Chapter I

Modeling and Querying Web Data: A Constraint-Based Logic Approach*

Evimaria Terzi
Purdue University, USA

Mohand-Saïd Hacid
LISI-UFR d'Informatique, France

Athena Vakali
Aristotle University of Thessaloniki, Greece

ABSTRACT

The efficient and sophisticated representation of the structure of the documents being circulated over the Internet allows for effective querying and reasoning over them. This is a major goal for large information resources like the World Wide Web (WWW). Constraints are a valuable tool for managing information. In this work, we consider how constraint-based technology can be used to query and reason about semistructured data represented using the constraint-logic implied representation models. The constraint system, $F\leq$, provides information-ordering constraints.

*This work is supported by the Platon 2002-2003 program for bilateral joint research work between France and Greece.

This chapter appears in the book, Information Modeling for Internet Applications, edited by Patrick van Bommel. Copyright © 2003, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
interpreted over feature trees. Based on this approach, we show how a generalization of \( FT \leq \) combined with path constraints can be used to formally represent and state constraints and reason over semistructured data. The proposed query language is extended to facilitate query relaxation when the exact solution to the query cannot be obtained from the data repository. The applicability of the above framework, proposed for semistructured data, is examined for a particular case regarding XML documents circulated and stored over the Web.

INTRODUCTION

New applications and the expanding rate of data being stored and circulated over distributed systems and the Internet have resulted in the development of new database technologies. The need for flexibility and accuracy in data representation and manipulation is vital in Web databases (Abiteboul & Vianu, 1997; Mendelzon & Milo, 1997). The most important characteristic of these applications is the diversity of the data being stored and the lack of fixed and rigid structure (i.e., schema)—semistructured data.

The core problem in semistructured data is that the structure is not fully known, and this leads to querying the data in a content-based fashion, as opposed to the structure-based queries over relational databases. The special features of semistructured data define a particularly interesting domain for query languages. Computations over semistructured data can easily become infinite, even when the underlying alphabet is finite. Query languages for semistructured data have been investigated in the context of algebraic programming (Abiteboul et al., 1997; Buneman et al., 1996). In this chapter, a different approach to modeling and querying semistructured and Web data is presented, based on Feature Logics (Aït-Kaci et al., 1994; Shieber, 1986; Smolka & Treinen, 1994). More specifically, we propose a rule-based constraint language for manipulating semistructured data. The proposed language has declarative and operational semantics based on fix-point theory, as in the typical logic programming (Lloyd, 1987). Moreover, the issue of relaxing queries by extending the proposed query language will be investigated. Finally, the chapter presents a special case study for the proposed framework (based on XML data structure).

Research Review

Semistructured and XML data modeling and retrieval earlier research efforts focused on query languages development, constraint-based approaches, query relaxation and XML-specific query languages:
Related Content

The Social Design of 3D Interactive Spaces for Security in Higher Education: A Preliminary View
www.igi-global.com/chapter/social-design-interactive-spaces-security/49518?camid=4v1a

Wireless Multimedia Content Distribution Architecture
Israel Pérez-Llopis, Carlos E. Palau and Manuel Esteve (2012). Next Generation Content Delivery Infrastructures: Emerging Paradigms and Technologies (pp. 78-104).
www.igi-global.com/chapter/wireless-multimedia-content-distribution-architecture/66994?camid=4v1a

Semantic Web Languages and Ontologies
www.igi-global.com/chapter/semantic-web-languages-ontologies/16897?camid=4v1a
A Unifying Framework Design for the Management of Autonomic Network Functions
[www.igi-global.com/chapter/a-unifying-framework-design-for-the-management-of-autonomic-network-functions/214427?camid=4v1a](http://www.igi-global.com/chapter/a-unifying-framework-design-for-the-management-of-autonomic-network-functions/214427?camid=4v1a)