The Impact of Artificial Intelligence:
Brave New World?

Gary Horrocks, Editor, eLucidate

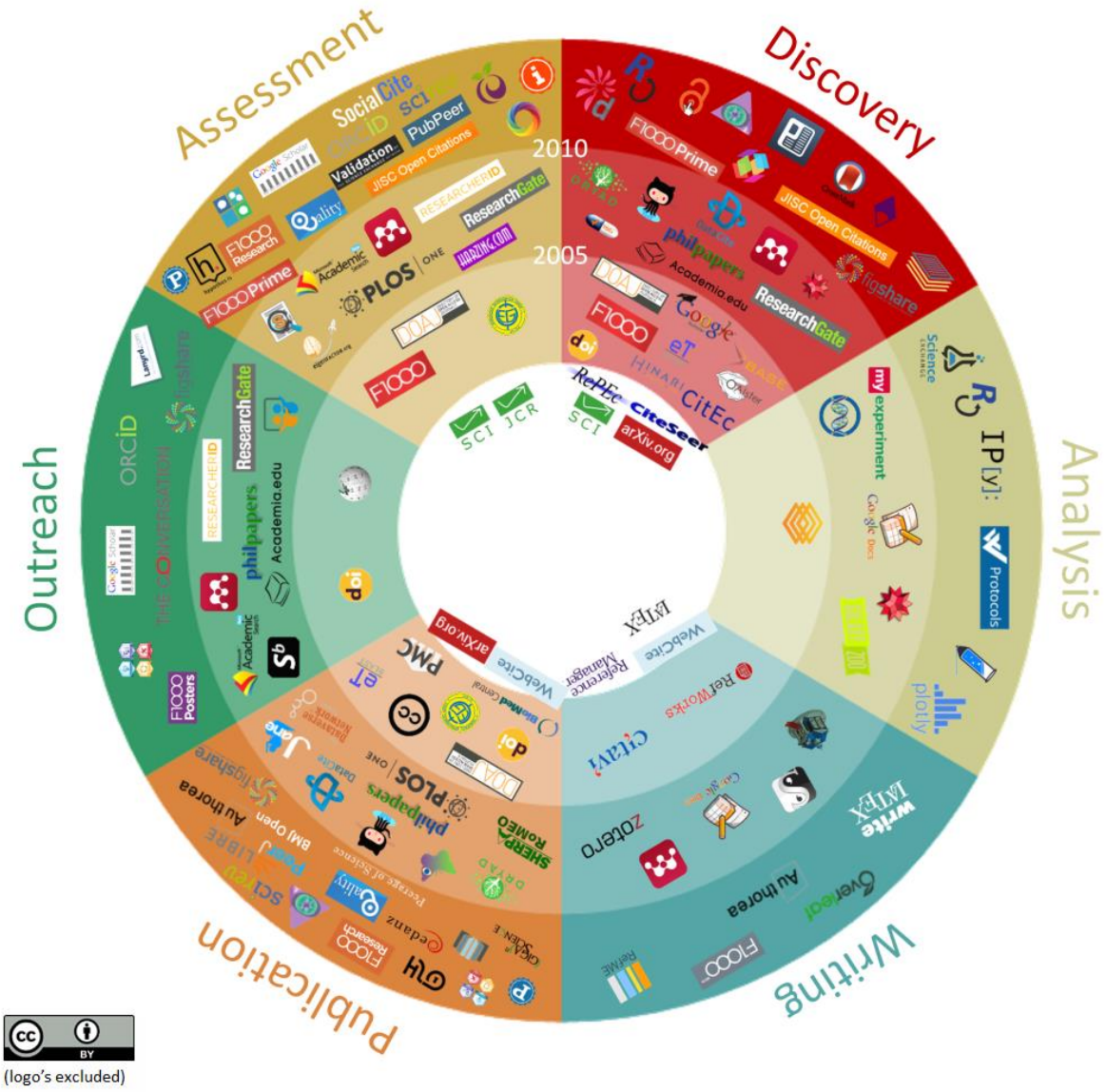
info.ukeig@cilip.org.uk

This year's UKeiG Members' Day, held at CILIP HQ on Friday 7th June 2019, was a great success attracting a diverse range of delegates from across the library and information profession. Several sectors were represented including academic, commercial, government and public. AI is all pervasive in the media and often perceived as a "black box problem" - mysteriously complex, ill understood and feared as it threatens a level of intelligent automation that will make the workforce redundant. On the other hand, it is sometimes dismissed as pie on the sky; another new-fangled idea that never quite materialises into reality.

Hype or game-changer?

Michael Upshall initiated the discussion with a fascinating presentation on the role of AI in digital academic publishing, content enrichment and knowledge identification. Far from being blue sky hype, AI was already a game changer. He articulated the work of the UNSILO project: "Rethinking publishing with AI". UNSILO goes beyond traditional string matching and keyword extraction using fully automated concept matching to extract meaning and context. The project utilises a mathematical algorithm to analyse a huge corpus of text identifying descriptive "significant phrases" within a document. It creates clusters of concepts and identifies semantic relationships by processing the proximity of words surrounding a term. The word "bridge", for example, has many meanings and synonymous alternatives. It could allude to a connecting structure, part of a ship, a partial denture or part of a stringed instrument. The terminology that surrounds it imparts context and meaning.

Upshall is using this approach to build "semantic profiles" of scholarly journals, linking the technology to the academic workflow; the complex "circle of scholarship" – six areas of activity in the research cycle, and hundreds of the tools that support the research process.



Bosman and Kramer 2015/16

Machine learning is crucial in scientific publishing where there are currently 24, 000 journals and 3, 000 papers published a day. Manual classification schemes, vocabularies, taxonomies and ontologies have always played an essential role in information retrieval but Upshall argues that they are expensive and fundamentally flawed; reactive not proactive. "They will never be complete. They will never be large enough." The pervasive controlled vocabulary MeSH (Medical Subject Headings) maps the paradigm of biomedical science - neologisms and synonymous relationships - but humans are required to build and maintain them. This human imposition of terminology can distort context and create an artificial language. Moreover, the multiplicity of controlled vocabularies, ontologies, cataloguing standards frameworks and classification schemes makes interoperability and translation between schemes incredibly difficult. UNSILO, Upshall argues, eliminates the ambiguity of human language. By linking data and analysing the proximity of phrases it enables the disambiguation of problematic terminological conundrums like abbreviations and synonymous phrases. "Semantic enrichment" is the way forward argued Upshall. More

controversially, in devil's advocate mode, he announced: "Why even bother building a taxonomy?"

UNSILO's approach is far from pie in the sky and already has several practical, real-life uses, enabling, for example, AI to build profiles of specific academics and researchers, journal and article level analytics. A corpus of 28 million abstracts from the PubMed database has been analysed to develop a "Reviewer Finder." Upshall noted that in 2016 26% of US academics declined requests to peer review papers as they were irrelevant to their research expertise. By identifying the juxtaposition and overlapping of concepts the project has made substantial inroads into improving the peer review workflow, much more easily identifying the most relevant organisations and researchers to submit papers to for review. Similarly, "Journal Analysis" supports the identification of the most relevant journals to publish in. AI is facilitating a much more sophisticated level of data analysis; the notion of "concept curation" way beyond information retrieval based on text. The technology is also capable of translating across subject domains even when there are significant differences and variations in terminology.

Information resource management

David Haynes, City, University of London, presented on the potential role for AI in information resource management (IRM) arguing that, if anything, ontological/typological models were on the ascendant, and that human intervention was key to the implementation of AI this this area. As Chair of ISKO UK, the UK Chapter of the International Society for Knowledge Organization, he referenced the recent July 2019 ISKO UK conference, "The Human Position in an Artificial World – Creativity, ethics and AI in knowledge organization." He also cited Synaptica Graphite KOS, a "powerful tool for creating and curating Knowledge Organisation Systems… based on Linked Data and Semantic Web standards [offering] speed and flexibility in the creation and management of various types of controlled vocabularies." He agreed that concept management was fundamental and extracting meaning and context from complex linguistic relationships and associations between ideas goes way beyond the traditional mapping of hierarchical associations between words.

The key consideration in David's discussion was the nebulous nature of AI across the library and information community. Was it synonymous with automation? Was it the replacement of cognitive processing by machines? There is a multiplicity of definitions that change every day, but the consensus is that AI equates to decision making capability utilising iterative systems that can learn and modify their behaviour.

In order to address the impact of AI on information resource management the first step is to articulate the IRM cycle.

- Identifying information needs
- Defining scope
- Collecting resources
- Organising the resources
- Storage and retrieval

- Make resources discoverable
- Feedback and evaluation
- Disposal

In a workshop format Haynes introduced a practical approach to assessing each aspect of the cycle. What role could AI play in each area of activity?

- Could AI enhance human activity in each area or not?
- Could AI replace this human activity?
- Is there anything uniquely human about these activities?

eLucidate readers are invited to answer some of these questions and email the editor with their thoughts.

Searching for meaning in text

David's presentation segued perfectly into the Director of UXLabs and Founder of 2dSearch Dr. Tony Russell Rose's informative and thought-provoking crash course in the fundamentals of natural language processing (NLP) - the terminology, techniques and applications - and how NLP interfaces with AI. There was, he argued, a significant overlap between the two, but NLP was a sub field of AI and closely aligned to computer science and programming. The primary objective of NLP is to disambiguate language and search for meaning in text. It is a major growth area, a multi-faceted field of research including basic text processing, text mining, the human computer interface, language modelling and lexical semantics.

NLP research faces monumentally complex obstacles confronted with linguistic phenomena. Language is ambiguous and the key tenet of NLP is resolving that ambiguity. Tony illustrated his case in point with some amusing examples of newspaper headlines.

"Prostitutes appeal to Pope."

"Drunk gets nine years in violin case."

"Miners refuse to work after death."

Tony articulated some of the linguistic dilemmas that make NLP so problematic:

- Polysemy, where a word maps to many different concepts - e.g.: Bat (sports), Bat (small animal with wings), BAT (British American Tobacco)
- Synonymy, where one concept maps to many different words – e.g.: Hardworking: diligent, determined, industrious, enterprising
- Word order – e.g.: Venetian blind versus blind venetian
- Stop word removal – e.g.: The Who, Take That, "To be or not to be"
- Stemming – e.g.: fish, fisher, fishing
- Parsing (analysing a string of text into logical syntactic components) – e.g.: "I saw the man on the hill with a telescope"

Tony argued that language is constantly changing. “I want to buy mobile” would have been meaningless twenty years ago, even meaningless today in the United States where “cell phone” is the popular parlance. How would you go about analysing sarcasm, irony, jargon and slang? Similarly, idiomatic language poses key problems. (He was a “dark horse”. She “threw in the towel”.) In a rapidly changing world neologisms are also prevalent. Social media alone has generated many: “Unfollow” and “retweet”, for example.

Tony went on to list some of the disciplines that are researching solutions to these problems, each approaching the challenges with different perspectives. Computer science underpins the foundations of all this research.

- Text analytics – linguistic, analytical and predictive techniques to extract structure and meaning from unstructured documents
- Computational linguistics – the use of computational techniques to study linguistic phenomena
- Cognitive science – research into human information processing
- Information science – the analysis, classification, retrieval, manipulation and dissemination of information

One fascinating field of research that I am keen to explore in a future issue of eLucidate is sentiment analysis; the identification and extraction of subjective information. How do we identify emotions in text; fact versus opinions?

Tony concluded by providing numerous examples of NLPtoolkits and applications including: spaCy software, TextBlob, Apache OpenNLP.

Challenges and opportunities

In January, CILIP CEO Nick Poole sent New Year’s greetings for 2019 to the membership. He cited Artificial Intelligence and machine learning. “If 2018 was a big year for Machine Learning and AI, 2019 looks to be even bigger as the technology continues to make its way into both public awareness and mainstream applications for consumers and the workplace.” AI is very much a reality. It is such an all-embracing term that it includes a multiplicity of technologies and applications at various stages of development. Some innovations and technologies may take years to come to fruition, others are very much impacting on resources and services here and now. Voice recognition, virtual assistants and chatbot services are a typical example.

There are huge challenges and opportunities for the knowledge, information management and library sector. The benefits are obvious, and this year’s UKeiG Members’ Day captured just a few examples of the huge potential that AI offers in transforming digital publishing and information retrieval with the development of analytical tools that identify, extract and analyse text. Information science is a key tenet of AI and the profession is well placed to lead in developments and research in this emerging discipline. Delegates were enthused by the day and eager to learn more. “Food for thought” was a common response.

"Challenging. Compelling". There is a demand to explore the practical implementation of AI in the workplace in much more depth.

The spectre of disintermediation, the impending fear of redundancy, has always haunted the profession. Online in the 1970s, CD ROM in the 1980s, the growth of the Internet in the 1990s, seemingly threatened to displace the library and information professional, but the sector has always risen to the challenge. A key concern about AI is the lack of trust; the potential for bias based on flawed algorithms. Cynicism about Google's search results provides a typical example of the pitfalls that lie ahead. The information professional is well-placed to question those algorithms, to identify and check bias and assure quality. The LIS workforce will be required to refine, review and evaluate AI applications and build business cases for them. AI is also an iterative technology, so will require always require human intervention and "training".

An exciting road lies ahead.

Michael Upshall, UNSILO mu@unsilo.ai - @michael_upshall
David Haynes, City, University of London - david.haynes@city.ac.uk
https://blogs.city.ac.uk/privacycalculus/
Dr. Tony Russell-Rose, UXLabs, 2dSearch - tgr@uxlabs.co.uk - @tonygrr
http://isquared.wordpress.com