Computers in Industry 65 (2014) 1264–1275

Contents lists available at ScienceDirect

Computers in Industry

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

A fast and economic ontology engineering approach towards improving capability matching: Application to an online engineering collaborative platform

Xiao Ma *, Jay Bal, Ahmad Issa

University of Warwick, WMG, International Institute for Product and Service Innovation (IIPSI), Coventry CV4 7AL, United Kingdom

ARTICLE INFO

Article history:
Received 2 July 2013
Received in revised form 17 January 2014
Accepted 27 May 2014
Available online 23 June 2014

Keywords:
Information systems
Collaboration platform
Knowledge management
Ontology engineering
Semantic interoperability
Social network analysis

ABSTRACT

Fulfilling needs through internal and external resources is a key business requirement. To better enable this, description of both needs and resources, using a common domain language is required. Using techniques from social network analysis (SNA) this paper describes a SENSUS-based methodology which generates domain ontology that can provide the breadth and depth of coverage required for automated need and resource matching systems. The mechanism described also enriches the semantic relationships in the generated ontology to form a network structure. This enables concept investigation to be undertaken from multiple perspectives, with fuzzy matching and enhanced reasoning through directional weight-specified relationships. The methodology was used to derive an ontology for engineering and tested against a traditionally derived and structured ontology. The methodology has the flexibility and utility to be of benefit in a wide range need and resource matching business applications.

1. Introduction

In this paper we describe a new and novel way to automatically generate ontology that can be used in information systems to reason and structure information. The application of the methodology is illustrated in an engineering sector case study, The West Midlands Collaborative Commerce Marketplace. This online portal helps match tender opportunities with companies that have the right capability, and can help form supply chains or consortia with all the capabilities required to enable collaboration to exploit an otherwise very difficult to address opportunity. WMCCM is a representative of a generalisable collaborative platform [1] or virtual organisation [2]. Many other "matching" type platforms exist in many sectors, ranging from personal "dating" to business "sourcing".

An ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. In information science it is used to reason about the entities within that domain, and may be used to describe the domain. The increasing need for information exchange within and between market sectors has driven the interest in ontology generation [3,4]. Ontology is increasingly used in knowledge management systems, medical and bio-informatics and play a key role in the semantic web and grid computing. Engineering was among the earliest sectors to benefit from ontology, and ontology in this sector is considered to be more mature than in others.

Engineering ontology is structured and populated to fit their special needs. Thus the way they are intended to be used determines how they are formed. Application orientation is also emphasised in the Developing Ontology-Grounded Methods and Applications (DOGMA) approach [5], where the ontology structure is designed as "double articulation" – a domain specific articulation and an application specific articulation. The practical requirement of the ontology application environment also drives the engineering ontology discussed to stretch the traditional ontology boundaries in terms of representation and weight specification.

Several ontology have been built by various organisations in the engineering sector, often in the form of industrial classifications to allow information exchange among organisations, such as United Nations Standard Products and Services Code[1] (UNSPSC) and UK

* Corresponding author. Tel.: +44 2476 524718.
E-mail addresses: x.ma@warwick.ac.uk (X. Ma), jay.bal@warwick.ac.uk (J. Bal), a.issa@warwick.ac.uk (A. Issa).

[1] http://www.unspsc.org/

http://dx.doi.org/10.1016/j.compind.2014.05.004
0166-3615/

Fig. 1. SENSUS approach to developing ontology.

Standard Industrial Classification (SIC) [6]. However, these classifications showed a lack of broad cover of the classes, especially with regard to the actual products and services, and insufficient relationships to demonstrate inheritance and commonality among classes [7]. In addition, the condensed classes produced by experts did not have enough attributive descriptions around concepts. In other words, there were a small number of words to cover a much larger generic keywords variation in natural language information. Finally, classes (concepts) proposed by such sources tended to stay at a higher level compared with the company/user proposed classes. The high level classes were found not to be specific or detailed enough to differentiate between the competences proposed by companies. These issues suggest that directly summarising ontology from existing sources (a single top-down procedure) may not satisfy the practical requirement of the collaboration platform for broad coverage and rich internal relationship.

The paper first investigates current methods for ontology generation and then goes on from an analysis of their shortfall for industrial applications, to describe a new methodology which addresses some of these key issues.

2. Related work

A review of ontology engineering methodologies, including Cyc Base [8], TOVE [9], On-To-Knowledge [10], METHONTOLOGY [11] and SENSUS [12], was conducted to assess their applicability to the notion of "economic, quick and reliable" ontology generation (Appendix B). Namely in this research, these criteria refer to a requirement for little or no reliance on domain experts, (fast) speed of corpus building and corpus structure analysis, and applicability to multiple (or cross) domains. The various ontology engineering methodologies were also evaluated on their coverage of the domain and the richness of the internal relationships.

Cyc methodology was applied to build the Cyc Knowledge Base, which is one of the top level ontology that SENSUS refers to. It was constituted by manually adding over a million pieces of consensus knowledge statements. Domain experts were the starting point for building the knowledge base. Most of the knowledge in the system would be based on the opinions of a group of experts. However this may not be sufficient to cover wider perspectives in the field and the common vocabulary of non-professionals. Domain experts were also needed in all of the later stages, resulting in a costly way of building such ontology.

TOVE's approach proposed a methodology in a linear process with detailed techniques at each stage. However, the technique details limited the methodology into wider application environment. For instance, using "first order logic" to specify the terms and relationships led to its inapplicability for developing ontology, which requires other types of binary relationship, i.e. semantic relationship. Although this relationship could be altered, it was bounded to TOVE's development environment, and any alterations might require much greater consideration so as to modify the remaining part of the methodology, for use in other projects.

On-To-Knowledge and KACTUS improved the linear process by suggesting a development cycle in order to enable knowledge reuse and continuous improvement (even for application in different domains). Researchers [11,13–15] have integrated formalised methodologies with ontology reuse methods, such as METHONTOLOGY. Despite a relatively comprehensive methodology with detailed techniques in ontology engineering, METHONTOLOGY did not appear to have the flexibility to rapidly respond to changes within the domain due to its manual corpus construction processes.

The proposed methodology for building the ontology is based on the principle that the ontology building should be initialised by linking specified keywords to the target source. SENSUS [12] constructs ontology for a domain from the foundation of a large knowledge base, or ideally, a previous large ontology. However, it does not engage in a traditional reusing or re-engineering process. It identifies key domain specific terms, a.k.a. seeding words, and then links them to the large ontology. Afterwards, the terms irrelevant to the new ontology can be pruned from the large source ontology. The processes undertaken in the SENSUS approach are shown in Fig. 1.

This approach contains unique characteristics that provide advantages over the other methodologies:

- It is an obvious improvement that SENSUS does not require constant input from domain experts: it only needs the initial seeding terms and their relationships to the knowledge base.
- SENSUS combines corpus construction and ontological analysis in one process, unlike others [8,11]. SENSUS thus ensures the terms collected are semantically connected to the seeding terms.
- SENSUS can act like a shared foundation to allow other ontology to be connected together and to share their terminology and relationships [12].
- Extracting related terms from the same sources through different seeding words is similar to seeing the same knowledge from different perspectives. This in theory could result in fuzziness around any given concept depending on the number of perspectives chosen. Thus the SENSUS ontology construction method may be capable of building cross-domain ontology.

Despite these benefits, it is difficult to apply SENSUS directly for our need: resource matching requirement, as there is insufficient detail on the techniques suggested to apply it. In addition, SENSUS did not propose any post-development stage, a development life cycle or project management mechanism which would help in industrial applications. Therefore, this research used the SENSUS approach as a foundation approach and developed techniques to formulate a new methodology that met the needs for quick, economical, reliable, and multi-domain ontology construction.

3. Research methodology

The SENSUS methodology recommended that the ontology building should be initialised by linking specified keywords to the target source.

3.1. Data source selection

Word clustering is a technique for partitioning the words that describe a domain into subsets of semantically similar words and is important in a number of Natural Language Processing tasks. The sets of words that describe the domain will be called 'keywords sets' hereafter. There are basically two main data sources (corpus) that could be used to generate these keywords: