Unsupervised Technique for Web Data Extraction: Trinity

Abstract

Search engine is a program which searches specific information from huge amount of data. So for getting results in an effective manner and within less time this technique is used. This article is having a technique which depends on two or more web documents which are generated from same server-side template. The technique does not provide any relevant data but searches for shared pattern and separates it into three sub parts then apply different ranking functions and stored it into database. When comparing our technique with other techniques we can see that input documents are not having any negative impact on its effectiveness, also it gives results in less time and in the exact form.

References

Unsupervised Technique for Web Data Extraction: Trinity

- V. Crescenzi, G. Mecca, and P. Merialdo, Road runner: Towards automatic data extraction from large websites, in Proc. 27th Int. Conf. VLDB, Rome, Italy, 2001, pp. 109118.
- C. -H. Chang and S. -C. Lui, IEPAD: Information extraction based on pattern discovery, in Proc. 10th Int. Conf. WWW, Hong Kong, China, 2001, pp. 681688.

Index Terms

Computer Science

Web Services
Keywords
Web Data extractor  Automatic Wrapper Generation  Wrapper  Unsupervised Technique