

When Artificial Intelligence meets Culture: The MESOC Serapeum

Dragan Čišić and Božidar Kovačić (University of Rijeka), Francesco Molinari (University of València)

Highlights

- The MESOC Serapeum is a collection of Artificial Intelligence tools offered to researchers and policy analysts interested in exploring the societal impacts of culture.
- The first working version took advantage of a repository of 573 free and open-source academic papers dealing with the topic and especially describing relevant case studies.
- The repository is being expanded to include grey (policy) literature texts, with a special focus on European cities.
- The areas of societal impact are those identified by the EU Agenda for Culture as “crossover themes”: health and well-being, urban renovation and regeneration, social cohesion and public participation.
- The contribution of cultural activities is defined according to the 10 domains introduced by the UNESCO Framework of Cultural Statistics.
- The URL is: <http://mesoc-serapeum.eu/>.

Background

Natural Language Processing (NLP) is the subfield of Artificial Intelligence (AI) that aims to process, analyse and ultimately understand natural language in written or spoken form.

Over the past few years, a major breakthrough in NLP has occurred with the introduction of the so-called transformers [Vaswani et al., 2017]. These are deep learning models – a class of machine learning algorithms inspired by the structure and function of the human brain – which have been proven to considerably outperform the more traditional NLP tools in a variety of key tasks, such as language translation, question answering, text summarization, semantic search and text classification.

Transformers use a novel mechanism, called attention, which enables them to capture long-range dependencies and contextual information at input level, making it possible to generate more coherent and meaningful outputs.

Several architectures developing the basic transformer model introduced in 2017 have been created and trained with textual data, which include, among others: BERT [Devlin et al., 2019], DistilBERT [Sanh et al., 2019], T5 [Raffel et al., 2019], GPT-2 [Radford et al., 2019], and the GPT-3 model [Brown et al., 2020] which is said to be superseded by GPT-4 during the first half of 2023.

Rationale

The MESOC project consortium has created a repository of 573 free and open-source academic papers dealing with the societal impact of culture and especially describing relevant case studies. The repository is now being expanded to include grey (policy) literature texts, with a special focus on European cities.

However, it is never easy nor practical for a researcher or a policy analyst to handle a huge and increasing number of literature texts. Just to make a single example, the number of sentences in the documents currently stored in the MESOC repository exceeds 300,000.

In these conditions, it is quite difficult to decide what is relevant to read and consider as a guideline or at least as an inspiration for future actions. Not to mention the possibility of not finding sufficient time to analyse the texts in question and highlight their key points or to establish connections with related documents having similar contents.

Solution

The MESOC Serapeum is a collection of Artificial Intelligence tools offered to researchers and policy analysts interested in exploring the societal impacts of culture.

The areas of societal impact are those identified by the EU Agenda for Culture as “crossover themes”: health and well-being, urban renovation and regeneration, social cohesion and public participation.

The contribution of cultural activities is defined according to the 10 domains introduced by the UNESCO Framework of Cultural Statistics.

The MESOC Serapeum is freely accessible from the following URL: <http://mesoc-serapeum.eu/>.

Available functionalities

The main features of the MESOC Serapeum so far developed include the following:

- Semantic search
- Document review
- Clustering and topic analysis
- Transition variable search and view
- Societal impact analysis
- Thesaurus and taxonomy

Semantic search

Semantic search is a technique that uses machine learning algorithms to understand the context, intent, and meaning behind a user's query, rather than relying solely on keyword matching. It could be defined as “search with meaning”, as it delivers more relevant and accurate results by analysing the relations between words and concepts and by understanding the user's intent based on search history and other contextual clues.

In the Serapeum, semantic search is implemented for all documents, abstracts and full texts.

Document review

The AI system reviews existing documents to assess the societal impact of culture. This activity includes document clustering, text summarizing, keyword analysis, and other analytical tools and methods.

There is also a document mapping function, presenting the geographical area an article is about.

Clustering and topic analysis

Clustering and topic analysis are two important techniques used in NLP for organizing and summarizing large collections of textual data.

Clustering is used to identify subgroups of similar documents, which can be helpful in information retrieval.

