Abstract

A huge portion of the web known as deep web is accessible via search interfaces to myriads of databases on the web. Deep web crawl is concerned with the problem of surfacing hidden content behind search interfaces on the web. Given the dynamic nature of the web, where data sources are constantly changing, it is crucial to discover these resources. The paper proposes a two level application namely deep web crawler for gathering relevant searchable forms. In the first level deep web crawler explores the forms based on reverse searching for a given seed site, ranking the sites to prioritize highly relevant sites and by extracting the links to find the forms. In the next level, it searches the forms based on preference and the result is enhanced by re-ranking, given the user feedback.

References


Index Terms
Computer Science  Information Sciences

Keywords
Deep web, adaptive learning, ranking