

Home | Computational Linguistics | List of Issues | Volume 30 , No. 4 | Fast Approximate Search in Large Dictionaries



Fast Approximate Search in Large Dictionaries

Stoyan Mihov and Klaus U. Schulz

Posted Online March 13, 2006

<https://doi.org/10.1162/0891201042544938>

© 2004 Association for Computational Linguistics

Computational Linguistics
Volume 30 | Issue 4 | December 2004
p.451-477



Download Options



Abstract Authors

The need to correct garbled strings arises in many areas of natural language processing. If a dictionary is available that covers all possible input tokens, a natural set of candidates for correcting an erroneous input P is the set of all words in the dictionary for which the Levenshtein distance to P does not exceed a given (small) bound k . In this article we describe methods for efficiently selecting such candidate sets. After introducing as a starting point a basic correction method based on the concept of a “universal Levenshtein automaton,” we show how two filtering methods known from the field of approximate text search can be used to improve the basic procedure in a significant way. The first method, which uses standard dictionaries plus dictionaries with reversed

Quarterly (March, June, September, December)

160pp. per issue

6 3/4 x 10

Founded: 1974

2018 Impact Factor: 1.319

2018 Google

Scholar h5-index: 32

ISSN: 0891-2017

E-ISSN: 1530-9312

Journal

Resources

[Editorial Info](#)

[Abstracting and Indexing](#)

[Release Schedule](#)

[Advertising Info](#)


Author

Resources

Submission Guidelines
Publication Agreement
Author Reprints

Reader Resources

Rights and Permissions
Most Read
Most Cited

More About Computational Linguistics 

Metrics 



- Twitter (2)
- Patents (1)
- Mendeley (2)
- CiteULike (6)




38 Total citations
3 Recent citations

5.44 Field Citation Ratio

n/a Relative Citation Ratio

Open Access 


 Computational Linguistics Computational Linguistics is Open Access. All content is


words, leads to very short correction times for most classes of input strings. Our evaluation results demonstrate that correction times for fixed-distance bounds depend on the expected number of correction candidates, which decreases for longer input words. Similarly the choice of an optimal filtering method depends on the length of the input words.


Forthcoming

Most Read

[See More](#)

 **Lexicon-Based Methods for Sentiment Analysis** (14087 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307


 **Computational Linguistics and Deep Learning** (10542 times)
Christopher D. Manning
Computational Linguistics
Volume: 41, Issue: 4, pp. 701-707


 **Near-Synonymy and Lexical Choice** (3675 times)
Philip Edmonds et al.
Computational Linguistics
Volume: 28, Issue: 2, pp. 105-144


(Note that the Most Read numbers are based on the number of full text downloads over the last 12 months.)

Most Cited

[See More](#)

 **Lexicon-Based Methods for Sentiment Analysis** (436 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307



 **A Systematic Comparison of Various Statistical Alignment Models** (174 times)
Franz Josef Och et al.
Computational Linguistics
Volume: 29, Issue: 1, pp. 19-51

 **Opinion Word Expansion and Target Extraction through Double Propagation** (147 times)
Guang Qiu et al.
Computational Linguistics
Volume: 37, Issue: 1, pp. 9-27

(Note that the Most Cited numbers are based on Crossref's [Cited-by service](#) and reflect citation information for the past 24 months.)

Download Options


Favorite  Sign up for Alerts 

Download Citation  RSS TOC 

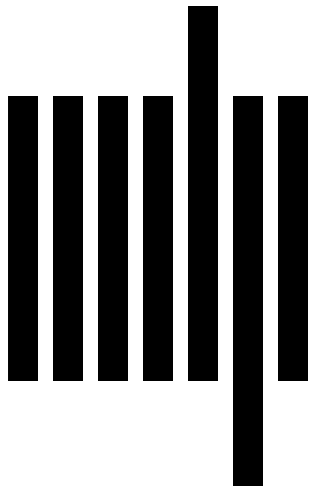
RSS Citation Submit your

freely available in electronic format (Full text HTML, PDF, and PDF Plus) to readers across the globe. All articles are published under a [CC BY-NC-ND 4.0 license](#). For more information on allowed uses, please view the CC license.

[Support OA at MITP](#)

 article

[Support OA at MITP](#) 



Journals

Books

US

One Rogers Street
Cambridge MA 02142-1209

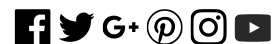
Terms & Conditions

UK

Suite 2, 1 Duchess Street
London, W1W 6AN, UK

Privacy Statement

Connect



Contact Us

© 2018 The MIT Press
Technology Partner:
[Atypon Systems, Inc.](#)
[CrossRef Member](#)
[COUNTER Member](#)
The MIT Press colophon is registered in the U.S. Patent and Trademark Office.
[Site Help](#)