

Coversheet for Pre-proposal



COVER SHEET FOR PROPOSAL TO SCIENCE FOUNDATION IRELAND

PROGRAMME NAME SFI PRINCIPAL INVESTIGATOR (PI) PROGRAMME	RESEARCH AREA BIO X ICT	INDICATIVE BUDGET DIRECT COSTS (euro) Euro 860000
AMS PREPROPOSAL ID: 33713	TITLE OF PROPOSAL (up to 30 words): PLINY - Processing Language, Images and Numbers	
<p>Signatures below confirm acceptance and agreement with the SFI grants and awards Terms and Conditions, and that the institution ensures the applicant meets eligibility requirements, and that the project is in full agreement with all legal and regulatory matters governing research in Ireland, and no aspect of this project is already being funded from another source and all details provided are correct.</p>		
<p>INSTITUTIONAL SIGNATORY AUTHORITY Name (print): Deirdre Caden Position: Research Projects Officer Email: cadend@tcd.ie Correspondence Address: TR&I, O'Reilly Institute, Trinity College, Dublin 2 Signed: <i>Deirdre Caden</i> Date: 15th January 2009</p>		<p>SIGNATURES: Signature of Lead APPLICANT: Khurshid Ahmad <i>Khurshid Ahmad</i> Date: 13 January 2009</p>

APPLICANT	
Name:	Khurshid AHMAD
Date of PhD (MM/YYYY):	March 1975
Gender:	Male
Position:	Professor of Computer Science
E-mail:	kahmad@cs.tcd.ie
Correspondence address:	Department of Computer Science, Lloyds Building, Trinity College, DUBLIN 2
Phone Number:	01 896 8429

CO-APPLICANT/ COLLABORATOR	COPY TABLE AS REQUIRED
Name:	
Date of PhD (MM/YYYY):	
Gender:	
Position:	
E-mail:	
Correspondence address:	
Phone Number:	

SECTION 1 – Required Details (up to 2 pages maximum)

<p>NAME AND CONTACT DETAILS</p> <p>Khurshid Ahmad Professor of Computer Science (1973) Department of Computer Science, Trinity College Dublin 2, IRELAND</p>
<p>CAREER PROFILE (Education and Employment)</p> <p><i>Fellowship of the British Computer Society (2004)</i> <i>Member of British Computer Society & Chartered Engineer (1987), UK Engineering Council</i> <i>PhD in Nuclear Physics (1975), Univ. Surrey, Guildford, Surrey, UK</i> <i>MSc in Nuclear Physics (1969), Univ of Karachi, Karachi, Pakistan (with Distinction)</i> <i>BSc (Hons) in Physics, 1st Class Honours, Univ of Karachi, Karachi, Pakistan</i> 2005- to date Professor of Computer Science, Trinity College Dublin 1999-2005 Professor of Artificial Intelligence, Dept. of Computing, Univ Surrey, Surrey, UK 1997-1999 Reader in Artificial Intelligence, Dept. Maths & Comp., Univ Surrey, Surrey, UK 1997-1998 Visiting Professor, Copenhagen Business School, Copenhagen, Denmark 1989-1997 Senior Lecturer, Dept. of Mathematics, Univ. Surrey, Surrey, UK 1977-1989 Group Leader, Knowledge Based Group, Comp. Centre, Univ. of Surrey 1974-1977 Research Fellow, Dept. of Physics, Univ Surrey, Guildford, Surrey, UK 1971-1974 Part time Demonstrator, Dept. of Physics, Univ. Surrey, Surrey, UK 1970-1971 Lecturer, Dept. of Physics, Univ Karachi, Pakistan</p>
<p>DETAILS OF MOST RELEVANT RESEARCH FUNDING AS LEAD/CO-APPLICANT</p> <p>REVEAL: Recovering Evidence from Video by Fusing Video Evidence Thesaurus & Video Meta-Data (2005-2008), UK EPSRC GR/S98450/01, £126,556 LIRICS: Linguistic Infrastructure for Interoperable Resources and Systems (2005-2007), EU-IST eContent 22236 £58,000 FINGRID: Financial GriID (2003-2004) UK ESRC RES-149-25-0028 £60,000 GIDA: Generic Information-based Decision Assistant (2001-2003) EU Fifth Framework (IST 2000-31123) €329,290 SOCIS: Scene of Crime Information System (1999-2002), EPSRC (GR/M/89041), £174,000 SALT: Standards Based Access to Lexica & Terminology (1999-2001), EU Fifth Framework (IST 1999-10951), €88,000 ACE: Analyst's Control Environment for Information Extraction and Prediction of Financial Markets (1996-1998), EU- ESPRIT 22271, €200,000 PLUS 13 More Projects funded by the UK EPSRC, EU-IST Programme totaling more than Euro 1 million since 1985</p>
<p>HISTORY OF MENTORING AND SUPERVISION</p> <p>I have supervised 36 PhD and 2 MPhil students to completion since 1985. In the same period 28 post-doctoral researchers had worked with me. The students and research fellows worked with me at the University of Surrey, England. Currently, I am supervising one PhD and one MSc student, and one post-doctoral fellow is working with me.</p>
<p>INNOVATION/COMMERCIALISATION ACTIVITY (e.g. invention disclosures, patents, spin-outs)</p> <p>I am of a firm view that computing is a professional subject and developments in its theoretical framework can and do come from application. A spin off company, Treocht, has been formed in Ireland for commercially exploiting forward my method for computing market sentiment and collating with other market indices. This work is being patented by Trinity College, Dublin. My work on automatic image annotation has drawn the interest of Irish Customs, Dublin Molecular Medicine Centre, and Roadstone. My work on terminology extraction from text lead to the formation of a start-up, InKe Ltd in the UK; I was its founder and technical director between 1994-1996.</p>
<p>OTHER INFORMATION AS APPROPRIATE (e.g. invited presentations....)</p>

I have established research in the areas of text based terminology and ontology extraction, automatic annotation of images, and learning in cross-modal environments. Both these projects were supported by competitive peer-reviewed project grants. The results of the research work were published in peer-reviewed journals and 36 doctoral dissertations.

I have recently organized two international workshops funded by Trinity's PRTL on Humanities: The first was on automatic annotation of fine art images (May 2008) organized in conjunction with the National Gallery of Ireland and national galleries in Sofia, Bratislava and Rome together with counterpart universities in these countries. We are developing a programme of research of an automatic annotation for curators of paintings and material collateral to the paintings – this programme is based on my work on automatic image annotation and on ontology generation. The second workshop – EMOT 2008, Emotion, Metaphor, Ontology and Terminology was held at the Language Resources and Engineering Conference (LREC) –the workshop was selected by the LREC Programme Organisers.

I have been invited to give keynote speeches at 15 international conferences. I have been on the organising committee of the following international conferences: International Joint Conference on Neural Networks, Terminology and Knowledge Engineering, European Symposium on Language for Special Purpose, International Data Engineering and Learning Conference, and EU-sponsored Language Resources and Engineering Conferences.

SECTION 2 – Publication Listing (up to three pages maximum)

FULL LIST OF PUBLICATIONS (or up to maximum space allowed)

Books

- * 2009 Khurshid Ahmad (Ed.), *Sentiment Analysis & Emotions and Metaphors: A Multi-disciplinary Perspective*, Heidelberg, Springer Verlag, 2009, 200 pp.
- 2006 Ahmad, K., Brewster, C., & Stevenson, M. (Eds.) *Words, Intelligence and Knowledge: A Festschrift for Yorick Wilks*. (2 Volumes). Berlin: Springer-Verlag.
- 2006 Ahmad, K. and Rogers, M. (Eds.) *Evidence-based LSP- Translation, Text and Terminology (Language for Special Purposes)*. Bern: Peter Lang. 535 pp.
- 1995 Kugler, M., Ahmad, K. & Thurmair, G. (Eds.). *Translator's Workbench: Tools and Terminology for Translation and Text Processing*. Berlin: Springer-Verlag.
- 1985 Ahmad, K., Corbett, G., Rogers, M. & Sussex, R. *Computers, Language Learning and Language Teaching*. Cambridge, UK: Cambridge University Press. (Reprinted 1987; 1989)

Refereed Journal Papers

- * 2008 Khurshid Ahmad, Edderkoppsspinn eller nettverk: News media and the use of polar words in emotive contexts. , *Synaps*, 21, 2008, p20 - 36
- *2008 Mesiar, Radko; Zemankova, Andrea; Ahmad, Khurshid, Level-dependent Sugeno integral, *IEEE Transactions on Fuzzy Systems*
- * 2006 Casey, M. & Ahmad, K. 'A competitive neural model of small number detection'. *Neural Networks* Vol. 19 (No. 10), pp 1475 - 1489
- 2006 Rafif Al-Sayed and Khurshid Ahmad. 'Knowledge Sharing in a Community of Practice: A Text-Based Approach in Emergent Domains'. *Electronic J. Knowledge Management*. Vol. 4 (No.2), pp 99-108.
- 2005 Taskaya-Temizel, T., Casey, M., & Ahmad, K. 'Pre-processing inputs for optimally configured time-delay neural networks.' *IEE Electronics Letters*. Vol. 41(4), pp 198-200.
- 2005 Gillam, L., Tariq, M., & Ahmad, K. Terminology and the Construction of Ontology. *Terminology* 11(1), pp55-81.
- 2005 Hippiseley, A., Cheng, D., & Ahmad, K. Head-Modifier Principle in Chinese. *Natural Language Engineering* Vol. 11 (2), pp 129-157.
- 2003 Al-Sayed, R. & Ahmad, K. Shared Languages and Shared Knowledge. *Electronic J. Knowledge Management*. Vol 1 (Issue 2), pp 1-16. (<http://www.ejkm.com/volume-4/v4-i2/v4-i2-art2.htm>)
- 2003 Ahmad, K & Musacchio, M. T. Enrico Fermi and the making of the language of nuclear physics. *Fachsprache*. Vol 25 (Nos 3-4). pp120-140.
- * 2002 Ahmad, K., Bale, T., & Casey, M. Connectionist Simulation of Quantification Skills. *Connection Science* Vol. 14 (No. 3). pp 165-201.
- 2002 Gillam, L., & K. Ahmad. Sharing the knowledge of experts. *Fachsprache*. Vol. 24 (1-2). pp 2-19.
- 2001 Ahmad, K. & Miles, L. Specialist Knowledge and its Management. *Journal of Hydroinformatics*. Vol. 24 (No.4, October 2001). pp215-230
- 2001 Ahmad, K., Vrusias, B., and Ledford, Anthony. Choosing Feature Sets for Training and Testing Self-Organising Maps: A Case Study. *Neural Computing & Applications*. Volume 10, pp 56-66.
- 2000 Ahmad, K., and Bale, T., Simulation of Quantification Abilities using a Modular Neural Network Approach. *Neural Computing & Applications*. Volume 10, pp 77-88.
- 2000 Ahmad, K. & White, D.C. 'Perl and Morphology'. *The PERL Journal – A Quarterly Magazin about Perl*. Vol5 (No.1), Spring 2000. pp 72-78.
- *1997 Abidi, S.S.R & Ahmad, K. 'Conglomerate Neural Network Architectures: The Way Ahead for Simulating Early Language Development'. *Journal of Inf. Systems Engineering*. Vol. 13 (No.2) pp 235-266.

Invited Contributions to Chapters in Books

- * 2007 Ahmad, K. Artificial Ontologies and Real Thoughts: Populating the Semantic Web?. In (Eds.) R. Basili and M.T. Paziienza Invited Talk at the *Annual Conf. of Italian Assoc. of Artificial Intelligence. AI*IA 2007, LNAI 4733*, Berlin & Heidelberg: Springer-Verlag. pp. 3–23.
- 2007 Ahmad, K. 'Being in Text and Text in Being: Notes on representative texts'. In Eds. G. Anderman and M. Rogers. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters. pp 60-94.
- 2006 Ahmad, K. 'Metaphors in the Languages of Science'. In (Eds.) V.K. Bhatia & M. Gotti. *Explorations in Specialised Genres* (Vol. 2). Bern: Peter Lang. pp 197-220.
- 2006 Ahmad, K & Musacchio, M. T). 'Citation Patterns in Nobel Prize Lectures in Economics'. In (Eds.) V.K. Bhatia & M. Gotti. *Explorations in Specialised Genres*. (Vol. 1) Bern: Peter Lang. pp 101-113
- 2006 Ahmad, K., D. Cheng, T. Taskaya, S. Ahmad, L. Gillam, P. Manomaisupat, H. Traboulsi and A. Hippiseley. 'The mood of the (financial) markets: In a corpus of words and of pictures.' In (Eds). A. Wilson, D. Archer and P. Rayson. *Corpus linguistics around the world*. Amsterdam/New York: Rodopi, pp 18-32.

- 2005 Ahmad, Khurshid., and Al-Sayed, Rafif. Community of Practice and the Special Language 'Ground'. In (Eds.) Clarke, S and Coakes, E. *Encyclopaedia of Knowledge Management and Community of Practice*. Hershey (PA): The Idea Group Reference. pp 77-88. (ISBN 1-59140-556-4).
- 2002 Ahmad, Khurshid 'Writing Linguistics: When I use a word it means what I choose it to mean'. In (Eds) Manfred Klenner and Henriëtte Visser. *Computational Linguistics for the New Millennium: Divergence or Synergy?* Bern: Publishing Group Peter Lang, pp 15-38.
- 2002 Ahmad, Khurshid 'Come nascono I termini: zeri, comportamenti e nuclei'. In (Eds). M.Magris, M T Musacchio, L Rega & F Scarpa. *Manuale di Terminologia: Aspetti Teorici, Metodologici e Applicativi*. Milano: Ulrico Hopeli Editore S. p. A pp63-81 (Zeroes, behavio(u)rs and nuclei: terms and their emergence). (ISBN 88-203-2943-3).
- 2001 Ahmad, Khurshid., and Rogers, Margaret A. (2001). 'Corpus Linguistics and Terminology Extraction'. In (Eds.) Sue-Ellen Wright and Gerhard Budin. *Handbook of Terminology Management (Volume 2)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. pp 725-760.
- 2001 Ahmad, Khurshid. (2001). 'The Role of Specialist Terminology in Artificial Intelligence and Knowledge Acquisition'. In (Eds.) Sue-Ellen Wright and Gerhard Budin. *Handbook of Terminology Management (Volume 2)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. pp 809-844.
- Refereed Conference Papers (Full Papers Refereed)**
- 2008 C. Zheng, D. Kelleher, K. Ahmad, A Semi-Automatic Indexing System for Cell Images, *2008 IEEE World Cong. on Computational Intelligence, Hong Knog, CHINA, June 1-6, 2008*,
- 2008 K. Ahmad, Emotion, Metaphor, Ontology & Terminology, Workshop , 27 May 2008, In: *2008 Language Resources and Evaluation Conf., 2008, Marrakesh, MOROCCO*. ELDA, Paris, 1-108
- 2008 A. Devitt and K. Ahmad, Sentiment Analysis and the Use of Extrinsic Datasets in Evaluation, *Language Resources and Evaluation Conference (LREC 2008), Marrakech, Morocco, 28-30 May, 2008*
- 2008 Ahmad, K. The 'return' and 'volatility' of sentiments: An attempt to quantify the behaviour of the markets?. In (Ed.) K. Ahmad. *Proc. of EMOT 2008: Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology. Workshop at the 13th Language Resources and Evaluation Conf., 27 May 2008. Marrakesh, Morocco.*
- 2007 Devitt, Ann., and Ahmad, Khurshid. 'Sentiment Polarity Identification in Financial News: A Cohesion-based Approach'. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. June 23-30, 2007, Prague, Czech Republic*. Stroudsburg, PA: Association for Computational Linguistics (ACL) pp 984-991. (ISBN 978-1-932432-88-6).
- 2007 Zheng, C., Ahmad, K., Long, A., Volkov, Y., Davies, A. and Kelleher, D. (2007) Hierarchical SOMs: segmentation of cell migration images. *Proceeding of 2007 International Symposium on Neural Networks* (July 2007, Nanjing, China) (Lecture Notes on Computer Science, Heidelberg: Springer Verlag.)
- 2006 Y. Almas & K. Ahmad. 'LoLo: A System based on Terminology for Multilingual Information Extraction'. In (Eds.) M E Califf, M A. Greenwood, M Stevenson & R Yangraber. *COLING ACL 2006: Workshop on Information Extraction beyond the Document, Sydney, Australia, 22 July 2006*. Association of Computational Linguistics. pp56 – 65.
- 2006 P. Manomaisupat, B. Vrusias & K. Ahmad (2006) 'Categorization of Large Text Collections: Feature selection for unsupervised and supervised neural networks.' In (Eds.) E. Corchado, H. Yin, V. Botti & C. Fyfe. *Proc. 7th Int. Data Engineering and Automated Learning Conf. (Lecture Notes on Computer Science - LNCS 4224), Burgos, Spain, 20th-23rd September*. Heidelberg: Springer-Verlag, pp1003 - 1013
- 2006 A. Nyamapfene & K. Ahmad. Unsupervised Multi-net Simulation: An Application to Child Development. , In *Proc. IJCNN 06: Int. Joint Conference on Neural Networks 2006 , Vancouver, CANADA, 16-22 July 2006*, Int. Neural Network Society & IEEE (Washington). pp2406 - 2411
- 2006 A. Popoola & K. Ahmad Testing the Suitability of Wavelet Pre-processing for TSK Fuzzy Models. . In *Proc. FUZZ-IEEE'06: Int. Conference on Fuzzy Systems Networks, Vancouver, CANADA, 16-22 July 2006*, IEEE Computational Intelligence Society, 2006, pp6655 - 6659
- 2006 K. Ahmad, D. Cheng and Y. Almas, 'Multi-lingual Sentiment Analysis of Financial News Streams.' In (Eds.) Stefano Cozzini, Stefano d'Addona and Rosario Mantegna. *Proc. 1st Int. Conf. on Grid in Finance (Palermo, February 2006)* (http://pos.sissa.it/archive/conferences/026/001/GRID2006_001.pdf)
- 2006 L. Gillam & K. Ahmad. 'Financial data tombs and nurseries: A grid-based text and ontological analysis.' In (Eds.) Stefano Cozzini, Stefano d'Addona and Rosario Mantegna *Proc. 1st Int. Conf. on Grid in Finance (Palermo, February 2006)* (http://pos.sissa.it/archive/conferences/026/002/GRID2006_002.pdf).
- 2006 A. Popoola, and K. Ahmad, "Fuzzy Models for Time Series Analysis: Towards Systematic Data Pre-processing ," *Proc. of the IEEE Int. Conf. on Engineering of Intelligent Systems (ICEIS 2006)*, Islamabad, Pakistan, April 2006
- 2006 A. Popoola, S. Ahmad, and K. Ahmad, "Multiscale Wavelet Preprocessing for Fuzzy Systems," *Proc. of the ICSC Congress on Comp. Intelligence Methods and Applications, Istanbul, Turkey, December 2005.*
- 2005 Manomaisupat, P. & Ahmad, K. Feature Selection For Text Categorisation Using Self-organising Map,. In (Eds.) M. Zhao & Z. Shi. *International Conference on Neural Networks and Brain, 2005. ICNN&B '05 13-15 Oct. 2005- Beijing, China*. Picatsway: IEEE Press. Volume 2, pp 1875 - 1880
- 2005 A. Popoola, S. Ahmad, and K. Ahmad, "A Fuzzy-Wavelet Method for Analyzing Non-Stationary Time Series," *Proc. of the 5th International Conference on Recent Advances in Soft Computing, Nottingham, UK, December 2004*, pp. 231-236. (ISBN: 1-84233-110-8)
- 2005 Al-Sayed, R and Ahmad, K. (2005). 'Knowledge Sharing in a Community of Practice: A Text-Based Approach in Emergent Domains'. In (Ed) Dan Remenyi, *Proceedings of the 2nd Int. Conf. on Intellectual Capital and Knowledge Management*, American University of Dubai, UAE, 21-22 November, pp 30-41.
- 2005 K. Ahmad and L. Gillam "Automatic Ontology Extraction from Unstructured Texts". In (Eds.) R. Meersman and

- Z. Tari. On the Move to Meaningful Internet Systems - OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005. Proceedings, Part II. Springer-Verlag Berlin Heidelberg. pp. 1330 – 1346.
- 2005 K. Ahmad & L. Gillam and D.Cheng. 'Society Grids'. In (Eds.) Simon Cox and David Walker. *Proceedings of the UK e-Science All Hands Meeting 2005. 18-21 September, Nottingham UK. Swindon: EPSRC Sept 2005.* pp 923-930. (ISBN 1-904425-53-4)
- 2005 K. Ahmad & L. Gillam and D.Cheng. 'Textual and Quantitative Analysis: Towards a new, e-mediated Social Science'. Proc. of the 1st Int. Conf. on e-Social Science (Manchester, July 2005).
- 2005 L. Gillam, K. Ahmad, G. Dear. 'Grid-enabling Social Scientists: The FINGRID infrastructure.' *Proc. of the 1st Int. Conf. on e-Social Science* (Manchester, July 2005).
- 2005 K. Ahmad, B. Vrusias, and M. Zhu. 'Visualising an Image Collection?' In (Eds.) Ebad Banisi et al. *Proceedings of the 9th International Conference Information Visualisation (London 6-8 July 2005).* Los Alamitos: IEEE Computer Society Press. pp 268-274. (ISBN 0-7695-2397-8).
- 2005 K. Ahmad and Yousif Almas. 'Visualising Sentiments in Financial texts'. In (Eds.) Ebad Banisi et al. *Proceedings of the 9th International Conference Information Visualisation (London 6-8 July 2005).* Los Alamitos: IEEE Computer Society Press. pp 268-274. (ISBN 0-7695-2397-8).
- 2005 Gillam, L., & Ahmad, K. 'Pattern Mining across Domain-specific Text Collections'. In (Eds.) P. Perner & A. Imiya. *International Conference on Machine Learning and Data Mining MLDM '2005.* (Lecture Notes on Artificial Intelligence). Berlin:Springer-Verlag. pp 570-579.
- 2005 Saragiotis, P. B. Vrusias, K. Ahmad. Learning to Classify a Collection of Images and Texts. In *European Symposium on Art. Neural Networks (Bruges, Belgium, 27-29 April 2005)*
- 2005 J. Mountstephens, C. Bennet, and K. Ahmad. 'Selective Visual Attention in Electronic Video Surveillance'. In (Eds.) H. Gamboa & A. Fred. *Proc. Of the 5th International Conference on Pattern Recognition in Information Systems (Miami, USA, May 2005).* Portugal: INSTICC Press pp 198-203.
- 2005 Andrew Salway, Andrew Vassiliou and Khurshid Ahmad (2005), 'What Happens in Films?', IEEE Conference on Multimedia and Expo, ICME 2005.
- 2004 Ahmad, K. & Ajala E., 'Acquiring and Filtering Knowledge: Discovery and Case-based Reasoning'. In (Ed.) *IASTED Conf. on Knowledge Sharing and Collaboration (KSCE 2004).* Pp 309-318.
- 2004 Ahmad, K., T. Taskaya-Temizel, D. Cheng, L. Gillam, S. Ahmad, H. Traboulsi and J.Nankervis. 'Financial Information Grid –an ESRC e-Social Science Pilot'. *Proceedings of the UK e-Science All Hands Meeting 2004. 31st August - 3rd September, Nottingham UK. Swindon: EPSRC.* (ISBN 1-904425-21-6)
- 2004 Ahmad, K. & Vrusias, B. 'Learning to Visualise High-Dimensional Data'. In (Eds.) E. Banissi et al. *Proc. of 8th Int. Conf. on Information Visualisation* (London, England, 14-16 July 2004). Los Alamitos: IEEE Computer Press. pp 507-512. (ISBN 0-7695-2177-0).
- 2004 Ahmad, S., Taskaya Temizel, T., and Ahmad, K. "Summarizing Time Series: Learning Patterns in 'Volatile' Series." Z.R. Yang, R. Everson, and H. Yin (Eds.), *Proc. of 5th Int. Conf. on Intelligent Data Engineering and Automated Learning* (Exeter, UK, 25-27 August 2004) (LNCS Vol. 3177). Heidelberg: Springer Verlag. 'pp 523-532.
- 2004 Casey, M. & Ahmad, K. "In-Situ Learning in Multi-net Systems". In (Eds.) Z.R. Yang, R. Everson, and H. Yin. *Proc. of 5th Int. Conf. on Int. Data Engineering and Automated Learning* (Exeter, UK, 25-27 August 2004) (LNCS Vol. 3177). Heidelberg: Springer Verlag. 'pp 752-757.
- 2003 Ahmad, K. *Notes on Learning to Compute and Computing to Learn.* (Invited Keynote Speech). International Conference on Machine Learning and Applications (ICMLA'03) Los Angeles, California, June 23-24, 2003. pp 107-117.
- 2003 Al-Sayed, R., & Ahmad, K. Shared knowledge: the role of special language. In (Eds.) McGrath, F. and Remenyi, D. Proc. of the 4th Euro Conf. on Knowledge Management.. Oxford, UK. (2003). pp 28-40.
- 2003 Ajala, E and Ahmad, K. Managing Knowledge: Nanotubes – A Case Study. In (Ed.) W. Chu. *Proc. 2nd IASTED Int. Conf. On Information and Knowledge Sharing.* Anaheim: ACTA Press. pp 39-44
- 2003 Ahmad, K., & Al-Thubaity, Abdul Mohsen. 'Can Text Analysis Tell us Something about Technology Progress'. In (Ed) M. Iwayama & A Fujii. *Proc. of the Workshop on Patent Corpus Processing.* East Stroudsburg, PA: The Ass. of Computational Linguistics. pp 46-55
- 2003 Al-Thubaity, Abdul Mohsen & Ahmad, K. (2003). 'Knowledge Maps as Lexical Signatures of Journal Papers and Patent Documents'. In (Eds.) Ebad Banissi et al. Proc. of 7th International Conference on Information Visualisation (London, England, 16-18 July 2003). Los Alamitos: IEEE Computer Press. pp 582-588. (ISBN 0-7695-1988-1).
- 2003 Taskaya, T., & Ahmad, K. (2003). 'Bimodal Visualisation: A Financial Trading Case Study'. In (Eds.) E. Banissi et al. *Proc. of 7th International Conference on Information Visualisation* (London, England, 16-18 July 2003). Los Alamitos: IEEE Computer Press. pp 320-326. (ISBN 0-7695-1988-1).

Project PLINY: Background and Significance

The information systems of the next decade and beyond will have to deal with and exploit information that is available for the same object or event in different modalities. What we now have are systems that can unimodally process information without reference to the interaction between the modalities. It appears that in human information processing this interaction plays a key role. Human decision making, either rule-based or learnt behaviour, involves the adoption of a utilitarian approach to a host of competing, collaborative and sometimes contradictory inputs. In information retrieval keywords are used to search for images and, in due course, images will be used to search for text documents; there are instances – in heritage management, in clinical diagnosis, in financial trading, and in crime prevention – where experts use both cues judiciously. Experts appear to aggregate this cross-modal information to optimise their search or make a decision. This ability of exercising judicious choice of cues of different physical origins, and of differing information content, is an emergent property of the animal brain; neurons in the visual areas interact with neurons in the language processing areas of the brain. Visual neurons have some capability of processing speech data and vice versa. And, there are bi-modal neurons that only fire when stimulated by two modalities. Multi-criteria decision making (MCDA), it appears, has a neural correlate.

In the rapidly developing field of behavioural economics similar cross-modal interaction appears to be one of the factors underscoring market volatility: the impact of news and rumours (expressed verbally or in writing) on prices and volumes of prices of shares, commodities and other financial instruments (expressed numerically) is an area of investigation that has been made more intense since the credit crunch. The ability to process verbal information (news, rumours) and then co-integrate with numerical information (prices, volumes) has its neural correlates: the development of numerosity from neonatal stages onwards relies on the cross-modal interaction of neurons in the spatial awareness areas and in the language processing areas. These inputs are sometimes vaguely defined and this situation is exacerbated due to uncertainty and imprecision.

Research aim, objectives, and hypotheses

It can be hypothesized that different criteria are applied to select one or more features of each of the interacting modalities – sometimes the features can be aggregated to achieve *super-addition*, such that the whole is greater than the sum of the individual features, and at other times some features can be relegated in importance such that the whole is less than the sum leading to *sub-addition*. Yet, sometimes a simple addition of the modalities suffices. The well-known cocktail party effect relies on the super-addition of low-level linguistic information with the visual information of facial changes that enables listeners to ‘listen’ in noisy environments. The collapse of enterprises and markets on rumours, despite encouraging quantitative information about the performance of their assets, is the sub-addition of linguistic information with numerical.

The aims of the project comprise the investigation of two generic and two specific questions: The generic questions I wish to ask are (i) can the cross-modal inputs be aggregated, if so how? (ii) how does the facility of aggregation evolve and what role does learning play in it. The objective is to explore, model and simulate *ab initio* learning of (a) how visual and linguistic information can be aggregated for each feature characterising each modality and then for both modalities, and (b) time-varying numerical and linguistic information can be aggregated individually and collectively. The project will seek to apply the findings in two main areas: First, in the critical area of large volume microscopic image annotation where an information system will use linguistic data collateral to an image with the visual features of an image; and second, exploring how market sentiments, qualitative information expressed verbally or through facial expressions can be aggregated with the quantitative information available in market prices.

Methodology

The proposed research question: how can a corpus of images and a corpus of time serial data, be constructed such that these corpora are as accessible to retrieval, browsing and summarization as text corpora are now? The intention is to create linguistic evidence data, essentially terminology of a specialist domain organised within an ontological framework, for annotating visual content or annotating time serial content. The aggregation techniques to be used include some founded on a new interpretation of Pareto’s utility theory by fuzzy logicians and fuzzy set theorists. We will invite senior research fellows in the areas of neural information processing, fuzzy logic, image interpretation and annotation, and behavioural finance, to work alongside PLINY research assistants and students.

There are four work-packages in the project PLINY:

- **Ontology and Terminology:** To develop models and algorithms for capturing the ontological and terminological information underpinning the work of imaging experts and that of finance experts involved in interpreting corpora of images or time series. Techniques to be used include corpus-based terminology and the use of terminological logics for generating ontologies (**15 person months**).
- **Aggregation:** To identify, refine and deploy algorithms for creating a preference structure for the features involved used in representing visual and graphic content and features representing the collateral linguistic

evidence for both. Algorithms to be studied and deployed include those used in the fuzzy integral formulation, and in fuzzy set formulation of MCDA(**15 person months**).

- Cross-modal Interactions: To develop methods for integrating the linguistic evidence (terms and conceptual structures) with visual content for images and numerical content for time serial data. This will be used to facilitate the retrieval of images and graphs from high-level queries including resemblance and affect based queries respectively. Algorithms to be tested include hierarchical self-organising maps for the linguistic evidence data and image (and numerical) content data, and a multi-net formulation for creating a well-grounded framework for interaction between modalities.(**45 person months**)
- Demonstration and evaluation: To build a prototype system that can be trained using corpora of images for medical applications, and corpora of time series for financial applications, and collateral texts. The evaluation will be carried out by external partners including Irish Customs, St James Hospital and Fairview Analytics for image retrieval. And, Irish Stock Exchange, The Irish Times and Treocht Ltd for selection of financial assets based on sentiment analysis. (**45 person months**)

Value to Ireland

The onset of the semantic web will require well-trained personnel in the areas of information retrieval from multi-modal corpora of texts, images, sounds, and time varying data emerging from interactions in the market, from laboratories and many other places yet to be identified. Financial transaction systems and health care systems will be the main beneficiaries here and there will be innovations. The R&D and training initiatives within Ireland are largely on uni-modal systems.

The PLINY project will be a fore-runner for a well grounded framework for building, deploying and testing such cross-modal systems where end-users work alongside innovators in ICT, neurosciences, critical care, and fuzzy logic and fuzzy set theory. We already have collaborations with universities in the EU and the USA. This project is a result of a large scale feasibility study (June 2006-Dec 2008) that has involved Trinity College's School of Business, School of Medicine/St James Hospital, and the School of Computer Science. We have a PhD programme in these areas and a number of peer-reviewed research papers have been published. The prototype systems for image annotation system are being used to annotate images for Irish Customs and the National Gallery of Ireland. The sentiment analysis system has been developed in conjunction with the Irish Stock Exchange and there are organisations that are negotiating licensing arrangements for the technology developed thus far.