Poster presented at the 4^{th} Annual IRIS/PRECARN Conference, Toronto, ON, June 21-23, 1994

FEATURES OF SIMILARITY

Igor Jurišica, Ph.D. Degree Student
Department of Computer Science
University of Toronto, Toronto, Ontario, M5S 1A4

Similarity plays a central role in theories of human problem solving and thus is important for research in artificial intelligence. Usually, an AI system assesses similarity by comparing symbolic knowledge representations. Intuitively, the goal is to compare those parts of knowledge so that the most similar cases will also be the most useful ones. Several approaches had been proposed, yet they are not flexible enough to handle complex domains.

In many information systems, much of the stored information is hard to retrieve since the need to do so was not foreseen at the time the information was created or a different use of the information was expected. One of the possible clues to help the retrieval process is to consider the kinds of activities in which the user was engaged in at the time of storage. Context is one such clue.

Often in these systems, it is unreasonable to expect to locate exact information so it is natural to look for some approximation of it, i.e., to look for something similar. In different situations (or for different individuals) different definitions of approximation may be valid. This may not be encoded into the system because one could not predict such a situation or one could not list all possibilities. Similarity judgment in such systems needs to be flexible. Context allows for this flexibility.

Even though similarity was studied by philosophers in the 18^{th} century, later by psychologists and then by computer scientists there still is no satisfiable theory of similarity. We present a novel theoretical foundation for flexible similarity assessment by defining context in similarity.

Based on this theoretical foundation we present the nature of transitivity, symmetry and monotonicity in similarity. Although these properties do not hold for similarity in general, they do hold under certain circumstances. Context plays a crucial role in determining these circumstances. We show that similarity is reflexive, symmetric and transitive and thus is and equivalency relation. Dissimilarity is not an equivalency relation because it is neither reflexive nor transitive.

Potential applications: similarity-based retrieval in software repositories, case-based reasoning systems, etc.