

SUBSCRIPTIONS

Natural Language Engineering (ISSN 1351-3249; electronic 1469-8110) is published six times a year in January, March, May, July, September and November. Six parts form a volume. The subscription price (excluding VAT and sales tax) of volume 28 (2022) (which includes postage) is £444 net (US\$744 in the USA, Canada and Mexico) for institutions print and electronic, institutions electronic only £362/\$597; £57 net (US\$90 in the USA, Canada and Mexico) for individuals ordering direct from the publisher and certifying that the journal is for their personal use. Single parts are £85 (US\$143 in the USA, Canada and Mexico) plus postage.

Orders, which must be accompanied by payment, may be sent to a bookseller, subscription agent or direct to the publisher: Cambridge University Press, Journals Fulfillment Department, UPH, Shaftesbury Road, Cambridge CB2 8BS, UK; or in the USA, Canada and Mexico: Cambridge University Press, Journals Fulfillment Department, 1 Liberty Plaza, Floor 20, New York, NY 10006, USA. EU subscribers who are not registered for VAT should add VAT at their country's rate. VAT registered subscribers should provide their VAT registration number. Japanese prices for institutions are available from Kinokuniya Company Ltd, PO Box 55, Chitose, Tokyo 156, Japan.

COPYING

This journal is registered with the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA. Organizations in the USA who are also registered with C.C.C. may therefore copy material (beyond the limits permitted by sections 107 and 108 of U.S. Copyright law) subject to payment to C.C.C. of the per-copy fee of \$16.00. This consent does not extend to multiple copying for promotional or commercial purposes. Code 1351-3249/2015.

ISI Tear Sheet Service, 3501 Market Street, Philadelphia, PA 19104, USA, is authorised to supply single copies of separate articles for private use only.

Organisations authorised by the Copyright Licensing Agency may also copy material subject to the usual conditions.

For all other use, permission should be sought from Cambridge or from the American Branch of Cambridge University Press.

INSTRUCTIONS FOR CONTRIBUTORS

Instructions for Contributors will be found on the Journal's webpage at www.cambridge.org/nle. This journal and other journals are included in the Cambridge Journals Online service which can be found at journals.cambridge.org

© Cambridge University Press 2022

This journal issue has been printed on FSC-certified paper and cover board. FSC is an independent, non-governmental, non-for-profit organization established to promote the responsible management of the world's forests. Please see www.fsc.org for information.

NATURAL LANGUAGE ENGINEERING

Volume 28 Part 1 JANUARY 2022

CONTENTS

ARTICLES

- Cluster-based mention typing for named entity disambiguation 1
Arda Çelebi and Arzucan Özgür
- Learning from noisy out-of-domain corpus using dataless classification 39
Yiping Jin, Dittaya Wanvarie and Phu T. V. Le
- Improved feature decay algorithms for statistical machine translation 71
Alberto Poncelas, Gideon Maillette de Buy Wenniger and Andy Way
- A clustering framework for lexical normalization of Roman Urdu 93
Abdul Rafae Khan, Asim Karim, Hassan Sajjad, Faisal Kamiran and Jia Xu

INDUSTRY WATCH

- \$NLP: How to spend a billion dollars 125
Robert Dale

BOOK REVIEW

- Text Mining with R: A Tidy Approach, by Julia Silge and David Robinson.
Sebastopol, CA: O'Reilly Media, 2017 137
Jianwei Yan

The cover illustration is computer generated from an engraving of the Tower of Babel in an eighteenth-century bible.

Cambridge Core

For further information about this journal
please go to the journal website at:
cambridge.org/nle



MIX
Paper from
responsible sources
FSC® C007785

CAMBRIDGE
UNIVERSITY PRESS