
Introduction

Nasrullah Memon* and Uffe Kock Wiil

The Maersk Mc-Kinney Moller Institute
University of Southern Denmark, Denmark
E-mail: memon@mmmi.sdu.dk
E-mail: ukwiil@mmmi.sdu.dk
*Corresponding author

Reda Alhajj

Department of Computer Science
University of Calgary, Canada
E-mail: alhajj@ucalgary.ca

Biographical notes: Nasrullah Memon is an Associate Professor at The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Denmark. He received a Master's degree in Applied Mathematics from the University of Sindh, Pakistan and a Master's degree in Software Development from the University of Huddersfield, UK. He holds a PhD in Investigative Data Mining from Aalborg University, Denmark. He is also affiliated with Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan and the Hellenic American University, Athens, Greece. He has published more than 50 research articles in international conferences and journals. His current research is on knowledge management, mathematical methods in counterterrorism, information extraction, and investigative data mining.

Uffe Kock Wiil is a Professor of Software Engineering and Technology at The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Denmark. He holds an MSc degree in Computer Engineering (1990) and a PhD degree in Computer Science (1993) both from Aalborg University, Denmark. His research interest includes software technology, knowledge management, hypertext, computer-supported cooperative work and distributed systems. These research interests are currently being applied in three overall areas: tools and techniques for counterterrorism, healthcare and planning. He has published more than 100 research papers, which have been cited more than 1000 times.

Reda Alhajj is a Professor in the Department of Computer Science at the University of Calgary, Alberta, Canada. He has published over 300 papers in refereed international journals and conferences. He served on the programme committee of several international conferences including IEEE ICDE, IEEE ICDM, IEEE IAT and SIAM DM; and as Programme Chair of OSINT-WM, ASONAM, CASON and IEEE IRI. Associate Editor of *IEEE SMC – Part C* and a member of the editorial board of *Information Assurance and Security*, *Data Mining and Bioinformatics* and *Data Mining, Modeling and Management*. His primary work and research interests are in the areas of computational biology and bioinformatics, data mining, multiagent systems, schema integration and reengineering, social networks and XML.

Open source data is increasingly important to support intelligence tasks. The increasing ability to successfully mine data from large, incoherent series of sources allows analysts to build up detailed composite pictures of their area of interest. Open source analysis is not a substitute for traditional classified work. The analyst can use the generated open source view to ascertain what additional information might be required on a particular subject. Having identified the gaps in his/her knowledge, the analyst can use increasingly scarce resources to target those gaps with whatever tools may be appropriate. Essentially, open sources provide for the foundation for classified collection management, while also providing cover for classified sources and methods and often giving evidence as to where new classified studies might be best targeted.

This special issue includes nine high-quality research papers. Four papers were selected based on invitations to submit significantly extended and enhanced versions of the best papers from the *International Symposium on Open Source Intelligence and Web Mining (OSINT-WM 2008)* held in July 2008 in London in conjunction with the *12th International Conference on Information Visualization (IV 2008)*. The remaining five papers were selected among several submitted papers based on an open call for papers for this special issue. All papers (invited or not) have gone through a rigorous peer-review process before being accepted for this special issue.

The nine papers cover an interesting range of topics related to open source intelligence and web mining, including social network analysis, visualisation, spatial hypertext, versioning, natural language processing, investigative data mining, ontology, information retrieval, fuzzy logic, web harvesting and many more.

The first paper by Fellman 'The complexity of terrorist networks' demonstrates how complexity science can offer a number of powerful tools to examine complex terrorist information. In particular, it is shown how tools such as social network analysis, dynamic network analysis, and NK-Boolean fitness landscape modelling can help map covert, dynamic terrorist networks.

Atzenbeck *et al.* propose in their paper 'Supporting reasoning and communication for intelligence officers' the use of spatial hypertext as a tool for intelligence analysts for representing, reasoning upon, and communicating information about the emergent nature of terrorist groups. Furthermore, they evaluate awareness and versioning as important features of their prototype tool 'Socs'.

The paper 'Stalker, overcoming linguistic barriers in open source intelligence' by Neri *et al.* describes a content enabling system (Stalker) that provides deep semantic search and information access to large quantities of distributed multimedia data. Stalker provides language independent search and dynamic classification features for data collected from several open sources in a number of cultural diverse languages.

The paper by Memon *et al.*, 'Harvesting covert networks: a case study of the *iMiner* database' describes the development of a database of terrorist related data harvested from publicly available authenticated websites. The *iMiner* prototype provides novel mathematical models and algorithms for analysis and destabilisation of terrorist networks based on investigative data mining techniques.

Kianmehr and Alhajj demonstrate in their paper 'A fuzzy prediction model for calling communities' how cluster analysis can be used to identify customer calling patterns. A classification algorithm generates a classifier model for predicting the calling communities of a customer. Two different classification methods are used: support vector machine algorithm and fuzzy genetic classifier.

The paper by Mahalakshmi and Sendhilkumar 'Context based citation retrieval' presents a system that can automatically suggest references to authors which are candidates for citation in a particular context in their paper. In the context based reference retrieval system, relevant references from a pre-collected corpus are filtered and ranked after mapping concepts from the paper with a domain ontology and document ontology.

The paper by Salome and Parthasarathi 'Mining ontological knowledge using Nyaya framework' presents a method to construct automatically ontology for any domain based on the Indian Philosophical system Nyaya Sastra. The categorisation provided by Nyaya acts as a framework for extracting ontological concepts and relations (using natural language processing techniques) from documents obtained from the web.

Mahalakshmi and Sendhilkumar propose in their paper 'Comparative evaluation of ontology-based automatic reference tracking' a system for automatic reference tracking to assist scholars in identifying closely related papers in two selected domains 'operating systems' and 'computer networks'. Automatic tracking is done by recursively extracting references listed at the tail end of a paper based on a generated ontology for the domains.

The last paper by Alshalalfa *et al.* 'Discovering cancer biomarkers: from DNA to communities of genes' uses techniques from data mining and social network analysis to generate a network of genes, which is used to discover communities of genes as well as the central genes within each community. These communities of genes are used as a basis for studying and identifying cancer biomarkers (and potentially biomarkers for other diseases also).

We would like to thank all the authors for submitting papers and for their efforts in enhancing their papers for this special issue based on the reviewer comments. The Editor-in-Chief, Professor Nilmini Wickramasinghe's guidance and support are particularly appreciated. Thanks also go to the external reviewers for their valuable contributions and assistance in making this special issue published in a timely fashion.

Finally, we hope that you will enjoy reading the papers as much as we did!