

AN ANALYTICAL STUDY ON NATURAL LANGUAGE PROCESSING

Anjali Saini¹, P.K.Yadav²

¹ Ph.D Scholar, Computer Science and Engineering, Singhania University, India

² Department of Computer Science & Mathematics, Singhania University, India

Abstract - Natural Language Processing is an important and innovative research field in database system and intellectualized GIS. It is an interesting, creative and difficult domain for developing and evaluating representation and reasoning theories. The interpretation of Natural Language (English) database questions formulated by users in database queries is an major issue in the field of Artificial intelligence(AI). This research paper aims at giving an overview in this area focusing and evaluating the current status and visualizing future trends of NLP.

Keywords - Natural language Processing(NLP), NQML, NLP Aspects

I INTRODUCTION

Natural Language Processing (NLP) is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human-computer interaction. Many challenges in NLP involve natural language understanding that is enabling computers to derive meaning from natural or human language input and others involve natural language generation. Natural language generation convert information from computer databases into readable human language. Natural language Understanding convert chunks of text into more formal representations like first order logic structures that are simple and easier for computer programs to manipulate. It involves the identification of the intended semantic from the multiple possible semantics which can be derived from a natural language expression that takes the form of organized notations of natural languages concepts.

Human understanding of language requires background or common sense knowledge of the world. Human consciousness is tightly coupled with both language and our internal models of the outer world. Indeed, many argue that it is our consciousness that creates our own world (i.e., we create the worlds that we live in). It makes little sense to assume that the real world is static and is not affected by conscious entities living in that world. So, in trying to understand life and consciousness, it is important to understand the context of experiences in the world. Children playing often make up new words spontaneously that for the children involved has real meaning in the context of their lives. There are two basic approaches depending on whether we want to write an effective “natural language front end” to a software system or if we are motivated to do fundamental research on minds and consciousness by building a system that acquire structure and intelligence through its interaction with its environment.

Parsers are defined with finite state machines that recognize word sequences as specific words, noun phrases, verb phrases, etc. The context free programming for NLP includes the following. Difficulty in dealing with different sentences structures that has the same meaning. Handling number agreement between subjects and verbs. Determining the deep structure of input texts.

The term morphological tags refers to labeling of words with parts of speech tags. Some of the examples are as follows.

- Noun – cat, dog, boy etc
- Pronouns – He, she, it
 - Relative Pronouns – which, who, that
- Verb – run, throw, see etc
- Adverbs
 - Describe how some thing is done – fast, well. Etc
 - Time after, soon, etc
 - Questioning – How, why, when, where
 - Place – down, up, here etc.
- Determiners
 - Articles – a, an, the
 - Possessives – my, your, theirs etc
 - Demonstratives – this, that, these, those

NLP based Information Extraction

It is a Web based system which works in a hierarchical manner. The system converges various collaborative working modules in a single package. The software is designed to work as a real time application. IE/IR is of the highest importance, in the art of detection, to be able to recognize out of a number of facts, which are, incidental and which are vital. The system is to extract & retrieve the data related to any field such as medical field, spatial etc according to the query.

The following diagram describes a common approach to information retrieval systems.

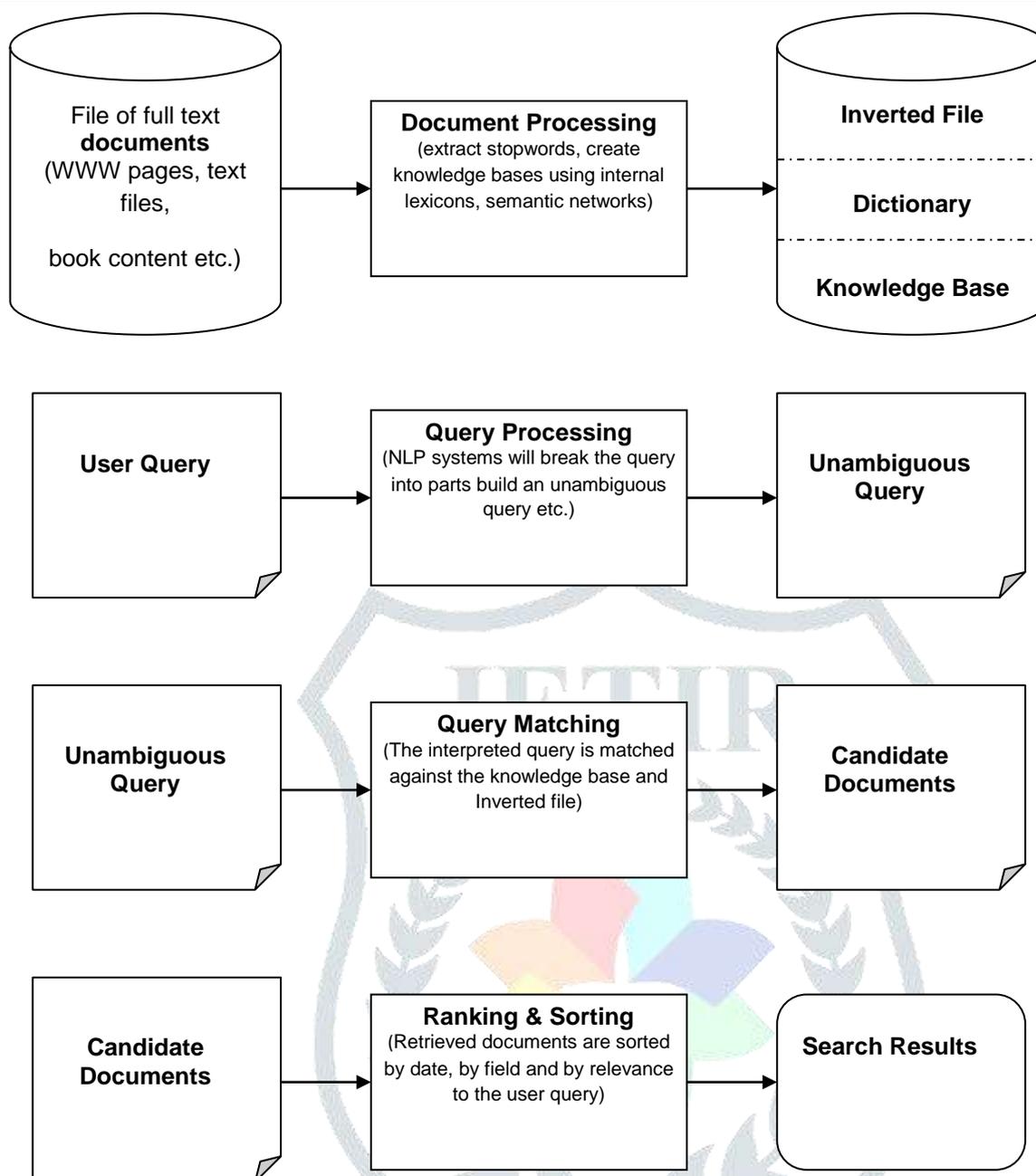


Figure 1 : Steps of NPL Search

II MAJOR ASPECTS OF NLP THEORY

Natural language processing is a combinatory discipline which combines linguistics, computer science and artificial intelligence in attempt to create an interactive and innovative system between human being and computer. The goal of NLP evaluation is to measure one or more qualities of an algorithm or a system ,in order to determine whether or what the system answers the goals of its designers or meet the requirements of users.

The important steps in building NLP technology are.

- To reduce domain of discourse to a minimum.
- Create a set of “use cases” to focus the effort in designing and writing ATNs, and to use for testing your NLP system during development.
- When possible capture text input from real users of your system, and incrementally build up a set of use cases that your system can handle correctly.
- Map indentified words / parts of speech to actions that system should perform.

There are three major aspects of any natural language understanding theory:

Syntax

The syntax describes the form of the language .It is usually specified by a grammar. Natural language is more complicated than other formal languages used for artificial languages of computer programs and logics.

Semantics

It provide the meaning of the utterances or sentences of the languages.Generally semantic theories exist,when we build a natural language understanding system for a particular application, we try to use the simplest representation.

Pragmatics

This component explains how the utterances relate to the world. For understanding language, an agent should consider more than the sentence; it has to take care about the context of the sentence, the state of world, the aim of speaker and listener etc.

III NATURAL QUERY MARKUP LANGUAGE(NQML)

Database systems are versatile and important technologies. A database is useful when users can easily retrieve data. It is extremely useful when people ask questions and queries to a database and get the answers. NQML is a mechanism that facilitates natural language frontend into relational databases. Natural query markup language is a system which tries to provide a solution to Question-Answering problem.

Some features of NQML are:

- It is compatible with XML {Extensible Markup language}
- NQML objects are human readable and clear.
- It is easier to write interpreters for NQML.

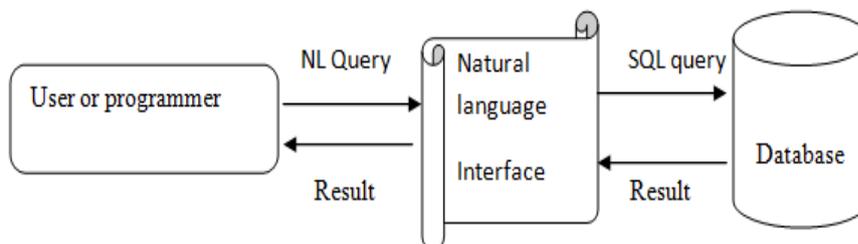


Figure 2: Data flow diagram of system

NQML system works as a mediator between user and database. The interpreter of NQML translates query from user's language into formal representations and hence translating the query. It then sends the translated query to the database and passes the result back to user.

IV LIMITATIONS AND FUTURE SCOPE

Here, we have assumed a simple form of natural language {NL}. This discussion of natural language processing assumes that natural language is compositional, the meaning of the whole can be derived or taken out from the meaning of its parts. We usually must know the context in the discourse and the situation in the world to discern what is meant by an utterance. Many types of ambiguity exist that can only be resolved by understanding the context of words. For example, we cannot always determine or tell the correct reference of a description without having knowledge of the context, conditions, and situations. A description does not always refer to a uniquely determined individual. NLP research is slowly shifting from lexical semantics to compositional semantics and further on narrative understanding. Human-level natural language processing, however, is an AI-complete problem. That is, it is equivalent to solving the central AI problem, i.e., making the computers as intelligent as people or strong AI. NLP's future is therefore tied closely to the development of Artificial Intelligence in general.

REFERENCES

- [1] Vesper Owei, "Natural Language Query Filtration in the Conceptual Query Language", 1060-3425/197, 1997 IEEE
- [2] Fangju Wang, "Parsing 'Grammatically Incomplete' Natural Language Queries to Spatial Databases", 0-7803-7078-3/01, 2001 IEEE.
- [3] Alberto Trombetta, "Equivalences and optimizations in an expressive XSLT fragment", Proceedings of the International Database Engineering and Applications Symposium (IDEAS'04) 1098-8068/04, 2004 IEEE
- [4] Alok Parlikar, "NQML: NATURAL QUERY MARKUP LANGUAGE", NLP - KE '05, 0-7803-9361-9/05, 2005 IEEE
- [5] WILLIAM LEIGH, "Interpretation of Natural Language Database Queries Using Optimization Methods", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, 0018-9472/86/0100-0040, D1986 IEEE
- [6] F. Siasar djahantighi, "Using Natural Language Processing in Order to Create SQL Queries", Proceedings of the International Conference on Computer and Communication Engineering 2008 978-1-4244-1692-9/08, 2008 IEEE
- [7] Yongping Du, "Pattern Optimization and the Application in Question Answering", Eighth International Conference on Intelligent Systems Design and Applications 978-0-7695-3382-7/08, 2008 IEEE
- [8] Radha Senthilkumar, "Nested XPath Query Optimization for XML Structured Document Database", ADCOM 2008, 978-1-4244-2963-9/08, 2008 IEEE
- [9] XU Jian-liang, "A Query Optimization Strategy Based on Domain Ontology", 2009 International Joint Conference on Artificial Intelligence 978-0-7695-3615-6/09, 2009 IEEE
- [10] Liguó DENG, "Query Processing for Fuzzy Temporal Queries Transformation", 978-1-4244-4507-3/09, 2009 IEEE