ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Closed Domain Keyword based Question Answering System for Legal Documents of IPC Sections & Indian Laws

Shubhangi C. Tirpude, Dr. A.S. Alvi

MTech Scholar, Department of CSE, Prof. Ram Meghe Institute of Technology & Research, Bandera, Amravati, MS India

Professor, Department of CSE, Prof. Ram Meghe Institute of Technology & Research, Bandera, Amravati,MS, India

ABSTRACT: Question Answering (QA), in information retrieval, is the task of automatically answering a question posed in natural language (NL) using either a pre-structured database or a collection of natural language documents. It delivers only the requested information unlike search engines which refers to full documents. As with the excessive information growth in the web, retrieving the exact fragment of information even for a simple query requires large and expensive resources. Additionally the need to develop exact systems gains more importance due to available structured knowledge-bases and the continuous demand to access information rapidly and efficiently. This paper describes the closed domain keyword based QA system. It extract the keywords from the query (structured/ unstructured), compares it with Indexed term dictionary, & finds the best suitable answer for the query

KEYWORDS: Question Answering, Information Retrieval, Natural Language, IPC Sections, Indian Laws, Jaccard Coefficient, TF-IDF

I. INTRODUCTION

The Text Retrieval Conference (TREC), a conference series co-sponsored by NIST, initiated the Question-Answering Track in 1999 which tested systems' ability to retrieve short text snippets in response to factoid questions revealed an increasing need for It revealed more sophisticated search engines able to retrieve the specific piece of information that could be considered as the best possible answer for the user question.

Question Answering (QA) is a fast growing research area that combines research from different, but related, fields which are Information Retrieval (IR), Information Extraction (IE) and Natural Language Processing (NLP). Question Answering (QA), in Information Retrieval delivers only the requested information unlike search engines which refers to full documents. QA research attempts to deal with a wide range of question types including: fact, list, definition, how, why, hypothetical, semantically constrained, and cross lingual questions.
There are two types of Question answering Systems

- Closed-domain question answering:
  Closed domain question answering deals with questions under a specific domain (music, weather forecasting, Tourism, Medical etc.) The domain specific QA system involves heavy use of natural language processing systems formalized by building a domain specific ontology.

- Open-domain question answering :
  Open-domain question answering deals with questions about nearly everything and can only rely on universal ontology and information such as the World Wide Web.(2).The domain specific Question Answering System gives more specific and correct answers than web based QA system as it is restricted for only one domain resource to answer

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

QA System answers NL questions from any type of document. Such as Structured, Heterogeneous, Text based, Web based on any type of domain such Open domain: newspaper articles (close corpora), the web, Close domain: Aerospatiale domain, medical domain, etc. The Question to be asked to the system is Factoid questions: who, when, where, how much, Definition, How, why questions, Opinion, comparative and evaluative questions, The Answers from the system are Exact answers, Passages, Multimedia, etc.

Question Answering System Overview

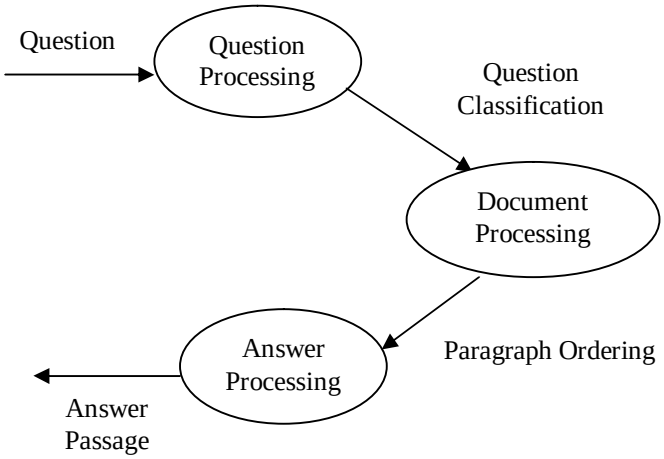

Figure 1: QA System Architecture

As shown in Figure 1, a typical QA system consists of three distinct modules, Query Processing Module" whose heart is the question classification, the "Document Processing Module" whose heart is the information retrieval, and the "Answer Processing Module" whose heart is the answer extraction.

Question Processing Module

Given a natural language question as input, the overall function of the question processing module is to analyse and process the question by creating some representation of the information requested.
The Question Processing has 3 tasks as:

- Determining the question type
- Determining the answer type
- Extract keywords from the question and formulate a query

Document Processing Module

From the keywords that selected, a query is formulated and is given to the component that is document retrieval. In this, all the passages are extracted that contains the selected keywords.

Answer Processing Module

In the answer extraction, the representation of the question and the representation of candidate answer bearing texts are matched against each other to give specific and correct answer. From this set of such candidate answers are produced and then ranked according to likelihood of correctness.

II. RELATED WORK

Open domain question answering using Wikipedia-based knowledge model [2]
It describes the use of Wikipedia as a rich knowledge source for a question answering (QA) system with multiple answer matching modules based on different types of semi-structured knowledge sources of Wikipedia, including

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

article content, info boxes, article structure, category structure, and definitions. These semi-structured knowledge sources each have their unique strengths in finding answers for specific question types, such as info boxes for factoid questions, category structure for list questions, and definitions for descriptive questions

For semi-structured knowledge-bases QA System [3], a new architecture to develop a factoid question answering system based on the DBpedia ontology and the DBpedia extraction framework. Dbpedia is a project that aims at extracting information based on the semi-structured data presented within the Wikipedia articles.

Authors in [4] given the system which finds answers of Malayalam factual questions by analyzing a repository of Malayalam documents for handling the four classes of factual questions in Malayalam for closed domain. The QA system is divided into three modules as Question Analysis, Text Retrieval and answer snippet extraction and Answer identification.

Jibin Fu in the paper [5] proposed a music knowledge question answering system on the ontology knowledge base through which the users can ask a question about music knowledge in natural language, and the system automatically extracts relative knowledge to give answer based on FAQ and ontology knowledge base.

Bhoir V & M. A. Potey [6] proposed solution of Question Answering system works for a specific domain of tourism. The crawler developed in the system gathers web page information which is processed using Natural Language Processing and procedure programming for a specific keyword. The system returns precise short string answers or list to natural language questions related to tourism domain like distance, person, date, list of hotels, list of forts, etc.

Brigitte Jörg and Ulrich Schäfer, (6) present an implemented approach for domain-restricted question answering from structured knowledge sources, based on robust semantic analysis in a hybrid NLP system architecture. The Author perform question interpretation and answer extraction in an architecture that builds on a lexical-conceptual structure for question interpretation, which is interfaced with domain-specific concepts and properties in a structured knowledge base. Question interpretation involves a limited amount of domain-specific inferences, and accounts for higher-level quantificational questions. Question interpretation and answer extraction are modular components that interact in clearly defined ways. They derive so-called *proto queries* from the linguistic representations, which provide partial constraints for answer extraction from the underlying knowledge sources. The search queries we construct from proto queries effectively compute minimal spanning trees from the underlying knowledge sources. The approach naturally extends to multilingual question answering, and has been developed as a prototype system for two application domains: the domain of Nobel prize winners, and the domain of Language Technology, on the basis of the large ontology underlying the information portal LT World.

Following table shows details about the Question Answering System & Their working.

QA system	Working
Open domain question answering using Wikipedia-based knowledge model. [2]	Resource: Wikipedia Questions: factoid, list and descriptive Answer Extraction: Answer merging strategy by matching modules
Dbpedia Based Factoid Question Answering System[3]	Resource: DBpedia ontology Questions: factoid Answer Extraction: SPRQL queries
A Natural Language Question Answering System in Malayalam Using Domain Dependent Document Collection as Repository[4]	Resource: Malayalam documents repository Questions: Malayalam factual Answer Extraction: Named Entity Recognition for snippet extraction
domain-restricted question answering from structured knowledge	Resource: the domain of Nobel prize winners, and the domain of Language technology Questions: large ontology underlying the information portal LT World

Table 1: Question Answering System & its working

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

III. PROPOSED ALGORITHM

In current scenario, there is No QA system for exactly answering the queries on legal documents of IPC & Indian laws. Also the answers differ from Wikipedia and other sites for legal documents for same question and the accuracy for unstructured query by users is very less. The proposed system for current scenario is to create such a closed domain keyword QA system that retrieve the answer related to question efficiently. For this the Corpus or Knowledge base is constructed from legal documents of Indian Laws. Then the matching of terms in queries with possible candidate terms in each document to retrieve the best answers stored in corpus. The system will extract and gives the specific and correct answer to query (structured/unstructured) by user
The Figure 2 illustrates Proposed Approach

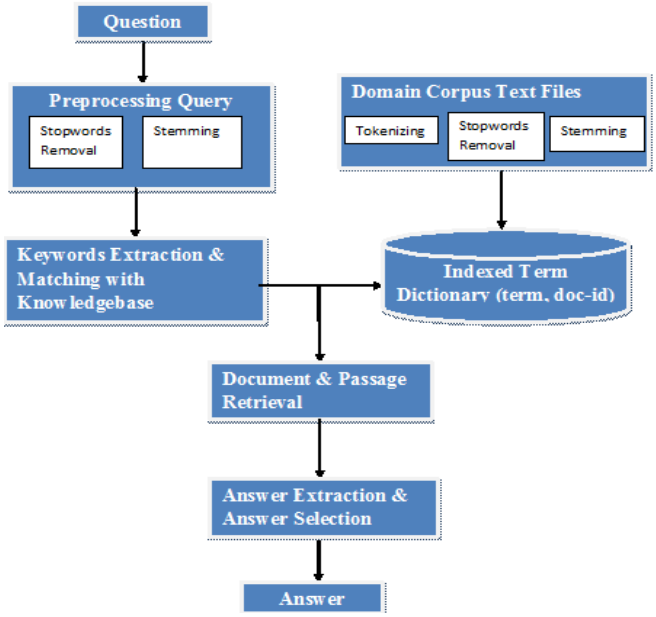

Figure 2: Proposed System Architecture

The question / query is supplied as an input to the system. The input query is processed by removing stop words, performing tokenization & Stemming. The keywords are extracted from the system. The keyword plays an important role in information retrieval. These keywords are matched with the Index term dictionary (Metadata). The Indexed term dictionary contains the term appears in all the documents, their count & the document ids. From the Indexed term dictionary it retrieves the documents that likely to contain an answer. Next the Answer extraction module performs the processing based on the type of question and retrieves the passage for exact answer.

- Total 600 files related to legal documents of IPC Sections & Indian Laws are used as corpus
- Input is given in the form of question/ query
- Output is generated as passage which contains the answer relevant to the query

The System is implemented with JAVA as front end & MYSQL as backend. The modules are as follows.

Module 1: Knowledge base pre-processing

- Input – Plain text documents
- Output – Pre-processed indexed term dictionary

In the first Module, the corpus stored as text files are preprocessed by performing various preprocessing operations and store all the keywords extracted from the documents in Indexed term dictionary.

Module 2: Question Type determination and Document Retrieval

- Input – User's query
- Output – Candidate answer's document s ids

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

In this module the input in the form of query/ question is given to the question answering system. The preprocessing of the query is done to get the refine query. The keywords are extracted from the query, matches against the terms in indexed term dictionary. From the Indexed term dictionary, retrieves the document id based on the occurrence of the term in the document. If more than one term is involved, it performs the intersection of the terms to find the document id where all the term is present.

Module 3: Answer Extraction

- Input – Documents from module 2
- Output – Best suitable answer according to question types from corpus

Once the documents are retrieved, QA system determines the type of answer based on the type of question. It determines the target of the question using the rules defined. Once the target is defined, it finds the best suitable answer.

IV. IMPLEMENTATION

The implementation details are as follows,

A Corpus:

The first phase of the closed domain Question Answering System is to create Corpus or Knowledge base. As we have to design the closed domain Question Answering System, the most important task is to decide the domain. There are so many QA systems already present for different closed domains. So we are dealing with the new domain for answering the user queries on IPC sections and some Indian Laws. The legal documents of IPC sections and laws like parent and company amendment are necessary to know in different ways for different users. On different websites the exact IPC sections and laws documents are available for interested users. But for exact answers on different questions related these documents can be given by using QA system. So we have gone through the different websites and taken the text data from the websites:
http://www.dabangvakil.com/indianpenalcode
http://indiacode.nic.in/coiweb/coifiles/amendment.htm

As these documents are authorized and highly sensitive therefore the legality is referenced by the lawyer. For each IPC sections and different laws, there is one text file, such for 511 IPC sections, 511 text files are stored and also for constitution amendment law 94 text files and each for parent amendment law and company amendment law are stored.

B Pre-processing:

Major tasks in pre-processing are tokenization, stop words removal and stemming

- Tokenization: Perform linguistic tokenization which includes each word and also the numbers as single token.
- Stop words Removal: Removing stop words reduce the dimensionality of term space. The most common words are in text documents are prepositions, articles, and pro-nouns etc that does not provide the meaning of the documents. Stop words are eliminated from documents because those words are not considered as keywords in Information Retrieval applications.
- Stemming: Stemming is most important process by which the different forms of word is replaced with basic root word.

C Indexed Term Dictionary.

The processing of text involves two main problems, first problem is the extraction of feature terms that become effective keywords in the training phase and then the second is actual classification of the document using these feature terms in the test phase. After pre-processing the TF-IDF is calculated for keyword term in a document [8]. The TF-IDF is stored as a table as the Term (keyword), count & document id. It represents that the term is appeared in the number of files & the count of the term in each file.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

	srno	terms	posting	
☐	25	issue	(s_1.txt,1),(s_2.txt,1),(...	133 b...
☐	26	law	(s_1.txt,1),(s_2.txt,1),(...	133 b...
☐	27	male	(s_1.txt,1),(s_8.txt,1),(...	36 b...
☐	28	man	(s_1.txt,2),(s_10.txt,2)	24 b...
☐	29	mean	(s_1.txt,1),(s_2.txt,1),(...	133 b...
☐	30	noncogniz	(s_1.txt,2),(s_2.txt,2),(...	133 b...
☐	31	noncompound	(s_1.txt,2),(s_2.txt,2),(...	133 b...
☐	32	offence	(s_1.txt,8),(s_2.txt,8),(...	135 b...
☐	33	penal	(s_1.txt,1),(s_2.txt,2),(...	133 b...
☐	34	permit	(s_1.txt,3),(s_2.txt,3),(...	133 b...
☐	35	police	(s_1.txt,2),(s_2.txt,2),(...	133 b...
☐	36	procedure	(s_1.txt,1),(s_2.txt,1),(...	133 b...
☐	37	register	(s_1.txt,2),(s_2.txt,2),(...	133 b...
☐	38	schedule	(s_1.txt,1),(s_2.txt,1),(...	133 b...
☐	39	section	(s_1.txt,2),(s_2.txt,1),(...	133 b...
☐	40	settle	(s_1.txt,2),(s_2.txt,2),(...	133 b...
☐	41	term	(s_1.txt,1),(s_2.txt,1),(...	133 b...

Table 3: Indexed Term Dictionary (Terms, Doc_id,freq)

D. Information Retrieval

The Information Retrieval has 3 components as:

1. **User Query**: user will enter the query related to domain in structured or in unstructured form.

2. **Extract Keywords**: From the user query, the keywords are extracted. These keywords are obtained by removing the symbols and stop words from user query. Also the stemming is applied on keywords so as to match with term indexed dictionary terms for document retrieval.

3. **Document and Passage Retrieval**: The keywords so obtained from query are matched in Term Indexed Dictionary to find the document ids where these keywords are present. For more than one keywords, it takes the intersections of all document ids where these terms are present so that where all the keywords are present only those document are to be retrieved for candidate answer passages. For Ex The term Law is present in file 1, 5, 8 & the term Murder is present in 1, 4, 6, 8. It retrieves all the files as file 1, 8 the keywords that obtained from user query are matched in documents. It can give one or more paragraphs which satisfies the keywords matching

E Question Dataset

As we are dealing with the closed domain Question Answering, the system need to answer the different questions related to domain. So the dataset of 300 different questions on IPC sections and amendment laws is maintained to train our QA system. These questions are taken from different users by survey, so the system can handle structured as well as unstructured queries from user. According to these questions, the QA system is designed to give the answers
The Examples of the Question dataset is given below.

1. What is the punishment for murder?
2. What is the punishment for rioting?
3. What is the vessel?
4. Who is the judge?
5. What is the court of justice?
6. What is the punishment for rioting armed with deadly weapon?
7. What is cognizable offence?
8. What is non-cognizable offence?
9. What is compoundable offence?
10. What is non-compoundable offence?

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

F. Finding Target from query terms according to answer type

Here we consider the "Query terms" to determine the "Target value" for classifying the questions according to the expected answer for certain query.

For example, the query terms like "what is the reason", "In case of", "suggest or tell me" gives answer as description. The terms like "define" or "what is meant by" refers the definition as answer. Also the terms like "list the IPCs" or "list the sections" give the expected answer as the list of related IPC sections for related crime. The questions started with the terms like "can" "Is", "whether" should give the answer as yes or no. Also if user wants the whole IPC section for reference the terms so considered as "which IPC" or "under which section" or the term IPC with digit value like "IPC 29". Also if user wants to know about exact punishment or charges for certain crimes it is referred with the query terms like "what is the punishment" or "charges for". As we are dealing with the domain of IPC sections, user need to deal with the different punishments and related IPCs mostly. Therefore this exact punishment question has more weight while designing and training the system.

We have different query terms according to the expected answers which are stored with certain target value in MySql table format. For example "X1" is the target value for expected answer as "description" or "definition", so the different possible query terms for this as show in table are stored for the same value as "X1" and same for the other answer types. Now, the user query that we have already used in Information Retrieval is going to get used again to deal with classifying the question and determining the target value.
Jaccard Coefficient mechanism is used for nearest matching of the two strings. It gives the relative score of matching the terms in string 1 with the string 2. Jaccard coefficient uses the formula to obtain the score given as:

$$\text{Score} = \text{Query} \cap \text{Query type} / \text{Query} \cup \text{Query type}$$

The query from user is matched with all the query terms from the table and gives the score by using the Jaccard formula above. So we have different scores for each query terms depending upon the matching keywords in both strings. To determine the target value for question classification here we considered the maximum score between the user query and query terms

Target	Answer Type	Query Terms
X1	Description or Defination	In case of, For What, What is the reason, Define, give the meaning of, what is meant by, tell me…
X1-1	List	List the ipcs, list the sections
X2	one word	what is the punishment, which ipc/ section
X3	Location, Place	where the act, In which
X4	Time, Duration	How long, What is period of
X5	Yes/ No	Can, will, Is, Whether
X6	Whole IPC	Under which section, which ipc
X7	Exact punishment or charges	Charges for, what is the punishment

Table 4: Finding Target from query terms according to answer type

Consider the following example.
Query: what is the punishment for murder?
Keywords are: [punish, murder]

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Selected Paragraphs: 303. Txt: [Punishment for murder by life-convict Whoever, being under sentence of1 [imprisonment for life], commits murder, shall be punished with death. , Punishment Death Cognizable Non-bailable Triable by Court of Session Non-compoundable.]
307. txt: [(a) A shoots at Z with intention to kill him, under such circumstances that, if death ensued. A would be guilty of murder. A is liable to punishment under this sections.]
304.txt: [Punishment for culpable homicide not amounting to murder Whoever commits culpable homicide not amounting to murder shall be punished with [imprisonment for life], or imprisonment of either description for a term which may extend to ten years, and shall also be liable to fine.
302. txt: [Punishment for murder Whoever commits murder shall be punished with death, or 1[imprisonment for life] and shall also be liable to fine. , Provisions of death sentence being an alternative punishment for murder is not unreasonable;

G. Answer Extraction

Information Retrieval module returns the documents from indexed term dictionary using keywords from users query. To retrieve the exact passage, it selects only that documents which contains the keywords in proximity value of 3. Then it extracts the passage which retains the proximity of 3 in keywords from users query according to the answer type using target value.
The various target types are defined.

- If the target type is definition or description it retrieves the first paragraph from the first file as the corpus is stored in sorted order.
- If the target type is List IPC, it retrieves the document id as the file is stored with its name
- If the target type is yes/no, if answer is yes, it retrieves the related document from IPC documents the file is stored with its name, if the answer is no it gives the message “No related IPC”
- If target type is whole IPC, it retrieves the whole file for related IPC
- If target type is punishment or charges, it retrieves exact one word answer using resource dictionary using punishment word

H. Answer Selection

Information Retrieval module returns the documents from indexed term dictionary using keywords from users query.. To retrieve the exact passage, it selects only that documents which contains the keywords in proximity value of 3. Then it extracts the passage which retains the proximity of 3 in keywords from users query according to the answer type using target value. The following figure illustrates the Answer extraction technique.

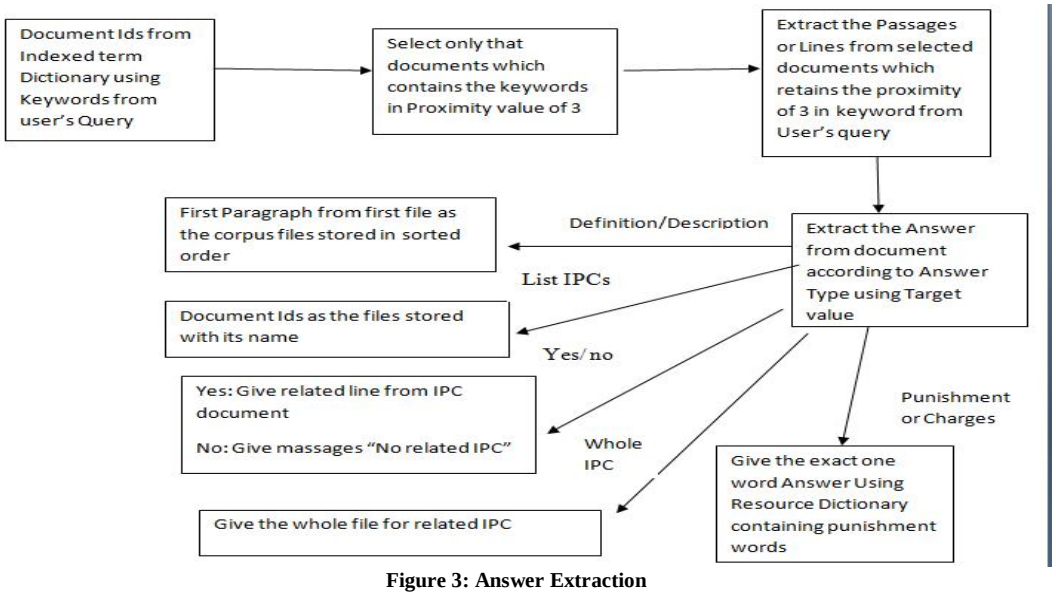

Figure 3: Answer Extraction

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

For different type of expected answer to the user query about the domain, we have defined some rules for selecting the answer. For definition or the description about any section or punishment, we are extracting the passage from first file retrieved in the Information Retrieval, as the corpus stored of text files in sorted ascending order so that most appropriate and matching answer and exact definition will obtain in first passage only. While listing the IPCs or sections for any crime, we only require the related IPCs number value. So, to give the related IPC list, only document ids are used, as the text files for related IPCs are stored with their section number. If the user will ask any question for which answer can be yes or no then the system will try to return the positive answer as related line from extracted passages if matches with the certain domain constraints otherwise give no answer. If the user wants the whole IPC as the answer then the whole text file related IPC section is retrieved, as the text files in corpus are stored with their names and each file has the title with its related contents

H. Domain Resource Dictionary:

As we are majorly dealing with the questions like "What is the punishment for" or "charges for". This question needs to give the answer as exact punishment or charges for certain crime. To deal with this type of question, additional knowledge about the answer is required. For that we are maintaining the domain resource dictionary. The Domain Resources Dictionary is the file in which all the possible punishments mentioned in all the sections are stored. For example: death, imprisonment for, liable to fine, solitary confinement etc. So while giving the answer to query about exact punishment or charges about any crime, from the passages or line that we have retrieved in the information retrieval model are searched to obtain such punishment terms stored in dictionary and give the exact answer to the user query about punishments.

I. Domain Concepts Dictionary

As we are dealing with the keyword based retrieval, therefore whole system depends upon the keywords obtained from user query and the keywords we stored in Indexed Term Dictionary from the corpus text files. The user can give the unstructured query or non question form query to the system. Also the common users are mostly unaware about the different terms which are exactly related to our domain that is IPC section and amendment laws. So we are maintaining the domain concepts dictionary in which the most commonly used words for queries are stored against the related keyword of our domain so that while dealing with the exact keywords it can be referred to the domain specific term.
For example:
Murder / kill
punish / charges/ guilty
fine / money/ amount
counterfeit / bogus/fraud

V. IMPLEMENTATION & TEST CASES

The Figure 4 illustrates the User Interface for Question Answering System. The user will enter the question in the text box. After clicking on 'Get answer' button, the exact answer retrieved in the answer box. After clicking on 'clear' button, the answer gets cleared

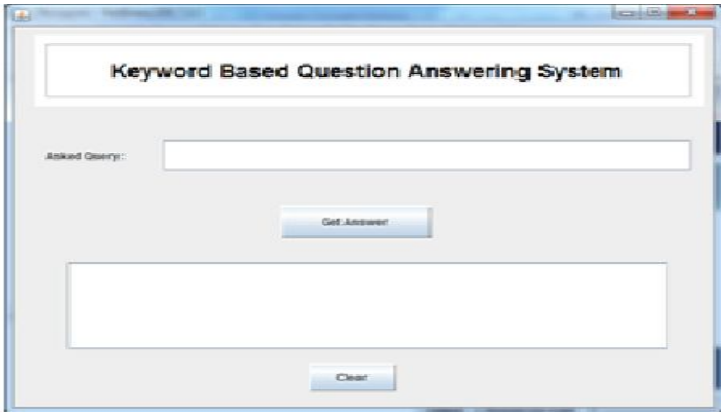

Figure 4: Main Window Screenshot

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

The various examples of various types are as follows

Implementation Example 1:

Question: What is punishment for murder?

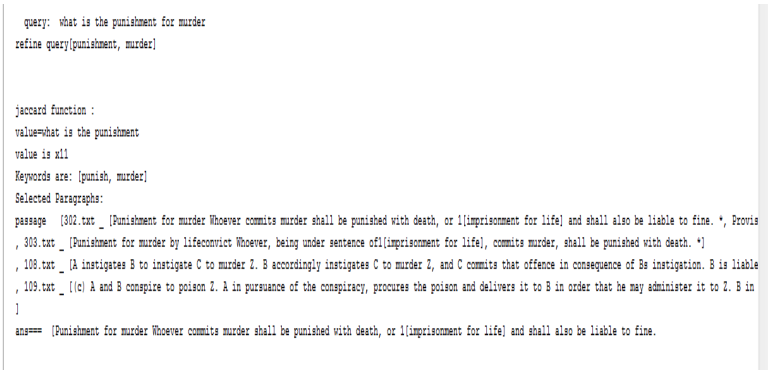

Figure 4 Implementation Example 1(Target Type is Exact Anwser)

Implementation Example 2:

Question: What is Murder?

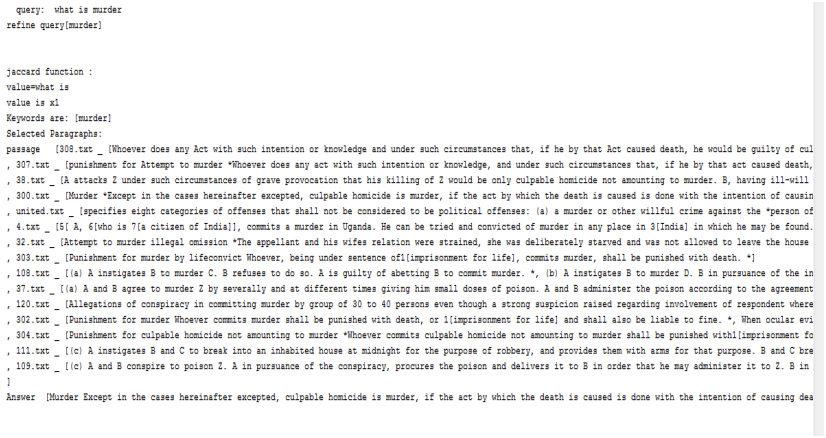

Figure 5: Imlementation Example 2(Target Type is Definition)

Implementation Example 3:

Question: If Ram killed Shyam then punishment to ram

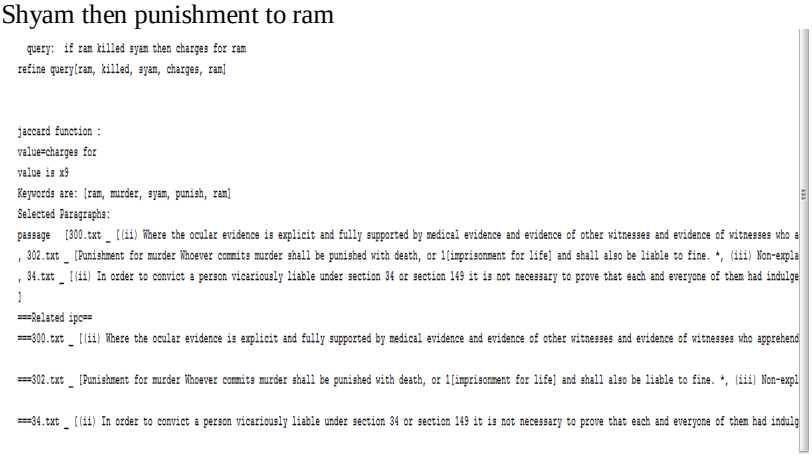

Figure 6: Implementation Example (Unstructured Query)

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Implementation Example 4:
Question: List the IPC for punishment for Murder

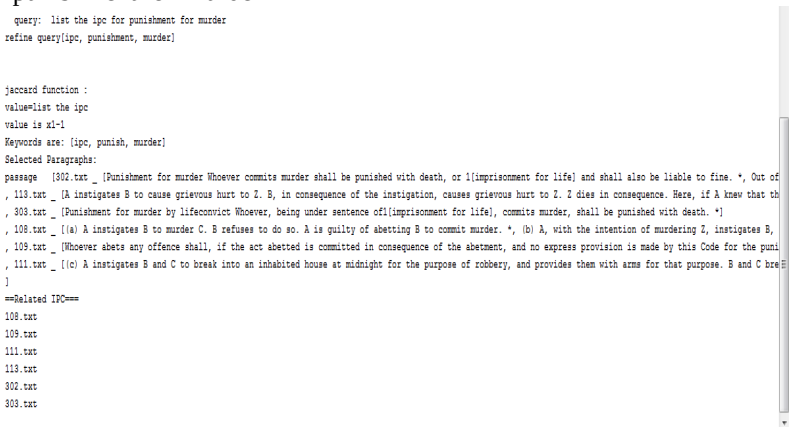

Figure 7: Imlementation Example 4 (Target Type is List)

VI. SIMULATION AND RESULTS

There are several evaluation metrics that differ from one QA campaign to another (e.g. TREC, CLEF, NTCIR, etc). Moreover, some researchers develop and utilize their own customized metrics.
The proposed system tries to find precise answers to factual questions and explanative answers for multiple answers, ranking is provided based on the semantic matching. Direct answers are generated using natural language generation techniques form the candidate set of answers.
The following measures are the most commonly used measures that are typically utilized for automated evaluation:

Precision, Recall and F-measure:

Precision & recall are the traditional measures that have been long used in information retrieval. While the F-measure is the harmonic mean of the precision and recall; these three metrics are given by:

Precision= number of correct answers number of / number of questions answered
Recall= number of correct answers / number of questions to be answered
F Measure = 2 Precision * recall / Precision + Recall

Experimental Results:

Total Questions	100
Response by the system	67
Correct answer	62
Precession	0.92
Recall	0.62

Table 5: Experimental Results showing precession & recall

The proposed system is tested with 200 different questions. These questions include various queries by the users about the all 511 IPC sections and Indian Amendment laws. These questions are mostly of structured question format. Out of 200 questions, nearly 100 are of exact punishment type, 45 are of description and definition type, 35 of list answers and nearly 20 are of yes or no type. For these 200 questions, system produces correct output with great accuracy as the queries are structured.. Now, we have again tested the system with different 100 random questions

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

related to our domain. These questions are majorly unstructured and non question form. For these questions the expected answers can be same as of trained questions. Here, we are getting less accuracy than in trained questions. Mostly the questions for yes or no answer have no response during testing because while dealing with keywords the "no" answer cannot be retrieved. Out of 100 questions, we got 67 responses from the system and out of which 62 are correct shown in table 5.

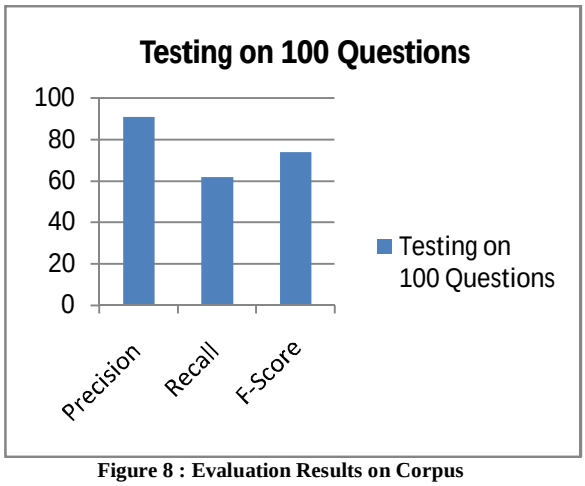

Figure 8 : Evaluation Results on Corpus

VII. CONCLUSION

Question Answering requires more complex NLP techniques compared to other forms of Information Retrieval. QA Systems can be developed for resources like web, semi-structured and structured knowledge-base. The Closed Domain QA Systems give more accuracy in finding answers but restricted to single domain only.

The proposed Question Answering system for closed domain gives accurate answer for the users query based on domain. The question classification module will determine the type of question. Based on the type of question, it determines the target type of the question so as to get the accurate answer. The system is tested based on the various evaluation parameters. The system is tested on 100 questions showing the accurate & precise results

The QA system for closed domain of legal documents of IPC sections and Indian Laws using machine learning approach and information retrieval is proposed to give the accurate and suitably more correct answers for user's structured or unstructured queries in efficiently.

VIII. FUTURE SCOPE

For future work we would like to extract the exact answer for user question instead of the whole paragraph so that it will increase the accuracy of the QA system.

In future more work may be done to improve answer generation. The generated answer should be as per the expectation of users. To meet this requirement many times language knowledge are not sufficient and it requires additional domain or world knowledge,

An authoring tool may also been developed to design QA System for different domains using the methodology described in the paper. This tool will provide the feature of domain portability. Using such tool QA system for different domain may be designed by replacing domain knowledge only without changing language knowledge and question processing module.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

REFERENCES

[1] Pum-Mo Ryu, Myung-Gil Jang and Hyun-Ki Kim "Open domain question answering using Wikipedia-based knowledge model" in Information Processing and Management 50 (2014) 683–692, Elsevier.
[2] Adel Tahri and OkbaTibermacine. "Dbpedia Based Factoid Question Answering System" in International Journal of Web & Semantic Technology (IJWesT) Vol.4, No.3, July 2013.
[3] Pragisha K. and Dr. P. C. Reghuraj, "A Natural Language Question Answering System in Malayalam Using Domain Dependent Document Collection as Repository", International Journal of Computational Linguistics and Natural Language Processing Vol 3 Issue 3 March 2014 ISSN 2279 – 0756
[4] Jibin Fu, KeliangJia and Jinzhong Xu, "Domain Ontology Based Automatic Question Answering", 2009 International Conference on Computer Engineering and Technology
[5] Bhoir V & M. A. Potey, "Question answering system: A heuristic approach" , Published in:2014**,** IEEE fifth international conference on Applications of Digital Information and Web Technologies .
[6] Anette Frank , Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg and Ulrich Schäfer, "Question answering from structured knowledge sources", In German Research Center for Artificial Intelligence, DFKI, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany Available online 27 January 2006
[7] Perera, Rivindu "IPedagogy: Question Answering System Based on Web Information Clustering", In Proceedings of the 2012 IEEE Fourth International Conference on Technology for Education (T4E '12). IEEE Computer Society, Washington, DC, USA
[8] Menaka S and Radha N. "Text Classification using Keyword Extraction Technique in International Journal of Advanced Research in Computer Science and Software Engineering", Volume 3, Issue 12, December 2013
[9] Matthew W. Bilotti and Eric Nyberg, "Improving Text Retrieval Precision and Answer Accuracy in Question Answering Systems", the 2nd Workshop on Information Retrieval for Question Answering (IR4QA), pages 1–8 Manchester, UK. August 2009
[10] Abdullah M. Moussa and Rehab F. Abdel-Kader, "QASYO: "A Question Answering System for YAGO Ontology", International Journal of Database Theory and Application, Vol. 4, No. 2, June, 2011
[11] Eric Brill, Susan Dumais and Michele Banko, "An Analysis of the AskMSR Question-Answering System", Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 257-264. Association for Computational Linguistics.
[12] Unmesh Sasikumar and Sindhu L. "A Survey of Natural Language Question Answering System" International Journal of Computer Applications Volume 108 – No 15, December 2014
[13] Ali Mohamed Nabil Allam1 and Mohamed Hassan Haggag "The Question Answering Systems: A Survey"
[14] Mengqiu Wang, School of Computer Science, Carnegie Mellon University "A Survey of Answer Extraction Techniques in Factoid Question Answering" Association for Computational Linguistics
[15] Athira P. M., Sreeja M. and P. C. Reghuraj "Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System", International Journal of Web & Semantic Technology (IJWesT) Vol.4, No.4, October 2013
[16] Lakshmi Palaniappan, Dr. N. Sambasiva Rao, "An Ontology-based Question Answering Method with the use of Query Template", *International Journal of Computer Applications Volume 9– No.9, November 2010*
[17] Lee, C., Hwang, Y.-G., Oh, H.J., Lim, S., Heo, J., Lee, C.-H., et al (2006). "Fine-grained named entity recognition using conditional random fields for question answering". In Proceedings of Asia information retrieval symposium (pp. 581–587).
[18]. ZHANG Yu, LIU Ting, WEN Xu, "Modified Bayesian Model Based Question Classificatio", 2005, vol.19, pp. 100-105.
[19]. Amit Mishra, Nidhi Mishra and Anupam Agrawal, "Context-Aware Restricted Geographical Domain Question Answering System", In 2010 International Conference on Computational Intelligence and Communication Networks
[20] Matthew W. Bilotti and Eric Nyberg," Improving Text Retrieval Precision and Answer Accuracy in Question Answering Systems", the 2nd workshop on Information Retrieval for Question Answering (IR4QA), pages 1–8 Manchester, UK. August 2008
[21] Jyothsna Cherapanamjeri, Lavanya Lingareddy, Himabind u. "Keyword based Question Answering System in Natural Language Interface to Database" International Journal of Advanced Research in Computer Engineering & Technology, December 2014
[22] http:// www.dabangvakil.com/indianpenalcode
[23] http:// indiacode.nic.in/coiweb/coifiles/amendment.htm

BIOGRAPHY

Shubhangi C. Tirpude is a MTech Scholar in the Comp Science & Engineering Department, Prof. Ram Meghe Institute of Technology & Research, Bandera- Amravati.

Dr, A. S. Alvi, Professor, in the Comp Science & Engineering Department, Prof. Ram Meghe Institute of Technology & Research, Badnera- Amravati.