

On the Issue of Aggregation in Information Retrieval

par Gabriella Pasi, University of Milano-Bicocca

Abstract

This lecture will address the issue of defining flexible approaches to Information Retrieval by means of the usage of aggregation operators in various phases of the retrieval process. In particular, these approaches rely on the interpretation of Information Retrieval as a Multi-Criteria Decision Making (MCDM) problem, from various perspectives. The first, more straightforward perspective, is to interpret the overall IR process as a MCDM process aimed at selecting the best alternatives (documents) based on the assessment of the performance of multiple criteria (the keywords specified in a user's query). Another and strongly related perspective is to see the assessment of the overall relevance estimate of a document (still an alternative) to a query as the process of evaluating the performance of several relevance dimensions (e.g., topicality, novelty, recency, etc.), which in this case represent the criteria to be aggregated. Another process that may require the application of appropriate aggregation operators is the indexing process, when applied to structured documents. Metasearch constitutes another interesting task that can be seen as an instance of a Multi-Expert Decision Making (MEDM) problem, also strongly relying on the appropriate choice of an aggregation operator. By this task a user query is separately evaluated by different search engines, each one providing its own relevance assessment of the considered documents. Metasearch aims to merge the ranked lists generated by the various search engines (experts) in response to a query, to the aim of providing a unique, consensual ranked list of results. A quite interesting aspect implied by the above interpretations of various phases of the IR process is that the choice of different aggregation operators can produce different results. In other words, the semantics of aggregation implies an interpretation of the affected process. For example, if considering the aggregation of different relevance assessments for a same query and the same documents, distinct rankings can be obtained by applying distinct aggregation strategies. Despite the potential impact of aggregation on the whole IR process, this aspect has not received the proper attention in the literature. Only recently some approaches have appeared demonstrating the importance of this issue, and its potential impact on the searching process. This lecture aims to shortly review the main contributions that in the literature have made use of aggregation operators in Information Retrieval.

Bio

Gabriella Pasi est Professeur à l'Université de Milano-Bicocca. Elle est un chercheur européen parmi les plus reconnus dans le domaine de la Re-

cherche d'Information. A ce titre, elle est régulièrement sollicitée par les institutions d'évaluation de la recherche de plusieurs pays (Italie, France, Suisse, Croatie, Espagne) et par la Commission pour l'évaluation de projets européens (elle a été longtemps membre du Panel of Experts de Computer Science pour les Starting Grants et les Consolidators Grants du Programme Ideas du European Research Council). Elle a été invitée en tant que conférencier dans une quinzaine de conférences internationales. Gabriella Pasi a beaucoup œuvré dans le rapprochement du monde industriel et de l'université et a été à l'initiative de plusieurs travaux et projets communs dont des projets européens (en particulier, elle a aussi coordonné un projet Européen). Elle est l'auteur de 36 chapitres dans des ouvrages internationaux portant sur des problématiques variées liées essentiellement à la Recherche d'Information. Elle est également auteur de 48 publications dans des journaux internationaux, et globalement l'auteur de plus de 200 publications scientifiques.