

The MIT Press

Journals

[Sign In / Register](#)
[Books](#)
[Journals](#)
[Digital](#)
[Resources](#)
[About](#)
[Contact](#)


Home | Computational Linguistics | List Article navigation
of Issues | Volume 26 , No. 3 | A
Compression-based Algorithm for Chinese
Word Segmentation



A Compression-based Algorithm for Chinese Word Segmentation

W. J. Teahan, Yingying Wen,
Rodger McNab and Ian H. Witten

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

Posted Online March 13, 2006

<https://doi.org/10.1162/089120100561746>

© 2000 Association for Computational Linguistics

Computational Linguistics

Volume 26 | Issue 3 | September 2000

p.375-393

 **Download Options** >

Journal Resources

[Editorial Info](#)

[Abstracting and Indexing](#)

[Release Schedule](#)

[Advertising Info](#)

Abstract Authors


Chinese is written without using spaces or other word delimiters. Although a text may be thought of as a corresponding sequence of words, there is considerable ambiguity in the placement of boundaries. Interpreting a text as a sequence of words is beneficial for some information retrieval and storage tasks: for example, full-text search,

Author Resources

Submission Guidelines
Publication Agreement
Author Reprints

Reader Resources

Rights and Permissions
Most Read
Most Cited

More About Computational Linguistics 

Metrics 



66 Total citations

3 Recent citations

11 Field Citation Ratio

n/a Relative Citation Ratio

Open Access 




Computational Linguistics
Computational Linguistics is Open Access. All content is freely available in


word-based compression, and keyphrase extraction. We describe a scheme that infers appropriate positions for word boundaries using an adaptive language model that is standard in text compression. It is trained on a corpus of presegmented text, and when applied to new text, interpolates word boundaries so as to maximize the compression obtained. This simple and general method performs well with respect to specialized schemes for Chinese language segmentation.


Forthcoming

Most Read

[See More](#)

 **Lexicon-Based Methods for Sentiment Analysis** (14129 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307


 **Computational Linguistics and Deep Learning** (10558 times)
Christopher D. Manning
Computational Linguistics
Volume: 41, Issue: 4, pp. 701-707


 **Near-Synonymy and Lexical Choice** (3688 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
Sign up for

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](#).

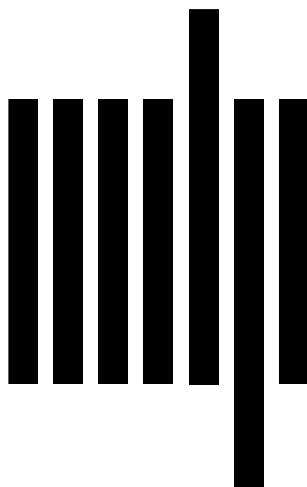
[Support OA at MITP](#)

[Favorite](#)  [Alerts](#) 

[Download Citation](#)  [RSS TOC](#) 

[RSS Citation](#)  [Submit your 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](#)