Ontology Matching for Geospatial Domain

Ujwala Bharambe ,Surya Durbha
Indian Institute of Technology Bombay. India
ujwala.bharambe@iitb.ac.in, sdurbha@iitb.ac.in

Abstract. Geospatial information is collected from different sources with disparate terminologies resulting in highly heterogeneous information thus making its compilation and retrieval difficult. Harmonization of these heterogeneities is needed for interoperability and seamless access to the information sources by providing a common platform for facilitating information exchange. This can be achieved by ontology matching. This research proposes a conceptual framework for Geo-ontology matching to tackle the two major issues for geo-information fusion: heterogeneity and uncertainty by using information theoretic approaches for representing features of ontology. Further, the multi-objective optimization technique such as Pareto Ranking technique is adapted for the optimization of the best match pairs to derive the correspondence between two Geo-ontologies.
Keywords: Interoperability, Geo-ontology Matching, Information Theory, Pareto ranking

1 Introduction

Interoperability in geospatial domain has garnered lot of interest in recent years with a lot of challenges faced towards the integration of diverse geospatial information sources. Several syntactic and structural approaches for integration are currently available, such as the Open Geospatial Consortium's (OGC) standards for geospatial web services. However, these standards only facilitate interoperability at the syntactic level, which impedes the seamless access of geospatial data. Building a semantic layer on top of the syntactic representation enables effective integration of the semantics of different geospatial data thereby removing gaps in the information acquisition and understanding of the integrated data, it's sharing and use [2]. The above can be effectively achieved by using ontologies [1] where each source is represented by an ontology that explicitly represents the implicit concepts of that domain. The aim of this research is to develop a framework for geospatial ontology matching to achieve semantic interoperability thus providing an integrated view of the geo-information sources.

2 Ontology Matching in the Geospatial Domain

The main objective of ontology matching is to resolve the discrepancy among the ontologies which are described by different communities using different vocabularies and different perspectives/contexts. The difficulties in Geo-ontology matching may also be aggravated due to peculiarities of geospatial information (i.e. topology, dimension and orientation, shape, size and location) and due to lack of widely accepted Geo-ontology model causing the Geo-ontologies to be at different granularity levels.

The proposed system is composed of a four components as shown in figure 1. The first component (Geo-ontology access and analysis) takes two input ontologies and performs parsing of the ontologies using OWL API. The second component (i.e. Geo-ontology Feature Extraction) is responsible for capturing all lexical, structural and instance based features of the given ontologies.

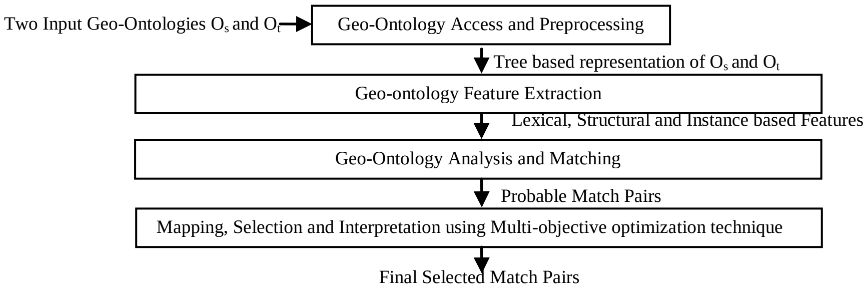

Fig. 1. Proposed System Architecture for Geo-ontology Matching

The third component (Geo-ontology Matching) comprises of the matching modules. The matching process is divided into three main modules: 1) Lexical Matcher: This is element based matcher (local). It uses different similarity metrics (string similarities: Jaro Wrinker and Wordnet Similarity) to handle the high terminological (lexical) heterogeneity of ontologies. 2) Structural matcher takes advantage of quantized features, which represent the structural characteristics of ontology. Ontology matching can be seen as an operation that takes two graph-like structures and produces a mapping between elements of the two graphs that correspond semantically to each other. Gaussian similarity measures (eq 1) are used for calculation of the structural similarity at the local el. $Sim_{\text{gauss}}(e1, e2) = e^{\left\{-\frac{d^2}{2\sigma^2}\right\}}$ (1)

Where d is Euclidean distance between relative entropy concepts [3] and σ is scaling factor. 3) Instance base Matcher: Cosine similarity measures are used for instance based matching. The final component is Best Match Pairs Selection and Interpretation where bi-objective optimization problem is solved using Pareto optimization technique wherein both the lexical and structural features are taken into equal consideration thereby achieving a most reasonable alignment and maximizing the number of correspondences in matching. The proposed framework of Geo-ontology matching is being implemented by using Java programming environment. The proposed framework for Geo-ontology matching would be evaluated in the context of rapid decision making for disaster management system especially flood management. The major focus is on developing a generalized adaptive multi-strategy Geo-ontology matching system.

References:

1. Kuhn, W.: Geospatial Semantics: Why, of What, and How? Journal on Data Semantics III. Springer Berlin Heidelberg. 3534: 1-24 (2005)
2. Lutz, M., E. Klien.: Ontology-based retrieval of Geographic Information. International Journal of Geographical Information Science 20(3): 233-260. (2006)