ABSTRACT

Information has a broad meaning. It can be interpreted as raw material that is directly acquired via observations. At a higher level, information is the results and opinions that are reached directly and/or indirectly through data processing and analysis. In fact, the level of analysis determines the information abstraction level. The semantic analysis can be defined as the methods by which the high-level information characteristics can be addressed in terms of the low-level, directly measurable information attributes. The semantic analysis is a suite of methods, there exists no unique semantic solution.

In this work, in order to represent semantic information, the information modeling requirements are considered.

Semantic information can be modeled on a manifold. Manifold is a generalized space than the metric space but through the local properties it provides specialization at the same time. The specialization comes from the position dependence by which curvature metric becomes available in addition to ordinary distance metric.

As part of the work, a manifold based semantic analysis method is proposed in order to support the idea of modeling semantic information on a manifold. The proposed approach is based on the concept of geodesic distance on a manifold. The aim is to utilize geodesic distances for making similarity measurements.

The proposed geodesic similarity measure is tested on the Wikipedia XML Corpus, which is a text-based linked data set. The clustering coefficient values that are computed using the link graph are substituted as curvature values by the assumption that clustering coefficients are rough estimates of curvatures. Then, the curvatures are combined with the Cosine text similarity measure in the formula of geodesic distance. The experimental work is conducted in the clustering context and clustering metrics are used in order to evaluate the experimental results.

Finally, the work's originality comes from the fact that it proposes the manifold structure for representing information, develops and tests a manifold-based semantic analysis method to justify the information manifold idea. Thus, it opens a new research area and motivates similar approaches for future research.

Keywords: Information model, semantic analysis, manifold, curvature, geodesic distance, similarity measure.

Short Biography

Selma Tekir is a researcher at Izmir Institute of Technology Department of Computer Engineering. Her research interests include data analysis, information/knowledge modeling, information warfare, intelligence analysis, and open source intelligence. In 2009, she worked as a visiting researcher at Faculty of Computer Science Chair for Databases, Data Analysis

and Visualization at University of Konstanz-Germany. Tekir received a PhD in computer engineering from Ege University, Turkey in 2010. Contact her at selmatekir@iyte.edu.tr.