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

Improving Machine Translation Performance by Exploiting Non-Parallel Corpora

Dragos Stefan Munteanu and Daniel Marcu

Posted Online March 13, 2006

https://doi.org/10.1162/089120105775299168

Computational Linguistics

Volume 31 | Issue 4 | December 2005

p.477-504

Abstract Authors

We present a novel method for discovering parallel sentences in comparable, non-parallel corpora. We train a maximum entropy classifier that, given a pair of sentences, can reliably determine whether or not they are translations of each other. Using this approach, we extract parallel data from large Chinese, Arabic, and English non-parallel newspaper corpora. We evaluate the quality of the extracted data by showing that it improves the performance of a state-of-the-art statistical machine translation system. We also show

Submission Guidelines
Publication Agreement
Author Reprints

Reader Resources

Rights and Permissions
Most Read
Most Cited

More About Computational Linguistics

Metrics

86 Total citations

14 Recent citations

17 Field Citation Ratio

n/a Relative Citation Ratio

Open Access

Computational Linguistics Computational Linguistics is Open Access. All content is freely available in electronic format (Full text HTML, PDF, and PDF Plus) to readers across the globe. All

that a good-quality MT system can be built from scratch by starting with a very small parallel corpus (100,000 words) and exploiting a large non-parallel corpus. Thus, our method can be applied with great benefit to language pairs for which only scarce resources are available.

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

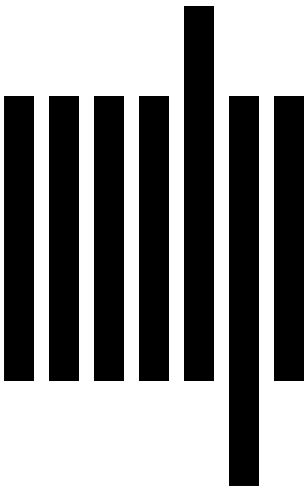
Opinion Word Expansion and Target Extraction through Double Propagation (147 times)
Guang Qiu et al.
Computational Linguistics
Volume: 37, Issue: 1, pp. 9-27

(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 article

Support OA at MITP



US

One Rogers Street
Cambridge MA 02142-1209

UK

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

Connect