

Developing an Arabic Infectious Disease Ontology to Include Non-Standard Terminology

Lama Alsudias^{1 2} and Paul Rayson¹

¹School of Computing and Communications, Lancaster University, UK

²College of Computer and Information Science, King Saud University, Saudi Arabia
lalsudias@ksu.edu.sa, p.rayson@lancaster.ac.uk

Abstract

Building ontologies is a crucial part of the semantic web endeavour. In recent years, research interest has grown rapidly in supporting languages such as Arabic in NLP in general but there has been very little research on medical ontologies for Arabic. We present a new Arabic ontology in the infectious disease domain to support various important applications including the monitoring of infectious disease spread via social media. This ontology meaningfully integrates the scientific vocabularies of infectious diseases with their informal equivalents. We use ontology learning strategies with manual checking to build the ontology. We applied three statistical methods for term extraction from selected Arabic infectious diseases articles: TF-IDF, C-value, and YAKE. We also conducted a study, by consulting around 100 individuals, to discover the informal terms related to infectious diseases in Arabic. In future work, we will automatically extract the relations for infectious disease concepts but for now these are manually created. We report two complementary experiments to evaluate the ontology. First, a quantitative evaluation of the term extraction results and an additional qualitative evaluation by a domain expert.

Keywords: Arabic, infectious disease, NLP, Ontology

1. Introduction

Ontologies are formal representations of concepts and their relationships in machine readable form. Ontologies can be built manually, semi-automatically or automatically. Generally, they fall into two categories: Upper level Ontology and Domain Ontology. There are many existing ontologies in English and other languages but for Arabic there has been little recent research. Developing ontologies from Arabic text is a difficult process due to the lack of existing resources and the nature of the Arabic language which has complex grammatical, morphological and semantic aspects. Existing NLP algorithms for other languages cannot simply be applied directly on Arabic. In addition, the Arabic language has many dialects which may complicate ambiguities for computational understanding of the meaning of the words and reduce the accuracy of tools created only for Modern Standard Arabic.

In this paper, we present a new Arabic Ontology from the infectious disease domain. The overall goal of this ontology is to be a key source of information related to infectious diseases in Arabic. We envisage that it can be used for many important applications in academia and the real world including Question Answering on the Semantic Web (AlAgha and Abu-Taha, 2015) and, in particular in our case, for online monitoring of health events (Collier et al., 2010).

The rest of paper is organised as follows. Section 2 covers related background work. Section 3 explains the requirement for an Infectious Disease Ontology in Arabic. The Ontology Engineering process includes Ontology Architecture, Data Gathering, Pre-processing, Term Extraction, Finding Synonyms, Adding informal Terms, Obtaining Concepts, and Defining and Updating Relations discuss

in Section 4. Section 5 clarifies the Ontology Implementation. Two types of ontology evaluation are performed in Section 6. Section 7 states conclusions and further work.

2. Related Work

One critical obstacle facing Arabic Ontology construction is the fact that there is no standard upper-level Ontology to act as a foundation from which to build and this causes a severe lack of coherence and consistency among the Arabic Ontologies. Most of the previous experiments in this field have structured the ontologies either fully manually or partially manually which leads to issues of high cost and a requirement for significant investment of time (Al-Zoghby et al., 2018).

In the case of building an ontology in the Arabic language there is some previous work related to specific topics such as the Islamic domain, Computing domain, News domain, and Legal domain (Al-Zoghby et al., 2018). Raisan and Abdullah (2017) implemented an ontology in the Arabic language related to the Iraqi News domain. They used a manual ontology development strategy to build the schema. A study by Albukhitan and Helmy (2013) produced an automatic ontology based annotation of Arabic Web resources related to food, nutrition and health domains. It used linguistic patterns to discover relevant relationships between the named entities in the Arabic Web resources.

Al-Safadi et al. (2011) presented a model for representing Arabic knowledge in the Computer Technology domain using ontologies. They stressed the importance of understanding the goal of the ontology and its end users during the development of the ontology. The study in (AlAgha and Abu-Taha, 2015) built an Arabic natural language interface to ontologies and RDF data (AR2SPARQL) by

translating the user query to RDF triple patterns. The study applied Arabic NLP techniques to link query terms to ontological entities, then used a set of grammar rules in the ontology to structure a SPARQL query by joining the ontological entities.

On the other hand, there are many studies related to Latin-character-set-based language ontologies in multiple topic areas. Here we focus only on the previous work related to the infectious disease domain. Cowell and Smith (2010) reviewed the vocabulary resources relevant to the infectious disease domain and emphasised that they are lacking in terms of their support of translational medicine and computational applications. The study in (Schriml et al., 2011) generated a disease ontology including 8,043 human diseases with the goal of providing consistent, sustainable, and reusable descriptions of human disease terms.

The BioCaster project is a multilingual ontology that aims to describe the terms and relations necessary to detect at an early stage public health events in the grey literature, which is information or research that has been published in a non-commercial way, (Collier et al., 2010). It supports 12 languages including Arabic, but employs a translation-based method that may potentially produce incorrect words. In other words, some diseases produce another meaning after translation due to the complexity of the disease name. For example, the disease (التهاب السحايا) which means Meningitis may be translated to Schizophrenia which is not correct. Moreover, using the abbreviations in English affect the results such as MARSAs, which is Methicillin resistant Staphylococcus aureus, and cannot be translated directly without the full meaning.

Critically, none of the above studies have developed an infectious disease ontology in the Arabic language. Moreover, the existing infectious disease ontologies in other languages were constructed with only formal language input. Therefore, the common problems include concepts having different surface forms between formal and informal variants, or professional and non-professional descriptions of disease terminology and symptoms, and Arabic dialect detection in non-standard settings.

3. Understanding the Needs of an Infectious Disease Ontology in Arabic

The Arabic infectious disease ontology that we wish to create is classified as a domain ontology, wherein the concepts and relationships belong to a specific domain, in our case infectious disease. In other words, it decreases the potential confusion of terms between users who share different kinds of information that relate to this domain (Al-Zoghby et al., 2018). Recently, ontologies have proved to be an effective approach to support data annotation to process data and information automatically. BioCaster is a notable example for using ontologies in detecting and tracking infectious disease (Cowell and Smith, 2010). Moreover, it has many advantages in data analysis of automated

reasoning applications and high-throughput technologies such as clinical decision support, biomedical research, biosurveillance applications, and analysing microarray data tools. These benefits led to increasing attention on developing vocabularies and reasoning algorithms that are used in ontologies (Cowell and Smith, 2010). As a result, we have conducted experiments to understand what needs to be done to improve the strictness and the structure of ontologies that are effective in supporting the computational reasoning, analysis and monitoring of infectious disease.

The large scale survey conducted by Joshi et al. (2019) suggested that there is need to improve the quality of the epidemic intelligence by the mitigation of false information. The survey also highlighted that it is important to include informal text such as social media in the medical ontologies which contain scientific and technical terms. Slang terms, which are the informal language used by the general public, pervade the language of social media since there are lower expectations of formality in such writing and the expressions of emotions are more frequent (Soliman et al., 2014). Using slang terms especially in the medical domain has the potential to cause health issues when a misunderstanding occurs. For example, the word (سيلان) may be understood as mucus or blood coming from the nose and this represents symptoms from multiple different diseases. The specific aims for our Arabic Infectious Disease Ontology are:

- To describe the terms and relations necessary to detect and analyse infectious disease.
- To bridge the gap between informal terms and scientific terms related to infectious disease.
- To be released in a form that is freely available.

4. Ontology Engineering

The process of Ontology engineering includes many phases. It starts with understanding the domain knowledge to be represented by the Ontology. The presentation of domain knowledge is represented by a basic classification of terminology in the domain and the relationships between the terms.

4.1. Ontology Architecture

Our proposed architecture for the creation of our ontology is based on a hybrid approach, which merges two techniques, manual ontology development with ontology learning. Figure 1 represents the proposed system architecture for the creation of an Arabic Infectious Disease Ontology.

4.2. Data Sources Selecting and Gathering

First, we defined a list of infectious diseases that typically spread in Arab countries based on information from the World Health Organization website¹. We have developed a web crawler that retrieves relevant documents from trusted websites, that represent Governments, Organizations, or Hospitals, on Arabic Infectious disease listed in Table 1 in

¹<https://www.who.int>

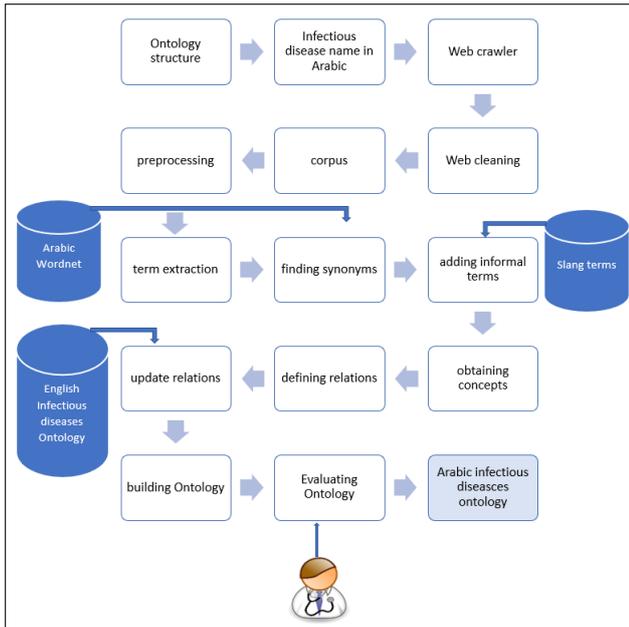


Figure 1: The proposed system architecture of Arabic Infectious Disease Ontology

order to generate a high quality corpus. We have 115 documents that covered around 25 infectious disease.

4.3. Pre-processing

We used standard approaches for pre-processing the text, applying the following processes using Python scripts² in sequence: Tokenization, Normalization, Stopword removal, POS Tagging, and Sentence Splitting in order to develop a high quality ontology as described in (Asim et al., 2018). The pre-processing steps are as follows:

Tokenization: Tokenization is the process of chunking the text into words. This is a vital first step as we will be working with domain terms.

Normalization: In normalisation, some Arabic letters are normalised such as (أ، آ، إ) are converted to (ا).

Removing Stopwords: Arabic stopwords were removed since they are less useful for ontology creation. There is also need to remove the non-Arabic words, special characters, and numbers.

POS Tagging: Part of speech tagging is the process of labeling corpus words with their corresponding part of speech tags. We used the Stanford Arabic Part of speech tagger in this step (Manning et al., 2014).

Sentence Splitting: To allow better extraction and linking of terms, we used (.) to split the sentences in our corpus.

4.4. Term Extraction

Terms are linguistic representations of domain-specific concepts. The target here is to discover a set of significant terms for concepts and relations (Cimiano, 2006). We

followed three statistical strategies for this step: TF-IDF, C-value and YAKE.

4.4.1. TF-IDF

The well known formula Term Frequency-Inverse Document Frequency (TF-IDF) is used to rank the candidate terms according to their importance in the corpus (Al-Thubaity et al., 2014). We apply unigrams, bigrams and trigrams and disregard terms that have a document frequency lower than two. Python's scikit-learn libraries 0.20.2 (Pedregosa et al., 2011) are used to apply TF-IDF on the data set.

4.4.2. Using C-value

C-value is a domain-independent method for automatic term recognition that aims to improve the extraction of nested multi-word terms (Frantzi et al., 2000). The algorithm first tags the words with their POS tags by Stanford Arabic Part of speech tagging (Manning et al., 2014). Then it extracts the strings using the linguistic filter, which is the type of terms extracted, and the stop words list. Then sequences that fall below the frequency threshold are removed. After that, we calculate C-value for each of the candidate strings. Finally, a list of candidate terms is built and ranked by their C-value.

4.4.3. Using YAKE

Yet Another Keyword Extractor (YAKE) is an Unsupervised Approach for Automatic Keyword Extraction using text features (Campos et al., 2019). YAKE has different steps including: pre-processing the text as before, and identifying the candidate term, feature extraction, computing term score, generating n-gram and computing candidate keyword score, and deduplicating and ranking the data. The ranking depends on the score for each term which is calculated from a defined set of features. These features are the casing aspect of a word, word position, word frequency, word relatedness to context, and the counts of word occurrences in different sentences. It is available online as an open source Python package. It includes support to extract keywords from Arabic language.

4.5. Finding Synonyms

Synonyms can be used to expand the terms that were not retrieved by web resources. We used Arabic WordNet (Black et al., 2006) to find synonyms of the term that will be used. For instance, the word (مراهم), ointments in English, has a synonym (دهان) and the two terms need to be added to ontology because they represent a treatment of Leishmania disease.

4.6. Adding Informal Terms

In order to find the slang terms related to infectious diseases, we consulted native speakers. This included eliciting the informal term equivalents of the infectious disease that may be used by the study participants when they write about disease in social media. A questionnaire was sent to five different types of people: academic, health professional, students, non-educational people and others as classified by Alsudias and Rayson (2019). Around

²<https://github.com/alsudias>

English Name	Arabic Name	Link
World Health Organization	منظمة الصحة العالمية	https://www.who.int
The Ministry of Health in Saudi Arabia	وزارة الصحة السعودية	https://www.moh.gov.sa
Wikipedia	ويكيبيديا	https://ar.wikipedia.org
Mayo clinic	مايو كلينك	https://www.mayoclinic.org/ar
WebTeb	ويب طب	https://www.webteb.com
Altibbi	الطبي	https://www.altibbi.com

Table 1: Some Arabic Web Sources Related to Infectious Disease.

100 people responded to the questionnaire with three informal names of each disease on average. For example, the disease (حمى الضنك), Dengue fever in English, has six informal names: (الدنك), (الدرنج), (أبو الركب), (الدنجية), (حمى تكسير العظام), and (حمى عدن).

In future, we plan to extend our approach by consulting these groups of people about the informal terms for the symptoms and then update the ontology with these terms. Words for symptoms are affected by the different dialects of the Arabic language since they depend largely on feelings and experiences. For instance, a person can discuss coryza using different informal words such as (رشح), (زكام), and (عطاس).

4.7. Obtaining Concepts

In order to decide which terms to be considered as a concept and included in the output ontology, we manually filter the terms that are meaningful. Table 2 shows the basic concepts of the proposed ontology. The ontology focuses on the infectious disease concept (المرض المعدى) due to the fact that it represents a central node in the ontology.

4.8. Defining Relations

This step is currently a manual one which we will aim to automate in future work. We defined appropriate relations for the infectious disease domain. Some of these relations are derived from the ontologies in (AlAgha and Abu-Taha, 2015) and (Schriml et al., 2011). The Other relations are defined after studying the goals of the ontology. For instance, the relations spread-in (ينتشر في) and infested-with (موبوء به) may help in finding the regions where the infectious disease typically spread. Also, the relation synonym (يعرف ، تعرف) is important to find all the slang names of the infectious disease. Table 3 represents the relations with Arabic examples and description of the relation.

4.9. Updating Relations

In order to expand our ontology with all terms and relations, we reviewed an English infectious disease ontology. For this, we used (Schriml et al., 2011) as infectious disease ontology and translated the concepts and relations to Arabic. Then, we updated our ontology with these components.

5. Ontology Storage and Representation

We stored the ontology in the Web Ontology Language (OWL) format in order to support sharing and reuse of this valuable resource. We used the Protégé tool³ version 5.5.0, a free open-source software tool typically used for building and maintaining ontologies. It supports many languages including Arabic. Protégé has particular benefits over other tools because it facilitates shareable access to structured knowledge for domain specialists, software designers, and knowledge engineers simultaneously (Kapoor and Sharma, 2010). Also, it is capable of supporting XML, RDF (S), XML Schema and OWL formats (Musen and others, 2015).

The classes have been organised in a hierarchical taxonomy as in Figure 2. The properties in OWL can be divided into Object property and Datatype property. As a result, in this study there are 21 Object properties that have been proposed in Table 3 and 11 Datatype properties that have been proposed as illustrated in Figure 3. The Object properties define the relations between classes in order to give a sufficient vision of how the components of the ontology interact with each other. There is a need to take into account the inverse characteristic for each property. For instance, ‘cause’ and ‘caused-by’ are two properties that work in opposite directions between the ‘infectious disease’ and ‘cause’ classes. In addition to the Object property, the Datatype property has been proposed to illustrate the type of data for individuals. In our case, the range for all Datatype property is a string.

Figure 4 shows a sub graph of the developed ontology with classes and their relationships, and Figure 5 illustrates a screen shot of the Measles disease (الحصبة) as an example.

6. Ontology Evaluation

Many techniques were used to evaluate ontologies. Here we conducted performance of the term extraction technique and a Domain Expert Evaluation approach.

6.1. Evaluating the Term Extraction Results

In order to evaluate the performance of the term extraction techniques, the terms in the extracted ontology are manually reviewed. Then, we compared the extracted terms by the three strategies, TF-IDF, C-value and YAKE, to the final concepts using the standard Precision, Recall,

³<https://protege.stanford.edu/>

Concept	Concept (in Arabic)	Description
Infectious disease	المرض المعدي	Any disease that classified as infectious disease.
Slang term of Infectious disease	المصطلح العامي للمرض المعدي	The informal or slang term of the infectious disease.
Symptom	عرض	Any symptom of the infectious disease.
Cause	سبب	Anything that cause the infectious disease.
Prevention	وقاية	The ways of prevention of the infectious disease.
Infection	عدوى	The ways that the infectious disease can infection between people.
Organ	عضو	The affected organ of the infectious disease.
Treatment	علاج	The methods taken to treat the infectious disease.
Diagnosis	تشخيص	The ways that help to define the infectious disease.
Place of the disease spread	مكان انتشار المرض	The place that the infectious disease may spread.
Infected category	الفئة المصابة	The most vulnerable category that may infected.

Table 2: Arabic Infectious Disease Ontology Concepts.



Figure 2: Arabic Infectious Disease Ontology Classes in Hierarchy

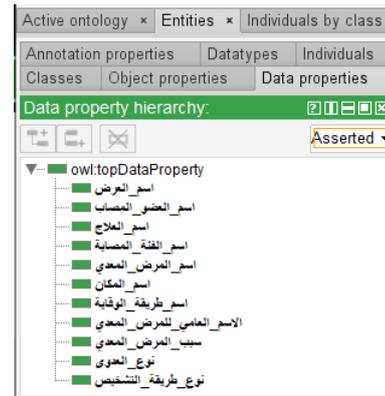


Figure 3: Arabic Infectious Disease Ontology Datatype properties (as translated above).

and F-Measure metrics. Table 4 summarises the results obtained by the system and Figure 6 visualizes the average results of the term extraction performance.

We can see that the lowest performance was found in the C-value method. This could result from the complexity of the information of infectious disease in Arabic language and the document size. The TF-IDF method achieved between 20.0% to 50.0% in their results due to the different types of documents included. In the YAKE method, the overall performance for the Precision is around 35.0% while the Recall is achieved up to 62.9% in Mumps documents. Moreover, the YAKE method supports multiword term extraction which is important in Arabic since many terms in infectious disease consisting of two or three words. For example, the

word (التهاب المتحمة), which means conjunctivitis in English, is a multi word that describes one symptom of the Zika virus disease.

6.2. Domain Expert Evaluation

We asked an expert in infectious diseases who is an Arabic native speaker to evaluate our ontology in terms of medical correctness. The evaluation included reviewing the concepts and the relations between the diseases. In other words, the expert needed to review every infectious disease and terms related to it and the correct connection between them. Also, we wished to check the accuracy of our ontology in terms of the individual infectious diseases. For example, is it correct that Malaria is infected by the Transfusion.

Relation	Relation (in Arabic)	Examples	Links Description	
			From	To
Is-a	هو	هو ، هي	infectious disease	definition (hyponym relationship)
Kind-of	نوع من	هو من ، هي من ، هو نوع من ، هي نوع من	infectious disease	type of the disease.
Synonym	مرادف	يعرف ، تعرف	infectious disease	informal term (slang one)
Caused-by	تحدث بسبب	تحدث بسبب، يحدث بسبب	infectious disease	cause class
Cause	يسبب	يسبب ، تسبب	cause class	infectious disease
Infected-by	يُعدى بـ	يُعدى بـ ، تُعدى بـ	infectious disease	infection class
Infect	يعدى	يعدى ، تعدي ، طرق العدوى	infection class	infectious disease
Prevent	يقي ، يمنع	يقي ، تقي	prevention class	infectious disease
Prevented-by	يقي بـ	يقي بـ ، يقي بـ ، طرق الوقاية	infectious disease	prevention
Spread-In	ينتشر في	ينتشر في ، تنتشر في	infectious disease	the place
Infested-with	موبوءه بـ	موبوءه بـ ، أماكن انتشار	the place	the infectious disease
Symptom-of	عرض لـ	أحدى أعراضها ، أحدى أعراضها	Symptom class	infectious disease
Has-symptom	له عرض	من أعراضه من أعراضها	infectious disease	Symptom class
Diagnoses	يشخص	طرق تشخيص	the Diagnosis	infectious disease
Diagnoses-by	يشخص بـ	لتشخيص	infectious disease	the Diagnosis
Infests	يصيب	يصيب ، تصيب ، يمرض ، يمرض	infectious disease	the organ
Infected-by	يصاب بـ	يصاب بـ ، تصاب بـ	the organ	infectious disease
Treats	يعالج	يعالج ، تعالج ، يداوي ، تداوي	treatment class	infectious disease
Treated-by	يعالج بـ	يعالج بـ ، تعالج بـ ، طرق العلاج	infectious disease	treatment class
Get-sick	تصيب بالمرض	تصيب بالمرض ، يصيب بالمرض	infectious disease	Infected category
Get-sick-by	تصاب بالمرض	تصاب بالمرض ، يصاب بالمرض	Infected category	infectious disease

Table 3: Arabic infectious disease Ontology relations.

In addition, we calculated the precision of the system after computing the average number of correct terms which was 88.4%. This result shows that our ontology has a high correctness since the expert agreed on the applicability of using it.

7. Conclusion and Further Work

In this paper, we have described a novel resource that integrates the scientific vocabularies of infectious diseases with their informal equivalents. The Arabic Infectious Disease Ontology is written in the Arabic language, and is the first ontology in Arabic specialising in the infectious

Disease Name	TF-ID			C-Value			YAKE		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
AIDS	32.0%	23.3%	26.9%	18.0%	12.4%	14.7%	40.0%	28.5%	33.3%
Avian influenza	21.0%	16.9%	18.7%	12.5%	12.8%	12.7%	28.0%	32.0%	29.9%
Coronavirus	25.0%	28.0%	26.4%	15.0%	18.6%	16.6%	32.0%	39.2%	35.2%
Dengue fever	30.0%	28.1%	29.0%	13.7%	23.2%	17.3%	34.0%	29.3%	31.5%
Elephant disease	36.0%	27.2%	30.9%	09.0%	06.0%	07.7%	33.0%	23.9%	27.8%
German measles	30.0%	33.6%	31.7%	18.0%	21.5%	19.6%	36.0%	42.6%	39.0%
HPV infection	30.0%	29.7%	29.8%	16.3%	15.0%	15.6%	27.0%	26.4%	26.7%
Leishmania	32.0%	32.4%	32.2%	21.0%	20.8 %	20.9%	38.0%	40.2%	39.1%
Lice	22.0%	23.9%	22.9%	13.0%	15.2%	14.0%	25.0%	28.4%	26.6%
Malaria	40.0%	53.1%	45.6%	21.2%	35.9%	26.7%	38.0%	51.4%	43.7%
Marsa	28.0%	29.6%	28.8%	12.0%	15.7%	13.6%	32.0%	38.8%	35.1%
Maltese fever	32.0%	45.2%	37.5%	09.0%	09.7%	09.3%	37.0%	50.7%	42.8%
Measles	43.0%	41.5%	42.3%	29.0%	27.6%	28.3%	46.0%	42.3%	44.0%
Meningitis	35.0%	35.3%	35.1%	16.0%	15.5%	15.7%	40.0%	34.5%	37.0%
Mumps	31.0%	49.0%	37.9%	16.3%	25.4%	19.8%	39.0%	62.9%	48.2%
Nipah virus	26.0%	40.4%	31.6%	15.0%	31.3%	20.3%	26.0%	44.4%	32.8%
Plague	28.0%	33.5%	30.5%	16.0%	17.8%	16.8%	34.0%	37.2%	35.5%
Rabies	31.0%	32.8%	31.9%	23.0%	24.6%	23.8%	33.0%	34.7%	33.8%
Schistosomiasis	22.0%	32.3%	26.2%	06.0%	08.4%	06.9%	33.0%	50.3%	39.9%
Seasonal flu	50.0%	52.6%	51.3%	11.7%	13.1%	12.3%	43.0%	47.4%	45.1%
Tuberculosis	34.0%	36.5%	35.2%	15.0%	13.9%	14.4%	32.0%	30.3%	31.1%
Yellow fever	25.0%	24.9%	24.9%	11.2%	14.1%	12.5%	34.0%	34.3%	34.2%
Zika virus	27.0%	23.3%	25.0%	20.0%	16.3%	17.9%	34.0%	28.7%	31.1%

Table 4: Term Extraction Performance.

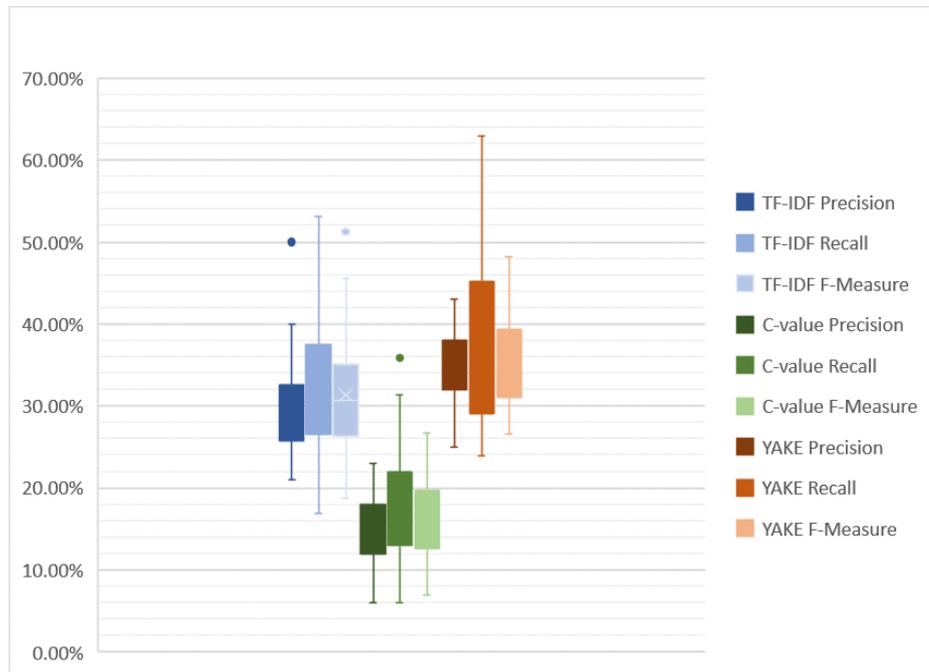


Figure 6: Average of Term Extraction Performance of Infectious disease

cross-lingual research by connecting the Arabic ontology to existing English ones. Finally, there is need to enhance the methods of Arabic Ontology learning of extracting terms and relations.

8. Acknowledgement

Our thanks to Dr. Saeed Dolgum, who is working as consultant medical doctor in pediatric infectious disease in Al Habib Hospital in Riyadh, Saudi Arabia, and has contributed to this paper by evaluating the Ontology.

9. Bibliographical References

- Al-Safadi, L., Al-Badrani, M., and Al-Junidey, M. (2011). Developing ontology for arabic blogs retrieval. *International Journal of Computer Applications*, 19(4):40–45.
- Al-Thubaity, A. M., Khan, M., Alotaibi, S., and Alonazi, B. (2014). Automatic arabic term extraction from special domain corpora. In *2014 International Conference on Asian Language Processing (IALP)*, pages 1–5. IEEE.
- Al-Zoghby, A. M., Elshawi, A., and Atwan, A., (2018). *Semantic Relations Extraction and Ontology Learning from Arabic Texts—A Survey*, pages 199–225. Springer International Publishing, Cham.
- AlAgha, I. M. and Abu-Taha, A. (2015). Ar2sparql: an arabic natural language interface for the semantic web. *AR2SPARQL: An Arabic Natural Language Interface for the Semantic Web*, 124(18).
- Albukhitan, S. and Helmy, T. (2013). Automatic ontology-based annotation of food, nutrition and health arabic web content. *Procedia Computer Science*, 19:461–469.
- Alsudias, L. and Rayson, P. (2019). Classifying information sources in Arabic twitter to support online monitoring of infectious diseases. In *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics*, pages 22–30, Cardiff, United Kingdom, 22 July. Association for Computational Linguistics.
- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., and Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. *Database*, 2018.
- Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C. (2006). Introducing the arabic wordnet project. In *Proceedings of the third international WordNet conference*, pages 295–300. Citeseer.
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., and Jatowt, A. (2019). Yake! keyword extraction from single documents using multiple local features. *Information Sciences*.
- Cimiano, P. (2006). Ontology learning from text. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*, pages 19–34.
- Collier, N., Goodwin, R. M., McCrae, J., Doan, S., Kawazoe, A., Conway, M., Kawtrakul, A., Takeuchi, K., and Dien, D. (2010). An ontology-driven system for detecting global health events. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 215–222. Association for Computational Linguistics.
- Cowell, L. G. and Smith, B. (2010). Infectious disease ontology. In *Infectious disease informatics*, pages 373–395. Springer.
- Frantzi, K., Ananiadou, S., and Mima, H. (2000). Automatic recognition of multi-word terms: the c-value/nc-value method. *International journal on digital libraries*, 3(2):115–130.
- Joshi, A., Karimi, S., Sparks, R., Paris, C., and Macintyre, C. R. (2019). Survey of text-based epidemic intelligence: A computational linguistics perspective. *ACM Comput. Surv.*, 52(6):119:1–119:19, October.
- Kapoor, B. and Sharma, S. (2010). A comparative study ontology building tools for semantic web applications. *International Journal of Web & Semantic Technology (IJWesT)*, 1(3):1–13.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.
- Musen, M. A. et al. (2015). The protégé project: a look back and a look forward. *AI matters*, 1(4):4.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Raisan, M. and Abdullah, A. M. (2017). *Building a Core Arabic Ontology for Iraqi News*. Noor Publishing.
- Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., Feng, G., and Kibbe, W. A. (2011). Disease ontology: a backbone for disease semantic integration. *Nucleic acids research*, 40(D1):D940–D946.
- Soliman, T. H., Elmasry, M., Hedar, A., and Doss, M. (2014). Sentiment analysis of arabic slang comments on facebook. *International Journal of Computers & Technology*, 12(5):3470–3478.