DATA MINING VII

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 2006 are archived in the WIT eLibrary in volume 90
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.comWIT Press publishes leading books in Science and Technology.
Visit our website for the current list of titles.
www.witpress.com

SEVENTH INTERNATIONAL CONFERENCE ON DATA MINING

DATA MINING VII

Conference Chairmen

A. Zanasi
TEMIS Text Mining Solutions, Italy

C. A. Brebbia
Wessex Institute of Technology, UK

N. F. F. Ebecken
Federal University of Rio de Janeiro, Brazil

International Scientific Advisory Committee

S. Ananyan
M. W. Berry
S. Bolasco
I. Caddy
C. Curotto
B. Drewes
A. G. Evsukoff
P. Giudici
M. Gottgtroy
J. P. Lawler
B. Little
D. Malerba
P. J-S. Pan
A. Sidhu
S. Sirmakessis
D. Sitnikov

Organised by
Wessex Institute of Technology, UK
Czech Technical University, Czech Republic
Federal University of Rio de Janeiro, Brazil

Sponsored by
WIT Transactions on Information and Communication Technologies

WIT Transactions on Information and Communication Technologies

DATA MINING VII

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
Federal University of Rio de Janeiro, Brazil

WIT*PRESS* Southampton, Boston

A. Zanasi
TEMIS Text Mining Solutions, Italy

C. A. Brebbia
Wessex Institute of Technology, UK

N. F. F. Ebecken
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: 1-84564-178-7
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.

Printed in Great Britain by Cambridge Printing.

Preface

This year the Data Mining Conference is held in the center of Europe: Prague, a symbolic door between West and East, North and South Europe. Prague is also a symbol of a new, unified Europe, a Europe, we hope, of peace, security and progress.

The Conference also, this year, remains faithful to the 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 many abstracts received, with participants coming from 17 countries, with a strong representation from USA, the Americas and Europe as well as many other continents, allowing a real multinational-multicultural exchange of experiences and ideas!

Last year we witnessed the explosion of interest in data mining applications to unstructured data and we forecasted[1] that, during the following year:

1. several applications dedicated to the analysis of unstructured content would have appeared;
2. the interest in unstructured data mining/text mining of researchers, OEM and system integrators working in sectors of information retrieval, semantic web, linguistics and knowledge management would have grown.

So, we were pleased during the last year to witness the appearance of IBM UIMA (Unstructured Information Management Architecture), launched by IBM with a group of sixteen worldwide partners, among them the big players of data/text mining: Clearforest, Inxight, SAS, SPSS and TEMIS, and the appearance of SAP TREX (Text retrieval and classification), containing text mining functions.

But probably our best forecasting was done again last year, when after having introduced the possibility of fighting terrorism using data and text mining capabilities, quoting the actions of CIA (which funded In-Q-Tel, a non-profit corporation designed to fuel private research, development and application in, amongst other things, data and text mining technologies[2]), and of the French government, which launched three initiatives, endowed with 250 million Euros, to support French start-ups working in this sector[3]. We wrote:

> *We expect that soon the European Union will decide to integrate these national interests in a common, European action.*

Our expectation was written in March 2005 and printed in May. We were pleased that, in June, the first meeting of ESRAB-European Security Research Advisory Board was held, just having been nominated by the European Commission to advise

in starting the ESRP-European Security Research Program, already endowed with almost four billion Euros to fund the development of security research, with particular attention given to data and text mining!

For a discipline which is central to research and also to business, and which generates interest 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 current analysis[4], of the growth rate of the text mining market, estimates it to be more than 22%, with a market share divided between organizations such as ClearForest, Inxight and Temis and the larger ones including IBM, SAP, SAS, SPSS and Google.

A Conference such as this can only succeed as a team effort, so we 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 conference will take place in the New Forest, UK in 2007. For further information visit www.wessex.ac.uk, or contact rgreen@wessex.ac.uk

The Editors
Prague, July 2006

[1] Preface in *Data Mining 2005* – Eds. Zanasi, Brebbia, Ebecken – WIT Press

[2] John Gannon-*The Strategic Use of Open-Source Information* –in Intelligence and the National Security Strategist: Enduring Issues and Challenges-Sherman Kent Center for Intelligence Studies National War College National Defense University (2004)

[3] *Guerre économique: la France passe à l'offensive* in "Le quotidian de l'Expansion" – 11-03-2005

[4] A.Weissman, Apax Partner – An investor's Perspective on Information Retrieval – Information Intelligence Summit, Phoenix, AZ - April, 10-13, 2006

Contents

Section 6: Web mining

Section 7: Customer relationship management

Section 8: Applications in science and engineering

Section 9: Applications in business, industry and government

Section 10: Information systems, strategies and methodologies