

DATA MINING VIII

DATA, TEXT AND WEB MINING AND THEIR BUSINESS APPLICATIONS

WIT*PRESS*

WIT Press publishes leading books in Science and Technology.

Visit our website for the current list of titles.

www.witpress.com

WIT*eLibrary*

Home of the Transactions of the Wessex Institute.

Papers presented at Data Mining VIII are archived in the WIT eLibrary in volume 38 of WIT Transactions on Information and Communication Technologies (ISSN 1743-3517).

The WIT eLibrary provides the international scientific community with immediate and permanent access to individual papers presented at WIT conferences.
<http://library.witpress.com>

WIT Press publishes leading books in Science and Technology.
Visit our website for the current list of titles.

www.witpress.com

EIGHTH INTERNATIONAL CONFERENCE ON DATA MINING

DATA MINING VIII

CONFERENCE CHAIRMEN

A. Zanasi

TEMIS Text Mining Solutions, Italy

C.A. Brebbia

Wessex Institute of Technology, UK

N.F.F. Ebecken

COPPEL/Federal University of Rio de Janeiro, Brazil

INTERNATIONAL SCIENTIFIC ADVISORY COMMITTEE

D. Anderson	C. Curotto	J.P. Lawler
V. Benson	S. de Rossi	B. Little
M.W. Berry	B. Drewes	P.J-S Pan
I. Caddy	A.G. Evsukoff	M.F. Santos
F. Camillo	M. Gottgtroy	S. Sirmakessis
M. Ceci	S. Larreina	D. Sitnikov

Organised by

*Wessex Institute of Technology, UK
Federal University of Rio de Janeiro, Brazil*

Sponsored by

WIT Transactions on Information and Communication Technologies

WIT Transactions on Information and Communication Technologies

Transactions Editor

Carlos Brebbia

Wessex Institute of Technology
Ashurst Lodge, Ashurst
Southampton SO40 7AA, UK
Email: carlos@wessex.ac.uk

Editorial Board

P L Aguilar

University of Extremadura
Spain

J J Casares Long

Universidad de Santiago de Compostela
Spain

A Davies

University of Hertfordshire
UK

A De Montis

Universita di Cagliari
Italy

G K Egan

Monash University
Australia

D J Evans

Nottingham Trent University
UK

A Genco

University of Palermo
Italy

D Goulias

University of Maryland
USA

J Jaafar

UiTm
Malaysai

C-H Lai

University of Greenwich
UK

J Lourenco

Universidade do Minho
Portugal

N Milic-Frayling

Microsoft Research Ltd
UK

J Rao

Case Western Reserve University
USA

M P Bekakos

Democritus University of Thrace
Greece

M Costantino

Royal Bank of Scotland Financial Markets
UK

N J Dedios Mimbela

Universidad de Cordoba
Spain

G Deplano

Universita di Cagliari
Italy

K-H Elmer

Universitat Hannover
Germany

U Gabbert

Otto-von-Guericke Universitat Magdeburg
Germany

P Giudici

Universita di Pavia
Italy

A Gualtierotti

IDHEAP
Switzerland

T Kobayashi

University of Tokyo
Japan

G Loo

The University of Auckland
New Zealand

D Malerba

Universita degli Studi
Italy

G Nakhaeizadeh

DaimlerChrysler Research & Technology
Germany

J Roddick

Flinders University
Australia

F Rodrigues

Poly Institute of Porto
Portugal

R Turra

CINECA Interuniversity Computing Centre
Italy

H Westphal

University of Magdeburg
Germany

N Zhong

Maebashi Institute of Technology
Japan

W Schreiber

University of Alabama
USA

D Van den Poel

Ghent University
Belgium

J Yoon

Old Dominion University
USA

H G Zimmermann

Siemens AG
Germany

DATA MINING VIII

DATA, TEXT AND WEB MINING
AND THEIR BUSINESS APPLICATIONS

Editors

A. Zanasi

TEMIS Text Mining Solutions, Italy

C.A. Brebbia

Wessex Institute of Technology, UK

N.F.F. Ebecken

COPPE/Federal University of Rio de Janeiro, Brazil

WITPRESS Southampton, Boston



Editors:

A. Zanasi

TEMIS Text Mining Solutions, Italy

C.A. Brebbia

Wessex Institute of Technology, UK

N.F.F. Ebecken

COPPE/Federal University of Rio de Janeiro, Brazil

Published by

WIT Press

Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK

Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853

E-Mail: witpress@witpress.com

<http://www.witpress.com>

For USA, Canada and Mexico

WIT Press

25 Bridge Street, Billerica, MA 01821, USA

Tel: 978 667 5841; Fax: 978 667 7582

E-Mail: infousa@witpress.com

<http://www.witpress.com>

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available
from the British Library

ISBN: 978-1-84564-081-1

ISSN: 1746-4463 (print)

ISSN: 1743-3517 (on-line)

*The texts of the papers in this volume were set
individually by the authors or under their supervision.
Only minor corrections to the text may have been carried
out by the publisher.*

No responsibility is assumed by the Publisher, the Editors and Authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. The Publisher does not necessarily endorse the ideas held, or views expressed by the Editors or Authors of the material contained in its publications.

© WIT Press 2007

Printed in Great Britain by Cambridge Printing.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Publisher.

Preface

This year the Data Mining and Information Engineering Conference comes back to the United Kingdom after 7 years away (it was held in Cambridge university in 2000) in the marvellous framework of The New Forest. The Conference remains faithful to its original idea of providing a platform to discuss theoretical and applicative aspects of data mining, with participants from all over the world, both from academia and from industry.

Its success is reflected in the papers received, with participants coming from several countries, with a strong representation from the USA, the Americas and Europe as well as many of the other continents, allowing a real multinational-multicultural exchange of experiences and ideas.

During this last year we witnessed the growth of the European Union interest in data mining, especially as applied to unstructured data (eg. text mining). The 7Framework Program (the program by which European Commission funds European Research) has a new line of funding: Security Research, where “business analytics” (as data and text mining) are well represented.

This is proof that data mining is seen not only as an exciting research area but also as a technology that may solve current European citizens’ concerns with problems such as security. The interest and appreciation of the European Union is also shown through the 2007 European ICT Prize that has been awarded for the first time to a text mining company.

For a discipline which is central to research and also to business, and which generates interests not only among software developers but also among large companies and government departments and agencies, it is important to look at the market and at its movements. A recent analysis about the growth rate of the text mining market, estimates it to be between 30% and 60% per year. This is a significant figure if we recall the estimate of IBM which expected that the text analytics market size was worth around \$5 billion in 2005).

A proof of the fact that although this market is still considered young and not yet consolidated, has been the vigorous wave of mergers and acquisitions that is transforming it. In December 2006 Lexis-Nexis, a powerhouse of information online, bought the French company Datops. Only some months later, in April 2007, its competitor Reuters, the global news provider, reacted by deciding to buy Clearforest,

an Israelian company considered one of the historic founders of the text mining market.

A Conference such as this can only succeed as a team effort, so the Editors want to thank the International Scientific Advisory Committee for their excellent work in reviewing the papers as well as their invaluable input and advice.

The next Data Mining and Information Engineering Conference will take place in Cadiz, Spain in 2008!

The Editors
New Forest, UK
2007

Contents

Section 1: Categorisation methods

A generalized algebraic approach to finding rough set approximations and generating logic rules <i>D. Sitnikov, O. Ryabov, O. Titova & O. Romanenko</i>	3
Implications of frequent subtree mining using hybrid support definition <i>F. Hadzic, H. Tan, T. S. Dillon & E. Chang</i>	13
A comparison of bio-inspired metaheuristic approaches in classification tasks <i>R. L. Oliveira, B. S. L. P. de Lima & N. F. F. Ebecken</i>	25

Section 2: Data preparation

Three dimensional terrain data compression using second generation wavelets <i>B. Pradhan & S. Mansor</i>	35
Identifying character non-independence in phylogenetic data using parallelized rule induction from coverings <i>J. Leopold, A. Maglia, M. Thakur, B. Patel & F. Ercal</i>	45
Implementation and performance assessment of a parallel solver for sparse linear systems of equations and rules for optimal solution <i>T. Grytsenko & A. Peratta</i>	55
Genetic Algorithms in a dynamically changing environment <i>B. Dilimulati & I. Bruha</i>	65
Defining logic structures in functional spaces <i>D. Sitnikov, N. Matski & O. Ryabov</i>	75

Performance assessment of parallel techniques <i>T. Grytsenko & A. Peratta</i>	83
---	----

Section 3: Enterprise information systems

A decision support methodology to rank Information Technology projects <i>S. F. C. Pinho & N. F. F. Ebecken</i>	95
The double wave data warehouse lifecycle model <i>D. P. du Plessis & T. McDonald</i>	105
A simple approach to temporal data in relational DBMS <i>M. K. Crowe</i>	115
SCADA “Supervisory Control And Data Acquisition” implementation in “Sui Northern Gas Pipelines Limited Pakistan” (SNGPL) <i>M. Shahid Khalil & A. Shahnaz Safdar</i>	127

Section 4: Mining environmental and geospatial data

Mining geospatial data in a transductive setting <i>A. Appice, N. Barile, M. Ceci, D. Malerba & R. P. Singh</i>	141
Remote sensing and US crop insurance program integrity: data mining satellite and agricultural data <i>B. Little, M. Schucking, B. Gartrell, B. Chen, S. Olson, K. Ross, C. Jenkerson & R. KcKellip</i>	151

Section 5: Text mining

Protein Ontology Project: 2007 updates <i>A. S. Sidhu, T. S. Dillon & E. Chang</i>	163
Internet search mechanisms and distortions of the semantic space: the scientific challenges facing the “Googles” <i>A. Linhares & C. W. Afonso</i>	173
Performance of information retrieval models using term co-occurrences <i>G. Desjardins, R. Godin & R. Proulx</i>	183

Data mining: an automatic tool for retrieval of data from a system's requirements <i>R. Ibrahim, N. Ibrahim & N. Ismail</i>	193
--	-----

Section 6: Applications in business, industry and government

Creating trading signals using swarm technology <i>N. F. F. Ebecken & B. S. L. P. de Lima</i>	205
--	-----

Knowledge representation in our systems <i>V. Aggelis</i>	211
--	-----

Knowledge discovery in a circle of trust <i>L. Peyton & J. Hu</i>	225
--	-----

Self-regulated learning strategies and computer self-efficacy in IT courses <i>F. Paraskeva</i>	235
--	-----

Physicians' Internet use: diffusion through compatibility and observability <i>F. Chew & W. Grant</i>	245
--	-----

A market-led approach to female student recruitment onto computing-related courses <i>S. Stocking, N. Stanton & L. Seldon</i>	255
--	-----

A re-examination of volatility spillovers in European government bond markets using a multi-objective artificial network <i>G. H. Dash, Jr. & N. Kajiji</i>	263
--	-----

Distributed knowledge-based medical system for data assimilation and analysis <i>V. Tomenko & V. Popov</i>	273
---	-----

Section 7: Customer relationship management

Estimation of the customer mean survival time in subscription-based businesses <i>Z. Mohammed, S. Maritz & D. Kotze</i>	285
--	-----

Grid-based data mining for market basket analysis in the retail sector <i>R. P. Singh, A. Turi & D. Malerba</i>	293
--	-----

Customer survival time in subscription-based businesses (case of Internet service providers)
Z. Mohammed, J. S. Maritz & D. Kotze303

Section 8: National security

Money laundering, terrorist financing and how to contrast them: data and text mining in business intelligence solutions
F. R. Federici315

A multilanguage platform for Open Source Intelligence
N. Baldini, F. Neri & M. Pettoni325

More security and autonomy for users: encryption system with evolutive key
E. Franti & M. Dascalu335

Author Index345