

THE EFFECT OF SIMILARITY MEASURES ON GENETIC ALGORITHM-BASED INFORMATION RETRIEVAL

MOHEB RAMZY GIRGIS, ABDELMGEID AMIN ALY & FATIMA MOHY ELDIN AZZAM

Department of Computer Science, Faculty of Science, Minia University, ElMinia, Egypt

ABSTRACT

Genetic algorithms (GAs) can be used in information retrieval (IR) to optimize the query "solution". This paper proposes a GA-based IR algorithm that adjusts the weights of keywords of a query in order to generate an optimal or near optimal query vector. In this algorithm, each query is represented by a chromosome. These chromosomes are feed into genetic operator process: selection, crossover, and mutation to get new population, then, to get better solutions, a local search procedure is applied on each individual in the new population. This process is repeated until an optimized query chromosome for document retrieval is obtained. The evolution of the possible solutions is guided by fitness functions that are designed to measure the goodness of those solutions. We used order-based fitness function with different similarity measures to study their effect on the quality of the generated solutions and decide which similarity measure leads to the best solution.

KEYWORDS: Information, Retrieval, Genetic Algorithm, Query Optimization, Similarity Measures