

LET'S NOT FORGET HOW TO WALK...

Last December Natural Language Processing and academic literature searching came into the spotlight in online conversations. Katherine Howard took a look at where and how this rapidly evolving technology fits with core competencies.

Natural Language Processing – or NLP – is not a new field of study but, according to a recent announcement from Swets, it has seen renewed attention since Yahoo acquired the NLP technology company SkyPhrase in early December last year.

NLP draws on concepts and research from multiple fields such as linguistics, including computer-mediated communication (CMC), human-computer interaction (HCI), artificial intelligence (AI), and those more closely related to LIS such as information retrieval (IR) and knowledge representation (KR).

Broadly, NLP is about developing a computer language that is capable of understanding natural human languages such as English, with all of its idiosyncrasies – synonyms, homonyms and homographs for example – that are largely incomprehensible to a computer. However, advances have been, and continue to be made with perhaps the best-known current application of NLP being spelling and grammar check.

In a series of posts on the Swets blog, the advantages of NLP to academic researchers in particular were discussed, given the requirement to “stay [...] abreast of recent developments” in their field. The case is made that while a researcher may be able to tell very quickly which search results are relevant, the sheer amount of information available makes it possible that other useful extant material may not be located by that researcher.

NLP seeks to understand the content/topic that the researcher is interested in by using ‘indicator phrases’ to contextualise the surrounding words and sentences in order to return more highly relevant results. Or, in ‘librarian speak,’ a higher precision to recall ratio, where the computer itself will be able to determine if the duck you want to carve is for dinner or a hobby.

This contextualisation is known as ‘semantic search’, and relates this concept to the (until now) hypothetical semantic web. The use of metadata as an aid to this semantic discoverability is important as NLP processing principles can be used to “automatically extract textual metadata and classify resources into established taxonomies”.

The premise underpinning this technological development is the realisation that “Familiar keyword search as in Google [...] is not always the ideal mechanism when looking for scholarly material [...]. Pure keyword search may well miss crucial subtleties of context in academic literature, fail to

return results rich in semantically similar keywords, or be incapable of tailoring results depending on whether the user is looking for discussion of methods, support for a hypothesis, background reading and literature review or a myriad of other specific uses for the material.”

As information professionals, we have known this for a while now. We have highly developed search skills that enable us to find the right information at the right time – but what do non-information professionals do? How do people obtain the skills to not only find information, but to evaluate it for its relevancy, accuracy and validity?

The necessity for these skills in today's information-rich environment, I believe, moves far beyond the academic researchers. Anyone who seeks information electronically – whether via the World Wide Web or via subscription databases – should have these skills.

This doesn't mean I'm arguing for Grandma to have a skill set equal to that of a qualified librarian – these skills can be learnt (and used) to varying degrees of expertise. I believe we, as a profession, need to advocate for these skills to be viewed as essential in today's world as both reading and writing.

“But the semantic web is coming soon – we won't have to think about how to structure a search statement, or to evaluate results for relevancy because it will all be done for us,” I hear you cry! As wonderful and exciting as these developments are – and I am looking forward to seeing the opportunities that NLP may bring to our profession – I am not sure that we should rely so heavily on technology to “determine whether some, all or none [of the search results] are relevant to the user's search”. It is another tool that can be used to enhance the service that we offer our users.

After all, most of us know how to drive a car, but that doesn't mean we have forgotten how to walk.

KATHERINE HOWARD

PhD Candidate, Queensland University of Technology
howard.katherine@gmail.com

YOU CAN READ THE FULL POSTS AT:

S.M. Das (2013, December 3). Natural language processing (NLP) and academic literature search (Part 1) [blog post]. Retrieved from <http://www.swets.com/blog/natural-language-processing-nlp-and-academic-literature-search-part-i#close>

S.M. Das (2013, December 17). NLP in academic literature search (Part 2) [blog post]. Retrieved from <http://www.swets.com/blog/nlp-in-academic-literature-search-part-ii#.UtyexCiZ420>