Lindsay Robbins  
Spring 2022

INFO 5200.001

Concept Briefing

**Natural Language Processing**

**Abstract**

Natural language processing (NLP) is an interdisciplinary concept that, within the context of information science, describes how computers use natural language search queries to retrieve information. This area of research undergoes constant development as it is closely involved with technology. This report will introduce the reader to NLP as a concept, its implications, benefits, drawbacks, and examples of implementation.

**Introduction**

This briefing examines the concept of Natural Language Processing (NLP) within the context of the information science disciplines. Through this briefing, the reader will gain understanding of the following: (1) what an NLP is, (2) ways an NLP can be used, and (3) what some potential implications and issues of an NLP are.

**Definition**

Natural Language Processing is the process by which a computer attempts to analyze a natural language query in order to retrieve information in a system (Joudrey & Taylor, 2018, p. 519). Natural language refers to the input provided by a user as they search an information system. NLP attempts to help improve a computer’s ability to understand natural language phrases and retrieve a greater number of results that are more relevant to a user’s information needs. NLP is a highly interdisciplinary subject as it requires knowledge related to linguistics as well as technological innovations such as artificial intelligence.

**Purpose and implications**

**NLP versus Controlled Vocabulary**

Language is a living concept which changes and adapts over time. This can be demonstrated by the colloquial use of the word Google as a verb to indicate the action of searching through Google, or through the use of Photoshop as a verb to describe editing a digital image. Users may begin a search for information, especially in online collections, through natural language. These queries are often conducted through text or voice and can be input in full sentences or through the use of keywords. As Hoenkamp and Bruza examine, natural language queries allow users to search for more conceptual or complicated information as opposed to something more definite or topic-specific (2015).

Natural language can pose issues in information retrieval, however, due to the nature of human language having different dialects, abbreviations and more. NLP systems often deal with verbal queries in addition to textual ones, however, many users may experience frustration in using verbal search systems. In a 2020 study, Sa and Yuan found that by introducing the ability to modify a vocal search query, users were more likely to use a voice-based system though very few systems have implemented this ability. This is in part why controlled vocabularies are created, in order to alleviate some of the issues associated with natural language. There are some issues that can create barriers to connecting users with information when using a controlled vocabulary, however, and there is debate whether or not information systems should do away with controlled vocabularies altogether (Gross et al., 2014).

**Benefits of NLP**

NLP can be beneficial in cases where search terms are variable. For example, in the cataloguing software, Libib, the use of a controlled vocabulary will only yield results for an exact spelling of a word such as “conversational” versus “conversations”. Systems utilizing NLP may be able to retrieve information that contains both of these terms, where Libib may only retrieve instances of one of the other, depending on the exact query.

**Challenges and Issues of NLP**

This is not a perfect system, though it has become more sophisticated within recent years. Joudrey & Taylor (2018) suggest that there has been some positive implementation of NLP in information systems with more narrow subject areas, however this may not be the case in some collections (p. 475). Among other potential issues, Callister (2020) expresses concerns with how in the context of a subject-specific collection of information—namely law, NLP can contribute to issues such as creating unintended biases in search algorithms.

**Examples**

**Chat Assistants**

Chat assistants, also referred to as chatbots or virtual assistants, are one example of NLP. Many users who have visited the website of a library or university website may have encountered a chat popup. These may be monitored by staff members or completely autonomous. The chatbot will retrieve information based on queries provided by the user, which can often be input through natural language in a conversational manner.

**Machine Translation**

Machine translation is an aspect of NLP which enables users to retrieve information from a system in several languages, which could help in countering language barriers in information access. In a 2016 study, it was determined that while this technology does not completely negate these barriers, it is still a beneficial tool (Nzomo et al.).

**Word Prediction Systems**

Word prediction is one way in which NLP may be demonstrated. When creating a query, especially through text, a prediction system will attempt to anticipate the following words entered, which are determined through lists of word frequencies. As demonstrated in a 2019 study, the Tigrigna Word Prediction System was found to be highly beneficial to users and was able to predict words with a 91.6% accuracy, leading to a smaller number of necessary keystrokes on the part of the user (Hayag & Nuruhsin).

**Conclusion**

As NLP is so intimately tied to technology, it is likely to change and become more sophisticated over time. In the future, other communication methods may even adapt NLP systems as well. However, as NLP deals with language, which is very flexible and human, it may not be possible to ever create a perfect system. In order to connect users more reliably with information, NLP likely needs to be implemented in conjunction with other systems such as those with controlled vocabularies.

**References**

Callister, P. D. (2020). Law, artificial intelligence, and natural language processing: A funny thing happened on the way to my search results. *Law Library Journal*, *112*(2), 161–212.

Gross, T., Taylor, A., & Joudrey, D. (2014). Still a lot to lose: The role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly, 53*(1), 1-39. https://doi.org/10.1080/01639374.2014.917447.

Hayag, M. R., & Nuruhsin, M. (2019). The implementation of statistical word prediction algorithm in Tigrigna word prediction system text entry on android phone. *Information Technologist*, *16*(2), 267–283.

Hoenkamp, E. & Bruza, P. (2015). How everyday language can and will boost effective information retrieval. *Journal of the Association for Information Science and Technology, 66*(8), 1546-1558. https://doi.org/10.1002/asi.23279

Joudrey, D. & Taylor, A. (2018). Systems for vocabulary control. In *The organization of information* (pp. 475-538). Libraries Unlimited.

Nzomo, P., Ajiferuke, I., Vaughan, L., & McKenzie, P. (2016). Multilingual information retrieval & use: Perceptions and practices amongst bi/multilingual academic users. *The Journal of Academic Librarianship, 42(*5), 495-502. https://doi.org/10.1016/j.acalib.2016.06.012.

Sa, N., & Yuan, X. (2020). Examining users’ partial query modification patterns in voice search. *Journal of the Association for Information Science & Technology*, *71*(3), 251–263. https://doi.org/10.1002/asi.24238