NLP Project: Paperwork

**TECHNICAL BACKGROUND**

**Natural Language Processing: The Basics**

Natural Language Processing (NLP), in its simplest sense, refers to the utilization of a computer’s capability to process sentences stated in natural human languages (such as English, Japanese, Korean, etc.) instead of specialized computing languages (Java, C++, C#, etc.). However, NLP still holds significant differences between natural human languages. The term “natural language” actually refers to a stricter subset of the said human languages. They do not function the same way as how people carry-on with their everyday conversations. In fact, a writer once said that “human languages allow anomalies that natural languages cannot allow” (Steedman, p. 229). The phrase “natural language processing” can cover a wide array of topics involving speech recognition, context reference issues, discourse planning and generation, and syntactic and semantic analysis and processing. The term “processing” in the similar phrase is not restricted to “interpretation” only. It pertains to a combination of both “interpretation” and “generation”, which is why systems possessing the said capability (NLP) are not limited to getting a grasp of the natural languages but are also given the ability to produce their own. A system’s skill that could count as an example of a natural language processing capability would be developing a decent conversation in pure human language. Additionally, computer systems that can convert human languages to computer languages and vice-versa are already existing. They provide assistance to people with disabilities such as blindness in terms of using personal computers. Translation programs were also made possible by NLP. There are applications in the market having the capability to translate one human language to another. Likewise, grammar and spelling checkers are also programmed following the NLP algorithm. Lastly, a computer that can read human languages (in publications such as books) is also a product of natural language processing. With respect to the interpretation side, NLP involves the following subtopics: signal processing, syntactic analysis, semantic analysis, and pragmatics. Signal processing is basically acquiring spoken words and turning them into text. Syntactic analysis, on the other hand, focuses on the grammar composition of the sentences. Semantic analysis deals with how the words and sentences refer to their corresponding worldly elements, or rather their own meaning in a particular spoken human language. Finally, the term pragmatics concerns how a specific word functions in an individual’s day-to-day conversations with one another. (Introduction to Natural Language Processing, 2006).

Public information dissemination across the Internet triggered the emergence of research areas under the field of NLP such as Information Extraction, Information Retrieval, and Question-Answering. (Mendes & Antunes).

**Products of NLP**

**Software Applications**

The following applications were developed based on NLP: Thinknowlogy – a grammar-based software, making use of the Natural Laws of Intelligence in grammar to be able to develop a certain intelligence through natural language in software, Bitextor – an application bearing the objective of generating translation memories with the help of multilingual Web sites as a source, TagAligner – a software used to generate translation memories from two XHTML tagged files, RelEx - is an English-language semantic dependency relationship extractor that can identify subject, object, indirect object, etc. between a group of words in a sentence, and Wintermute – a framework of applications using neural networking to learn about its host. (13 Projects Tagged "Natural Language Processing", 2015).