Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

A semantic approach for web reasoning

Fei Liu *

Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, VIC 3086, Australia

A R T I C L E I N F O

Article history:
Received 2 February 2012
Received in revised form 21 February 2014
Accepted 17 April 2014
Available online xxxx

Keywords:
Web reasoning
Set calculation
PT-resolution

A B S T R A C T

The paper explores web reasoning based on Resolution with Partial Intersection and Truncation (PT-resolution). Instead of the traditional reasoning mechanism which is based on back-tracking and pattern matching, PT-resolution reasons based on set calculations. It prevents a derivation on a finite logic program from infinite looping and therefore, is ideal for web reasoning.

1. Introduction

The rapid development of the World Wide Web has significantly changed our expectations of it and its search engine. We are no longer satisfied with merely searching for a telephone number or looking for a road map on the web, we expect information presented on the web to be machine-processable and we expect its search engine to be intelligent (i.e. not only can the search engine extract information, but it can also derive information) [8,13,15,25,26,29]. This requires an effective mechanism on web reasoning – reasoning based on information that is extracted from the web.

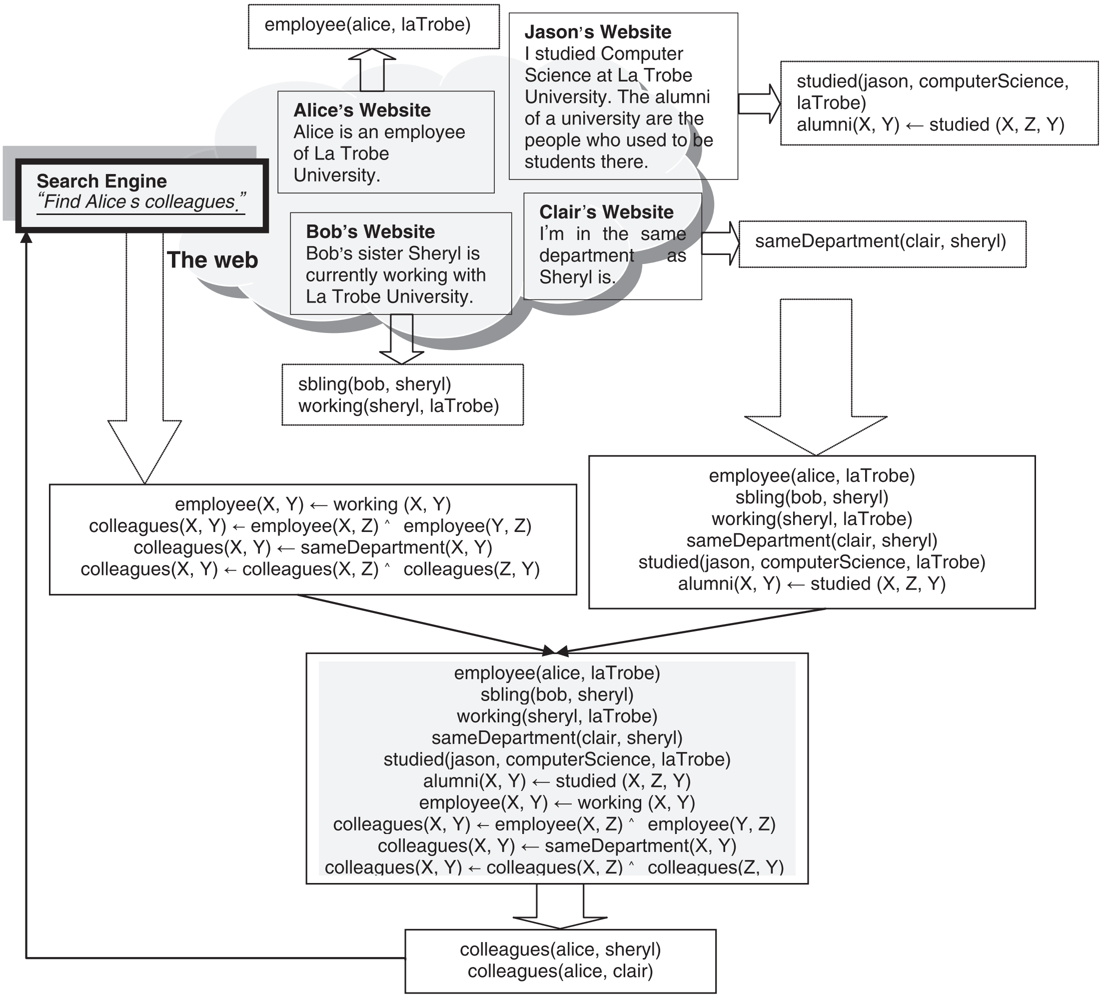
Although rule-based reasoning (logic programing) has been a major research field in artificial intelligence for more than a half of a century, reasoning in the web environment (web reasoning) is still in its infancy. Theoretically, all reasoning should be based on the same principles, regardless of the source of information; technically, however, there are significant differences between the two types. Web reasoning refers to the situation where reasoning is conducted, based on information extracted from more than one source. Precisely, the logic program is formed using clauses from different locations [5,22,27,28]. Normally, a logic program is formed in response to a given goal (i.e. a query being entered into the search engine). Research in this area concentrates not only on the mechanisms for deriving information, but also on the principles and practices of forming the logic program itself. Apparently, web reasoning requires substantial web searching, information extraction and ontology mapping; hence the manner in which information is processed will significantly influence the performance and outcomes of the derivation. The scenario below illustrates this.

Fig. 1.1 illustrates the architecture of web reasoning in response to a user's query. Web information is extracted; Clauses (facts and rules) are created based on extracted information; and hence web reasoning can be conducted. Outcomes of reasoning are returned to the search engine to answer the query.

Various resolution strategies have been developed over the last five decades, and these strategies can be approximately classified into two categories – syntactic approaches and semantic approaches [7,15,28]. The first category [1,3,24] is based on the syntax of First Order Logic (Proof Theory). They start the derivation from the top – the goal, and backtrack down to the

* Tel.: +61 3 94791949.
E-mail address: f.liu@latrobe.edu.au

http://dx.doi.org/10.1016/j.ins.2014.04.035
0020-0255/© 2014 Elsevier Inc. All rights reserved.



Q6 **Fig. 1.1.** Formation of a logic program for web reasoning.

bottom – the clauses of the logic program. The derivation is mainly based on pattern matching and backtracking. As a result, syntactic approaches are also called top-down approaches. The second category [4,11,17,24] is based on the semantics of First Order Logic (Model Theory). They define the logic program as a mapping from the Herbrand base to the Herbrand base. They repeatedly apply the mapping till a Herbrand model is found. Semantic approaches are also named bottom-up approaches.

Although syntactic approaches have always been the favourite in logic programming, we argue that semantic approaches are probably more appropriate for web reasoning. Reasons are:

(1) The logic program which is formed based on information being extracted from the web in responding to a query is not necessarily a complete definition of a problem [18]. A logic program is a complete definition of a problem if all its predicate instantiations have a truth value "true" or "false". If a logic program is not a complete definition of the problem, then the "closed world assumption" [1] does not hold. As a result, "Negation as Failure" cannot be applied to handle "negation" during its derivation.

Below is a simple example that demonstrates the point.

Example 1.1. Assume that following information is extracted from the web:

$$\text{like}\,(a,\text{iphone}) \leftarrow$$
$$\text{like}\,(b,\text{galaxy}) \leftarrow$$
$$\text{like}\,(c,\text{iphone}) \leftarrow$$
$$\text{inStock}\,(\text{galaxy}) \leftarrow$$
$$\text{onlinePurchase}\,(X,Y) \leftarrow \text{like}\,(X,Y) \wedge \neg\text{inStock}\,(Y)$$