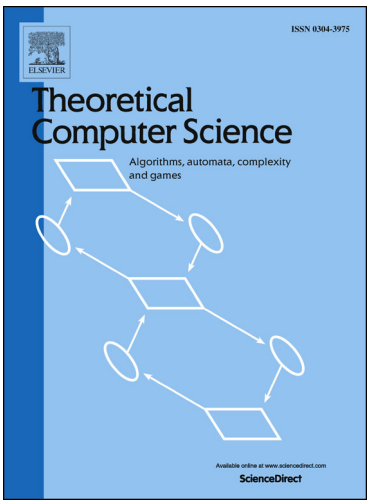
Accepted Manuscript



Ranked Document Retrieval for Multiple Patterns

Sudip Biswas, Arnab Ganguly, Rahul Shah, Sharma V. Thankachan

PII: S0304-3975(18)30445-6
DOI: https://doi.org/10.1016/j.tcs.2018.06.029
Reference: TCS 11652

To appear in: *Theoretical Computer Science*

Received date: 8 September 2017
Revised date: 10 May 2018
Accepted date: 15 June 2018

Please cite this article in press as: S. Biswas et al., Ranked Document Retrieval for Multiple Patterns, *Theoret. Comput. Sci.* (2018), https://doi.org/10.1016/j.tcs.2018.06.029

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Ranked Document Retrieval for Multiple Patterns[☆]

Sudip Biswas[a,*], Arnab Ganguly[b,**], Rahul Shah[c], Sharma V. Thankachan[d]

[a]*Facebook Inc., USA*
sudipbiswas@fb.com
[b]*Department of CS, University of Wisconsin - Whitewater, USA*
gangulya@uww.edu
[c]*School of EECS, Louisiana State University, USA & National Science Foundation, USA*
rahul@csc.lsu.edu, rahul@nsf.gov
[d]*Department of CS, University of Central Florida, USA*
sharma.thankachan@ucf.edu

Abstract

Let $\mathcal{D} = \{\mathsf{T}_1, \mathsf{T}_2, \ldots, \mathsf{T}_D\}$ be a collection of D documents having n characters in total. Given two patterns P and Q, and an integer $k > 0$, we consider the following queries.

- top-k *forbidden pattern query:* Among all documents containing P, but not Q, report the k documents most relevant to P.
- top-k *two pattern query:* Among all documents that contain both P and Q, report the k documents most relevant to P.

For the above two queries, we provide a linear space index with $O(|P| + |Q| + \sqrt{nk})$ query time, under some standard relevance functions such as

[☆]A preliminary version [5] appeared in the 26th Annual Symposium on Combinatorial Pattern Matching (CPM) 2015.

[*]Part of the work was done while a graduate student at LSU, and was supported by NSF Grants CCF–1017623, CCF–1218904, and CCF–1527435.

[**]Part of the work was done while a graduate student at LSU, and was supported by an LSU Dissertation Fellowship.

Download English Version:

https://daneshyari.com/en/article/10225760

Download Persian Version:

https://daneshyari.com/article/10225760

Daneshyari.com