

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/94148>

Please be advised that this information was generated on 2021-01-25 and may be subject to change.

Topical Language Model for Snippet Retrieval

Rongmei Li and Theo van der Weide

Radboud University, Nijmegen, The Netherlands

Our previous study found that topical language model improves document ranking of ad-hoc retrieval. In this work, our attention is paid on snippets that are extracted and generated from the provided ranked list of documents.

In our experiments of the Snippet Retrieval Track, we hypothesize that the user recognizes certain combinations of terms in created snippets which are related to their information needs. We automatically extract snippets using terms as the minimal unit. Each term is weighted according to their relative occurrence in its article and in the entire Wikipedia. The top K scoring terms are chosen for inclusion in the snippet. The term-extraction based snippets are then represented differently to the user. One is a cluster of words that indicate the described topic. Another is a cluster of semi-sentences that contains the topic information while preserving some language structure.