Arguments extraction for e-health services based on text mining tools

Dena Kadhim Muhsen[1], Sura Mazin Ali[2], Rana M. Zaki[3], Amjed Abbas Ahmed[4]

[1]Computer Science Department, University of Technology-Iraq, Baghdad, Iraq
[2]College of Political Sciences, AlMustansiriyah University-Iraq, Baghdad, Iraq
[3]Computer Science Department, University of Technology- Iraq, Baghdad, Iraq
[4]Imam Al-Kadhum College(IKC)/Iraq

ABSTRACT

The task of recognizing arguments and their components in text is known as argument extraction. Most arguments might be broken down into a petition and at least one premise that support it. A method to extract arguments is suggested in this work. The major words which are of high importance in arguments extraction were included in the suggested method on the basis of Arabic lexicon. The lexicon tool was used to apply classic text mining stages. The dataset, which includes over 3000 petitions, was collected from the Citizen Affairs Department in the Ministry of Health-Iraq. In addition, the experimental results exhibit that the suggested method extracts arguments from collected dataset with a 93.5% accuracy ratio.

Keywords: Argument Extraction, Lexicon-Based, Text mining

Corresponding Author:

Dena Kadhim Muhsen
Computer Science Department, University of Technology
Baghdad, Iraq
Dena.k.Muhsen@uotechnology.edu.iq

1.   Introduction

E-government is the application of information technologies to act as a bridge between individuals, government agencies, and non-governmental organizations. The previously mentioned technologies might be highly significant in providing individuals with various government services, establishing interactions with different sectors, facilitate individual's access to the information they are permitted to have, and ultimately increasing government administration efficiency. Fighting corruption, decreasing waste of money and time, achieving transparency, providing comfort to citizens, and increasing state revenues are some of the advantages of using e-government[1]. E-government has been developed as the most important factor in providing citizens with distinct, dynamic, effective, clear, and accountable public services. The major aim of e-government is providing services to public via extensive use of networks and IT, while also taking into account the development of community trust, which leads to an improvement in the government-citizen interaction. Citizens are provided by e-government with a number of significant advantages, one of which is to reduce the time and cost of completing various government transactions, as well as to make them available seven days a week and around the clock[2, 3]. Understanding the requirements of citizens via governments is the major node in the process of progressing in the service type provided through e-government, which is heavily impacting the reduction in expenditures. Regrettably, gathering such vital information is difficult[4]. Argumentation between groups and individuals is utilized in everyday life for disproving or proving many arguments using a slew of evidence and facts. The number of individuals utilizing social media has constantly increased recently, with more individuals presenting and arguing many arguments to impact each other's opinions, making evaluating the arguments in such texts a critical issue. The argument might be indicated briefly and simply as a petition (or premise) backed up via the supporting evidence [5, 6]. Argument's mining can be defined as the process for exploring arguments and their major components within text. Many algorithms specializing in argument mining make efforts to recognize the major argument' elements, such as premises and petitions, in texts written in various languages and disciplines. Along with the foregoing, the methods of sentiment analysis, which were rapid in determining the human feelings' direction regarding certain issue, might extract a great deal of useful

information from various texts. The significance of understanding public opinions on a certain issue in general, and particularly in e-government and policy-making isn't hidden to any one [7]. Certain approaches that work on textual data are used for extracting information from a collection of texts. Text Mining (TM) can be defined as the collection of such methods which combines them with a few procedures of Natural Language Processing (NLP) which are specialized in text structures. TM methods impose certain limitations on the tools of NLP because they need large amounts of textual data. TM actions on textual bases include knowledge or information clustering and summarization, exploration of structures between various groups and elements, discovery of hidden links between textual components, along with a general review regarding a large document[8-13]. TM employs a variety of procedures for locating connections and structures within text. A few TM methods need target text classification before applying classical Data Mining (DM) methods, whereas others are exploring the whole target text [11]. There is a collection of data extracted with the use of TM methods, like a set of words serving as tags for each one of the documents, and extracting word sequences from a target document [12, 13]. In the presented work, we suggested a method for extracting arguments from citizen petitions that is on the basis of Arabic lexicon that includes the key words that are significant in arguments extraction. The lexicon tool has been used to apply the traditional TM stages. The remaining sections of this work are organized in the following manner. The second section discusses related works. A brief explanation related to argument extraction, along with its principles is provided in section three. The suggested system, which includes implementation and design, is presented in section 4. Section 5 describes the results and experimental setup. In section 6, we wrap things up with a summary.

2. Related works

There are some studies related to arguments extraction; we have selected a few of these researches:

Jasim et al., suggested the concept of argument extraction and differentiation based on supervised learning, while their prototype is applied to Arabic texts containing legal information, making it one of the first in the field. The arithmetical model utilized in this study is collect and collaborates among the fundamental bulk of Arabic legal text[14]. In [15], the authors suggested a concept of unsupervised learning-based information extraction and analysis, with their prototype being applied to Arabic texts containing legal information. The extracted information from legal Arabic text is categorized into 2 types; the first is referred to as valuable qualities, and the second is referred to as worthy information. Palau et al. suggested a process of categorization for the process of finding arguments in legal texts. The model utilized in this study analyses the relation between sentences as a foundation for high accuracy argument exploration, resulting in an increase in overall accuracy of approximately 8% when put to comparison with other models[16]. Moens et al., suggested a process of classification for extracting arguments from legal texts, which they applied to a set of explained arguments. Syntax, linguistic vocabulary, semantics and conversation are among the properties of legal texts that are assessed[17]. Poudyal et al., suggested a method for extracting arguments from legal texts using a new clustering algorithm. The major flaw in such model is that a sentence's argument might be associated with and a part of another argument. Fuzzy logic is utilized for solving such problem, which is indicated via Fuzzy c-means (FCM) clustering algorithm[18].

3. Argument extraction

Personal conviction is a complex and spiky phenomenon that is formed by a collection of human conclusion factors, including perspectives, evaluation and opinions on various topics. Worse, human conclusions are not often possible to determine whether they are false or true. In any case, individuals are more likely to accept other person's conclusions if they were backed up by compelling evidence. We can see why people are interested in extracting arguments from various text types [19]. The basic components related to argument are conclusion and construction, and argument extractions were concerned with discovering, indicating, and mining the so-called components from natural language text. Argument extraction is a step in a machine's comprehension of natural human languages[20]. Arguments can be found in individuals' everyday lives, and they are using logical approach for the purpose of reaching conclusions and understand their causes [6, 21]. Arguments are widely used in a variety of fields, including law, economics, sports, politics, and culture. Politicians, their assistants and advisors, for instance, spend a lot of time to search for and analyzing various types of data in the hopes of coming up with an adequate set of arguments for supporting their petitions or refuting the petitions of others in certain cases. As a result, the argument extraction suggested to simplify such difficult issues by focusing on the

fundamental components of arguments (petitions and premises) and their relationships. Also, politicians might be persuaded for accepting a policy via extracting arguments from public statements and using them to refute or support a few issues which were incompatible or compatible with the policy[22-26]. In a different context, and particularly in field of law, specialists (lawyers and judges) devote a significant amount of time and efforts to searching for and analyzing information relevant to defendants' cases for assisting them in extracting convincing arguments that refute or support their petitions. The argument extraction, on the other hand, is utilized for simplifying complex issues. In addition, legal professionals might be assisted in rejecting or accepting the collection of arguments [14, 27].

4. Proposed system

The suggested system involved the implementation and creation of a website that would receive citizen petitions and process them automatically for extracting arguments from each one. The process of extracting arguments from the application is critical, as it provides evidence and credibility of citizen's requirements to carry out his petition. A significant feature for decision-makers and for strengthening e-governance is arguments' extraction from the texts of citizens' petitions. The process of extracting arguments from citizen petitions' texts isn't simple, yet it was made possible thanks to a number of factors, including a few lexicon style capabilities, the support of TM approaches and lastly the implementation of a set of logical steps.

Figure 1 shows a block diagram related to the suggested system for extracting arguments from citizen petition texts. In addition, pre-processing is the first stage, and it consists of some steps like tokenization, normalization, removing punctuations, and removing stop-words. The Features Extraction stage is the second stage, and it entails extracting the significant properties which are vital in the argument's extraction stage. The third stage, Arguments Extraction, is the major one since it is representing the result of previous stages and determines the failure or success regarding the whole process on the basis of what is acquired. Lastly, there is the system evaluation stage, which evaluates the entire process. The aforementioned stages will be discussed in greater depth in the next sections.
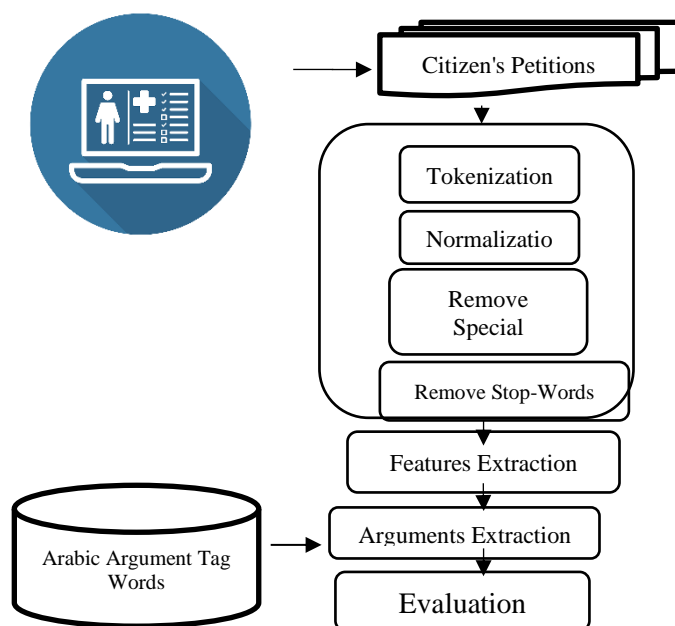

Figure 1. Proposed system

4.1 Pre-processing phase

This phase is a necessary preparatory procedure that involves many steps and is carried out at the start of the algorithm. Tokenization is the first step, which divides the input text into smaller chunks, such as words. The second step entails the removal of punctuation. This step is critical because it removes all punctuation types from

the tokenized text, including (:"!? ; - "/*/...[ ],). The unwanted words, as shown in Table (1), will be removed in the third step with the use of a list of regular expressions.

Table 1. List of regular expressions

No	Regular Expression	Results
1	[a-zA-Z]+	Remove English characters
2	[0-9]+	Remove English numbers
3	[0-9]+	Remove Arabic numbers
4	#$%|@^~(&*)+	Remove another

Lastly, the fourth step entails Arabic words' normalization, which is the major step in pre-processing phase due to the fact that the Arabic language has numerous writing shapes. For performing normalization, the suggested approach employs 3 operations: Diacritics removal, Tatwheel removal, and Latter removal. Starting with the removal of diacritics from Arabic word and progressing through the conversion of an alphabetic word to another. Figure 1 shows an instance of such procedure (2). If these diacritics aren't removed, we'll end up with a lot of different shapes for the same words, which is going to make the vector very large and requires more articles to build, which is going to take longer.

Figure 2. The diacritics have to eliminate

In Arabic, a tatwheel is applied for making characters appear longer than others. The primary reason for the use of tathweel is making the word's shape more attractive. This, yet, is an issue since the same words might be written in a variety of ways. The final step in the normalization process will result in a distinctive letter that corresponds to a set of letters. After completing all the above steps, remove any words which are only 2 characters long, as 2-character words in Arabic don't make sense. Lastly, any token which matches the word in the stop words list can be removed from an article to remove the Arabic stop word. The NLTK library includes a stop words list. All words in NLTK are sign words.

4.2 Features extraction phase

What remains after removing stop-words and other parts indicated in the Pre-processing Phase are clear and pure words like (adjectives, verbs, nouns and adverbs). The words containing the properties of the argument are going to be extracted at this point. The major significant words handling arguments were extracted from expert opinion in Arabic language, and a distinctive dictionary of words was utilized. In addition, a few Iraqi dialect words related to argument sources were added. The lexicon of arguments is critical in specifying the beginning of arguments-related causes. An arguments lexicon is useful for determining the beginnings of sentences that contain an argument. The sentence regarding the arguments in the texts of citizens' petitions might be predicted via identifying the words associated with arguments. From the foregoing, we can see how important this stage is in extracting arguments from the texts of citizen petitions, as well as the majority role it plays.

4.3. Arguments extraction phase

At this stage, words associated with arguments were identified as well as extracted, while the argument sentence in the text, which indicates the citizens' petition, is specified and extracted. It is possible to determine the end of the sentence representing the argument and whose beginning has been indicated from the properties and which was indicated in previous stage by relying on a few NLP principles. In addition, the sentence which is the argument in text is determined and extracted at such stage via specifying the words that are and extracting the words associated with arguments. It is possible to determine the end of the sentence which indicates

the argument and whose beginning has been indicated from the properties and which was represented in the previous stage using a few NLP principles.

In the Arabic language, a group of words which indicated the argument has been utilized. In the Arabic language, Table (2) is showing the examples of words associated with argument. These were the words which appear the most frequently in the argument we're trying to extract. Table (2) contains keywords that were utilized for extracting the argument from citizen petitions. As a result, the suggested study's foundation is lexicon-based, with these words serving as the lexicon's foundation.

Table 2. Samples of Arabic argument tag words (translated to English)

That leads	That causes	To be	considered as	Because
Whereas	To our need	For being	Then	It said

The words in Table (3) have been verified on the basis of petitions received from sources interested in the subject, also the assistance of Arabic language experts and those interested in Arabic language affairs, particularly in the area of argument. If citizens use words not found in the dictionary, new words can be added to the proposed lexicon, particularly if the dialects utilized in the petitions differ.

5. Experimental results

Our dataset has been collected from the Citizen Affairs Department in the Ministry of Health-Iraq. More than 3000 Petitions have been collected about several affairs (Chronic disease treatments, rare disease treatments, Cancer treatments, prosthetic limbs, Special medical supplies and Treating victims of terrorist bombings), Table (3) includes the total number of petitions for each department. Several words of the Iraqi colloquial dialect were taken from[28].

Table 3. Types and numbers of citizen's petitions

No	Department	No. of Petitions
1	Chronic Disease Treatments	575
2	Rare Disease Treatments	540
3	Cancer Treatments	760
4	Prosthetic Limbs	340
5	Special Medical Supplies	115
6	Treating Victims of Terrorist Bombings	890
	Total	3220

According to many of citizens' petitions who neglected the arguments when it's wrote the petition led to lack arguments for various reasons and cause the major problem. the most significant reasons in the E-health applications such as the citizens do not have the following:

- argument for his petition.
- Knowledge about the argument formulation of his petition.
- the citizen inadvertently neglected writing the argument for his petition.

The group of the 3220 applications were collected from Citizen Affairs department in the Ministry of Health-Iraq shown in Table (3). there were only 1771 applications containing arguments approximately 45 % and 1449 applications that did not contain arguments about 45%. Additionally, among 1771 applications containing

arguments, the accuracy percentage of valid arguments was (93%), meaning (1647) applications. Table (4) shows the numbers and types of applications that contain arguments and the number that the system extracted from those petitions.

Table 4. Argument extraction from citizen's petitions

No	Department	No. of Petitions	No. of Petitions That Including Argument	No. of Automatic Argument Extraction
1	Chronic Disease Treatments	575	315	297
2	Rare Disease Treatments	540	206	193
3	Cancer Treatments	760	387	348
4	Prosthetic Limbs	340	183	175
5	Special Medical Supplies	115	71	53
6	Treating Victims of Terrorist Bombings	890	609	581
	Total	3220	1771	1647

Table (4) shows the number of applications from which the arguments were extracted and their total was 1647 out of 1771 with a success rate of 93% as an accuracy ratio. As a result, the number of applications containing arguments is relatively small compared to the number of applications. the following equation represent the accuracy ratio computation:

$$Accuracy = \frac{No.\,of\ Corrected\ Extracted\ Arguments}{Total\ No.\,of\ Arguments} \, x100 \ \dots\dots\dots(1)$$

6. Conclusion

The proposed system extracts arguments from citizens' petitions submitted to service foundations and others. The basis for the proposed system is the lexicon, which contains the basic words that play an important role in extracting arguments from the texts. The system was applied on real data taken from the Citizen Affairs Department in the Ministry of Health-Iraq. The system achieved a accuracy rate of 93% The most important reason influencing the data collected is that the citizen does not write the arguments on which his application is based, and as we mentioned for that, there are several reasons. The dictionary that was adopted for the words of the arguments played a big role in extracting those arguments correctly. There is no doubt that the issue of extracting arguments from the texts is not an easy task with the different dialects in the texts of Iraqi citizens. As a future work, we suggest using deep learning to explore words that are more profound and effective in extracting arguments from Arabic texts.

References

[1] M. Katsonis and A. Botros, "Digital government: a primer and professional perspectives," Australian Journal of Public Administration, vol. 74, no. 1, pp. 42-52, 2015.

[2] A. M. Samsor , "Challenges and Prospects of e-Government implementation in Afghanistan," International Trade, Politics and Development, 2020.

[3] I. Astawa and K. Dewi, "e-Government facilities analysis for public services in higher education," in Journal of Physics: Conference Series, 2018, vol. 953, no. 1, p. 012061: IOP Publishing.

[4] J. C. Bertot, P. T. Jaeger, and C. R. McClure, "Citizen-centered e-government services: benefits, costs, and research needs," in Proceedings of the 2008 international conference on Digital government research, 2008, pp. 137-142.

[5]   P. Poudyal, Matthias Thimm, "Automatic Extraction and Structure of Arguments in Legal Documents," Sarah A. Gaggl, Matthias Thimm (Eds.), p. 19, 2016.

[6]   E. Florou, S. Konstantopoulos, A. Koukourikos, and P. Karampiperis, "Argument extraction for supporting public policy formulation," in Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, 2013, pp. 49-54.

[7]   C. Sardianos, I. M. Katakis, G. Petasis, and V. Karkaletsis, "Argument extraction from news," in Proceedings of the 2nd Workshop on Argumentation Mining, 2015, pp. 56-66.

[8]   J. Abutridy, "Text mining: principles and applications," Revista Facultad de Ingeniería, no. 7, pp. 57-62, 2000.

[9]   D. Tao, P. Yang, and H. Feng, "Utilization of text mining as a big data analysis tool for food science and nutrition," Comprehensive reviews in food science and food safety, vol. 19, no. 2, pp. 875-894, 2020.

[10]   S. Dang, P. Ahmad, "Text mining: Techniques and its application," International Journal of Engineering & Technology Innovations, vol. 1, no. 4, pp. 22-25, 2014..

[11]   N. J. Kadhim and H. Saleh, "Improving Extractive Multi-Document Text Summarization Through Multi-Objective Optimization," Iraqi Journal of Science, pp. 2135-2149, 2018.

[12]   H. A. Alatabi and A. Abbas, "Sentiment analysis in social media using machine learning techniques," Iraqi Journal of Science, pp. 193-201, 2020.

[13]   W. Daelemans, T. Du Plessis, C. Snyman, and L. Teck, "Multilingualism and Electronic Language Management'. Proceedings of the 4th International MIDP Colloquium (Studies in Language Policy in South Africa 4)," Studies in Language Policy in South Africa, no. 4, 2005.

[14]   K. Jasim, A. T. Sadiq, and H. S. Abdullah, "A Framework for Detection and Identification the Components of Arguments in Arabic Legal Texts," in 2019 First International Conference of Computer and Applied Sciences (CAS), 2019, pp. 67-72: IEEE.

[15]   K. Opasjumruskit, S. Schindler, P. M. Schäfer, and L. Thiele, "Towards Learning from User Feedback for Ontology-based Information Extraction," Towards Learning from User Feedback for Ontology-based Information Extraction, 2019.

[16]   R. Mochales-Palau, and M. Moens, "Study on sentence relations in the automatic detection of argumentation in legal cases," Frontiers in Artificial Intelligence and Applications, vol. 165, p. 89, 2007.

[17]   M.-F. Moens, E. Boiy, R. M. Palau, and C. Reed, "Automatic detection of arguments in legal texts," in Proceedings of the 11th international conference on Artificial intelligence and law, 2007, pp. 225-230.

[18]   P. Poudyal, T. Gonçalves, and P. Quaresma, "Using Clustering Techniques to Identify Arguments in Legal Documents," in ASAIL@ ICAIL, 2019.

[19]   E. Cabrio and S. Villata, "Five Years of Argument Mining: a Data-driven Analysis," in IJCAI, 2018, vol. 18, pp. 5427-5433.

[20]   M. Moens "Argumentation mining: How can a machine acquire common sense and world knowledge?," Argument & Computation, vol. 9, no. 1, pp. 1-14, 2018.

[21]   B. H. Majeed. L. Jawad, and H. TH. ALRikabi, "The Impact of CATs on Mathematical Thinking and Logical Thinking Among Fourth-Class Scientific Students," International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 10, 2021.

[22]   D. L. Weimer and A. R. Vining, Policy analysis: Concepts and practice. Taylor & Francis, 2017.

[23]   L. F. Jawad. B. Majeed, and H..Salim,"The impact of teaching by using STEM approach in the Development of Creative Thinking and Mathemati-cal Achievement Among the Students of the Fourth Sci-entific Class," International Journal of Interactive Mobile Technologies (iJIM), vol. 15, no. 13, 2021.

[24]   D. Al-Malah, and H. K. Jinah, "Enhancement of educational services by using the internet of things applications for talent and intelligent schools," Periodicals of Engineering and Natural Sciences (PEN), vol. 8, no. 4, pp. 2358-2366, 2020.

[25]   B. K. Mohammed, M. B. Mortatha, and A. S. Abdalrada, "A comprehensive system for detection of flammable and toxic gases using IoT," Periodicals of Engineering and Natural Sciences (PEN), vol. 9, no. 2, pp. 702-711, 2021.

[26]   N. F. AL-Bakri, A. F. Al-zubidi, A. B. Alnajjar, and E. Qahtan, "Multi label restaurant classification using support vector machine," Periodicals of Engineering and Natural Sciences (PEN, vol. 9, no. 2, pp. 774-783, 2021

[27]   N. A. jassim. H. TH.,"Design and Implementation of Smart City Applications Based on the Internet of Things," iJIM, vol. 15, no. 3, 2021

[28]   A. A. Tuama and A. Sadiq, "Twitter Sentiment Analysis and Events Prediction Using Lexicon Sentiment Analysis for Iraqi Vernacular," Solid State Technology, vol. 63, no. 6, pp. 10007-10015, 2020.