NLP and AI techniques for matchmaking between user requests and scientific papers / courses

This project aims at the definition of a matchmaking algorithm capable of mapping user requests in textual form to a ranked list of relevant Artificial Intelligence (AI) assets. These assets are represented by scientific papers and AI-related material like courses and corpora. In particular, the devised matchmaking algorithm must be able to digest both structured and unstructured information types, such as template-based textual user requests, document meta-descriptors, higher-level annotation layers and plain textual documents. Additionally, the algorithm will require advanced language understanding capabilities to guarantee high-quality matching. Thus, a robust understanding of research fields like Natural Language Processing (NLP), Machine Learning (ML) and, more broadly, AI are required.

The project is part of the StairwAI European project, financed by the European Union (call H2020-ICT-2020-2). The general objective of the StairwAI project regards the enhancement of the existing AI4EU online platform via advanced and state-of-the-art AI methods. These improvements address the development of interfaces that allow users of all backgrounds to search and retrieve AI assets that match their requirements. The presented matchmaking project is part of the Work Package (WP) 5 of the StairwAI project, centered on the horizontal matchmaking between platform users and available AI assets.

Proposed matchmaking solutions for the purpose must be capable of handling textual data in several formulations. In particular, the abundance of plain textual documents, such as scientific papers, will represent one of the major challenges. To this purpose, efficient methods that scale up to large textual documents will be preferred. Parallelly, such advanced information retrieval and language understanding techniques must be juxtaposed to transparent methods that map user requests to retrieved information. Therefore, techniques that move in both directions, such as memory-augmented neural architectures, represent a viable solution.

Advanced annotation layers derived from NLP research areas like Argument Mining (AM), Sentiment Analysis (SA) and from external ontologies will be applied as well. These additional resources open up high-level reasoning processes over available data that can potentially ease parts of the matchmaking algorithm.