

[[Pagewise preview](#)]

Category	Value
Available via	http://dbpubs.stanford.edu/pub/1999-65
Submitted on	31st of October 2001
Author	Brin, Sergey
Title	Extracting Patterns and Relations from the World Wide Web.
Date of publication	11th of November 1999
Published in	WebDB Workshop at EDBT'98
Citation	Brin, Sergey. Extracting Patterns and Relations from the World Wide Web., WebDB Workshop at EDBT'98
Number of pages	12
Language	English
Project	Digital Libraries
Type	Conference or Journal Paper
Subject group	Digital Libraries
Abstract	<p>The World Wide Web is a vast resource for information. At the same time it is extremely distributed. A particular type of data such as restaurant lists may be scattered across thousands of independent information sources in many different formats. In this paper, we consider the problem of extracting a relation for such a data type from all of these sources automatically. We present a technique which exploits the duality between sets of patterns and relations to grow the target relation starting from a small sample. To test our technique we use it to extract a relation of (author, title) pairs from the World Wide Web.</p>
Notes	Previous number = SIDL-WP-1999-0119
Fulltext source	<ul style="list-style-type: none"> ● Postscript (ps, ps.gz, ps.zip) ● PDF (pdf, pdf.gz, pdf.zip) ● Plain text (text, text.gz, text.zip)

[Management
of the
document by](#)

siroker@db.stanford.edu

[[Pagewise preview](#)]

[Stanford Database Group Publication Server](#)