Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computer and System Sciences 70 (2005) 1–25
JOURNAL OF COMPUTER AND SYSTEM SCIENCES
www.elsevier.com/locate/jcss

Partially ordered regular languages for graph queries

Sergio Flesca, Sergio Greco*,[1]

DEIS Department, Università della Calabria, 87030 Rende, Italy

Received 19 January 2000; received in revised form 28 March 2003

Available online 7 October 2004

Abstract

In this paper we present an extension of regular languages to support graph queries. The proposed extension is based on the introduction of a partial order on the strings of the languages. We extend regular expressions and regular grammars by introducing partial orders on strings and production rules, respectively. The relations between regular expressions and regular grammars are analyzed. We show how partially ordered languages can be used to define path queries to search graph databases and present results on their computational complexity. Finally, we present an application of partially ordered languages for searching the Web.

Keywords: Databases; Graphs; Regular languages; Path queries

1. Introduction

Graph data is an emerging model for representing a variety of database contexts ranging from object oriented databases to hypertext and semistructured data [18,19,7,16,22,25,26,28]. Also many of the recursive queries that arise in relational databases are in practice graph traversals [2,12,13,27,29]. Recently, several languages and prototypes have been proposed for searching graph-like data such as the

* Corresponding author.
E-mail address: greco@si.deis.unical.it (S. Greco).
[1] Supported by ISI-CNR. A preliminary version of this paper appeared in [15].

0022-0000/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcss.2003.04.003

Web [1,3–5,14,20,21,23,24]. All these languages permit us to express (declarative) navigational queries, called path queries, by means of regular expressions denoting paths in the graph [3,8,9,11,13,22].

Path queries permit us to express queries of the form "find all objects reachable from a given node by paths whose labels form a word in r" where r is a regular expression over an alphabet of labels. However, path queries are not satisfactory since the result is the complete set of answers. In practical applications, where the number of answers can be large, the result should be a (limited) list of answers ordered with respect to some criteria specified by the user [10]. This is confirmed by current index servers which enable the user to search documents in the Web, by defining criteria based on the content of documents [6]. Generally, the answer of index servers is a list of documents, ordered on the basis of the criteria specified by the user.

Thus, in order to better capture the navigational aspects of graph-like data, we introduce partially ordered regular languages, an extension of regular languages where strings are partially ordered. Two strings s_1 and s_2, such that $s_1 > s_2$, denote two paths in the graph with the constraint that the path s_1 should be preferred to the path s_2. Thus, in terms of grammars generating (partially ordered) strings, the relation $s_1 > s_2$ implies that the string s_1 should be generated before the string s_2. In terms of paths, this means that the paths spelling the string s_1 should be navigated before the paths spelling the string s_2.

Whenever the user wants a limited set of answers, the whole graph does not need to be explored and the possibility of expressing a preference between paths should be given to the user. This is a very common situation in searching graph-like data such as the Web, where the number of paths and solutions can be extremely high. Thus, partially ordered languages capture the important aspect of navigational queries such as "search first the paths starting with the edge e_1 and next, in case the solution is not complete, those starting with the edge e_2".

1.1. Contributions

The main contributions of this paper are

- the extension of regular expressions by introducing a partial order between 'alternative' strings;
- the extension of regular grammars to generate sets of partially ordered regular strings;
- results on the complexity of and algorithms for graph searching by means of partially ordered path queries.

A practical contribution is the application of partially ordered languages, in query languages for searching graph-like data such as the Web.

1.2. Organization of the paper

The rest of the paper is organized as follows. In Section 2, we present basic definitions on regular languages and regular grammars. In Section 3, we introduce partially ordered languages. More specifically, we introduce, in Section 3.1, extended regular expressions and then, in Section 3.2, we present partially ordered regular grammars. In Section 4, we show how partially ordered regular languages can be used to search graph data and present an algorithm to compute path queries. In Section 5, we show how partially