

Stopword Removal: Why Bother? A Case Study on Verbose Queries

ABSTRACT

Stopword removal has traditionally been an integral step in information retrieval preprocessing. In this paper we question the utility of this step for verbose queries on standard datasets. We show in seven FIRE test collections in four languages – Bangla, Hindi, Gujarati and Marathi, that stopword removal does not lead to noticeable difference in retrieval performance as opposed to not removing stopwords. However, for European languages like English (TREC678 Ad Hoc) and French (CLEF 2005 to 2007), stopword removal leads to a statistically significant drop in performance.

CCS CONCEPTS

• Information systems → Information retrieval;

KEYWORDS

Stopword removal, retrieval, test collections

ACM Reference format:

. 2017. Stopword Removal: Why Bother? A Case Study on Verbose Queries. In Proceedings of ACM CIKM conference, Pan Pacific, Singapore, November 2017 (CIKM'17), ?? pages.

DOI:

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CIKM'17, Pan Pacific, Singapore

© 2017 Copyright held by the owner/author(s). ...\$15.00

DOI: