Abstract

In present days finding relevant and desired information in less time is very crucial however problem is that very small proportion data on internet is interpretable and meaningful and need lot of time to extract. The paper provides solution to problem by extracting information from top-k websites, which consist top k instances of a subject. For Examples "top 5 football teams in the world". In comparison with other structured information like web tables top-k lists contains high quality information. It can be use to enhance open-domain knowledge base [which can support search or fact answering applications]. Proposed system in paper extract the top k list by using title classifier, parser, candidate picker, ranker, content processor.

References

2. Z. Zhang, K. Q. Zhu, and H. Wang, “A system for extracting top-k lists from the web” in
KDD, 2012.
6. F. Fumarola, T. Weninger, R. Barber, D. Malerba, and J. Han, ”Extracting general lists from web documents: A hybrid approach,” in IEA/AIE (1), 2011, pp. 285294.
8. A. Angel, S. Chaudhuri, G. Das, and N. Koudas, ”Ranking objects based on relationships and fixed associations,” in EDBT, 2009, pp. 910921.
14. K. Chakrabarti, V. Ganti, J. Han, and D. Xin, ”Ranking objects based on relationships,” in SIGMOD, 2006, pp. 371382.

Index Terms

Computer Science Information Sciences

Keywords

Data extraction, Structured information, top k list, top k web pages, web parser