Qualitative and Quantitative Methods in Libraries (QQML) 12,1: 153-166,2023

Applications of Artificial Intelligence Technology in the Field of Scientific Information

Ioanna Tsimara[1], Ageliki Oikonomou[2] Anthi Katsirikou[3]

[123]University of Piraeus Library

Abstract

It is beyond dispute that today's world is characterized by direct communication and internet access to vast data storage with information becoming more and more available. However, this information that is disseminated through various information channels and especially in the field of scientific information is often scattered, fragmented and unconnected, resulting in losing the correlation of its concepts and thus hinders the path to knowledge. Search engines as the right tool for researching and retrieving these data retrieve results that are displayed sorted by the relevance of their content to the question.

However reliable search engines are, they fail to produce results on a query that can be "translated" into more than one concept. This lack and the need for machines to mimic the function of human brain resulted in the creation of the Semantic Web. While critics of the Semantic Web have questioned its appropriateness, proponents say its applications in libraries and information science, industry, biology, medicine and humanities research have already proven their validity and their contribution to providing a more interactive user experience.

The purpose of this article is to describe the use of Semantic Web and Artificial Intelligence technologies in libraries as important tools to enhance access to the ever-increasing need for full text retrieval and how the University of Piraeus Library utilizes them for an effective information retrieval of the sources available to its users.

Keywords: Semantic Web, Artificial Intelligence, Library catalogs, Information retrieval

Received: 9.2.2023 Accepted:23.3.2023 ISSN 2241-1925

1. Introduction

Everyone uses the World Wide Web as their primary source of information, including students, researchers, professors, professionals from every industry, information managers like librarians, and members of the general public who regularly look up news, product reviews, entertainment options, and book information.

This information is shared between distant systems via the Internet, a worldwide system of interconnected computer networks, and is retrieved via programs such as search engines (such as Google).

The resulting search results are displayed and sorted by relevance to the query content, and searching for them is done through a process that involves identifying (related) objects often in the form of sentences containing verbs and words (e.g. The sun is a star).

However, because such phrases lack specific language descriptions or a structure that can be processed by computers directly, the data generated may not be sufficient to grasp more than one subject (Martinez-Rodriguez, Lopez-Arevalo & Rios-Alvarado, 2020). The Semantic Web was developed as a result of this deficiency and the necessity for machines to mimic how the human brain understands natural language.

The Semantic Web, also known as Web 3.0, is a project of the W3C (World-Wide Web Consortium) that was launched in 2001 and was motivated by the idea of Tim Berners-Lee, the man who invented the World Wide Web, who envisioned it as a more adaptable, integrated, and automated web.

The Semantic Web concept was not new since linguists and cognitive scientists have been interested in how computers might grasp human language since the early 1960s.

For instance, the linguist M. Ross Quillian created a model based on human language behavior in 1967 that was embedded into a computer and was intended to encode data from a common dictionary into a complex network of components and linking compounds of English words (Quillian, 1967). Although the term was first used by Berners-Lee and his associates. They claimed in Scientific American journal, in May 2001, that "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling

computers and people to work in cooperation." (Hendler, Lassila, and Berners-Lee, 2001).

1.1 Web 2.0 - Web 3.0

The conventional Web (also known as Web 1.0 or Document Web) contains nothing but information resources. Websites offer information on the material that users have recently seen or downloaded. In other words, Web 1.0 is a "Read Only" web (DiNucci, 1999). Before 1999, on the internet data was provided as HTML, pdf, etc. the files and their links were based on human-readable hyperlinks. In contrast, Web 3.0 is a web of connected data, a single global database that enables machines to draw conclusions about the world without needing explicit input from humans.

In particular, with Web 2.0 the interaction of users and the Web begins where people create blogs, Wikis (e.g. Wikipedia), social networks, Instant Messaging applications, etc. transforming the Web into a "Participatory Web". In contrast, Web 3.0 has enabled contextual ideas, personalized search, and inductive reasoning through the use of ontologies, semantic search, thesauri, and knowledge bases (Douligeris et al., 2017). The Semantic Web, according to Park & Kipp (2019), is "a technology that enables the sharing of machine-readable data using metadata schemas and ontologies to assign meaning to the data".

1.2 Artificial Intelligence and Libraries

Along with the mentioned growth of the Internet, Artificial Intelligence has also been produced and has a significant impact on the development of several fields, including medical, finance, education, industry, etc. The most well-known Artificial Intelligence systems that integrate expert systems with fuzzy logic, according to Georgouli (2015), include intelligent search machines, data mining systems, recommender systems and ambient intelligence systems.

Robotics, chatbots, natural language processing, Big Data, and Text Data Mining are all examples of how Artificial Intelligence is making its way into the library environment as a result of the growth of the Internet and the abundance of information (Ali, Naeem, & Bhatti, 2020).

The reason for this is that, in accordance with Bohyun Kim, chief technology officer at the University of Rhode Island Libraries in Kingston (Eberhart, 2018), with artificial intelligence:

- Indexing services are much faster and more accurate.
- Information retrieval offers a wider range of answers to user queries. For instance, Bohyun highlights Yewno, which uncovers connections that the researcher may not have been aware of, and Quartolio (https://quartolio.com), a platform that ties to several research points.
- Referral service is enhanced by the use of chatbots, such as through Amazon's Alexa virtual assistant.

2. Yewno

Yewno (https://www.yewno.com/) is a semantic search engine that recognizes relationships between ideas to automatically extract concepts, allowing users to explore sources beyond full-text items (Schreur, 2019). Yewno's mission, according to the information provided on its website, is to mine knowledge and documents from a huge amount of data.

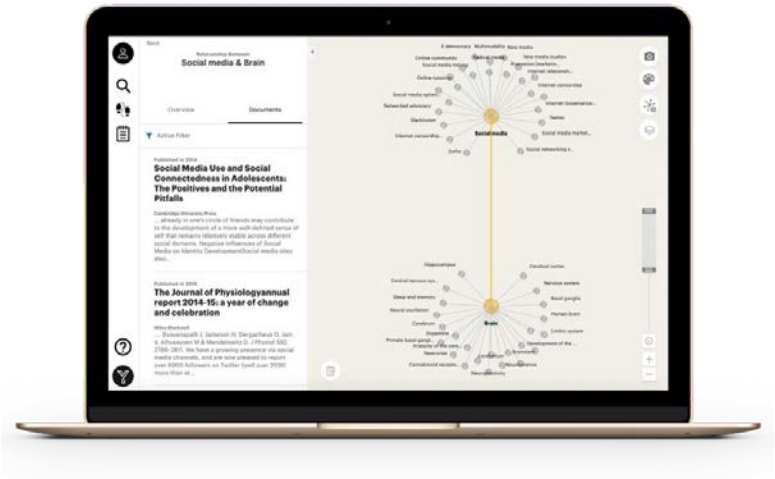
Yewno offers five platforms:

1. Yewno Discover, a knowledge search tool used in the field of education,
2. The artificial intelligence-based investing research platform (Yewno Edge),
3. Yewno Unearth, which is used in the publishing industry,
4. Yewno Biomedical, which is utilized in life science sectors, and
5. Yewno Government, which specializes in government information.

In this article, we will deal more extensively with Yewno Discover which is related to the education sector and is used by many libraries.

Knowledge Search - Yewno Discover

Therefore, Yewno Discover is a conceptual search engine that allows users to discover relationships between ideas and visualizes information using full-text analysis and machine learning. Users may now explore ideas and connections, learning from resources they would have otherwise missed. Through an interactive web, concepts are connected to other relevant concepts. The links between concepts are graphically represented using Yewno Discover. The user may understand how closely connected the concepts are to one another by looking at this visual map (Image 1.) showing the links between individuals, events, and concepts. The software extracts concepts from journal publications, books, and dissertations.



Yewno.com

Image 1.

A completely new approach to knowledge and information, supplemented by a visual style and the most recent developments in Artificial Intelligence, sets it apart from typical search products. Yewno Discover, which was introduced in 2016 and has already been adopted by both major research institutions and minor colleges in the US, Germany, and the UK, is intended to encourage users to explore new ideas.

Launch of Yewno Discover

Yewno was inspired by the need to make sense of the ever-growing data sources that are fragmented and scattered across the World Wide Web. Yewno was founded in Redwood City, California in 2014. Yewno is a word play on the phrase "you know." It also makes reference to the yew tree, a tree that represents rebirth and transformation.

Yewno's idea was first applied by Yewno CEO Gramatica Ruggero to the life sciences industry when a Swiss biotech business company expressed interest in investigating how it can reuse current biological molecules to treat uncommon disorders. Ruggero used an algorithmic approach to help. Using an analytical framework, he developed the original concept into a rich biomedical knowledge network by receiving and processing roughly 23 million medical papers.

From there, an algorithm revealed the biological mechanism of actions associated with the potential repositioning of drugs. Today's Yewno Discover platform is the result of this concept.

Michael Keller, a librarian at Stanford University, approached Ruggero in 2013 with the idea, thinking it might be used to create a new method for academics and students to uncover previously unknown ideas and make better use of the collection. Using Artificial Intelligence Ruggero and his colleagues developed a solution. After that, they understood that the concept could be used to the advantage of other institutions, universities, and businesses (Ruggero and Pickering, 2017).

How Yewno Discover works

Yewno reads full-text digital information from approximately 200,000,000 sources, including books, academic articles, and digital library repositories. Once the content is read, it is stored in the knowledge base graph. Users access the content and then type in the concept they would like to research. The user is given the chance to further define the results once the topic has been revealed; for example, if the concept was "key," they might indicate whether they are seeking a key. Once this is done, Yewno Discover displays the concept in the form of a knowledge graph with links to other related concepts (Image 2.). The researcher chooses which course to follow and how thoroughly he wants to go into certain concepts, similar concepts, or the connections between them. The system scans and saves entire text rather than simply keywords in this way, increasing the opportunities for connections.

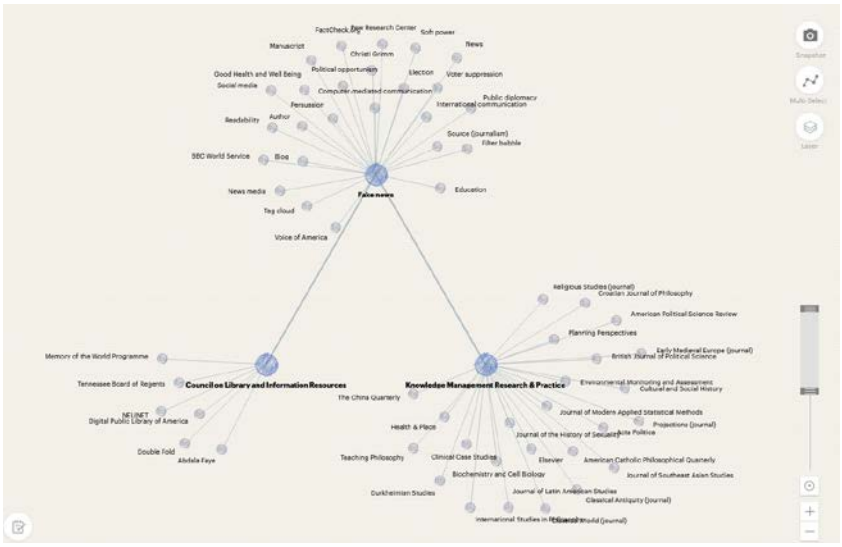


Image 2.

Yewno and Education

Yewno offers many benefits in the fields of education and research. The time savings from using journal articles, scientific publications and books for data collection is one of the main benefits.

It is also much easier to reach them. Almost 600 million semantic relationships between ideas are mapped by Yewno Discover using data taken from scholarly sources. Through these Yewno has a lot to offer in terms of research and education. The time savings from using books, journal articles, and scientific publications for data collection is one of the main benefits. It is also way easier to reach them. Almost 600 million semantic relationships between ideas are mapped by Yewno Discover using data taken from scholarly sources. More than 120 million scholarly journals publications, books and database resources may be accessed through these connections. Users have access to an increasing amount of materials every day.
It has received widespread recognition and is swiftly gaining ground in colleges all around the world that offer higher education.

Despite the complex technology it uses, Yewno's search for information is simple. When users have identified the concepts they are most interested in, they can link to the details where they instantly reset all relevant information. Medicine, Social Sciences, Business, Humanities, Engineering and Technology, Education, Life Sciences, Computer Science, Agriculture, Communications,

Natural Sciences, Law, Architecture, and Mathematical Sciences are some of the content categories on Yewno for Education.

3. Yewno and Libraries

Yewno offers a powerful addition to more conventional library methods of resource discovery. Users can read the entire text of publications directly since it incorporates the licensing of library resources. With a broader search and more trustworthy results, libraries are able to support the complete research process for students and researchers more successfully.

Yewno also features repository visualization on the Yewno Discover platform, providing a way for universities to visualize their own repositories. According to the article "Start-up story: Yewno: an AI-driven path to a knowledge-based future" by Ruggero Gramatica and Ruth Pickering, partner libraries have found that if they give library users access to Yewno Discover for a specific task , nine out of ten will reach a different conclusion than they would have without access to the platform (Ruggero and Pickering, 2017).

Yewno is already utilized by a number of libraries internationally, including: Stanford University (https://library.stanford.edu/search-services/yewno), University of Nottingham (https://www.nottingham.ac.uk/library/collections/yewno.aspx), National University of Singapore (https://libguides.nus.edu.sg/yewno) and University of Denver (https://libguides.du.edu/yewno). The reason for this, in accordance with Schreur (2019), is that libraries enable their catalog metadata to be connected with the richness of data on the open web by releasing them from a format that can only be understood by libraries (i.e., their catalog metadata in, for example, MARC format).

3.1 Yewno Discover at Piraeus University Library

The Piraeus University Library began using Yewno by combining its institutional repository DIONI (http://dione.lib.unipi.gr/xmlui/), research repository PANDORA (http://pandora.lib.unipi.gr), journal SPOUDAI (https://spoudai.unipi.gr/index.php/spoudai), and subscription and open access databases.

Search using Yewno Discover

When a user searches for a phrase after signing in, the initial results show at the bottom as a list (Image 3.).

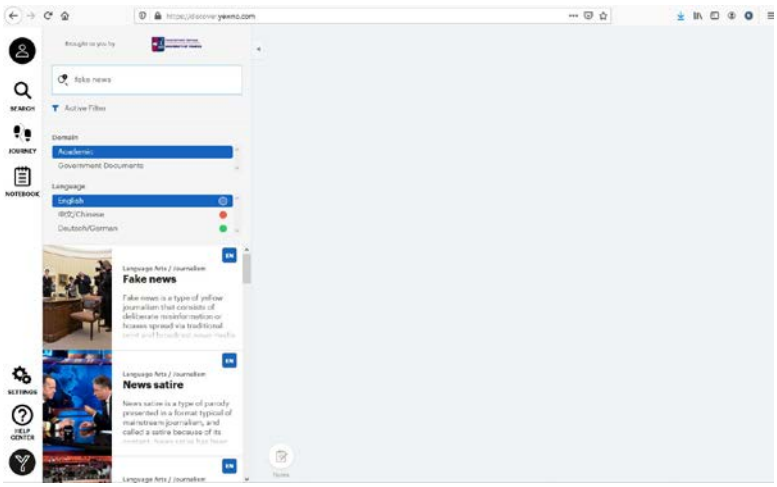


Image 3.

The knowledge map is displayed after clicking on a result (Image 4.).

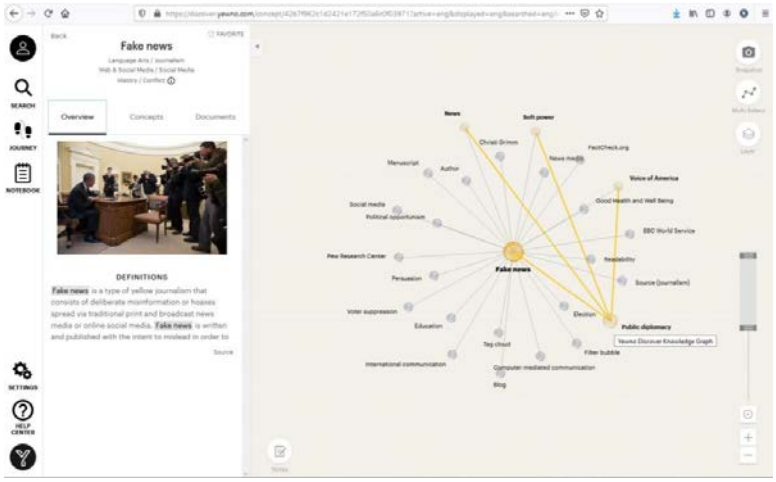


Image 4.

The results display offers the option to examine the results as a list (concepts) (Image 6), aggregated by subject and type of document (documents) (Image 7.) and with a brief description for each individual result (overview) (Image 5.).

Overview

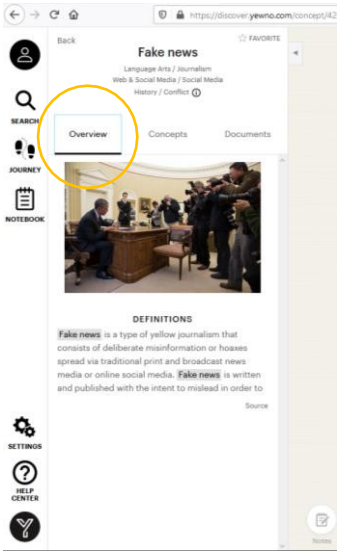


Image 5.

Concepts

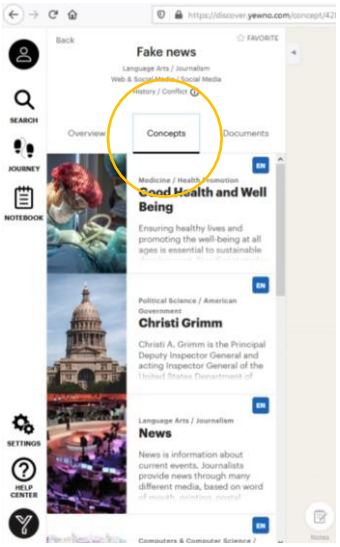


Image 6.

Documents

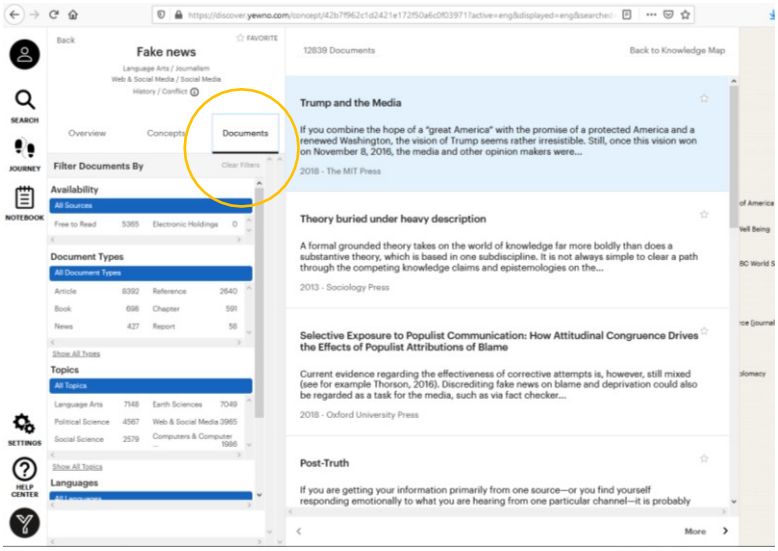
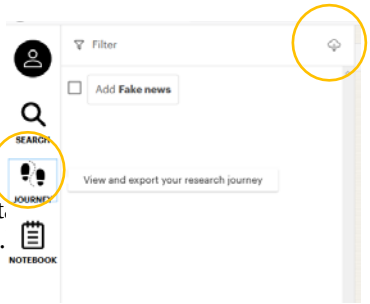
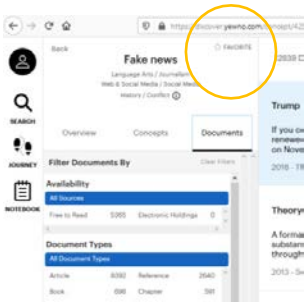


Image 7.



Additionally, it is possible to bookmark cert
history, and export search history (Image 9.



Add to favorite

Image 8.

History/ Export history

Image 9.

Each subsequent search that the user makes is conceptually linked to the previous one (Image 10. and 11.).

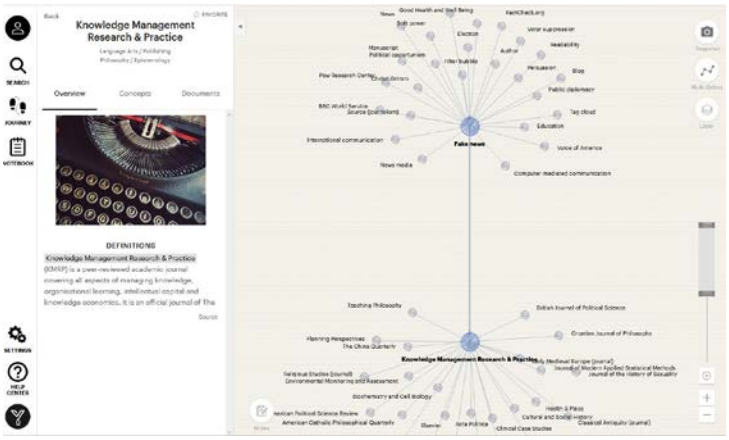


Image 10.

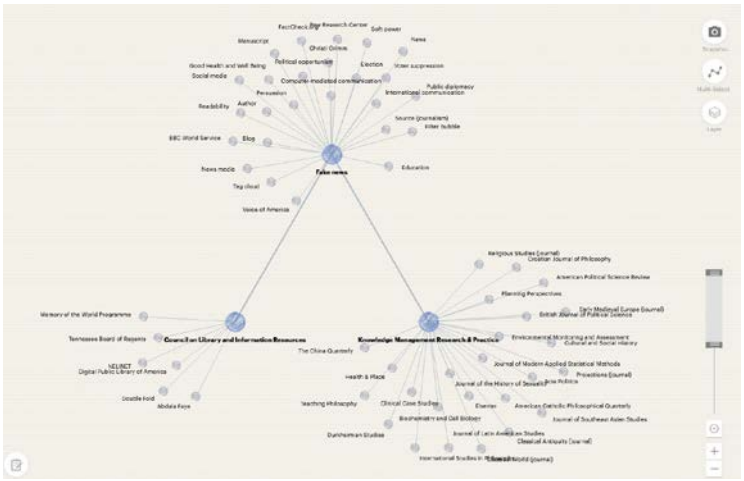


Image 11.

4. Conclusions

In conclusion, the usage of Semantic Web and Artificial Intelligence technology provides librarians with a great potential to influence how their users view libraries.

Users have the capacity to significantly enhance their research and learning settings and handle information more extensively and in-depth by taking advantage of the entirely new means of accessing knowledge that this new technology offers.

But as Bohyun Kim said when asked by George M. Eberhart and American Libraries Magazine (2018): «Are libraries ready for intelligent machines that directly interact with human users? There are certainly opportunities to improve existing services and programs. The more AI improves and the more things get delegated to machines, the more risks increase. How do we know the results are fair or unbiased? ».

Technology undoubtedly has its problems, but problems are a necessary element of evolution. However, in any case, the effectiveness of these tools is obvious, and that is why libraries adopted them, as long as we have control and vigilance.

References

Ali, M. Y., Naeem, S. B., & Bhatti, R. (2020). Artificial intelligence tools and perspectives of university librarians: An overview. *Business Information Review*,37(3), 116–124. https://doi.org/10.1177/0266382120952016

Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43. https://www.jstor.org/stable/26059207?seq=1

DiNucci, Darcy (1999). Fragmented Future. *Print.* 53(4), 221–222.

Douligeris, X., Mavropodi, R., Kopanaki, E., Karalis, A., (2017). Technologies and Programming on the World Wide Web. Athens: New Technologies Publications.*(in Greek)*

Eberhart, G. M. An AI Lab in a Library. *American Libraries Magazine*, 26 Ιούνιου 2018. https://americanlibrariesmagazine.org/blogs/the-scoop/ai-lab- library/
Georgouli, A., (2015). *Artificial Intelligence. [elec. bibl.] Athens: Association of Hellenic Academic Libraries. (in Greek)* http://hdl.handle.net/11419/3381

Martinez-Rodriguez, J. L., Lopez-Arevalo, I., & Rios-Alvarado, A. B. (2020). Mining information from sentences through Semantic Web data and Information Extraction tasks. *Journal of Information Science*, 1-8. https://doi.org/10.1177/0165551520934387

Park, H., & Kipp, M. (2019). Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*, 57(5), 261-277. https://www.tandfonline.com/doi/full/10.1080/01639374.2019.1641171

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basicsemantic capabilities. *Behavioral science*, 12(5), 410-430. https://doi.org/10.1002/bs.3830120511

Ruggero, G. and Pickering, R. (2017). Start-up Story: Yewno: An AI-driven path to a knowledge-based future. *Insights*, 30(2), 107-11. http://doi.org/10.1629/uksg.369

Schreur, P. (2019). Yewno: Transforming data into information, transforming information into knowledge, *IFLA WLIC 2019 - Athens, Greece - Libraries: dialogue for change*, Session 114 - Knowledge Management with Information Technology and Big Data. http://library.ifla.org/id/eprint/2538

The present paper presented at 26ο Panhellenic Conference Academic Libraries «The role of Libraries during special conditions», December 17-18, 2020, National Technical University of Athens Library & Information Center.