Cambridge Core

The new home of Cambridge Journals

cambridge.org/core
Computer Science
Books and Journals from Cambridge University Press

Cambridge publish in a number of areas of Computer Science including artificial intelligence, programming languages, robotics and the application of mathematics to computer science

For further details visit: cambridge.org/core-computer-science
Natural Language Engineering meets the needs of professionals and researchers working in all areas of automatic language processing, whether from the perspective of theoretical or corpus linguistics, translation, lexicography, computer science or engineering. Its aim is to bridge the gap between traditional computational linguistics research and the implementation of practical applications with potential real-world use. As well as publishing original research articles on a broad range of topics - from text analysis, machine translation, information retrieval, speech processing and generation to integrated systems and multi-modal interfaces - it also publishes special issues on specific natural language processing methods, tasks or applications. The journal welcomes survey papers describing the state of the art of a specific topic. The Journal of Natural Language Engineering also publishes the popular Industry Watch and Emerging Trends columns as well as book reviews.

EDITORS
Ruslan Mitkov (Executive Editor), University of Wolverhampton, UK. Email: R.Mitkov@wlv.ac.uk
John I. Tait, john@tait.net Ltd., UK. Email: john@john@tait.net
Branimir K. Boguraev; IBM Thomas J. Watson Research Center, USA. Email: bran@us.ibm.com

ASSOCIATE EDITOR
Diana Inkpen, University of Ottawa, Canada. Email: Diana.Inkpen@uottawa.ca

BOOK REVIEW EDITOR
Diana Maynard, University of Sheffield, UK. Email: d.maynard@sheffield.ac.uk

INDUSTRY WATCH EDITOR
Robert Dale, Language Technology Group, Australia. Email: rdale@language-technology.com

EMERGING TRENDS EDITOR
Kenneth Church, IBM, NY, USA. Email: kwchurch@us.ibm.com

FOUNDEd EDITOR
Roberto Gattigiano, SenseGraph Ltd., UK.

EDITORIAL BOARD
Kevin Brétar, University of Colorado, USA
Ondřej Bojar, Charles University, Czech Republic
Gregory Greiner, Florida Institute for Human & Machine Cognition (IHMC), USA, France
Tracy Holloway King, A9 Amazon, USA
Sandra Kuebler, University of Indiana, USA
Lott Lamel, LIMSI, France
Suresh Manandhar, University of York, UK
Aleksandro Moschitti, University of Trento, Italy
Pavlík Nakov, Qatar Computing Research Institute, HBKU, Qatar
Vivi Nastase, University of Heidelberg, Germany
Roberto Navigli, Sapienza University of Rome, Italy

EDIToRAL ASSISTANTS
Martina Cotella, University of Wolverhampton, UK
April Harper, University of Wolverhampton, UK

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 24 (2018) (which includes postage) is £387 net (US$632 in the USA, Canada and Mexico) for institutions print and electronic, institutions electronic only (321 /528, 46 €net (US$75 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 £71 (US$119 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, Shaftsbury 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. IS1 1469-8110

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

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 2018

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

Printed in the UK by Bell & Bain Ltd, Glasgow
Cambridge University Press
Journals Fulfillment Department, UPH, Shaftsbury Road, Cambridge CB2 8BS, UK
1 Liberty Plaza, Floor 20, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
https://www.cambridge.org/core/terms. E&O E. © Cambridge University Press 2018

Dock House, The Waterfront, Cape Town 8001, South Africa
CONTENTS

Editorial note 1
Prof Ruslan Mitkov, Executive Editor

ARTICLES

To use or not to use: Feature selection for sentiment analysis of highly imbalanced data
Sandra Kübler, Can Liu and Zeeshan Ali Sayyed
3

A Semi-automatic and low-cost method to learn patterns for named entity recognition
M. Marrero and J. Urbano
39

CCG supertagging with bidirectional long short-term memory networks
Rekia Kadare, Yu Zhang, Wenan Zhang and Ting Liu
77

New treebank or repurposed? On the feasibility of cross-lingual parsing of Romance languages with Universal Dependencies
Marcos García, Carlos Gómez-Rodríguez and Miguel A. Alonso
91

A survey of diacritic restoration in abjad and alphabet writing systems
Franklin Oladipú Asahiah, Odejániyi Ajádi Oléjóbí and Emmanuel Rótímí Adágúnódó
123

Emerging trends: A tribute to Charles Wayne
Kenneth Ward Church
155

Book Review
Formalizing Natural Languages: The NooJ Approach
Max Silberztein
161

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