Information Retrieval and Question Answering NLP Approach: An Artificial Intelligence Application

Mukul Aggarwal

Abstract-Natural Language Processing (NLP) is a major area of artificial intelligence research, which in its turn serves as a field of application and interaction of a number of other traditional AI areas. Until recently, the focus in AI applications in NLP was on knowledge representation, logical reasoning, and constraint satisfaction-first applied to semantics and later to the grammar. In the last decade, a dramatic shift in the NLP research has led to the prevalence of very large-scale applications of statistical methods, such as machine learning and data mining. Naturally, this also opened the way to the learning and optimization methods that constitute the core of modern AI, most notably genetic algorithms and neural networks. In this talk I will give an overview of the current trends in NLP and discuss the possible applications of traditional AI techniques and their combination in this fascinating area. Question-Answering(QA)aims at delivering concise information that contains answers to user questions. Intelligent Questing Answering Systems (IQASs) aim to help students become better readers. The computational challenges involved are (1) to assess the students' natural language inputs and (2) to provide appropriate feedback and guide students through the IQASs.

Index Terms—Natural Language Processing; NLP, information Retrieval, Laboratory, Framework.

I. INTRODUCTION

Artificial Intelligence is a branch of Science which deals with helping machines find solutions to complex problems in a more human-like fashion. This generally involves borrowing characteristics from human intelligence, and applying them as algorithms in a computer friendly way[2][8]. A more or less flexible or efficient approach can be taken depending on the requirements established, which influences how artificial the intelligent behavior appears.

We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does

Manuscript received May 27, 2011.

when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?

II. NATURAL LANGUAGE PROCESSING

NLP techniques are used in applications that make queries to databases, extract information from text, retrieve relevant documents from a collection, translate from one language to another, generate text responses, or recognize spoken words converting them into text. From the QA point of view, NL interfaces to databases [3] and information extraction [7] are most interesting. A common feature of NLP systems is that they convert text input into formal representation of meaning such as logic (first order predicate calculus), semantic networks, conceptual dependency diagrams, or frame-based representations [2]. Since the early days of NLP, (QASs) systems simulated human intelligence within the NL understanding research field. They worked as NL front-end to databases [48], dialogue systems [11] or story comprehension systems [2-3].

The linguistic front-end parses and analyses the user input in NL before translating it into an intermediate logical query. The logical query is then converted into a database query language, supported by the underlying database management system. The domain dependent knowledge contains information specific for the domain of interest: a lexicon and a world model. The lexicon contains admissible vocabulary words from the knowledge domain, often divided in word stems and word affixes (i.e. suffixes or prefixes), plus a representation of how they fit together [16]. The world model describes the structure of the domain of interest, i.e. the hierarchy of classes of the domain objects, plus the properties and the constraints that characterize the relationship between them.

NLP-based (QASs) systems may utilize machine learning to improve their syntax rules [2], lexicon [4], semantic rules [5], or the world model [4].



Mukul Aggarwal, Assistant Professor, Department of Information Technology, Krishna Institute of Engineering & Technology, Ghaziabad, Uttar Pradesh, India (e-mail: mukul.digital@gmail.com, mukul@kiet.edu: fisrtauthor@gamil.com).

Information Retrieval and Question Answering NLP Approach: An Artificial Intelligence Application

III. INFORMATION NEEDS

A. Question Answering

Intelligent Questing Answering Systems (IQASs) that are designed to interact meaningfully with, and adapt to, the users' input[5],. Different (IQASs) use different Natural Language Processing (NLP) techniques in their system. NLP systems may be structural, i.e., focused on grammar and logic, or non-structural, i.e., focused on words and statistics.

Limitations:

- Answers questions indirectly
- Does not attempt to understand the "meaning" of user's query or documents in the collection



IV. INFORMATION RETRIEVAL CYCLE

IR systems:

- Use statistical methods
- Rely on frequency of words in query, document, collection
- Retrieve complete documents

V. LABORATORY PLAN FRAMEWORK

NLP Laboratory section defines some features of NLP Lab that were described in proposed plan, specifically for the Language Processing Laboratory. These requirements lead to build a framework for NLP IQASs. This means, satisfying those requirements and describing solution for IQASs.

VI. IDEA OF IQASS

Determine the semantic type of the expected answer Who won the Nobel Peace Prize in 1991?" is looking for a PERSON. Retrieve documents that have the keywords "won", "Nobel Prize", and "1997

Look for named-entities of the proper type near keywords Look for a PERSON near the keywords "won", "Nobel Prize", and "1997"

VII. ARCHITECTURE

QA Architecture:



VIII. ANALYSIS THE QUESTION

- Question word signals
 - Who → person, organization, location (e.g., city)
 - When \rightarrow date
 - Where→location
 - What/Why/How $\rightarrow ??$
- Head noun signals
 - What city, which country, what year...
 - Which astronaut, what blues band, ...
- □ Scalar adjective signal
 - How long, how fast, how old etc...

Information Extraction (IE) from large amounts of unstructured text became especially important along with the expansion of WWW. It has moved the interest of the QA community to a more shallow approach to NLP [8]. Shallow NLP does not imply text understanding, i.e. semantic analysis of NL input. Instead, it focuses on extracting text chunks matching patterns or entities, thus providing the answer to user questions [2-3].



retrieval and natural language processing

- A real-world application of NLP technologies
- □ The dream: a vast repository of knowledge we can "talk to"
- □ We're a long way from there...

REFERENCES

- P. J. Hancox, A brief history of Natural Language Processing, [Online] http://www.cs.bham.ac.uk/~pjh/sem1a5/pt1/pt1_history.html.
- [2] H. Feili, Natural Language Processing Projects, [PowerPoint] Sharif UT, Tehran, Iran, 2003.
- [3] Androutsopoulos, G. D. Ritchie, P. Thanisch, "Natural language Interfaces to Databases: An Introduction", Journal of Natural Language Engineering, 1 (1), 1995.
- [3] S. Wintner, Laboratory in Natural Language Processing, [PDF Document] 2008.
- [4] A. Kirschenbaum, S. Wintner, "Minimally supervised transliteration for machine translation", Proceedings of The 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09), April 2009.
- [5] E. Sneiders, Automated Question Answering: Template-Based Approach, PhD thesis, Stockholm University / KTH press, Sweden, 2002.
- [6] William B. Frakes and R. Baeza-Yates (Eds.): Information Retrieval : Data Structures & Algorithms, New Jersey, Prentice Hall,1992.
- [7] D. Molla et al., "NLP for Answer Extraction in Technical Domains", Proc. of EACL 2003, Morgan Kaufmann, USA, 2003.
- [8] van Rijsbergen C J 1979 Information Retrieval Butterworths.
- [9] [31] D. Ravichandran and EH. Hovy, "Learning Surface Text Patterns for a Question Answering System", Proc. of ACL-2002, ACL press, USA, 2002.
- [10] B. Katz et al., "Omnibase: Uniform access to heterogeneous data for question answering", Proc. Of NLDB'2002, Springer-Verlag, Germany, 2002.
- [11] P.I. Sosnin, "Question-answer processor for cooperative work in human-computer environment", Proceedings of 2nd International IEEE Conference on Intelligent Systems, Varna, Bulgaria, June 22-24, 2004, Vol. 2, pp. 452 - 456.
- [12] G Huang, and H. Yao, "A system for Chinese question answering", Proceedings of IEEEIWIC International Conference on Web Intelligence, Halifax, Canada, Oct. 13-16,2003,pp.458 -461.
- [13] Y. Liu, S. Li, Y. Cao, C. Lin, D. Han and Y. Yu, "Understanding and summarizing answers in community-based question answering services," In COLING 2008, pages 497–504, Manchester, UK.
- [14] G. Erkan, D.R. Radev, "LexRank: graph-based lexical centraility as salience in text summarization," Journal of Arifical Intelligence Research 22.



Mr. Mukul Aggarwal has received his B.E. degree in Computer Science and Engineering in 2004 and M.Tech in Computer Science and Engineering from Rajiv Gandhi Proudyogiki Vishvavidyalaya,Bhopal, Madhya Pradesh in 2007. His areas of interest are Artificial Intelligence, Data Mining and Soft Computing. He is working as Assistant Professor in department of of Information Technology, Krishna Institute of Engineering & Technology, Ghaziabad, (Uttar Pradesh), India.



IX. REASON FOR SUPPORTING IQASS



X. TESTING OF IQASS BASSED ON INFORMATION RETRIEVAL

- Collection of documents (quantity)
- Set of information needs (topics)
- Sets of documents that satisfy the information needs (relevance judgments)

Three components of a test collection:

- Exactness
 - Recall
- Other measures derived there from

XI. FUTURE TRENDS

IQASs, simple word matching against various benchmarks seems adequate to provide satisfactory results especially when aggregated over several explanations. It is also demonstrates that a combination of techniques produces better results than using one technique on its own. Standardization needs will be studied in research and standards units, and development and evaluation groups will be announced for what standards they should be considered in developing and evaluating components.

XII. CONCLUSION

- □ Question answering is an exciting research area!
 - Lies at the intersection of information