

Snippet Matching for Click-Tracking Blocking - Abstract

Yana Kulizhskaya

This dissertation proposes a solution to the privacy issues regarding user click-tracking by search engines by a means of a URL resolution service. The proposed solution would leverage the search result snippet data to be used for look up of the respective search result URL in a user-collected snippet-URL repository. In order to examine the viability of the proposed solution, the dissertation investigates whether search result snippets can be used to successfully infer search result URLs.

A search result dataset is collected from the Google search engine and a snippet matching algorithm is developed. An examination of the search snippet dataset indicates that snippets consist of a highly diverse content, making them easily identifiable. The error rate evaluation of snippet matching yields an error rate of 3.48%, indicating that search result snippets can be used for URL resolution with great accuracy, provided a priori knowledge of the snippet-URL mapping exists. Additional investigation into snippet stability and the snippet matching algorithm's resilience to modification of input further support the viability of a click-tracking blocking system that employs snippet matching for URL resolution as snippets are mostly stable over time and the matching process is robust to input change.