LawBot - A Community Based Chatbot to Provide Legal Assistance
Project ID: 18-010

Final Report (Draft)

B.Sc. Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

03 September 2018

Using NLP to analyze and answer user's questions
Project ID: 18-010

Final Report (Draft)

Bhanuka Hingalagoda

IT1513224

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Signature: ……………………. 　　　　　　Date: ………………………

Signature of the supervisor: ……………………. 　　　　Date: ………………………

Abstract

Today lot of people face the question of, from where and whom to find the legal assistance for the different issues they face in life. To find the accurate and contemporary legal solutions, most people need to go through lot of hassles wasting lot of their time, money and effort. To overcome this issue, we have come up with 'Lawbot', a legal assistant chatbot which helps people with their legal issues.

It will be developed using Natural Language Processing (NLP), which is an artificial intelligence field concerned with interactions between human languages and computers, adding the human element to the chatbot. A k-means clustering algorithm would be used to map the ambiguous words and phrases in the user's input to the pre-defined legal areas in the knowledge base. Lawbot will be made only for three legal areas which were finalized through results of a survey done for community. Those areas are 'Road Accidents', 'Citizen's Registrations' and 'Employment Information'.

Lawbot will be different from so many other currently available chatbots which serve in the legal related areas.

Lawbot will be available all 24/7, providing legal solutions for user's question asked either as a text or voice. Unlike messaging apps, lawbot can answer user's question instantly without any delay, so that the user doesn't need to wait for the response. While almost all the legal assistant chatbots available today provide their service through contacting an attorney which involves money, lawbot does the same, free of charge. And also, lawbot has the ability to provide directions for the legal places and get the user download any document he needs. So far, there are no any legal assistant chatbots in Sri Lanka which have been made specifically for Sri Lankan laws. Lawbot will be the first Sri-Lankan legal assistant chatbot.

Acknowledgement

Table of Contents

List of Figures

List of abbreviations

1. Introduction

I.   Problem to be addressed

Majority of the people today has a very limited knowledge in the existing laws and legal systems.. When people encounter a situation in life where they need legal advice, most people have no idea of the prevailing laws related to that situation. They go through lot of hardships to get the necessary, accurate legal advice. In such situations, what people mostly do is go to a place where they think they can get the service done or they hire a lawyer. This way involves lot of time, cost and of course it's inefficient. Maybe they have come to the wrong place seeking for information. And what often happens is people assume things and jump into conclusions. "This might be right. Let's do it" is the approach taken by most people in situations like this. This way they may end up with false and misleading information.

Or there are legal information providing websites and legal assistant apps where people needs to login first, search for the necessary information, filter out what they actually need and then logout. This way is again time consuming and not to mention how inefficient they are. Above all, most of them are not free information providing sources. In the case of legal assistant apps, user needs to wait till he gets the response for the question asked. This involves lot of delay.

What if there's that one source which is available all 24/7, provides you with the legal information you need instantly and free of charge. How easy and efficient that is? To address this requirement, we are proposing a legal assistant chatbot, a platform designed to understand, learn and converse like a human and answer ad-hoc queries in real time. It will assist people in their legal issues related to 3 specific areas which were finalized through a survey conducted on the community. They are,

 i) Road accidents - pedestrian and vehicle accidents

ii) Citizen's Registrations - obtaining new NIC, obtaining certified copy of marriage, birth, death certificate, voter registrations
iii) Employment Information - about pensions, EPF, ETF

II.   Background Context

When compared with the existing ways of finding legal solutions to the problems people encounter in day today life chatbots are the most appropriate and trending way to address that. Because chatbots provide answers instantly without any delay. Unlike legal assistant apps available, chatbots are efficient and easy. Since people like to experience a human-like conversation experience, chatbots have replaced lot of things in today's world.
There are some chatbots available giving you legal assistance but have several downsides. One of the biggest downside is they don't allow user to ask any kind of legal related question as they wish. They restrict user's natural way of conversation by commanding user to select an option out of the pre-defined options. This is more like navigating through a normal app with some radio buttons and dropdowns. They always expect the user to follow their predefined procedure. This is totally contrasting to the way a chatbot should work.

If you have some experience using any of these available chatbots, you must have noticed what they actually do is instead of providing the necessary information, they direct us to a attorney working on that area of concern and leave us a with a message saying that attorney will contact you in future. This is more similar to booking a lawyer. And not to mention any of these services are not free and you need to wait till a lawyer gets back to you. And none of these available chatbots support voice.

There are lot of downsides in the existing systems and through Lawbot, we have proposed a solution addressing these downsides of the existing systems for the legal assistance.

III.   Research Gap and Research Questions

Lawbot is designed as a web application. It can be used from desktop, laptop or smartphone. User's questions are directed to the server. Within the server user's state is maintained. Question is analysed within the server using NLP techniques like sentimental analysis, tokenization, named entity recognition and normalization.

Sentimental analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. It aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional

reaction to a document, interaction, or event. In the case of lawbot, sentimental analysis is used to train the lawbot to user's different attitudes.

Tokenization is the process of breaking up the given text into units called tokens. The tokens may be words or number or punctuation mark. Tokenization does this task by locating word boundaries. Ending point of a word and beginning of the next word is called word boundaries. Challenges in tokenization depends on the type of language. Languages such as English and French are referred to as space-delimited as most of the words are separated from each other by white spaces. Tokenization is also affected by writing system and the typographical structure of the words.

The task of Named Entity Recognition and Classification can be described as the identification of named entities in computer readable text via annotation with categorization tags for information extraction. Ex: extracting names of people, organizations, locations, or other entities. Lawbot will have the ability to identify the keywords through this method.

Finally in order to carry out processing on natural language text, we need to perform normalization that mainly involves eliminating punctuation, converting the entire text into lowercase or uppercase, converting numbers into words, expanding abbreviations, canonicalization of text, and so on.

2. Body of the Report

I.   Addressing the Literature

Researches have been done on chatbots in different areas focused on community services such as virtual teaching assistants for open e-learning platforms, intelligent web-based voice chatbots, question answering bots for API documentation and chatbots made for banking purposes. There is a chatbot that shows how integration of Artificial Intelligence (AI) can help students as well as tutors across the learning process within an open source e-learning platform. Especially, we present T-BOT and Q-BOT, a couple of chatter bots capable of tutoring and evaluating students using open platforms as Moodle or Caroline. APIBot can answer API questions given API documentation as an input.

Also, we have found some papers that discuss about Artificial Intelligence via chatbots, their framework design, capabilities, utilization and future scope.

But we came across only one research done for chatbots made specifically for legal assistance. That is 'A multiagent assistant for legal research' which helps to collect and organize the statutes and case histories relevant to a legal search. That's for easy search of legal information for public.

II.   Methodology

Lawbot is designed as a web application. It can be used from desktop, laptop or smartphone. User's questions are directed to the server. Within the server user's state is maintained. Question is analysed within the server using NLP techniques like sentimental analysis, tokenization, named entity recognition and normalization.

To get the answers for user's questions, information source is needed.Law grows in volume and complexity in the process of enactment. To get the information in Lawbot ,the technique that we have use is creating an ontology. An ontology is a way to take a set of concepts and organize it in a formalized way. To create an ontology first ,have to identify the main classes that have to implement.After identifying the main classes (Ex.Area of Law; Type of Law,Final Product and Legal Action) what we have done is coming up with the terms for the ontology .To implement the ontology what we have used is Protege.(use an .owl extension).

To map the ambiguous words in the user's question, they are mapped using text clustering. Text clustering is the task of grouping a set of texts in such a way that texts in the same group (called a cluster) are more similar to each other than to those in other clusters. Our approach in using text clustering is to use semantic clustering in sentence level. Since semantic clusters group sentences having almost the same meaning and conveying the same information unlike thematic clustering which regroups sentences that deal with the same topic/event, semantic clustering was chosen to fine grain the grouping of user's input to the pre defined clusters in the knowledge base. To produce the semantic clusters, Cosine similarity of tf-idf (term frequency-inverse document frequency) vectors, a 'Text Similarity Measure', which is the tf-idf weight of a word w in a document d belonging to a corpus is the ratio of the number of times w occurs in the document d to the number of documents in which w occurs at least ones. As the suitable text clustering algorithm,   k-means clustering using tf-idf vectors with the scikit-learn implementation, a machine learning algorithm is used. [8]

K-means clustering is a technique used to uncover categories. In our case, it can be used to categorize user's questions into clusters. k represents the number of categories(clusters) identified, with each category's average (mean) characteristics being appreciably different from that of other categories.[9] According to the lawbot, categories would be Road accidents, Citizen's Registrations and Employment Information. This is done for effective querying of the knowledge base. So that those ambiguous words can be mapped to the words already defined in the knowledge base.

Finally, the question which is analysed using NLP and mapped using ML is directed to the knowledge base to query and find the appropriate answer. Answer is again converted to human understandable format using NLP.

If there are documents involved with that user's specific legal issue, user is asked if he needs the documents and if the user gave the approval, he is directed to the necessary documents.

If user requests for directions of the locations related to his legal issue, Lawbot provides the directions using google map API.
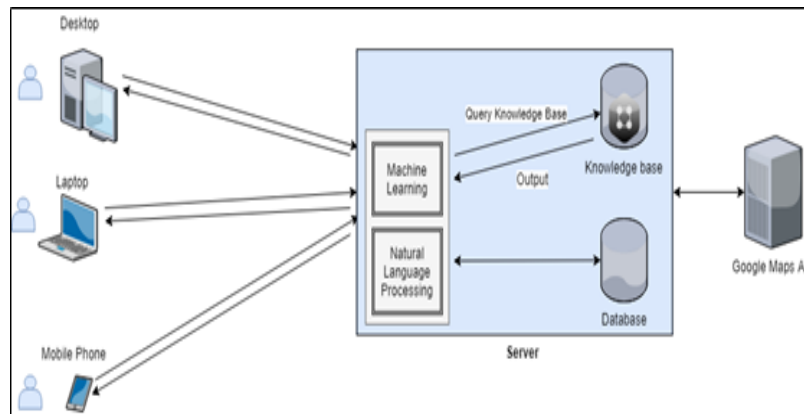

Figure 1: System overview

3. Results & Discussion

II.   Discussion

Lawbot has filled so many existing gaps in the available chatbots and researches that have been done in the context of chatbots.
Unlike most available chatbots, lawbot is not structure based. It's more conversation based which makes it more like a normal conversation with a human. User won't feel like he's chatting with a machine. NLP possess the power to add the human element to the chatbot.

Lawbot is trained to possess the ability to answer any unexpected questions providing more examples to the chatbot, giving it a wider basis with which it can interpret and answer questions

and commands. Even though the user's question is indirect and ambiguous, with the use of clustering, lawbot is able to map the user's question to the way it can answer.

For the clustering, K-means is chosen since in the context of text data, it is suitable and fast when we have an approximation of the clusters number and when the similarity measure is not expensive in terms of computation time.

Unlike most of the existing chatbots made for legal assistance which are made specifically for law firms, lawbot will provide answers for user's question instantly without any delay. In other chatbots, user will be directed to contact a lawyer by providing lawyer's details. And also they charge some fee for legal assistance.

Lawbot is the only legal assistant chatbot available which supports voice as well. Therefore a disabled person also has the ability to use the lawbot with voice.

4. Conclusion

Our solution is to develop a chatbot which assists people with their legal issues. With this, we hope people will be more aware of the existing laws and legal system in Sri Lanka and they will be able to find information related to legal issues easily without going through much hardships.

Lawbot has the following constraints at the moment.
Lawbot will not support any other languages than English. This is intended for the use of only Sri Lankan citizens and tourists who would like to get an idea about the current laws in Sri Lanka pertaining to a legal scenario. All the solutions and answers that provided by the Lawbot are in the context of Sri Lankan laws and might be different to a law which is applied in a different country pertaining to the same legal scenario. Lawbot is created with the laws and legal solutions that were available during the time information were gathered. Update of current laws will not be reflected in the Lawbot at the same time the laws were amended until an admin manually update the knowledge base. And Lawbot will assist you only in the legal areas that is it currently supporting and with the user requirement, developers of the Lawbot will decide of the new areas to be supported by the Lawbot.

We hope to improve the Lawbot by supporting other different legal areas which are frequently asked by people other than the areas we are currently supporting. And also, we hope to add a social platform to the Lawbot where people can share and discuss their ideas, incidents related to different legal areas. Lawbot will be integrated with messaging platforms like facebook, viber etc.

5. References

[1][Online]. Available: https://medium.com/@Legal.io/meet-lara-your-legal-assistant-a579dac68cf. [Accessed 2 05 2018].
[2][Online]. Available: https://www.ailira.com/. [Accessed 04 05 2018].
[3][Online]. Available: https://www.paperstreet.com/blog/chatbot-law-firms/. [Accessed 03 05 2018].
[4][Online]. Available: http://www.expertsystem.com/examples-natural-language-processing-systems-artificial-intelligence/. [Accessed 06 05 2018].
[5][Online]. Available: https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68. [Accessed 7 05 2018].
[6][Online].Available:https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
[7] [Online]. Available: https://chatbotslife.com/nlp-nlu-nlg-and-how-chatbots-work-dd7861dfc9df
[8] [Online]. Available:http://www.lumenai.fr/blog/quick-review-on-text-clustering-and-text-similarity-approaches
[9] [Online]. Available:https://www.quora.com/What-is-the-k-Means-algorithm-and-how-does-it-work

6. APPENDICES

I. Appendix A: Work Break Down Structure