International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Using Ontology for Revealing Authorship Attribution of Arabic Text

Abeer H. El Bakly, Nagy Ramadan Darwish, Hesham A.Hefny

Abstract*: *Authorship attribution analysis is a research field that assigns an author to an unknown text based on writing features. These features reflect the author's gender, age, religion, education, job, motivation or ideology. It has several types of features such as character, lexical, Syntactic, Structural and Semantic. This research proposed using Arabic ontology as a semantic feature in authorship attribution through a proposed new model. In the Islamic society, there is a problem in detecting unknown fatwa to specific jurisprudence doctrine so this research proposed a new model for detecting unknown fatwa to specific jurisprudence doctrine. This model depends on a new corpus which is manually collected and annotated fatwas from books of Islamic jurisprudence doctrines. This corpus is called ElWafaa LlFokahaa. It includes the fatwas of traveller's prayer for main Islamic doctrines (Hanfi, Shafie, Malki, and Hanbali). The proposed model used Arabic ontology for traveller's prayer in each Islamic doctrine which is established with protégé framework. It is divided into a training set in 70% of fatwas (known fatwas the owing Islamic jurisprudence doctrines) and 30% testing set (unknown fatwas of Islamic jurisprudence doctrines). For evaluating the proposed model, it is used the proposed evaluated method which is 90% with final experiments.

Keywords* : *Artificial intelligence, fiqh, similarity, feature selection, ontology, authorship attribution.

I. INTRODUCTION

Authorship attribution (AA) is the process of assigning specific text to the specific author [1]. It started as a solving problem in the 19th century by statistical analysis methods such as using Bayesian statistical analysis of the frequencies of a small set of common words (e.g., 'and', 'to', etc.) and produced significant discrimination results between the candidate authors. Recently, Artificial intelligence techniques are used in solving this research problem [9][2]. These techniques include machine learning, information retrieval, and natural language processing learning. The researchers start the solution of this problem by determining the different features of the text content which should reflect the author's gender, age, religion, education, job, motivation or ideology. These features have many types of features such as character, lexical, syntactic, structural, content-specific, language-specific, and semantic [5]. This research used ontology for the first time as semantic features in authorship attribution. Ontology is included in domains of artificial intelligence, knowledge system and information system. Ontology provides the relationships between different lists of words in a particular domain. It includes object type, concept, attributes and relationships [10]. In Islamic society, people would like to know if a specific fatwa is owned to specific Islamic doctrine or not. This research used a proposed model for knowing who wrote the text (fatwa) of the main Islamic doctrines (Hanfi, Malki, Shafie, and Hanbali) by using ontology as a semantic feature. The rest of the paper is organized as follows: Section 2 presents the background about architecture of authorship attribution and definition of ontology; Section 3 presents related works which include semantic features in authorship attribution and Arabic authorship attribution; Section 4 presents a proposed model; Section 5 discusses experiments and results of a proposed model, and Section 6 shows the conclusion and future work.

Revised Manuscript Received on July 22, 2019.
* Correspondence Author
Abeer H. El Bakly *, Information system and technology department, Faculty of Graduate Studies for Statistical Research, Cairo University,. Email: abeerhassan012@gmail.com
Nagy Ramadan Darwish, Information system and technology department, Faculty of Graduate Studies for Statistical Research, Cairo University,.. Email: nagyrd@cu.edu.eg
Hesham A.Hefny, computer system department, Faculty of Graduate Studies for Statistical Research, Cairo University,

II. BACKGROUND

This section includes 2 subsections, the first one is authorship attribution architecture and the definition of ontology.

A. Architecture of Authorship Attribution

Figure 1 presents the architecture of solving this research problem that includes a set of steps as follows:

The first step is a corpus or dataset which may be a set of documents or novels or books or articles … etc. It should be chosen with high care to reflect the features of the author.

The second step is the extraction of the features from the author's text. These features have several types such as lexical features (a sequence of tokens which are grouped into sentences, each token corresponding to a word, number, or punctuation mark), character features (a sequence of characters), syntactic features and semantic features. Semantic features depend on the production of semantic dependency graphs that contain two kinds of information: binary semantic features and semantic modification relations. The tools can be used for extracting these features such as splitting sentence, POS (part of speech) tagging, text chunking, partial parsing ….etc (e.g., a nominal node with a nominal modifier indicating location).

To attribute these features there are two approaches for treating the dataset [9][2].

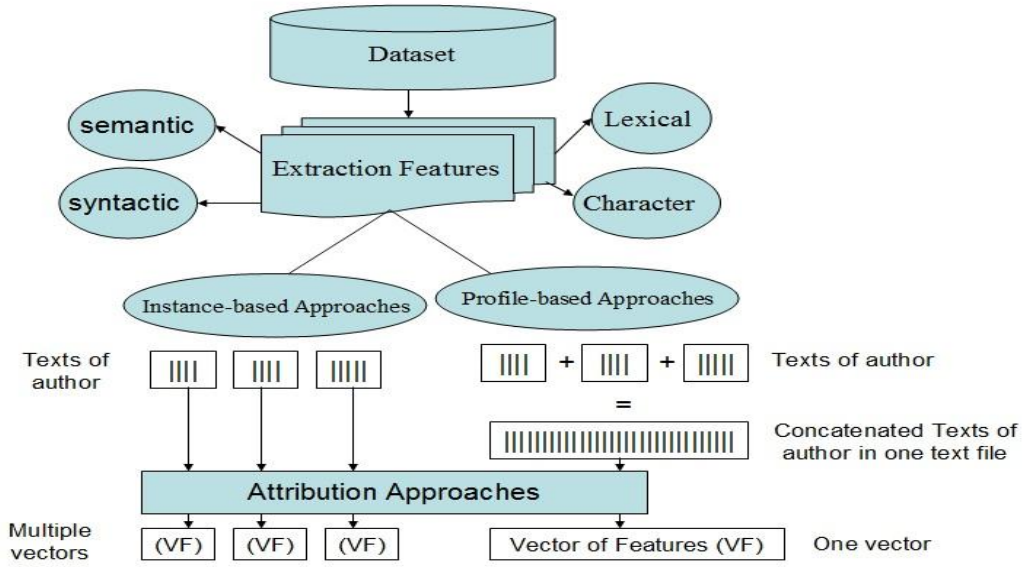

Fig. 1. Architecture of AA with extraction in main two attribution dataset methods

The first approach is profile-based which takes the documents of the author and merge them in one file then extraction the features from this file. The production of this approach is one vector of features. The second approach is instance-based which takes the documents of the author as one by one so the production of this approach is a set of vectors.

The last step of the architecture may take two choices, one of them using feature selection techniques to reduce the features to effective only then using the similarity methods. Or, it could use the similarity methods only such as Euclidean Distance, The Jaccard's Co-efficient, The Dice's Co-efficient and Cosine Similarity [9][2].

B. The Definition of Ontology

Ontology is a branch of philosophy which is considered the science of what is a thing. It includes the kinds and structures of objects, properties, events, processes, and relations in each area of reality. Philosophers used ontology as a synonym of 'metaphysics' (a label meaning literally: 'what comes after the Physics') [11].

Ontology is defined as O = {C, R, F, A, I}, where C is class or set of concepts, c is a concept (c ∈ C), which refers to everything such as work specification, function, behavior, strategy and reasoning process; R is a set of relationships, the interaction between concepts in domain, defines a subset of n-dimensional Cartesian product formally: R: C1 × C2 × • • • × Cn, r is relationship (r ∈ R), basic relationships contain: subclass of, part-of, kind-of, and attribute-of; F is function, a kind of special relationship. Formally F: C1 × C2 ×• • •×Cn−1 → Cn, such as Mother-of is a function, Mother-of (x, y) means y is mother of x; A is axiom, represents tautological assertion, like concept B, belongs to the range of concept A; I is set of instances, i is an instance (i ∈ I). Figure 2 presents a sample of ontology [10].

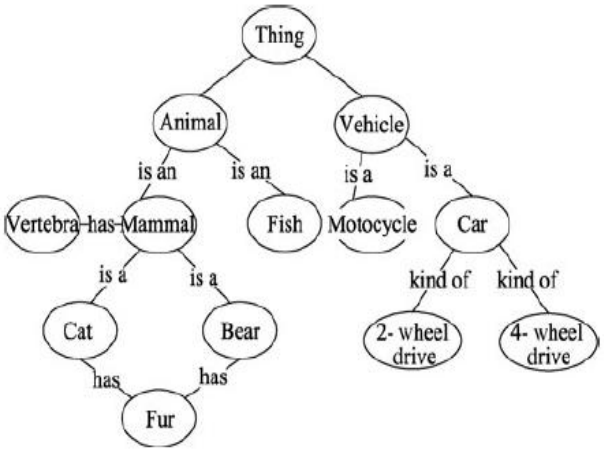

Fig. 2. The sample of ontology [10].

III. RELATED WORKS

Through reviewing the literature to stand on what other researchers have reached in this research area, the proposed approach of authorship attribution depends on Arabic language and semantic feature so the related studies are divided into two directions:

Arabic AA has several issues such as Omer and Oakes used the computer stylometric techniques of Hierarchical Cluster Analysis, Principal Component Analysis and Machine Learning to determine the argument about some chapters of book "The Liberation of Women" which is normally attributed to Qassim Amin and disputed chapters of this book were written secretly by Mohammad Abdu which were assessed by Mohamed Emara. They used the books (The Liberation of Women, The New Women and Mohammed Abdu's Opinion on Women) as a corpus. The features were extracted which were Character n-grams.

The results showed that the disputed texts were more similar to Qassim Amin's style than Abdu's style [12]. Al-Falahi et al measured the impact performance of Naïve Bays, Support Vector Machine and Linear discriminant analysis for Arabic poetry authorship attribution using text mining classification. Several features were extracted such as lexical features, character features, structural features, poetry features, syntactic features, semantic features, and specific word features are utilized as the input data for text mining, using classification algorithms Linear discriminant analysis, Support Vector Machine and Naïve Bays by Arabic Poetry Authorship Attribution Model (APAAM). The dataset of Arabic poetry is divided into two sets: known poetic in training dataset texts and anonymous poetic texts in a test dataset part. In the experiment, a set of 114 random poets from entirely different eras are used. The highest performance accuracy value is 99% for the Linear discriminant analysis [4].

Ahmed et al presented a solution for determining who the poet wrote an unknown text (Arabic poetry) by using style markers to identify the author by machine learning. They proposed public features in poetry such as characters, poetry sentence length; word length, rhyme, meter and the first word in the sentence were used as input data for text mining classification algorithms Naïve Bays (NB) and Support Vector Machine (SVM). They used 73 poets and it is considered little, also they need other features such as synonyms of words [3].

Using semantic feature type in authorship attribution is little. Zhang et al proposed a semantic association model based on word dependency relations, voice, and non-subject stylistic words for representing the writing style of different authors. They developed an unsupervised approach for extracting the word dependencies and patterns of semantic structures of a sentence. The different words or different syntactic patterns may have the same patterns of semantic structures. The types of semantic association features are confined neither to specific lexicons, phrases, and part-of speeches, nor to specific domains, topics and contents of texts. They developed a uniform vector space model to represent the semantic patterns of sentences then solving the problem of the independence of different dimensions to some extent. They used the context-free grammar for the language model which cannot represent the lexical and semantic dependencies between words in a sentence [13].

L´opez-Monroy et al presented a new method which is called Document Author Representation (DAR) for representing and classifying documents for AA. They proposed using the lexical richness of documents and relationships among terms, documents and authors for improving the representative. In this way, they were interested in relationships between authors and their terms, to define how a document is related to its author. In the DAR, document vectors in a space of authors build and the dimensionality will be limited by the number of authors. Besides, they proposed using the vocabulary richness in documents because the authors tend to write their documents with similar term repetition rates [14].

Al-Azani proposed using syntax and rhetorical Arabic styles as semantic features on Arabic AA. He proposed a set of 39 semantic features which consisted of the most popular grammatical and rhetorical Arabic styles. He built a new Arabic corpus which included selected newspapers' articles published in Alriyadh, Alhayat and Shorouk newspapers during the period from 2011 to 2013 written by a total of 20. The semantic features were extracted features and evaluated on this corpus. Then, the semantic features are tested by using different classification methods (ED, K-NN, MLP, LS-SVM, and SMO) [15].

IV. PROPOSED MODEL

The main problem of this research is assigning the unknown fatwa to one of the main Islamic jurisprudence doctrines (Hanfi, Malki, Shafie, and Hanbali). This research solves this problem by proposing a model using ontology of main jurisprudence doctrines as a semantic feature. Using an ontology as a semantic feature that is the pioneer in authorship attribution so it is considered the first contributions of this research. The second contribution is establishment a new corpus which includes the traveller's Prayer fatwas of main Islamic doctrines.

Figure 3 shows the main phases of the proposed model as follows:

In the first phase, the new corpus is divided into 70% for the training dataset and 30% for the test dataset.

In the second phase, using loop to process and extract features from each doctrine. The steps of this phases as follows: The first step starts with building ontology by training dataset for each doctrine (Hanfi, Malki, Shafie and Hanbali). In the pre-processing step, there are many tools are used for each process such as the Arabic token tool for tokenization, snowball stemmer for stemming the words, and number filter. In the next step, the processed text is used for extracting features such as a bag of words that is used to transfer text to terms then term frequency is used for calculating several occurrences of terms. In the last step, the document vector is used for transferring texts to vectors. The results of this phase are four vectors, each vector is specified for doctrine (Hanfi, Malki, Shafie, and Hanbali) with the same order as figure 3.

In the third phase, using the pre-processing tools for the testing set (unknown fatwas) such as Arabic token tool for tokenization, snowball stemmer for stemming the words, and number filter. In the feature extraction step, the bag of words is used for transferring each sentence to terms then the term frequency is used to calculate occurrence for each term. The last of this step using a document vector for transferring terms to vectors. The results of this phase are a vector of features that is used as a query by using row filter.

In the fourth phase, using a loop to detect the similarity between the vector of features which is resulted from the third phase (query) and each doctrine vector which is resulted from the second phase.

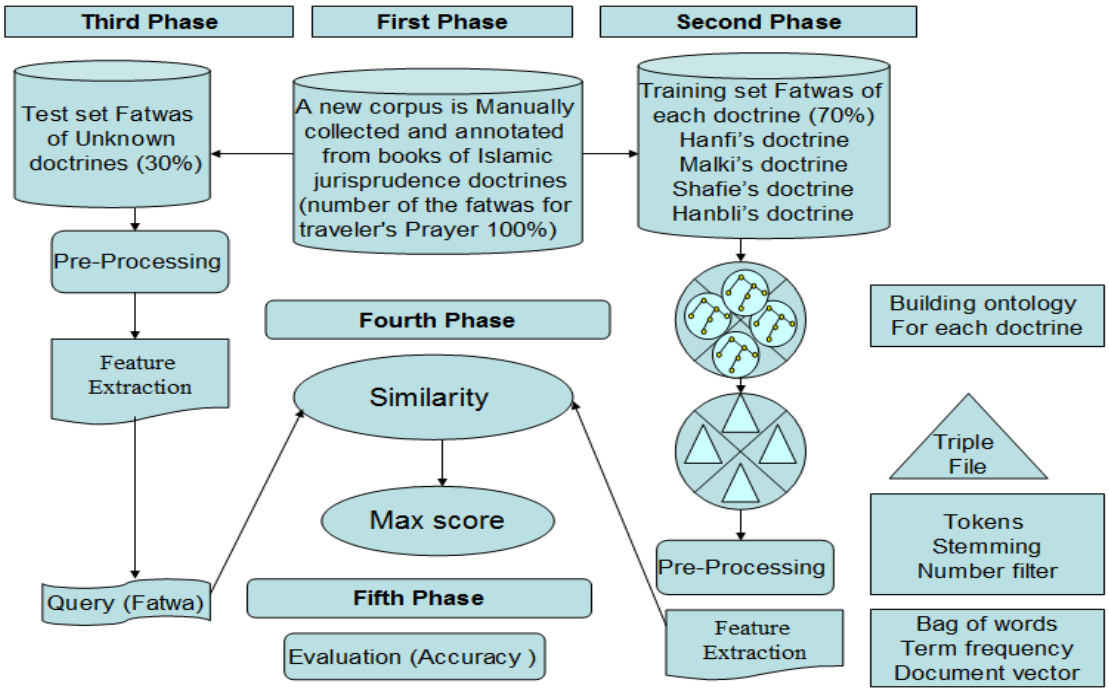

Fig. 3. Proposed model for assigning unknown fatwa to specific doctrine

The resulted of this step is four doctrine scores then selected the maximum score. The maximum score of similarity reflected the doctrine's name.

The fifth phase includes evaluation the proposed model by calculating proposed evaluated method.

Table. I. an algorithm of detecting author with ontology as semantic feature

Phase no.	Input	Process	Output
Phase I	The books of Islamic jurisprudence doctrines (Hanfi, Malki, Shafie and Hanbli)	Manually collected and annotated the fatwas	The number of fatwas of traveller's Prayer for each doctrine (100%)

Phase II	Training set Fatwas of each doctrine (Hanfi, Malki, Shafie and Hanbli) (70% of corpus)	Step1: building ontology for each doctrine (Hanfi, Malki, Shafie and Hanbli) with the same classes.
		// i is the number of doctrines
		Step2.1: for i in 0 ,1 … 3 do
		Step2.2: if (i<4)
		// taking each ontology
		Step 2.3: transferring ontology into triple file
		// preprocessed the triple file
		Step2.4: tokenize, stemming and number filter (the triple file)
		// transferring processed documents to features
		Step2.5: Bag of words, Term frequency and (processed documents) (processed documents)
		Step3: vector of features for Hanfi's doctrine, Malki's doctrine, Shafie's doctrine and Hanbli's doctrine.
			Four vectors of features.

Phase III	Test set Fatwas of Unknown doctrines (30% of corpus)	// preprocessed the csv file
		Step1: tokenize, stemming and number filter (the csv file)
		// transferring processed documents to features
		Step2: Bag of words, Term frequency and Document vector (processed documents)
		Step3: vector of features and raw filter for query of fatwa.
			Vector of features.

Phase IV	Vector of features for Hanfi's doctrine, Malki's doctrine, Shafie's doctrine and Hanbli's doctrine.	// i = 0, the name of doctrine Step1: for i in 0 ,1 … 3 do Step2: if (i<4) Step3:cosine similarity (query of fatwa , doctrine name) Step4: score of Hanfi's, Malki's, Shafie's and Hanbli's doctrine Step5:max(score of Hanfi's, Malki's, Shafie's and Hanbli's doctrine).	The name of doctrine.
Phase V	General queries of fatwas.	No. right answer of queries divided on number of all queries.	

Algorithm:

1. Get input unknown text(fatwa)(query)
2. Convert unknown text to vector of features.
3. Build ontology for each doctrine (known text).
4. Convert each ontology to vector of features.
5. Compute cosine similarity between unknown text and known text to get score for each doctrine.
6. Get maximum score from four values from step3 then get the name of doctrine.
7. Evaluate this model by dividing right answer over all queries.

V. EXPERIMENTS AND RESULTS

In these experiments, using two software as follows:
Protégé framework (https://protege.stanford.edu/) is used to establish ontology which is saved as a triple file.
KNIME platform (https://www.knime.com/) used the triple files which are created by protégé in other phases of the proposed model.

A. The corpus details

In this research, our corpus is established and it includes 1073 fatwa (255 fatwa for Hanfi, 269 fatwa for Malki, 279 fatwa for Shafie and 270 fatwa for Hanbali). These fatwas are extracted from books of each doctrine for the traveller's Prayer [6][7][8] … etc which were downloaded from the web site of el maktba el shamla(http://shamela.ws/). In addition, it is freely accessible and downloaded from this link https://www.dropbox.com/s/h9s9vr7aeknu8i2/ElWafaa%20LlFokahaa.zip?dl=0. This corpus is called ElWafaa LlFokahaa. It is split for each doctrine to a training set and test set with a percentage equal to 70% for training and 30% for the testing set. The results of pre-processing are as table 1.

Table. II. The details of ElWafaa LlFokahaa corpus

Doctrine	No. fatwas	No. training term	No. testing term
Hanfi	255	7461	3198
Malki	269	7693	3298
Shafie	279	7750	3320
Hanbli	270	6114	2621

Figure 4 presents an example of fatwas for a doctrine which is the opinion of Hanfi's doctrine. It includes one of the conditions for getting travel prayer license which means the travel should have a good or bad purpose for getting travel prayer license as the highlight red circle.

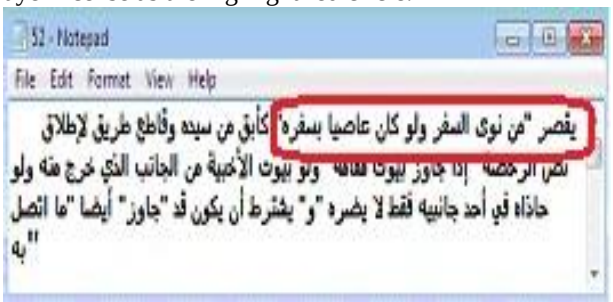

Fig. 4. The fatwa for Hanfi's doctrine

B. Pre-processing of corpus

In the proposed model, using KNIME for preprocessed the corpus in a set of steps:

1) Tokenization: splitting each sentence in fatwas to tokens (small units) such as words or characters.
2) Stemming: extracting the base of words to stem or root word such as ("يسافر", "المسافر","سافر") to "سفر"
3) Number filter: it is used Standard Arabic so number filter is mandatory for removing numbers.

C. Ontology as Semantic Feature

The concept of ontology is originated in philosophy. It is considered a semantic graph with a semantic relationship. The domain of ontology is "فقه صلاة المسافر " (Travel Prayer Jurisprudence) which is Arabic ontology. This research used top-down approach in building Arabic ontology and established it manually by helping domain experts in the field of " فقه الصلاة " (Prayer Jurisprudence) and the field of " الفقه المقارن " (Comparative Jurisprudence). This research built ontology for each doctrine so it includes four ontology (Hanfi, Malki, Shafie, and Hanbali). The protégé is used for building each ontology. The number of classes and subclasses of each doctrine is 19. Figure 5 shows the classes of Shafie's doctrine ontology with protégé.

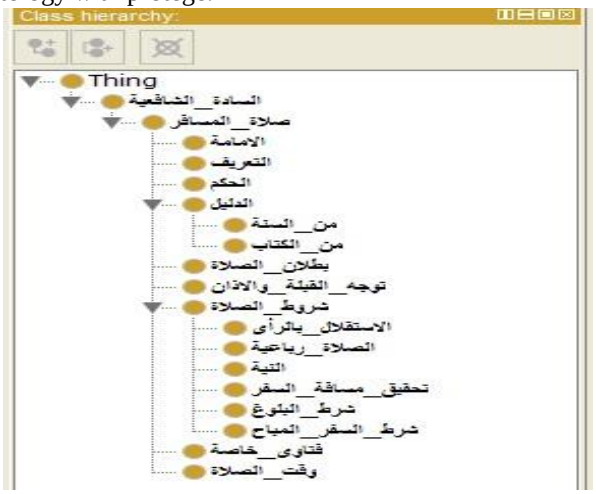

Fig. 5. The classes of Shafie's doctrine in protégé

The four ontology for doctrines includes the same classes and subclasses. The differences between four ontology in instances and relations between classes and instances as figures 5 and 6.

Fig. 6. Ontology instances of Shafie's doctrine in protégé

Figure 6 clarifies the part of instances of Shafie's doctrine in protégé, the highlight of instance presents ruling on praying "القصر رخصة وليس بعزيمة" "alkaser rokhasa wa laisa bazima". This means alkser is a license and not wajib.

Fig. 7. Ontology instances of Hanfi's doctrine in protégé

Figure 7 clarifies the part of instances of hanfi's doctrine in protégé, the highlight of instance presents ruling on praying "القصر عزيمة" "alkaser azima". This means alkaser is wajib.

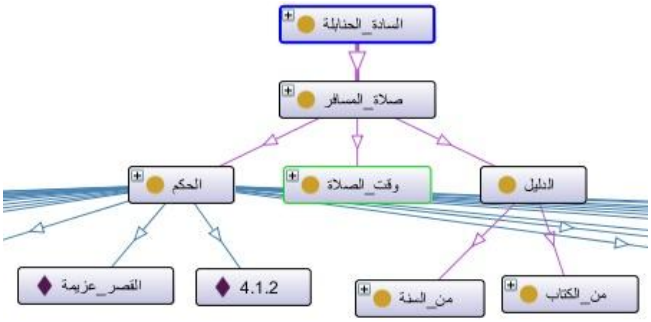

Fig. 8. Part of Ontology Concepts and Instances for Hanbli's doctrine.

The research proposed stored fatwas as instances of ontology in codes or short term. Figure 8 shows the code of fatwa which is presented as 4.1.2 that it means the doctrine number 4 (Hanbali), 1 is the number of class and 2 is several fatwa while the short term for the class is "القصر عزيمة" "alkaser azima".

Fig 9. fatwa as comment of instance which has code as a name.

The details of coding the fatwa are presented in figure 9. Code 4.1.1 refers to 3 parts, the first part is the Hanbali's doctrine, a second part is a class "الحكم" "alhokm" and the third part is the fatwa number … and so on. Each code includes a fatwa of doctrine which is stored in the comment of instance. This comment is "يجوز للمسافر إن شاء قصر وإن شاء اتم الصلاة" "yagoz llmosafer en shaa kasr wa en shaa atm alsalat.

D. Feature Extraction in Proposed Model

Figure 10 presents processing the ontology .ttl files by using KNIME. Each ontology file is represented as numerical vectors of features by using BOW which is a classical and common model for feature extraction. The bottom part of figure 11 presents the steps of extraction. The first highlight box is BOW which produces the processed file that includes a frequency of occurrence for each term. The highlight box in figure 11 is TF which is defined as the relative frequency of a term in the [16]. Then, using a document vector to change the processed file to term vectors.

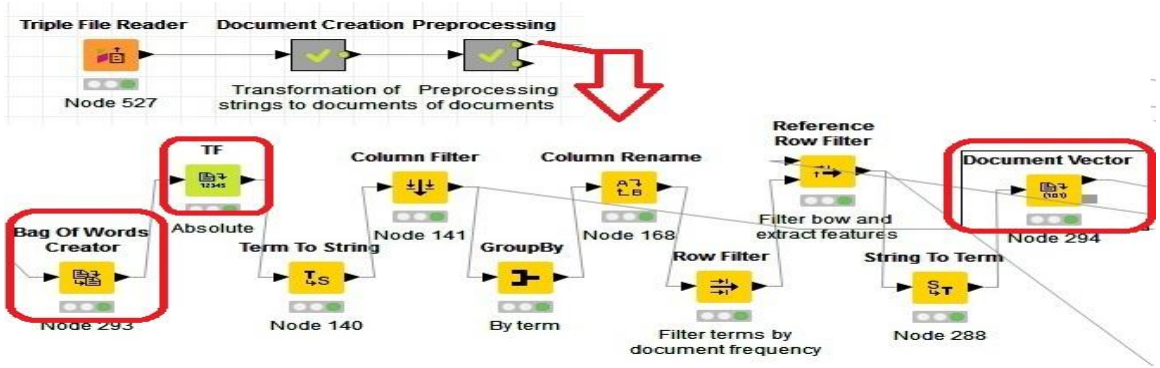

Fig. 10. The processed of feature extraction in KNIME

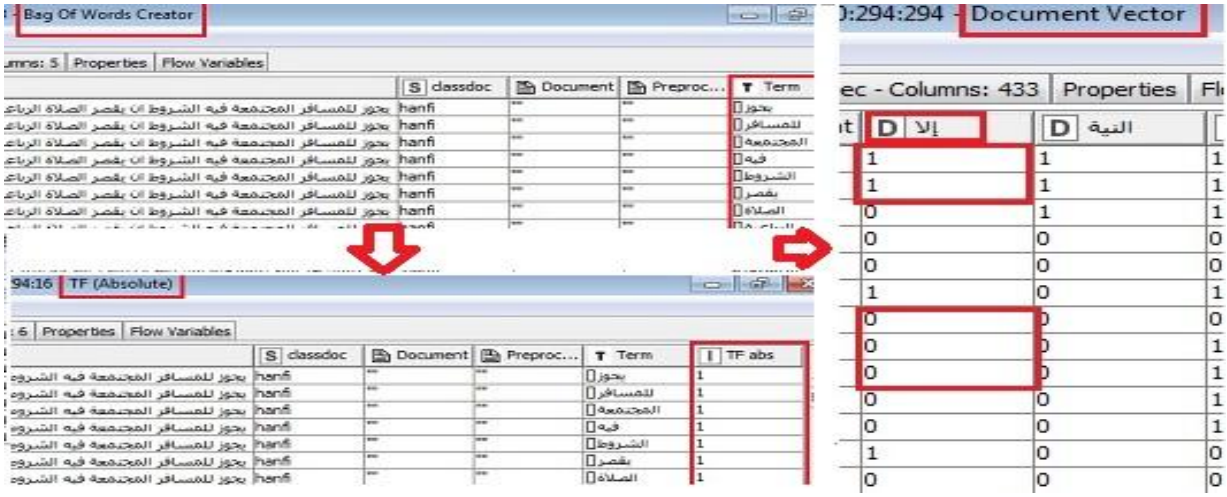

Fig. 11 The steps of feature extraction in KNIME

E. The Similarity Algorithm

Cosine similarity is used for measuring the similarity between two vectors of an inner product space that measures the cosine of the angle between them. It is calculated by using KNIME. This distance function is 1 – cosine similarity [17].

The configuration of cosine similarity is as follows: The neighbour selection is nearest (most similar) and the number of neighbours is 3 which is selected the maximum from them for a calculated score for similarity. The similarity (1-distance) – only for Tanimoto is used for coefficient type.

Table-III (A): The results of queries for proposed evaluation

Q no.	Hanfi	Malki	Shafie	Hanbli	Q A V	Q A	Q A S
q 1	0.648	0.509	0.591	0.394	0.648	hanfi	TRUE
q 2	1.000	0.539	0.866	0.553	1.000	hanfi	TRUE
q 3	0.981	0.538	0.401	0.655	0.981	hanfi	TRUE
q 4	0.816	0.585	0.401	0.521	0.816	hanfi	TRUE
q 5	0.505	0.434	0.408	0.587	0.587	shafie	FALSE
q 6	1.000	0.378	0.482	0.542	1.000	hanfi	TRUE
q 7	1.000	0.430	0.447	0.504	1.000	hanfi	TRUE
q 8	1.000	0.577	0.564	0.474	1.000	hanfi	TRUE
q 9	1.000	0.437	0.581	0.577	1.000	hanfi	TRUE
q 10	0.583	0.649	0.627	0.533	0.649	malki	FALSE
q 11	1.000	0.507	0.569	0.451	1.000	hanfi	TRUE
q 12	1.000	1.000	0.589	0.452	1.000	hanfi	TRUE
q 13	1.000	0.445	1.000	0.545	1.000	hanfi	TRUE
q 14	0.507	0.504	0.603	0.514	0.603	shafie	FALSE
q 15	1.000	0.478	0.480	0.744	1.000	hanfi	TRUE
q 16	1.000	0.514	0.527	0.749	1.000	hanfi	TRUE
q 17	1.000	0.535	0.512	0.558	1.000	hanfi	TRUE
q 18	1.000	0.423	0.476	0.459	1.000	hanfi	TRUE
q 19	1.000	0.621	0.495	0.515	1.000	hanfi	TRUE
q 20	1.000	0.500	0.577	0.516	1.000	hanfi	TRUE
q 21	1.000	0.500	0.471	0.494	1.000	hanfi	TRUE
q 22	1.000	0.486	0.591	0.499	1.000	hanfi	TRUE
q 23	1.000	0.494	0.418	0.600	1.000	hanfi	TRUE
q 24	1.000	0.567	0.566	0.434	1.000	hanfi	TRUE
q 25	1.000	0.566	0.559	0.669	1.000	hanfi	TRUE

Q A V(query answer value)

[a] Q A(query answer)

[b] Q A S(query answer status)

Table-III (B): The results of queries for proposed evaluation

Q no.	Hanfi	Malki	Shafie	Hanbli	Q A V	Q A	Q A S
q 26	0.481	0.516	0.492	0.435	0.516	malki	FALSE
q 27	1.000	0.436	0.436	0.651	1.000	hanfi	TRUE
q 28	1.000	0.620	0.474	0.568	1.000	hanfi	TRUE
q 29	1.000	0.615	0.612	0.454	1.000	hanfi	TRUE
q 30	0.627	0.542	0.542	0.525	0.627	hanfi	TRUE
q 31	1.000	0.481	0.462	0.476	1.000	hanfi	TRUE
q 32	1.000	0.463	0.577	0.476	1.000	hanfi	TRUE
q 33	1.000	0.546	0.554	0.398	1.000	hanfi	TRUE
q 34	0.630	0.568	0.572	0.477	0.630	hanfi	TRUE
q 35	1.000	0.538	0.390	0.617	1.000	hanfi	TRUE
q 36	1.000	0.463	0.499	0.553	1.000	hanfi	TRUE
q 37	1.000	0.556	0.468	0.490	1.000	hanfi	TRUE
q 38	1.000	0.459	0.445	0.490	1.000	hanfi	TRUE
q 39	1.000	0.584	0.471	0.417	1.000	hanfi	TRUE
q 40	1.000	0.462	0.408	0.543	1.000	hanfi	TRUE

a. Q A V(query answer value)

b. Q A(query answer)

c. Q A S(query answer status)

Table 2 (part A and B) presents 40 queries (unknown fatwas), it is applied in the proposed model for each query then the answer value is the name of doctrine. The details of table 2 are as follows:

The first column presents the number of queries.

The second column presents the similarity between unknown fatwa (query) and hanfi's doctrine, the third column presents the similarity between unknown fatwa (query) and malki's doctrine, the fourth column presents the similarity between unknown fatwa (query) and shafi's doctrine, the five-column presents the similarity between unknown fatwa (query) and hanbli's doctrine.

The sixth column (query answer value) presents the maximum value of four doctrines, the seven column (query answer) presents the answer of query which is doctrine's name.

The eighth column presents value query answer status which takes two values true or false according to the opinion of a domain expert in Islamic jurisprudence doctrines. A number of all queries is 40, the true answer of queries is 36, the false answer of queries is 4.

The proposed evaluated method = the true answer of queries/ number of all queries (1)

The proposed evaluated method = (36/40)x100 = 90%.

Table-III: The evaluation for proposed model

No. query	No. right answer	No. wrong answers	Proposed evaluated method(%)
40	36	4	90%

VI. CONCLUSION

The main problem of this research is who wrote the unknown text (fatwa) from the known text (main Islamic jurisprudence doctrines). So this research proposed a new model for solving this problem using ontology as a semantic feature in the authorship attribution which is the main contribution in this research. In addition, this research presents a new corpus is called ElWafaa LlFokahaa which is another contribution. It is manually collected and annotated the fatwas of traveller's prayer from books of Islamic jurisprudence doctrines. An ontology for each Islamic jurisprudence doctrines is built from the corpus by helping the expert domain by using protégé framework. Then, using the KNIME to process the other steps in experiments. The experiments include four scores from the cosine similarity between two vectors of features which are a query (unknown text (fatwa)) and known text then getting the maximum score which has the name of doctrine. To evaluate the proposed model testing 40 queries and check the truth the answers by expert domain. The proposed method for evaluation is 90%.

In the future work, ElWafaa LlFokahaa dataset will be expanded to include the rest of all jurisprudence doctrines then, a proposed model is applied in the expanded dataset.

REFERENCES

1. A. Gungor,. "Benchmarking Authorship Attribution Techniques using over a thousand books by Fifty Victorian Era Novelists Investigating ", (Master degree). Purdue University, Indianapolis, Indiana, 2018.
2. K. Shaker, "Investigating Features and Techniques for Arabic Authorship Attribution", (Doctoral dissertation). Heriot-Watt University, Malaysia, 2012.
3. A. Al-Falahi, M. Ramdani, M. Bellafkih, (2017). Machine Learning for Authorship Attribution in Arabic Poetry. International Journal of Future Computer and Communication(IJFCC), 6(2), , pp. 24-46.
4. A. Al-Falahi, M. Ramdani, M. Bellafkih, (2019). Arabic Poetry Authorship Attribution using Machine Learning Techniques. Journal of Computer Science, 15(7), pp. 1012-1021.
5. A.S. Altheneyan, M.E. Menai, (2014). Naı¨ve Bayes classifiers for authorship attribution of Arabic texts. Journal of King Saud University, Elsevier, 26(1), pp.473-484.
6. A. Abd AlRahman, *كتاب الفقه على المذاهب الاربعة*.[Book of Jurisprudence (Fiqh) on the Four Doctrines]. part one, second edition, Dar El Kotb Elalmia, Labnan, Bairot, 2003.
7. A.M.E.E.Al Kortaby,*المجتهد ونهاية المقتصد بداية* [Badiat Al Mogthd wa Nahiet Al Moktsd]. Part one, six edition, Dar El Marafa, Bairot, Lebanan.
8. M.A.E.E.Ebn Kodama, *المغنى*. [Al Moghni]. A brief explanation of the immortal. part one, 1223.

9. E. Stamatatos, (2009). A Survey of Modern Authorship Attribution Methods. Journal of the American Society for Information Science and Technology (JASIST), 60(3), PP. 538-556.
10. L. Lei, Y. Feng., Z. Peng, W. Jing-Yi, H. Liang, (2012) .SVM-based Ontology Matching Approach. International Journal of Automation and Computing, 9(3), pp. 306-314.
11. B. Smith, "Ontology" in L. Floridi (ed.), Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, 2003, pp. 155–166.
12. A.I.A.Omer, M.P.Oakes, "Stylometric Comparison of Writings by Qassim Amin and Mohammed Abdu on Women ' s Rights", In Proceedings of the 3rd Workshop on Arabic Corpus Linguistics, Cardiff, United Kingdom, 2019, pp. 1-6.
13. C. Zhang., X. Wu, Z. Niu, W. Ding, (2014). Authorship identification from unstructured texts. Knowledge-Based Systems, Elsevier, vol. 66, pp. 99-111.
14. A.P. López-Monroy, M. Montes-y-Gómez, L. Villaseñor-Pineda, J.A. Carrasco-Ochoa, J.F. Martínez-Trinidad, "A New Document Author Representation for Authorship Attribution", Dig. Conf Mexican on Pattern Recognition. Lecture Notes in Computer Science, (7329), Berlin, Heidelberg: Springer, 2012.
15. S. H. M. Al-Azani,, "Authorship Attribution of Arabic Texts." (Master degree). KING FAHD UNIVERSITY of PETROLEUM & MINERALS, SAUDI ARABIA, 2014.
16. G. Tripathi, S. Naganna, (2015). Feature Selection and Classification Approach for Sentiment Analysis. Machine Learning and Applications, Machine Learning and Applications: An International Journal (MLAIJ), 2(2).
17. https://www.knime.com/, last accessed 1/1/2020

AUTHORS PROFILE

Abeer Hassan received her MCs degree in Department of Information Systems and Technology from the Institute of Statistical Studies and Research, Cairo University, Egypt. She is PHD student in Information & Technology Systems, Faculty of Graduate Studies for Statistical Research, Cairo University.

Nagy Ramadan Darwish received his PhD. in Information Systems from Faculty of Computers and Information, Cairo University, Egypt. He is an Associate Professor and Acting Head of Department of Information Systems and Technology, Faculty of Graduate Studies for Statistical Research, Cairo University. He is a reviewer in many national and international conferences and Journals such as: IJCSIS, IJACSA, IJARAI, and IJST. He is an editorial board member of Circulation in Computer Science. He published about 90 papers in International Journals and conferences. He is a Consultant of Software Project Management, Software Quality, Business Information Systems, Quality of Education, and Institutional Development.

Hesham Hefny received his Bsc, MSc and PhD degrees all in electronics and communication engineering from Cairo University in 1987, 1991, 1998, respectively. Currently, he is a Professor of computer science and the Vice Dean for Graduate Studies at Faculty of Graduate Studies for Statistical Research, Cairo University. His research of interest include fuzzy systems, artificial neural networks and granular computing.