Information Sciences 192 (2012) 152–173

Contents lists available at ScienceDirect

Information Sciences

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

Discrete particle swarm optimisation for ontology alignment

Jürgen Bock [a,*], Jan Hettenhausen [b]

[a] FZI Research Center for Information Technology, Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Germany
[b] Institute for Integrated and Intelligent Systems, Griffith University, Nathan Campus, Brisbane, Australia

ARTICLE INFO

Article history:
Available online 15 August 2010

Keywords:
Discrete particle swarm optimisation
Ontology alignment
Ontology mapping
Ontology matching
Semantic integration
Semantic Web

ABSTRACT

Particle swarm optimisation (PSO) is a biologically-inspired, population-based optimisation technique that has been successfully applied to various problems in science and engineering. In the context of semantic technologies, optimisation problems also occur but have rarely been considered as such. This work addresses the problem of ontology alignment, which is the identification of overlaps in heterogeneous knowledge bases backing semantic applications. To this end, the ontology alignment problem is revisited as an optimisation problem. A discrete particle swarm optimisation algorithm is designed in order to solve this optimisation problem and compute an alignment of two ontologies. A number of characteristics of traditional PSO algorithms are partially relaxed in this article, such as fixed dimensionality of particles. A complex fitness function based on similarity measures of ontological entities, as well as a tailored particle update procedure are presented. This approach brings several benefits for solving the ontology alignment problem, such as inherent parallelisation, anytime behaviour, and flexibility according to the characteristics of particular ontologies. The presented algorithm has been implemented under the name MapPSO (ontology mapping using particle swarm optimisation). Experiments demonstrate that applying PSO in the context of ontology alignment is a feasible approach.

1. Introduction

Particle swarm optimisation (PSO) [17,27] is a biologically-inspired optimisation meta-heuristic which has continuously gained momentum in recent years. It is generally applicable for problems, where the global optimum of an objective function is to be found in a multi-dimensional search space. Although originally developed for continuous optimisation problems, a number of modifications have been proposed to make PSO also applicable to discrete optimisation problems [17,2,3]. Convergence and run-time behaviour of PSO cannot be trivially determined, however, increasingly numerous case studies and applications show that it usually performs better than non-heuristic optimisers.[1]

The emergence of intelligent information systems bears the need for information and knowledge to be represented in machine readable form. Ontologies in the context of computer science have been formally described by Gruber [11] in 1993 as "an explicit specification of a conceptualisation"—a description which has frequently been refined and reinterpreted since. In particular the attribute "formality" has been added to the definition, since formality is a key feature for ontologies in order to be machine processable and hence usable in automated systems. Making real-world concepts explicit in a machine readable form now allows for the modelling of knowledge bases in order to make *knowledge* processable by intelligent systems. Informally, ontologies formalise concepts, individuals, and relations among them, in order to describe real-world entities in a certain domain of interest.

* Corresponding author.
E-mail addresses: bock@fzi.de (J. Bock), j.hettenhausen@griffith.edu.au (J. Hettenhausen).

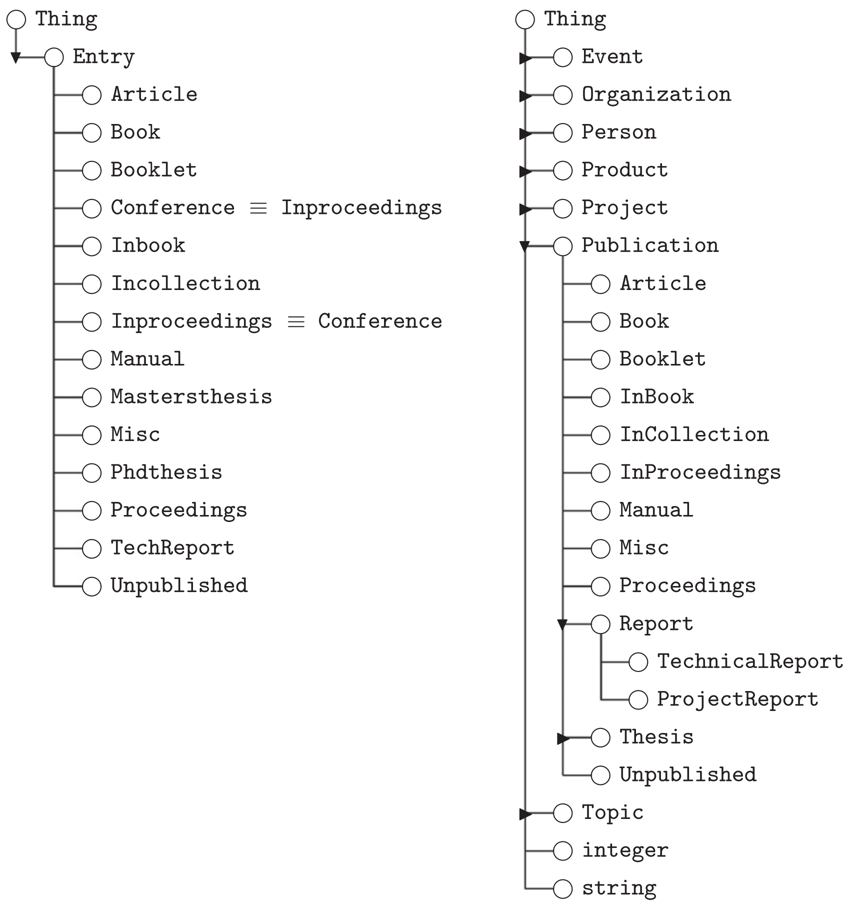
[1] For some variations of PSO such analyses were performed. They are, however, not trivially transferable to the proposed DPSO algorithm.

0020-0255/$ - see front matter
doi:10.1016/j.ins.2010.08.013

From a number of approaches to standardise a language for ontologies, the Web Ontology Language (OWL) [13] became widely accepted and is standardised by the W3C. Large extents of OWL are based on Description Logics [1]. This logical underpinning enriches OWL ontologies with a formal semantics allowing them for being used in so-called *semantic applications* or the Semantic Web.

One of the key benefits of using ontologies in semantic applications apart from the ability to infer implicit information, is the automated processing of knowledge that is formally described in ontologies. This allows for, e.g., integrating knowledge to be used in semantic application or the Semantic Web from different sources, i.e., ontologies. This problem is tackled in the discipline of *ontology alignment*. It is based on the observation, that ontologies to be integrated are heterogeneous models, often representing a similar or equal domain of interest, and hence have a certain overlap. As an example consider the two ontologies about bibliography presented in Fig. 1. The ontologies originate from different providers (MIT and University of Karlsruhe) and an intelligent information system might want to refer to the literature which is annotated by either of the two ontologies in an integrated application. As one can easily see, there is a significant overlap in these two ontologies, which needs to be identified by sophisticated ontology alignment systems. Other examples for the need of ontology alignment can be found in the context of information sharing among peers in distributed environments, such as peer-to-peer systems or grid environments [25,15,30]. A use case for ontology alignment would also become apparent in a medical information system which needs to incorporate knowledge from a disease ontology, as well as from an ontology about human anatomy. In order to retrieve information about which body parts are affected by a certain disease, both ontologies need to be consulted. Since the disease ontology contains information about anatomy in much less detail, a small overlap between the two ontologies exists. Again an ontology alignment system is required to identify this overlap and provide an alignment, which can be used by the medical information system.

An alignment is defined as a set of correspondences between ontological entities, i.e., classes, properties, and individuals, from two ontologies. It is agreed that finding a unique best alignment of two ontologies is an inherently difficult task, which



(a) MIT BibTEX ontology (b) Karlsruhe BibTEX ontology

Fig. 1. Two example ontologies about the domain of bibliography. The figures show the class hierarchy of each ontology, i.e., indented classes are subclasses of their parent, which denotes an *is-a* relationship between sub- and superclasses. Note that there is only a partial overlap between the two ontologies, since ontology 1b covers a wider domain.