Knowledge-Based Systems 55 (2014) 101–113

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Towards the estimation of feature-based semantic similarity using multiple ontologies

Albert Solé-Ribalta *, David Sánchez, Montserrat Batet, Francesc Serratosa

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Avda. Països Catalans 26, 43007 Tarragona, Spain

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 11 October 2013
Accepted 12 October 2013
Available online 20 October 2013

Keywords:
Feature-based semantic similarity
Multiple ontologies
WordNet
MeSH
Ontologies

ABSTRACT

A key application of ontologies is the estimation of the semantic similarity between terms. By means of this assessment, the comprehension and management of textual resources can be improved. However, most ontology-based similarity measures only support a single input ontology. If any of the compared terms do not belong to that ontology, their similarity cannot be assessed. To solve this problem, multiple ontologies can be considered. Even though there are methods that enable the multi-ontology similarity assessment by means of integrating concepts from different ontologies, most of them are based on simple terminological and/or partial matchings. This hampers similarity measures that exploit a broad set of taxonomic evidences of similarity, like feature-based ones. In this paper, we tackle this problem by proposing a method to identify *all* the suitable matchings between concepts of different ontologies that intervene in the similarity assessment. In addition to the obvious terminological matching, we exploit the ontological structure and the notion of concept subsumption to discover non-trivial equivalences between heterogeneous ontologies. Our final goal is to enable the accurate application of feature-based similarity measures in a multi-ontology setting. Our proposal is evaluated with regard human judgements of similarity for several benchmarks and ontologies. Results shows an improvement against related works, with similarity accuracies that even rival those obtained in an ideal mono-ontology setting.

1. Introduction

General-purpose ontologies such as WordNet [12], or domain-specific ones like MeSH [23], provide a formal and machine readable representation of knowledge that can be used in a variety of tasks in which a semantic interpretation of text is required, such as information extraction [37] and retrieval [25], semantic data mining [2] or privacy-preserving methods [20,36]. The main use of ontologies, on which most of the above tasks rely, is the computation of the *semantic similarity* between textual terms. Semantic similarity measures quantify the degree of taxonomic resemblance between a pair of terms (e.g. *flu* and *bronchitis* are similar because both are *disorders of the respiratory system*), by analysing taxonomic relationships modelled in an ontology.

Throughout the years, many ontology-based similarity measures have been developed. These can be classified into different families according to their theoretical principles. *Edge-counting measures* quantify similarity according to the *length of the shortest path* defined by the taxonomic relationships that separate two concepts in an ontology [29,43], as exemplified in Fig. 1. Since these measures do not evaluate paths other than the shortest one, their accuracy is limited [35]. To benefit from additional knowledge that ontologies provide, *feature-based measures* quantify similarity according to the amount of common and/or non-common taxonomic subsumers of the compared concepts [19,35], as shown in Fig. 2. Thanks to the evaluation of additional taxonomic knowledge than edge-counting measures, feature-based measures achieve a higher accuracy [35].

However, these measures are limited by the coverage and detail of the input ontology [38]. If any of the compared terms cannot be found in the input ontology, the similarity cannot be assessed. As acknowledged by several authors [1,4,26,38], this limitation can be overcome by exploiting *multiple ontologies*. In this scenario, each term belongs to a different ontology. Thus, the basic idea is to discover *common taxonomic subsumers between ontologies* to later apply similarity measures like in a mono-ontology setting [31]. Most works framed in the multi-ontology scenario rely on terminological matching to discover common taxonomic subsumers [1,4,26,31]. However, since ontologies rarely model concepts in the same way, or refer to them by using the same label (due to synonymy), strategies based on terminological matching omit part of the equivalent concepts, a circumstance that tends to cause an underestimation of similarity. Moreover, most authors solely focus on discovering a *unique* common subsumer (i.e. the least common one) [1,38], so that the shortest path between concepts can be computed to apply

---

* Corresponding author. Tel.: +34 977 558676.
E-mail address: albert.sole@urv.cat (A. Solé-Ribalta).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.10.015

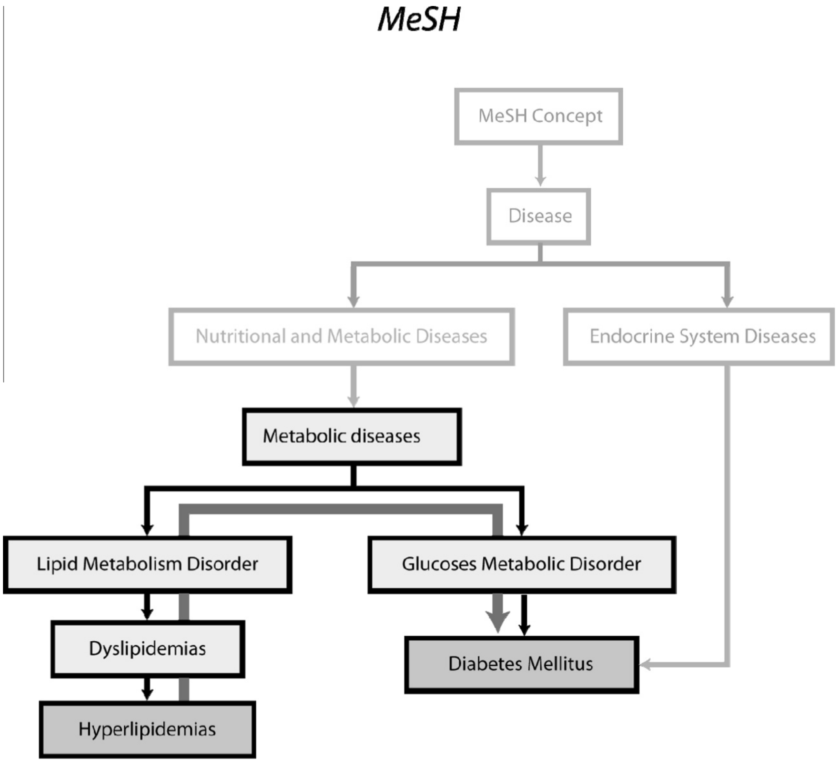

Fig. 1. Similarity assessment for *Hyperlipidemias* and *Diabetes Mellitus* in MeSH ontology with an *edge-counting* similarity measure. The thick arrow shows the taxonomic path evaluated in the similarity assessment.

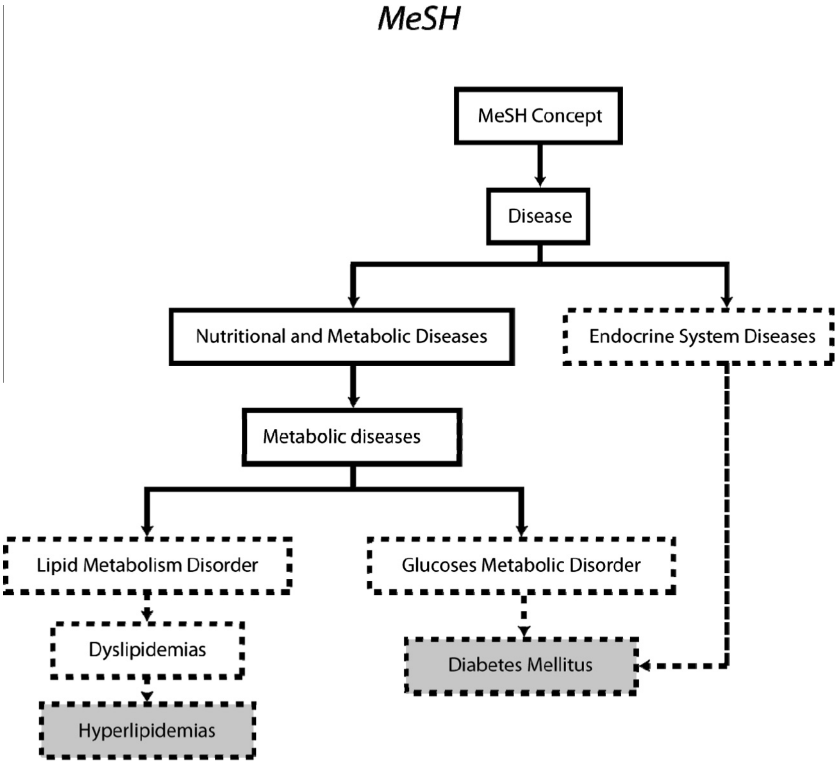

Fig. 2. Similarity assessment for *Hyperlipidemias* and *Diabetes Mellitus* in MeSH ontology with a *feature-based* similarity measure. Concepts within dashed squares correspond to taxonomic subsumers that the compared concepts do not have in common, whereas concepts within solid squares represent common subsumers.

edge-counting measures. This fact hampers feature-based measures because they rely on additional taxonomic evidences whose correspondences across the different ontologies will remain unknown.

To tackle these problems, in this paper we propose a method for enabling similarity assessments from multiple ontologies, which offers the following contributions: