

Ontology Alignment Techniques

Marcos Martínez Romero

University of A Coruña, Spain

José Manuel Vázquez Naya

University of A Coruña, Spain

Javier Pereira Loureiro

University of A Coruña, Spain

Norberto Ezquerro

Georgia Institute of Technology, USA

INTRODUCTION

Sometimes the use of a single ontology is not sufficient to cover different vocabularies for the same domain, and it becomes necessary to use several ontologies in order to encompass the entire domain knowledge and its various representations. Disciplines where this occurs include medical science and biology, as well as many of its associated subfields such as genetics, epidemiology, etc. This may be due to a domain's complexity, expansiveness, and/or different perspectives of the same domain on the part of different groups of users. In such cases, it is essential to find relationships that may exist between the elements of a specific domain's different ontologies, a process known as *ontology alignment*.

There are several methods for identifying the relationships or correspondences between elements associated with different ontologies, and collectively these methods are called *ontology alignment techniques*. Many of these techniques stem from other fields of study (e.g., matching techniques in discrete mathematics) while others have been specifically designed for this purpose. The key to successfully aligning ontologies is based on the appropriate selection and implementation of a set of those ontology alignment techniques best suited for a particular alignment problem.

Ontology alignment is a complex, tedious, and time-consuming task, especially when working with ontologies of considerable size (containing, for instance, thousands of elements or more) and which have complex relationships between the elements (for example, a particular problem domain in medicine). Furthermore, the true potential of ontology alignment is realized when different information-exchange processes are integrated

automatically, thereby providing the framework for reaching a suitable level of efficient interoperability between heterogeneous systems. The importance of automatically aligning ontologies has therefore been a topic of major interest in recent years, and recently there has been a surge in a variety of software tools dedicated to aligning ontologies in either a fully or partially automated fashion. Some of these tools—generally referred to as *ontology alignment systems*—have been the result of well known and respected research centers, including Stanford University and Hewlett Packard Laboratories, for instance. In Shvaiko & Euzenat, 2007, updated information is given regarding the currently available ontology alignment systems.

Each ontology alignment system combines different alignment approaches along with its own techniques, such that correspondences between the different ontologies can be detected in the most complete, precise, and efficient manner. Since each system is based on its own approximation techniques, different systems yield different results, and therefore the quality of the results can vary among systems. Most of the alignment systems are oriented to solving problems of a general nature, since ontologies associated with a single domain share certain characteristics that set them apart from ontologies associated with other domains. Recently, some systems have emerged that are designed to align ontologies in a specific domain. An example is the SAMBO alignment system (Lambrix, 2006) in the biomedical domain. These and other domain-specific systems can produce excellent results (when used for the domains for which they were designed), but are generally not useful when applied to other domains.

This article presents a classification of the most commonly used, recently developed alignment techniques, supported by simple examples to illustrate the specific techniques underlying different systems. Future directions in ontology alignment are also examined.

BACKGROUND

The key to ontology alignment is to find those entities in one ontology that may correspond to other entities in another ontology. Basically, this can be viewed as finding a similarity measure between elements (or so-called entities) associated with different ontologies, and subsequently selecting the set of correspondences that produce the strongest measures of similarity. There are, however, different ways to compute similarity measures; there are various studies dedicated to the classification of these techniques (Rahm & Bernstein, 2001, Euzenat & Valtchev, 2004, Euzenat et al., 2004, Shvaiko & Euzenat, 2005).

Following these classification schemes (especially those undertaken by Euzenat and Valtchev (Euzenat & Valtchev, 2004) and based on Euzenat et al., 2004), the next section will introduce an abbreviated classification of those ontology alignment techniques that are most commonly utilized by current ontology alignment systems. This condensed classification is centered on the type of element being manipulated by the alignment technique, and complements the taxonomy proposed by Rahm and Bernstein (Rahm & Bernstein, 2001), and—for the purpose of clarity and brevity—summarizes only those alignment techniques that compare on an individual basis a single element in one ontology with another element associated with another ontology (known as local alignment techniques, as in Euzenat et al., 2004).

ONTOLOGY ALIGNMENT TECHNIQUES

Ontology alignment techniques can be classified according to the following (please refer to Figure 1):

1. **Terminological techniques.** These calculate the similarity between text strings and describe several elements in the ontologies (names, labels, and/or comments). There are two types of terminological

techniques: those based on text strings and those based on the language.

1.1. Terminological techniques based on text strings.

These are based on the idea of comparing the structure in text strings, which are viewed as sequences of characters. These techniques consider that the similarity between two terms increases when the similarity between their corresponding text strings also increases, but without considering the underlying semantics in the terms. In this manner, the application of a technique of this type to the terms *Apple* and *Apples* would yield a relatively high measure of similarity, whereas the application of the same technique to the terms *Apple* and *Orange* would yield a lower degree of similarity (or a lower similarity measure), since in the second case the text strings are quite different. The isolated use of these techniques is usually not recommended, since it is preferable to use them in conjunction to other, more powerful alignment techniques; these can be easily illustrated with the following example: it would be erroneous to conclude that the terms *Cream* and *Scream* are highly similar (although their meanings are very different), or that the terms *Student* and *Pupil* are very distinct or dissimilar (although the semantic concepts are generally the same). Some examples of terminological techniques based on text strings are the distance measure proposed by *Hamming* (Hamming, 1950), which counts the number of different characters in two different text strings; the distance measure suggested by *Levenshtein* (Levenshtein, 1966), which examines the minimum number of operations (insertions, deletions and/or substitutions) that are necessary to transform one text string into another; and the distance measure *Jaro* (Jaro, 1989), which analyzes the number and order of two common characters in two text strings.

1.2. Terminological techniques based on language.

These techniques are more complex but more reliable than those previously discussed, and do not treat terms as simple sequences of characters that are independent of one another. Rather, these techniques view terms as groups of elements with meaning (lexima and morphema, i.e., prefixes and suffixes). The main objective of these techniques is to discover the similarity that may exist between terms associated with one concept, although the relationships can be formed by strings of characters that are very different. In other words, these techniques attempt to obviate the different termino-

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/ontology-alignment-techniques/10406

Related Content

Hybrid Energy Storage Systems: AI for Capacity Planning and Charge Management

Sashi Tarun (2026). *AI-Driven Solutions for Solar Energy Efficiency, Irradiance Modeling, and PV Forecasting* (pp. 253-282).

www.irma-international.org/chapter/hybrid-energy-storage-systems/387929

Online Surveillance of IoT Agents in Smart Cities Using Deep Reinforcement Learning

Ahmad Alenezi (2024). *International Journal of Intelligent Information Technologies* (pp. 1-15).

www.irma-international.org/article/online-surveillance-of-iot-agents-in-smart-cities-using-deep-reinforcement-learning/349942

The Impact of AI on Higher Education Trends and Educational Horizons

Ting Li, Sanitah Binti Mohd Yusof, Hassan Abuhassnaand Qi Pan (2026). *AI in Education, Governance, and Leadership: Adoption, Impact, and Ethics* (pp. 1-32).

www.irma-international.org/chapter/the-impact-of-ai-on-higher-education-trends-and-educational-horizons/384251

A Systematic Review: Role of Artificial Intelligence During the COVID-19 Pandemic in the Healthcare System

Shazia Iqbal, Shahzad Ahmad, Bushra Bano, Khalid Akkour, Manal Abdullah Ali Alghamdiand Alreuf Mohammed Alothri (2021). *International Journal of Intelligent Information Technologies* (pp. 1-18).

www.irma-international.org/article/a-systematic-review/272005

Voyage Virtuoso: Artificial Intelligence in Transforming Tourism

Geetha Manoharan, C. V. Guru Rao, Sunitha Purushottam Ashtikar, Sanjeev Kumarand M. Nivedha (2024). *Utilizing Smart Technology and AI in Hybrid Tourism and Hospitality* (pp. 79-97).

www.irma-international.org/chapter/voyage-virtuoso/341537