercredi, vendredi)
2	TB-MR/RR	Début de la deuxième phase ou...	7è au 9è mois	Moxifloxacine; Ethambutol; Pyrazinamide; Clofazimine	
3	TB-MR/RR	Début de la deuxième phase ou...	Semaine 3 à 16	Moxifloxacine; Ethambutol; Pyrazinamide; Clofazimine	
4	TB-MR/RR	Début de la prémière phase	14 premiers jours du tra...	Isoniazide_à_haute_dose; Bédaquiline; Prothionamide; Moxifloxac...	Bédaquiline 4 cp/jour (2 cp matin et 2 cp so ir)
5	TB-MR/RR	Fin de la prémière phase	Semaine 3 à 16	Isoniazide_à_haute_dose; Bédaquiline; Prothionamide; Moxifloxac...	Bédaquiline 2 cp 3 fois par semaine (lundi, mercredi, vendredi)
6	TB-MR/RR	Fin de la prémière phase	7è au 9è mois	Moxifloxacine; Ethambutol; Pyrazinamide; Clofazimine	
7	TB-MR/RR	Fin de la prémière phase	Semaine 3 à 16	Moxifloxacine; Ethambutol; Pyrazinamide; Clofazimine	
8	Tuberculose phar...	Phase Intensive ou prémière Ph...	Deux mois	Ethambutol; Rifampicine; Isoniazide; Pyrazinamide	
9	Tuberculose phar...	Phase de continuation ou deuxi...	Quatre mois	Rifampicine; Isoniazide	

Fig. 5. Result of the query on the phases of tuberculosis (TB) treatment

The results allow us to extract information on the different phases of tuberculosis treatment from an RDF database. These details, such as treatment durations, antibiotics used, and resistant drug dosages, are essential for proposing a treatment plan for a tuberculosis case.

Figure 6 is a query that interrogates an RDF database to retrieve information on antibiotics, their major side effects, and the associated management guidelines.

```
PREFIX a: <http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#>

SELECT ?Antibiotique ?effetMajeur ?conduiteATenir
WHERE {
  ?a a:nomAntibiotique ??Antibiotique.
  ?Antibiotique a:_a_pourEffetMajeur ?effectMajeur.
  ?effectMajeur a:_a_Pour_ ?conduite.

  BIND(STR(?effectMajeur) AS ?effetMajeur)
  BIND(STR(?conduite) AS ?conduiteATenir)
}
```

Fig. 6. Query: Antibiotics, Side Effects, and Management Guidelines

<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Isoniazide>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj2	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem2
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Isoniazide>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj3	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem3
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Rifampicine>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj7	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem7
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Rifampicine>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj6	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem6
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Ethambutol>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj1	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem1
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Pyrazinamide>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj4	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem4
<http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Pyrazinamide>	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#Effetmaj5	http://www.semanticweb.org/asus/ontologies/2023/8/untitled-ontology-3#conduitem5

Fig. 7. Result of the query to display antibiotics, their side effects, and management guidelines

The query results are presented in the form of URIs, which is essential for uniquely identifying each entity and establishing relationships between them in the RDF database. This query helps provide information on the expected reactions to antibiotics and the actions to take in case of adverse effects.

5 Discussion

This study aimed to develop a model using large language models (GPT-3 DaVinci) to automatically generate ontologies dedicated to the management of multidrug-resistant tuberculosis (MDR-TB). The objective was to optimize clinical decision-making by minimizing human intervention while providing reliable and tailored treatment recommendations.

The results show that the automatic approach enabled the creation of a rich ontology, comprising 158 classes, 55 object properties, and 57 data properties. These results surpass those obtained by the semi-automatic method, which requires more human intervention and produces less accurate results. The use of GPT-3 has thus proven its effectiveness in automating the generation of OWL axioms.

Our results align with research on tuberculosis management while providing significant advancements.

Raymond Ondzigue's work: Raymond Ondzigue's thesis focuses on the epidemiological surveillance of tuberculosis in Gabon, using a multi-agent simulation model (SMA-TB) and spatio-temporal analysis to model the evolution of the disease. Our approach stands out through the complete automation of ontology generation for clinical decision-making. While Ondzigue's approach focuses on monitoring and modeling epidemiological data, our work aims to optimize the clinical recommendation process for MDR-TB based on textual data, reducing reliance on human expertise.

Kumar Abhishek's work: Kumar Abhishek's article proposes an ontology-based decision support system for the Revised National Tuberculosis Control Program (RNTCP) in India, using SWRL rules. Although this work offers a formalized approach to managing drug-sensitive tuberculosis, it requires significant human intervention to manually define relationships between concepts. Our method goes further by fully automating the creation of OWL axioms from natural language sentences, enabling the handling of more complex cases, such as MDR-TB, which were not sufficiently covered in Abhishek's work.

The strengths of our approach are multiple:

- **Complete automation**: The use of GPT-3 DaVinci allows the generation of OWL axioms without significant human intervention, improving the efficiency and speed of the process.
- **Improved efficiency in terms of time and resources**, which is particularly relevant in the context of Burkina Faso.

However, this study presents certain limitations:

- **dependence on data quality**: the GPT-3 model relies on the quality of the training data. Ambiguous or poorly structured sentences can affect the quality of the generated axioms;
- **human supervision**:although the approach is automated, supervision is still necessary to validate the clinical relevance of the axioms, especially for complex cases;
- **lack of clinical validation**: the treatment recommendations generated still need to be validated in a clinical setting to assess their real impact on decision-making.

Our study makes a unique contribution to the field of automatic ontology generation applied to tuberculosis management, particularly for MDR-TB cases. Unlike previous work, such as that of Kumar Abhishek or Raymond Ondzigue, our approach offers an automated solution for generating OWL axioms from natural language sentences, significantly simplifying the ontology creation process. This approach allows for the standardization of treatment recommendations in local contexts, while adapting to regional specificities.

6 Conclusion

This study set out to develop a model leveraging large language models, specifically GPT-3 DaVinci, for the automatic generation of ontologies aimed at supporting the clinical management of multidrug-resistant tuberculosis (MDR-TB). Through an automated approach, we successfully created a rich ontology comprising 158 classes, 55 object properties, and 57 data properties. This approach surpassed the semi-automatic method in terms of efficiency, accuracy, and the need for minimal human intervention.

The findings demonstrate that GPT-3, when fine-tuned for domain-specific tasks, can effectively automate the generation of OWL axioms from natural language sentences, which is essential for building robust medical decision support systems. By automating the process, our work not only enhances the speed of ontology generation but also improves the personalization and relevance of treatment recommendations for MDR-TB cases.

Comparisons with previous work, such as the epidemiological surveillance system for TB in Gabon developed by Raymond Ondzigue, and the ontology-based decision support system for tuberculosis in India by Kumar Abhishek, highlight the novelty of our approach. While their systems focus on specific aspects of TB management, such as epidemiological surveillance and decision-making based on predefined rules, our model automates the entire ontology creation process. This automation allows for greater flexibility and scalability in clinical decision-making, particularly in the context of MDR-TB treatment, which remains a critical challenge in regions like Burkina Faso.

Looking forward, future work will focus on extending the ontology to cover additional aspects of tuberculosis management, including diagnostics and predictive models for patient outcomes. Moreover, clinical validation of the treatment recommendations generated by our system will be a key area of exploration, ensuring that the theoretical advancements achieved translate into tangible benefits in real-world healthcare settings.

Acknowledgments

This work was conducted as part of the Artificial Intelligence for Development in Africa (AI4D Africa) program, with the financial support of Canada's International Development Research Centre (IDRC) and the Swedish International Development Cooperation Agency (Sida).

References

1. Abhishek K, Singh M. An ontology based decision support for tuberculosis management and control in India. International Journal of Engineering and Technology. 2016;8(6):2860-77.
2. Du Y, Ranwez S, Sutton-Charani N, Ranwez V. Apports des ontologies aux systèmes de recommandation : état de l'art et perspectives. In: 30es Journées Francophones d'Ingénierie des Connaissances, IC 2019. AFIA. Toulouse, France: AFIA; 2019. p. 64-77. Ffhal-02329644.
3. Menad S, Abdeddaïm S, Soualmia LF. Fusion d'ontologies biomédicales par des modèles siamois et validation par modèles de langue. In: 35es Journées francophones d'Ingénierie des Connaissances (IC 2024) @ Plate-Forme Intelligence Artificielle (PFIA 2024). La Rochelle, France; 2024. Ffhal-04636888f.
4. Gadicherla S, Krishnappa L, Madhuri B, Mitra SG, Ramaprasad A, Seevan R, et al. Envisioning a learning surveillance system for tuberculosis. PLoS One. 2020;15(12):e0243610.
5. Olukunle Ayodeji Ogundele AWP Deshendran Moodley, Seebregts CJ. An ontology for factors affecting tuberculosis treatment adherence behavior in sub-Saharan Africa. Patient Preference and Adherence. 2016;10:669-81. Available from: `https://www.tandfonline.com/doi/abs/10.2147/PPA.S96241`.
6. Ondzigue Mbenga R, Devogele T, Maghendji S, Peralta V, Ngoungou EB. Un système d'information géographique décisionnel pour la surveillance épidémiologique de la tuberculose en Afrique subsaharienne : Cas du Gabon; 2019. Poster. Francophone International Conference on Spatial Analisys and Geomatics (Sageo 2019). Available from: `https://hal.science/hal-02375898`.
7. Pellison FC, Rijo RPCL, Lima VC, Crepaldi NY, Bernardi FA, Galliez RM, et al. Data integration in the Brazilian Public Health System for Tuberculosis: use of the semantic web to establish interoperability. JMIR Medical Informatics. 2020;8(7):e17176.
8. Al-Aswadi FN, Chan HY, Gan KH. Automatic ontology construction from text: a review from shallow to deep learning trend. Artificial Intelligence Review. 2020;53:3901-28.
9. Elnagar S, Yoon V, Thomas MA. An automatic ontology generation framework with an organizational perspective. arXiv preprint arXiv:220105910. 2022.
10. Wohlgenannt G, Minic F. Using word2vec to Build a Simple Ontology Learning System. In: ISWC (Posters & Demos); 2016. p. 1-7.

11. Mateiu P, Groza A. Ontology engineering with large language models. arXiv preprint arXiv:230716699. 2023.
12. Chokshi HJ, Panchal R. Using Apache Jena Fuseki Server for Execution of SPARQL Queries in Job Search Ontology Using Semantic Technology. International Journal of Innovative Research in Computer Science & Technology. 2022;10(2):497-504.