INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Virtual question answering

Manda Swapna Reddy
mandaswapnareddy@gmail.com
Matrusri Engineering College, Hyderabad, Telangana

ABSTRACT

Question Answering (QA) system in information retrieval is a task of automatically answering a correct answer to the questions asked by the human in natural language using either a pre-structured database or a collection of natural language documents. It presents only the requested information instead of searching full documents like a search engine. As information in the day to day life is increasing, so to retrieve the exact fragment of information even for a simple query requires large and expensive resources. This is the paper which describes the different methodology and implementation details of question answering system for general language and also proposes the closed domain QA System for handling documents related to education acts sections to retrieve more precise answers using NLP techniques.

Keywords— *Virtual question answer, Natural language processing, Taggers, Sample stories, Rule based queries, Passage search*

1. INTRODUCTION

We are developing a Question Answering (QA) System for English sentences.

The user should be able to access answer of their questions in a user-friendly way, that is by questioning the system from the given English paragraph and the system will return the intended answer by searching in context of the paragraph using the repository of the English dictionary.

A novel strategy, in addition to conventional search and NLP techniques, will is used to construct the QA system. The focus is on context-based retrieval of information.

To extract passages from the collection of documents an information retrieval engine is needed which can analyze the keywords and passages in detail.

The answers to a query each sentence of the passages are converted into the Predicate Logic using Artificial Intelligence. And then answer is retrieved from the passages using resolution. In predicate logic, we can represent real-world facts as statements written as Well Formed Formula (WFF).

2. LITERATURE SURVEY

It also provides a good way of reasoning with knowledge. NLP research dates back to the late 1940s with Machine Translation (MT) being said to be the first computer-based application related to natural language.

It was Weaver and Booth who started one of the earliest MT projects in 1946, on computer translation based on expertise in breaking enemy codes during World War II.

However, a general agreement was made that, Weaver’s memorandum of 1949 has brought the idea of MT to general notice and had inspired many projects.

Weaver suggested using ideas from cryptography and information theory for language translation.

According to Liddy earliest works in MT followed the basic view, that the only difference between languages was vested in their vocabularies and the permitted word orders.

Hence systems which were made from this perspective basically used dictionary-lookup (for appropriate words for translation and reordering of the words after translation to fit the word-order rules of the target language). This was done without considering the lexical ambiguity inherent in natural language.

This generated poor results and called for researchers to come up with a more sufficient theory of language.

It was the Chomsky's 1957 publication of the syntactic structures which introduced the idea of generative grammar, which gave the linguistic a better understanding of how they could help the machine translation. Subsequently, other NLP application areas began to emerge, such as speech recognition.

3. PROBLEM DEFINITION and STATEMENT

Current systems have limited discourse capabilities that are almost exclusively handcrafted. Thus current systems are limited to viewing interaction, translation, and writing text as processing a sequence of either isolated sentences or loosely related paragraphs. Consequently, the user must adapt to such limited discourse.

Domains must be narrow enough so that the constraints on the relevant semantic concepts and relations can be expressed using current knowledge presentation techniques that is, primarily in terms of types and sorts. Processing may be viewed abstractly as the application of recursive tree re-writing rules, including filtering out trees not matching a certain pattern.

Handcrafting is necessary, particularly in the grammatical components of systems (the component technology that exhibits the least dependence on the application domain). Lexicons and axiomatizations of critical facts must be developed for each domain, and these remain time-consuming tasks.

The user must still adapt to the machine, but, as the products testify, the user can do so effectively.

4. METHODOLOGY

Step 1: To store the whole story into a file.
Step 2: To split each sentence of the story and save it in a different file.
Step 3: Now each sentence is converted into predicate logic and then clause form.
Step 4: Now question is to be asked and converted into predicate logic and then clause form.
Step 5: Finally the clause form of the question is resolved to final answer using resolution and unification algorithm.

5. ALGORITHMS

5.1 Algorithm: Unify (L1, L2)

1. If L1 or L2 is a variable or constant, then:
   (a) If L1 and L2 are identical, then return NIL.
   (b) Else if L1 is a variable, then if L1 occurs in L2 then return FAIL, else return {(L2/L1)}.
   (c) Else if L2 is a variable, then if L2 occurs in L1 then return FAIL, else return {(L1/L2)}.
   (d) Else return FAIL.
2. If the initial predicate symbols in L1 and L2 are not identical, then return FAIL.
3. If L1 and L2 have a different number of arguments, then return FAIL
4. Set SUBST to NIL.
5. For i 1 to number of arguments in L1:
   (a) Call Unify with the ith argument of L1 and the ith argument of L2, putting result in S.
   (b) If S = FAIL then return FAIL.
   (c) If S is not equal to NIL then: i. Apply S to the remainder of both L1 and L2. ii. SUBST: = APPEND(S, SUBST).

Return SUBST.

5.2 Algorithm: Convert to Clause Form

1. Eliminate, using: a b= ~a v b.
2. Reduce the scope of each to a single term, using:~ (~p) = p demerging's laws:
   ~(a ^b) = ~a V ~b
   ~ (a V b) =~ a^~ b
   x P(x) = x P(x) x P(x) = x P(x)
3. Standardize variables.
4. Move all quantifiers to the left of the formula without changing their relative order.
5. Eliminate existential quantifiers by inserting Skolem functions.
6. Drop the prefix.
7. Convert the expression into a conjunction of disjuncts, using associativity and distributivity.
8. Create a separate clause for each conjunct.
9. Standardize apart the variables in the set of clauses generated in step 8, using the fact that: (x: P(x) Q(x)) = x: P(x) x: Q(x) C.

Basis of Resolution and Herbrand's Theorem Herbrand's Theorem: To show that a set of clauses S is unsatisfiable, it is necessary to consider only interpretations over a particular set, called the Herbrand universe of S.

A set of clauses S is unsatisfiable if and only if a finite subset of ground instances (in which all bound variables have had a value substituted for them) of S is unsatisfiable.

5.3 Algorithm: Resolution

1. Convert all the propositions of F to clause form.
2. Negate P and convert the result to clause form. Add it to the set of clauses obtained in 1.
3. Repeat until either a contradiction is found, no progress can be made, or a predetermined amount of effort has been expended.
   (a) Select two clauses. Call these the parent clauses.
   (b) Resolve them together.

The resolve will be the disjunction of all the literals of both parent clauses with appropriate substitutions performed and with the following exception: If there is one pair of literals T1 and T2 such that one of the parent clauses contains T1 and the other contains T2 and if T1 and T2 are unifiable, then neither T1 nor T2 should appear in the resolve. If there is more than one pair of complementary literals, only one pair should be omitted from the resolve. c) If the resolve is the empty clause, then a contradiction has been found. If it is not, then add it to the set of clauses available to the procedure.

6. CONCLUSION

NLP is a very difficult task because human beings have good common sense and reasoning mechanism which they use to answer the question. Had the task of NLP be easy then 15 computer would have been told the story, ask questions and computer would have given answers and this preparation of the program would have required very less time.

7. REFERENCES

[1] Agius, H.W., and Angelides, M.C. “Developing knowledge-based intelligent multimedia tutoring systems using semantic content-based modelling”, Artificial Intelligence Review, 13, 1999, pp 55-83.
[2] O'Connor, C., Sceiford, E., Wang, G., and Foucar- Szocki, D. “Departure, Abandonment, and Dropout of E-learning: Dilemma and Solutions”, TechLearn 2003 Conference.
[3] Lempert, R. J., Popper, S. W., Bankes, S. C. “Shaping the next one hundred years: new methods for quantitative, long-term policy analysis, RAND”, Santa Monica, CA, 2003.
[4] Voorhees, E. M., Tice, D. M. “Building a question answering test collection”, Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece,2000.
[5] Harabagiu, S.,“FALCON: Boosting Knowledge for Answer Engines”. In Proceedings of the Text Retrieval Conference (TREC-9), 2000.
[6] Soubbotin, M., and Soubbotin, S. “Use of patterns for detection of likely answer strings”, A systematic approach. In the Proceeding of TREC 2002.
[7] Roussinov, D., and Robles, J. “Web Question Answering: Technology and Business Applications”. In the proceedings American Conference on Information Systems. August 6– 8, 2000, New York, NY.
[8] Dumais, S., Banko, M., Brill, E., Lin, J., and Ng, A. “Web Question Answering: Is More Always Better?” Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 2002.
[9] Yang, H, Chaisorn, L., Zhao, Y., Neo, S., and Chua, T. “VideoQA: Question Answering on News Video”.In Proceedings of the ACM conference on Multimedia (Multimedia’03), Berkeley, CA, November 2-8, 2003.
[10] Miller, G. “WordNet: An On-line Lexical Database”. In International Journal of Lexicography, 3-4, 1990.
[11] Jawad Makki, Anne-Marie Alquier, and Violaine Prince, “Ontology Population via techniques in Risk Management”, ICSWE: Fifth International Conference on Semantic Web Engineering, Germany, v.1, 2008.
[12] Farid Meziane, Nikos Athanasakis, Sophia Ananiadon, “Generating Natural Language specifications from UML class diagram” Springer-Verlag London Limited 2007.
[13] Pawan Kumar, Rashid Ahmad, Arun Kumar Rathaur, M.K.Sinha, R.Singhal: “Reengineering Machine Translation System through Symbiotic Approach”, Accepted in the 3rd International Conference in Contemporary computing, Noida India, August 9-11, 2010. Proceedings to be published by Springer in Communications in Computer and Information Sciences, ISSN: 1865-0929.
[14] G.Anthes: Automated Translation of Indian Languages, CACM, and Vol.53 (1), 2010.
[15] Artificial Intelligence by Elaine Rich, Kevin Knight and Shivashankar B Nair.
[16] Natural Language Understanding by James Allen. 17. Neural network, Fuzzy System and Genetic Algorithm by S. Rajasekharan and G.A.Vijaya Lakshmi Pai.

