Frontiers in Artificial Intelligence and Applications

ROB MILNE:

A Tribute to a Pioneering AI Scientist, Entrepreneur and Mountaineer

Edited by
Alan Bundy
Sean Wilson

IOS Press

ROB MILNE: A TRIBUTE TO A PIONEERING AI SCIENTIST, ENTREPRENEUR AND MOUNTAINEER

Frontiers in Artificial Intelligence and Applications

FAIA covers all aspects of theoretical and applied artificial intelligence research in the form of monographs, doctoral dissertations, textbooks, handbooks and proceedings volumes. The FAIA series contains several sub-series, including "Information Modelling and Knowledge Bases" and "Knowledge-Based Intelligent Engineering Systems". It also includes the biannual ECAI, the European Conference on Artificial Intelligence, proceedings volumes, and other ECCAI – the European Coordinating Committee on Artificial Intelligence – sponsored publications. An editorial panel of internationally well-known scholars is appointed to provide a high quality selection.

Volume 139

Recently published in this series

Vol. 138. Y. Li et al. (Eds.), Advances in Intelligent IT – Active Media Technology 2006
Vol. 137. P. Hassanaly et al. (Eds.), Cooperative Systems Design – Seamless Integration of Artifacts and Conversations – Enhanced Concepts of Infrastructure for Communication
Vol. 136. Y. Kiyoki et al. (Eds.), Information Modelling and Knowledge Bases XVII
Vol. 135. H. Czap et al. (Eds.), Self-Organization and Autonomic Informatics (I)
Vol. 134. M.-F. Moens and P. Spyns (Eds.), Legal Knowledge and Information Systems – JURIX 2005: The Eighteenth Annual Conference
Vol. 133. C.-K. Looi et al. (Eds.), Towards Sustainable and Scalable Educational Innovations Informed by the Learning Sciences – Sharing Good Practices of Research, Experimentation and Innovation
Vol. 132. K. Nakamatsu and J.M. Abe (Eds.), Advances in Logic Based Intelligent Systems – Selected Papers of LAPTEC 2005
Vol. 131. B. López et al. (Eds.), Artificial Intelligence Research and Development
Vol. 130. K. Zieliński and T. Szmuc (Eds.), Software Engineering: Evolution and Emerging Technologies
Vol. 129. H. Fujita and M. Mejri (Eds.), New Trends in Software Methodologies, Tools and Techniques – Proceedings of the fourth SoMeT_W05
Vol. 128. J. Zhou et al. (Eds.), Applied Public Key Infrastructure – 4th International Workshop: IWAP 2005
Vol. 127. P. Ritrovato et al. (Eds.), Towards the Learning Grid – Advances in Human Learning Services

ISSN 0922-6389

Rob Milne: A Tribute to a Pioneering AI Scientist, Entrepreneur and Mountaineer

Edited by

Alan Bundy
School of Informatics, University of Edinburgh, UK

Assistant Editor

Sean Wilson
School of Informatics, University of Edinburgh, UK

IOS Press

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

ISBN 1-58603-639-4
Library of Congress Control Number: 2006929170

Publisher
IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

Distributor in the UK and Ireland
Gazelle Books Services Ltd.
White Cross Mills
Hightown
Lancaster LA1 4XS
United Kingdom
fax: +44 1524 63232
e-mail: sales@gazellebooks.co.uk

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

PRINTED IN THE NETHERLANDS

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Foreword

This festschrift celebrates the life of a remarkable man.

Rob Milne died while climbing Mount Everest early on 5th June 2005 Nepal Time. He was 48. He is survived by his wife Val and his two children Alex and Rosemary.

His untimely death was a tragedy, but Rob packed 96 years of living into his 48 years of life. In any one of his three "careers" — as a hi-tech entrepreneur, as an AI researcher and as a mountaineer — his achievements would have been enough for most ordinary mortals. But Rob combined world-class success in all of them. This book covers all these facets of his life. Each chapter has been contributed by one or more of his close collaborators as their tribute to Rob and to his legacy.

Rob's ascent of Everest was to have been the culmination of a lifetime's ambition to climb the highest summits in each of the world's seven continents. Everest was the last of these seven summits. He was only 400 metres from the top when he died from a sudden and massive heart attack. He had been an ambitious and successful mountaineer since his childhood in Colorado. As Val, said in a radio interview, "Rob died at the top, doing what he loved". This was true not just of his mountaineering, but in all the spheres of his life.

I first met Rob in 1978 in Boston, Massachusetts. He was just finishing his undergraduate degree at MIT and had applied to be a PhD student at Edinburgh under my supervision. I was visiting MIT that Summer, so we met to discuss his research project. I was quickly introduced to his climbing expertise. He showed me how to climb a vertical brick wall using the gaps in the bricks as hand and foot holds. He invited *me* to try; I declined. He came to Edinburgh that Autumn to work on machine understanding of mechanics problems written in English as part of our Mecho project: one of the first non-US expert systems. In 1980, Rob initiated his project to conquer the seven summits by climbing Denali (Mt McKinley) in Alaska. I remember getting vertigo just by reading his subsequent article in our University Bulletin. Rob met Val at Edinburgh, and they married in 1981.

Climbing Denali required determination and persistence. Rob exhibited these qualities in everything he did, which is why he achieved so much. But it wasn't always the best approach. When it came to his PhD viva, Rob somehow got the misconception that to concede to his examiners on any point, however minor, would destroy his chances. He, therefore, fought every step of the way. The viva lasted eight hours! He obtained his PhD in 1983.

In 1982, Rob returned to the USA, where he worked first at the Wright-Paterson Airforce Base and then at the Pentagon. At the Pentagon, he introduced AI research into the US Army by founding the Army AI Center, in which he was the Chief AI scientist. Returning to Scotland in 1986, he founded Intelligent Applications: one of the first non-US expert system companies. After experimenting with various AI products, IA focused on turbine monitoring with its 'Tiger' product.

Most entrepreneurs running innovative, hi-tech companies have little time for extra-curricular activity, but Rob found time both for his mountaineering and his AI research.

He continued to publish in the top journals and conferences, authoring over 75 papers on knowledge-based systems, data-mining, qualitative reasoning, etc. He was a popular speaker, giving many invited talks and tutorials at major conferences. He acted as a bridge between academia and industry, for instance, frequently talking to academics on the technology transfer process.

Rob was a natural leader; he tirelessly and selflessly gave his time to help organise the activities in which he was involved. For instance, in both the British and European AI communities, he regularly served on conference committees, having been chair of both the British Computer Society's Specialist Group on AI Conference and of the European Conference on Artificial Intelligence. He was an officer of both organisations, including being the President of ECCAI 2000-04. He was also the inspiration behind bringing IJCAI-05 to Edinburgh, being the Local Arrangements Chair until his death. Rob played a key part in setting up the European Network of Excellence MONET (Model Based and Qualitative Systems), and in a second phase its Task Group BRIDGE (Bridging AI and Control Engineering model based diagnosis approaches) that focused on diagnosis.

Rob played an active role in the Scottish Software Federation (which merged to form ScotlandIS in 2000): the industry bodies for IT and software companies in Scotland. Rob was a director of each organisation 1997–2002. He was a mentor to a number of start-up companies and guided other entrepreneurs in their efforts to establish successful businesses. In recognition of his academic and industrial achievements, he was elected a Fellow of the Royal Society of Edinburgh in 2003. He was also active in Scottish Mountaineering Club, being Convener of the publications sub-committee and co-authoring a book on the Corbetts (the 219 Scottish hills between 2500ft and 3000ft high). In 1997, Rob became only the 1860th person to have climbed all the Munros (the 284 Scottish mountains over 3000ft high). He was a keen winter climber, helping to establish a number of high-grade new climbs throughout Scotland.

Rob summed up his attitude to life in a radio interview, by saying that it was important to wake up every morning with an exciting challenge in mind. He always set himself ambitious goals then attained them by persistence and determination. Ambition, persistence and determination are qualities sometimes associated with people who are difficult to get on with. Not so Rob. He was one of the most pleasant and easy-going people it has been my pleasure to work with. We already miss him.

Alan Bundy
May 25, 2006

Contents

This page intentionally left blank

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Model Based Thoughts

CA~EN and TIGER then and now

Louise TRAVE-MASSUYES[1]
LAAS–CNRS, France

Abstract. This paper is in memory of Rob Milne. It recalls the scientific experience of the TIGER and TIGER SHEBA European projects that he led through the presentation of CA~EN, a model-based diagnosis software that was devised during the projects and integrated with the commercial tool TIGER™. Several lines of more recent work that all build on CA~EN are also presented. These developments are closely related to the Qualitative Reasoning and Model Based Diagnosis communities, represented by the MONET network of Excellence in Europe. Their theories inspired this work and I like to believe that this work inspired the communities in return. Rob Milne and myself were active members of MONET.

Keywords. Model-based diagnosis, causal models, interval models, industrial applications

Introduction

I met Rob Milne for the first time in 1988 during the *12th IMACS World Congress on Scientific Computation* in Paris, France, where I had to present a paper in the Second Generation Expert Systems session. It was the time of the *International Workshop on Expert Systems and their Applications* series and I got to meet Rob again in Avignon, France. Avignon surroundings being the French paradise for limestone climbers, I discovered that Rob and I not only shared a common passion for new technologies but also for climbing and mountaineering. However I could not imagine at that time that this was the start of a partnership that would stongly mark my life both professionally and personally. This paper recalls the professional part of it[2].

The real kick-off event of the TIGER and TIGER-SHEBA projects was a phone call from Scotland to Brazil in 1991. I was on sabbatical at Vitoria University and this little university offered restricted facilities. In particular, I had no phone in my room, not even in my building. So, I scrambled down the stairs of my building two per two as

---

[1] LAAS-CNRS, 7 avenue du Colonel Roche, 31077 Toulouse, France; Email: louise@laas.fr.

[2] I would like to flash some of the climbs that we partnered in the most mythical mountains all around the world: North Face of Half Dome, Yosemite, USA, 1993; Major route at Mont Blanc, Alps, France, 1993; Spigolo of Petite Aigille d'Ansabère in the Pyrénées, France, 1994; several climbs in Ben Nevis, Cairn Gorns, Scotland; and Aconcagua (6962m), Argentina, 2003.

soon as one of my brazilian colleagues reported that the phone was ringing for me in the other building. This phone call was followed by eight years of fruitful collaboration. The team of Intelligent Applications Ltd was very dedicated and very efficient in making the bridge between researchers and end-users. At the lead of this team, Rob was an outstanding manager: he knew how to keep control of the most delicate situations, and had the art to make complex decisions rapidly. Above all, Rob was very focused and he knew exactly what he wanted. However, his management style was to listen first.

In the context of the two TIGER projects, Rob and I stretched the collaboration to a higher level by co-authoring 6 journal papers and 14 international conference papers, giving 5 joint tutorials on *Applied Qualitative Reasoning* in major international conferences (*IJCAI'93*, Chambéry (France), August 1993; *AAAI'93*, Washington DC (U.S.A), July 1993; *AI for Applications*, Orlando (U.S.A), March 1993; *Expert Systems Conference*, Cambridge (U.K), December 1992; *ECAI'92*, Vienna (Austria), August1992) and two seminars addressing industry (*UNICOM*, London (UK), November 1995; *SITEF*, October 1995, Toulouse (France)).

It was a time of intense activity for the Qualitative Reasoning (QR) and the Model Based Diagnosis (known as DX) communities. Rob and I were active members of these communities and played a key part in setting up the European Network of Excellence MONET (Model Based and Qualitative Systems), and in a second phase its Task Group BRIDGE that focused on diagnosis.

The developments of the two TIGER projects are closely related to QR and DX communities. Their theories inspired this work and I like to believe that this work inspired the communities in return. In this paper, I do recall the scientific experience of the TIGER and TIGER SHEBA European projects through the presentation of CA~EN, a model-based diagnosis software that was devised during the projects and integrated with the commercial tool TIGER™. More importantly, I subsequently present several lines of more recent work that all build on CA~EN, showing that the work that Rob and I initiated is still very much alive.

1. The TIGER and TIGER SHEBA European Projects

The TIGER and TIGER SHEBA European Projects were both leaded by Intelligent Applications Ltd and managed by Rob Milne. The consortium included end users[3] and a research team leaded by myself composed of LAAS-CNRS, Toulouse (France) and UPC, Terrassa (Spain). TIGER, full name "Real-Time Assessement of Dynamic, Hard to Measure Systems", ran within the European ESPRIT program from 1992 to 1995 (Project n°6862). This project set the bases of the TIGER™ tool, commercialized by IA Ltd.

TIGER-SHEBA, full name "TIGER with Model Based Diagnosis", succeeded to TIGER in the ESPRIT TRIAL Applications European program starting in 1998 and ending in 2000 (Project n°27548). TIGER SHEBA enhanced TIGER™ by integrating the model-based diagnosis system CA~EN.

[3] Exxon Chemical (UK), John Brown Engineering (UK) and Dassault Aviation (France) in TIGER, and Kvaerner Energy Ltd (UK) and National Power Cogen (UK) in TIGER SHEBA.

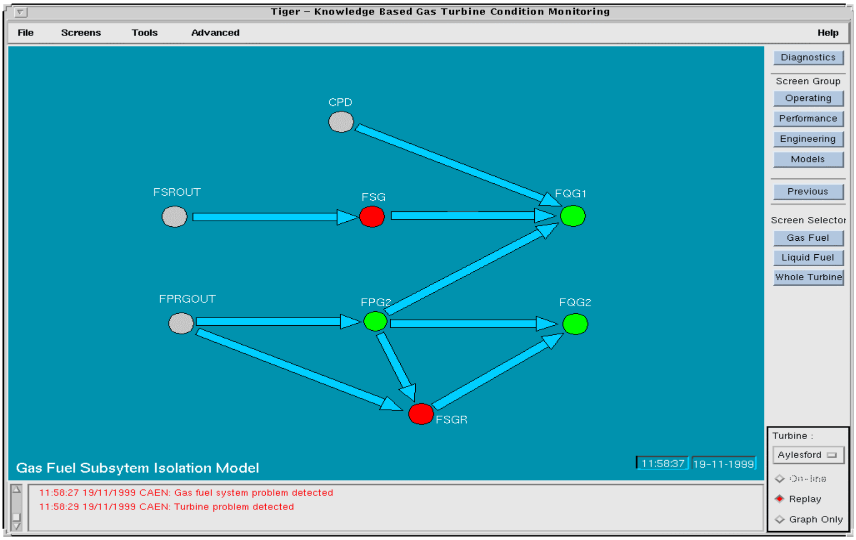
The design of TIGER™, the CA~EN software and its integration in TIGER-SHEBA have led to an innovative product for gas turbine monitoring and diagnosis that has been commercialized all over the world, and to the dissemination of scientific ideas in recognized scientific journals and conferences, with a Best Application Paper Award in the *Expert Systems Conference ES'95*, Cambridge, UK in 1995 [27]. TIGER-SHEBA was recognized in 2001 with an award from the European Commission among the 25 best European ESPRIT projects and received the Research & Technology Duran Farell Award from Universitat Politecnica de Catalunya sponsored by the *Gas Natural* Coorporation in 2002 [37].

The first prototype of the TIGER system was built with Exxon Chemical in Scotland. This then led to the TIGER project to develop the initial TIGER prototype [26][27][28][45]. The objective of the TIGER™ system was to provide users of gas turbines with an advanced knowledge based condition monitoring system, capable of detecting and diagnosing faults experienced during operation. This was achieved by continuous on-line monitoring: reasoning, interpreting and predicting expected behavior and faults, utilizing its internal knowledge.

The core of TIGER is a temporal rule-based system but there is an important class of faults to detect and isolate which require a model-based approach, that is, the predicted value for a measurement depends on a complex combination of several inputs and reasoning on the basis of the model is necessary to isolate faults. One area where this is particularly important is predicting the expected power output of a gas turbine, giving a combination of atmospheric air conditions, fuel setting, and steam injected for pollution control.

The focus of the development performed within TIGER SHEBA was to use the CA~EN software including a qualitative interval predictor based fault detection and a fault isolation based on an abstraction of the models in terms of a timed causal structure [43][44]. This combines model-based fault detection ability by predicting what expected outputs should be with a model-based fault isolation mechanism providing this additional leverage over rule-based implementations [43][44]. Qualitative reasoning techniques are used to generate the expected value of key measurements, such as a fuel valve position, and to abstract the model in a suitable form for reasoning about the relations between variables so the tracing of faults is made possible.

The CA~EN system was used on critical sub-systems of several gas turbines such as the gas and liquid fuel systems.



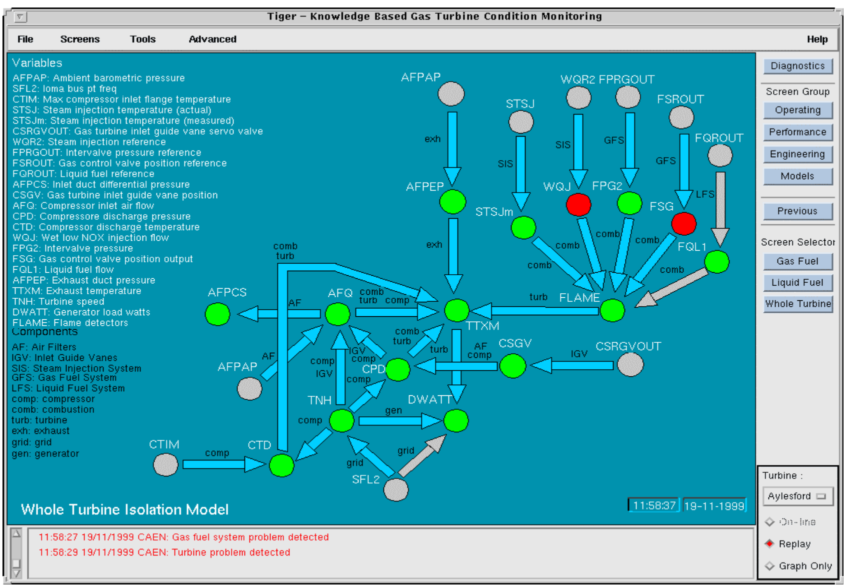


Figure 1 and 2. Turbine incidents reported in red by TIGER-SHEBA: CA~EN Activation Graph of (1) the Gas turbine Fuel System (2) the whole turbine.

The combination of TIGER with CA~EN provided a much more sensitive fault detection capability for the gas fuel and other sub-systems. In addition, its built-in fault isolation mechanism was able to follow the physical operation of the gas fuel system and its detailed components to identify which are working properly and which are not.

The integration of CA~EN with TIGER was performed at several user interface levels. The top-level display showed the activation conditions and inputs to the model. The variables used within the models were highlighted as nodes on the display: *green* meaning 'not detected as misbehaving', *red* meaning 'detected as misbehaving'. The screen snapshots in Figures 1 and 2 show the activation display of CA~EN when a fault is detected. The arcs between the nodes represent the components that provide the translation of the input of a component to its output. Note, that because of the resolution of the sensors available, very often there are several components and sub-systems, and possibly the actions of the controller on each arc (cf. section 2.3).

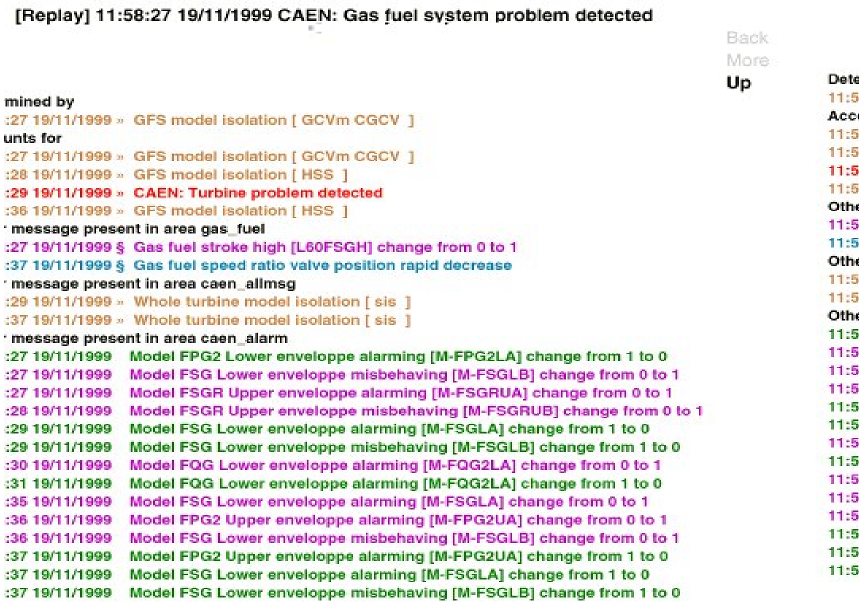


Figure 3. TIGER Fault Manager Screen.

The diagnoses from the model-based system were fully integrated with the output of the other TIGER diagnostic modules by the TIGER fault manager. In the above situation, the diagnoses are as shown in Figure 3. When looking at a TIGER high-level rule conclusion the messages from CA~EN are colour coded brown so that it is easy to visually see the contribution of the models to the high-level rules.

The CA~EN diagnoses often produced one or more sets of possible faulty components, these were displayed by clicking on the diagnostic message within the fault manager. For each sub-system a schematic of the sub-system was then displayed, highlighting the possible faulty components. Figure 4 is a display showing the gas fuel sub-system (GFS) with two faulty component candidates highlighted.

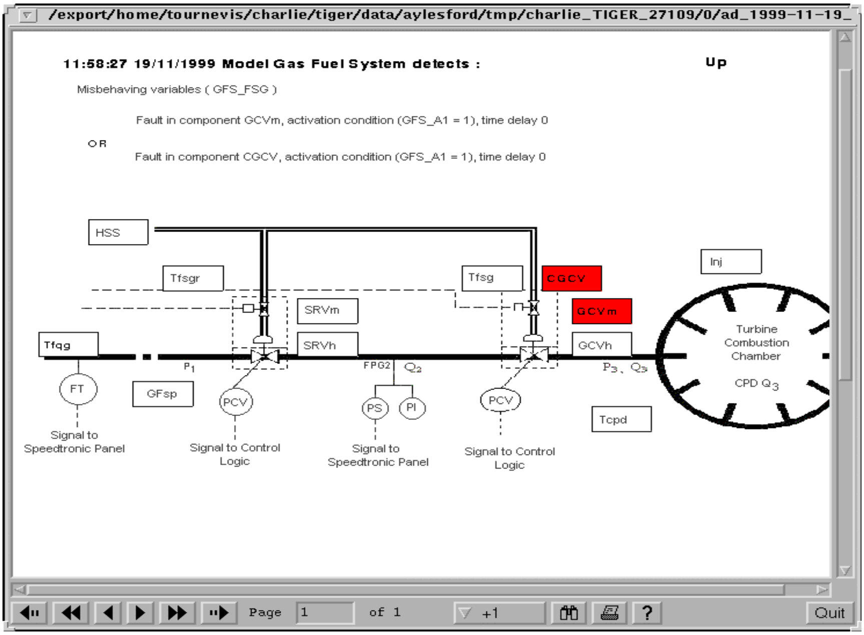


Figure 4. Fault Manager of CA~EN Diagnosis.

All this is displayed on the model-based predictions from CA~EN as illustrated on the graph in Figure 5. The channels with the prefix "M-" are from the models.

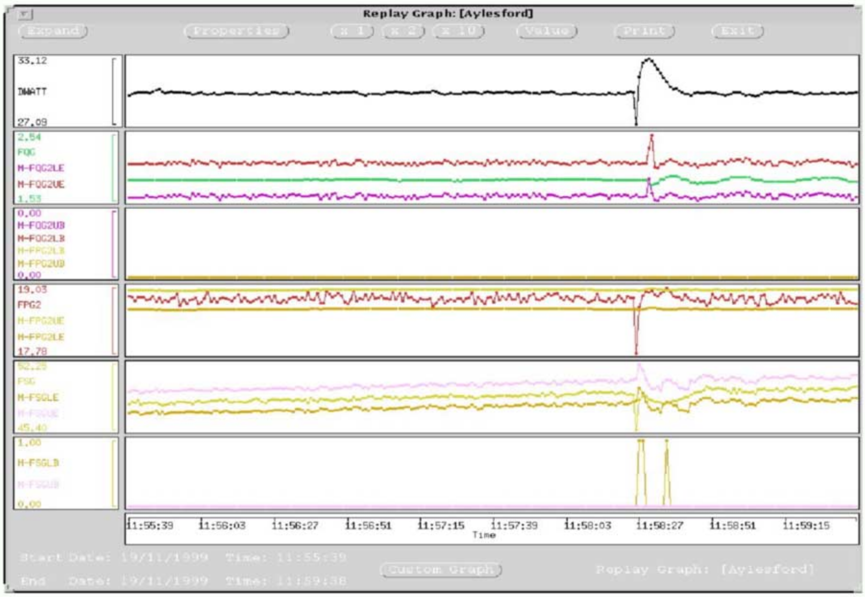


Figure 5. Detection envelopes reported by TIGER SHEBA.

The fault predictions of CA~EN are based on envelope predictions which were fully integrated with the TIGER data stream, TIGER graphs, and trending mechanisms. The envelopes show the expected high and low value, as well as the actual value, for each variable. In addition, digital indications of whether the actual value is alarming or misbehaving relative to the envelopes were available and could be displayed.
In the framework of the TIGER project (1992-1996), CA~EN was successfully tested on several subsystems of a 28-MW General Electric Frame 5 gas turbine operating at the Exxon Chemical Fife Ethylene Plant (UK) and of a Dassault Aviation Auxiliary Power Unit manufactured by MicroTurbo (F). In the TIGER-SHEBA project, extensive testing of the TIGER /CA~EN system was carried out. Numerous test runs were conducted varying from 2 weeks to 2 months at a time covering key periods of operation for the considered gas turbine. In addition, the combined system run continuously on-line on the Frame 6 gas turbine of the National Power's cogeneration plant at Aylesford (UK).

2. The CA~EN diagnosis system

2.1. CA~EN history

CA~EN (like Causal Engine) is rooted in the Qualitative Reasoning (QR) and Model based Diagnosis (MBD) communities [42][54][56] and it exemplifies some of the bridges that exist between the MBD Control and AI approaches. QR allows one to perform symbolic processing while still keeping strong semantics related to numerical properties. These are useful to achieve predictive and explanatory tasks from an incomplete and imprecise representation of the domain in terms of approximate relations. The evolution which has taken place in CA~EN reflects the evolution of these communities [33][36].

The pionneering studies in QR [2] tried to keep both the representation and the reasoning at a pure symbolic level, elaborating most of the conclusions from the knowledge of order relations. Because of the obvious limitations of this approach (knowledge about physical systems, even incompletely known, is never purely symbolic), there were evolutions towards integrating quantitative knowledge in qualitative reasoning systems. Efficient semi-qualitative systems, like CA~EN, were proposed [14][55].

The first version, CA~EN 1.0, was devised and implemented for the biotechnology application domain to be part of a model-based supervision system [6][7]. Fermentation processes present non linearities in the dynamics, the biological reaction is non stationary and there is uncertainty associated with the sensor technology. The influence of control is very difficult to model analytically and it is generally not accounted for in the existing analytical models. Therefore, qualitative approaches seemed most appropriate. CA~EN 1.0 was built following pure qualitative lines: the value set of the physical quantities are discretized into a discrete quality space, and the reasoning makes use of the fact that a quantity can only have predefined qualitative values. The output of CA~EN 1.0 provides the behavior of the process' variables across time according to the behavior of the exogenous variables. This is given in the form of the « most probable » behavior expressed by a sequence of dated qualitative states. This approach belongs to the family of approaches that produce a behavior tree (e.g. QSIM [14]), although the application requirements led us to adopt a probability measure to force the produced behavior to be unique.

In the TIGER project, our task was to build a model-based diagnosis system based on the CA~EN modeling formalism. After analyzing the gas turbines requirements and the type of faults to be diagnosed, it was concluded that pure qualitative models would not be sufficient for fault detection [44][45].

The CA~EN formalism was hence extended to cope with more numeric features and revised prediction algorithms achieved the completeness property required for diagnosis applications[4] [7]. In the revised formalism, which is presented in section 2.2, a quantity can take its value in any subinterval (not predefined) of its value set. All the reasoning mechanisms are numerically grounded in the interval algebra and the numeric knowledge is easily representable. The influences between two variables are all brought back to linear (algebraic or differential) propagation functions. These are possibly complemented by static constraints that encode non linearities.

2.2. The CA~EN (2.0) knowledge representation formalism

CA~EN is a causal and constraint-based reasoning system whose representation formalism includes:

1. A causal model in the form of an influence graph in which the links represent influences between pairs of variables, referred as the *local level*,
2. A constraint level which allows one to represent non linearities, referred as the *global level*.

Both levels can manage imprecise knowledge. A CA~EN program represents a formal model of the physical system built from knowledge about the physics underlying the behavior of the system. Time is dealt with explicitly with a logical clock which delivers a sampled version of continuous time of period δt.

Causal influences allow one to represent causal dependency type knowledge. Given two variables x and y, the CA~EN formalism offers five types of influences:

- $I_{D1}(x, y, K, cond, T_r, d)$, dynamic 1st order influence denoted by $x \text{ —D1}\rightarrow y$,
- $I_{D2}(x, y, cond, K, \omega, \xi, d)$, dynamic 2nd order influence denoted by the $x \text{ —D2}\rightarrow y$,
- $I_I(x, y, cond, K, d)$, integral influence denoted by the symbol $x \text{ —I}\rightarrow y$,
- $I_S(x, y, cond, K, d)$ static influence denoted by the symbol $x \text{ —S}\rightarrow y$,
- $I_C(C, y, cond)$, constant influence denoted by the symbol $C \text{ —C}\rightarrow y$.

[4] A concurrent version, called BioCaen was developed from CA~EN 1.0 in the framework of a join project with a neuropsychologists group of INSERM. The language was significantly augmented: a structural network description level was added and the influence descriptions were enriched (for example, the influence gain which was a constant was made variable, and the influence condition syntax more complex). These features provided BioCaen with the expressiveness needed for brain activation and other highly complex time-varying systems modeling [32][33].

Causal influences are characterized by several parameters:

- K is the parameter *gain*, that represents the static gain of the influence,
- T_r is the parameter *response time,* that represents the time (given as the number of sampling periods) needed by y to get a new equilibrium state after having been perturbed,
- d is the parameter *delay*, that represents the time (given as the number of sampling periods) needed by y to react to x,
- ξ and ω are the *damping ratio* and the *undamped natural frequency* for 2nd order transfer functions,
- C is a *constant,*
- *cond* is a parameter which specifies the logical conditions under which the influence is active.

All parameters but the delays can be given an interval value when known with imprecision or a real value otherwise. Influences are labeled by the underlying component(s).

The following CA~EN example states that the variable `GFS_FQG2` is influenced by the variable `GFS_FPG2` through a static influence, which is active when condition `GFS_A1=1` is true. The gain interval value is `[-0,3630,-0,3411]`and the delay is `5` sampling periods. The underlying components are SRVh (Stop Ratio Valve hydraulics), Tfqg (*fqg* transducer) and GFsp (Gas Fuel supply system).

```
SRVh_Tfqg_GFsp:
       condition (GFS_A1=1)
       {GFS_FPG2 –S–> GFS_FQG2:
       gain in [-0,3630,-0,3411], delay=5;}
```

Following a component-oriented modeling approach, the CA~EN language allows the user to specify generic models, which can be invoked and instanciated on request.

The global constraint level is composed of functional numeric constraints associated with interval domains, e.g. constraints arising from physical laws. In other words, a global constraint is any algebraic equation, which may be non-linear, in which unknowns are assumed to take on interval values. The global constraints are expressed by means of traditional arithmetic operators: +, -, *, / and **. These operators are interpreted in the interval algebra.

Given that variables and parameters take interval values, one can easily adapt the model's granularity to the requirements of the faults. Hence CA~EN has a wide coverage of faults, from those radically changing the behaviour of the physical system to those causing smooth deviations.

2.3. CA~EN Modeling

The CA~EN modeling methodology takes advantage of automated modeling methods. In particular, it applies the causal ordering method for hybrid systems, i.e. systems that have multiple operation modes, that was proposed in [48] as an extension of [14]. The work in [48] and the corresponding software CAUSALITO were developed during the TIGER project to handle the complexity of the gas turbine subsystems that had to be dealt with. In this respect, the CA~EN formalism was ahead of its time, dealing with hybrid models when they were not yet popular in the diagnosis community. Given a component-oriented equation model of the system, CA~EN modeling includes three steps: the automatic generation of the causal structure, the determination of the causal influences' parameters and, the derivation of the operational detection models by an aggregation operation.

Referring to the generation of the causal structure, the originality of our approach is that, rather than generating a new causal structure for every mode, CAUSALITO performs in an incremental way. It is able to compute the minimal changes in a causal structure that are implied by a mode transition. After considering all possible mode transitions, the influences of the final causal structure are hence labeled with an activation condition reflecting the mode of the system, more precisely of the system components'modes. CAUSALITO can as well be used on line to determine the dynamic causal structure when a mode transition occurs but this is not the way CAUSALITO was used in the TIGER project.

CAUSALITO takes as input the occurrence matrix of the system in which element m_{ij} is 1 if the variable x_j appears in the equation e_i or 0 otherwise. CAUSALITO builds a bipartite graph G in which each variable x_j and equation e_i are represented by a node, and it exists an edge between x_j and e_i if m_{ij} is 1. Finding the corresponding causal structure stands in first place on computing a perfect matching in the graph G. The perfect matching defines a one-to-one association between variables and equations such that every equation can be interpreted as a causal mechanism that solves for one variable. The detailed procedure is presented in [48]. Figure 6 shows the causal interpretation of every equation e_i represented by the circles, and the underlying causal structure among variables for the Frame 6 turbine GFS used in the TIGER project [49]. It also shows how the components of the system, identified in Figure 4, are associated to every equation.

Whereas the component-oriented model and its associated causal structure, are suitable for fault isolation, the models that can be used for fault detection, i.e. the *operational detection models*, require the determination of the causal influences'parameters. Indeed, the CA~EN model prediction algorithm supporting fault detection requires to know the interval values of the influences'parameters. However in

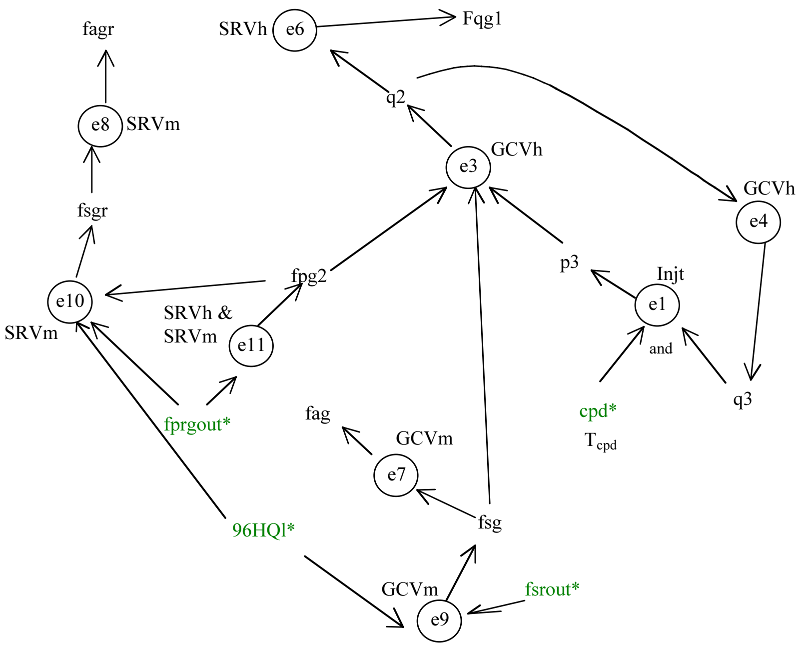


Figure 6. GFS causal structure.

many applications, none of the parameters involved in the relations of the component-oriented models is known; i.e. their values are not even known to the user. Hence, the only relations that can be operationally used for fault detection are the ones whose parameters can be estimated from the data. This rests on the condition that the variables appearing in the relations are measured or can be considered as constant.

An aggregated causal structure can be automatically derived from the original causal structure by removing all the non measured nodes. In some cases, it might be more suitable to use the aggregated causal structure for fault isolation as well. In this case, the component labels to be associated to every operational relation must be aggregated as well. The aggregated causal structure of the Frame 6 turbine GFS that was used for fault detection in the TIGER project is shown in Figure 1.

3. CA~EN diagnosis algorithms

The internal structure of CA~EN incudes two processing modules corresponding to the main tasks to be performed in fault diagnosis:

- A fault detection module based on a causal interval prediction mechanism,
- A fault isolation module based on an abstraction of the models in terms of timed causal models.

Let's notice that CA~EN fault detection and fault isolation can both be used on their own. For instance, fault isolation can take as input the discrepancies indicated by another fault detection mechanism. In this case, the second and third modeling phases which consist in informing the influences with their underlying propagation functions and appropriate parameters is not required.

3.1. Fault detection

In model-based diagnosis, decision about the existence of a fault is based on comparing the expected and the observed behaviour and exhibiting discrepancies. In this section, we present how CA~EN model predictions are checked against observations to obtain robust decisions about the existence of faults.

The prediction module can be used on its own, i.e. in an "open-loop mode" as a pure simulator, or coupled with the fault detection module, in which case it is used in a *Semi-Closed Loop* (SCL) mode and takes the measured variable values into account in real time.

The input data are the causal model - including initial conditions - and the evolution of the exogenous and other measured variables over time, acquired synchronous to the data acquisition clock. The output of the system is the trajectory of each process variable [25][46].

In CA~EN, two steps are executed to predict the variable values:

1. At the local level, from the superposition theorem that applies to linear systems, the computation of the updated value of a variable y consists of processing the sum of the activated influences having exerted on the variable during the last time-interval. The result is an interval.
2. At the global level, the variable updated values are refined with the global constraints by performing a tolerance propagation algorithm [14] on the set of variables. The effect of the tolerance propagation algorithm is to filter (reduce) for consistency the variable interval values using the global constraints.

The simulation results produced by the CA~EN prediction module are envelopes for the variables of interest (cf. screen from the TIGER system in Figure 7). The envelopes provide upper and lower bounds of the variable values at each sampled instant, which can be taken as *adaptive thresholds* for checking consistency against the measured signals. As a consequence of the interval-based reasoning used in CA~EN, the results are complete but not correct [1], i.e. the predicted intervals at each time step are guaranteed to include all the possible real values but they include spurious values as well. In the interval model-based fault detection approach, predicting the variable values is one of the most critical steps. Envelops need to be tight enough to be sensitive to faults, but not too tight so as to avoid generating false alarms [40].

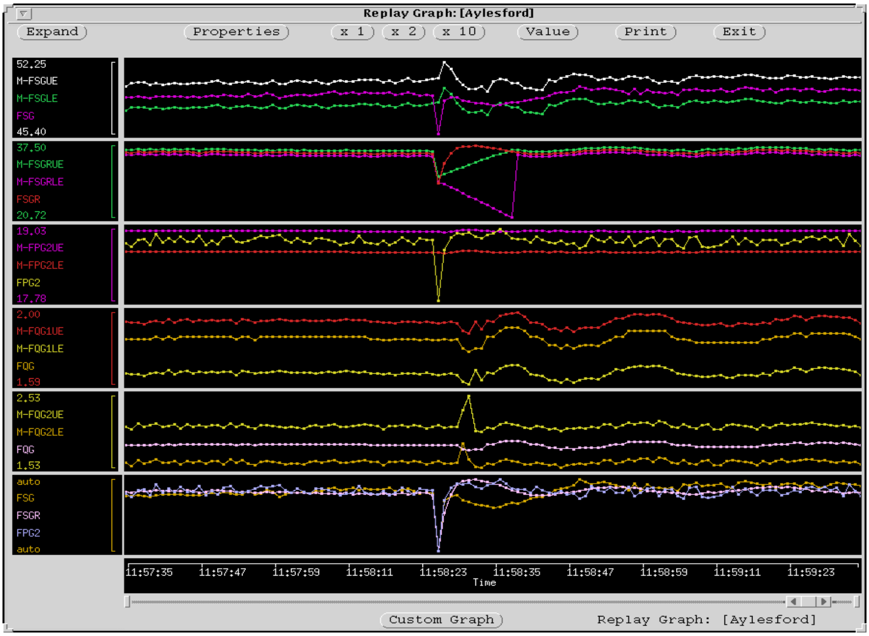


Figure 7. CA~EN enveloppes.

On fault detection mode, the observations acquired on measured variables are used to determine the estimates of their causally downstream variables. For example, for a static influence $I_S(x, y, cond, K, d)$, we have:

$$y_{pred}(t+1) = Kx_{meas}(t-d) \tag{7}$$

where the suffixes *"pred"* and *"meas"* stand for "predicted" and "measured", respectively.

At each instant t, if y is a measured variable, CA~EN checks whether the measured value $y_{meas}(t)$ (a real number) is consistent with the predicted value $y_{pred}(t)$ (an interval). If not, variable y is said to be *alarming* at time t. The set of alarming variables is denoted by AL. This is equivalent to the calculation of an interval residual:

$$r_y(t) = y_{pred}(t) - y_{meas}(t) \tag{8}$$

for which $0 \not\subset r_y(t)$ in the faulty case and $0 \subset r_y(t)$ otherwise. y being alarming interprets as its observed trajectory going out of its predicted envelope at time t as shown in Figure 8.

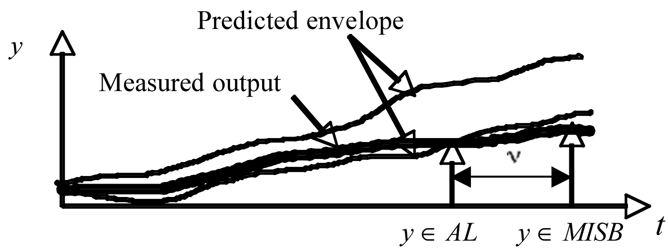


Figure 8. CA~EN Fault detection is based on adaptive thresholds.

In practice, noise in the measurements or other kind of disturbances call for a more robust indicator than just alarming variables. A variable is required to remain alarming during a time interval T of length significantly greater than the sampling period before being set to *misbehaving* (cf. Figure 8), then indicating a fault. The length of T is a multiple of the sampling period, i.e. $T=\nu\delta t$, and its tuning is left to the user. ν is called the *alarming-misbehaving threshold*.

More formally, the set *MISB* of misbehaving variables can be defined as follows:

$$y \in MISB \text{ at time } t \text{ if } y \in AL \text{ since } t-\nu, \text{ i.e. } 0 \not\subset r_y(t-i) \text{ for } \forall i = 0,\dots,\nu \tag{9}$$

The CA~EN fault detection strategy is a mixed strategy which combines an observer type strategy (closed-loop mode) with a simulation strategy (open-loop mode) to determine the residuals. We call this strategy a *Semi-Closed Loop(SCL) strategy* [14]. SCL is relevant to influences with memory (1st and 2nd order), whose estimation at current time of the output (influenced variable) needs previous values of the output. The closed-loop mode uses measured values for the previous values whereas open-loop mode uses predicted values. The mode control strategy depends on whether the observed value of y is in the predicted envelope (normal situation) or out of the predicted envelope (alarming situation) as illustrated below for a 1st order influence $I_{D1}(x, y, K, cond, T_r, d)$:

If $y \notin Al$ then closed-loop mode, then

$$y_{pred}(t+1) = a_D y_{meas}(t) + b_D x_{meas}(t-d) \tag{10}$$

If $y \in Al$ then open-loop mode, then

$$y_{pred}(t+1) = a_D y_{pred}(t) + b_D x_{meas}(t-d) \tag{11}$$

where $a_D \in [a_{D\min}, a_{D\max}]$ and $b_D \in [b_{D\min}, b_{D\max}]$.

The intuition behind the SCL strategy is related to two issues:

- The closed-loop mode performs successive one step ahead predictions, obtaining this way tighter interval estimates [1].
- As soon as the variable becomes alarming, running on a closed-loop mode would drive the prediction to "follow the fault", turning the fault detection procedure insensitive to the fault.

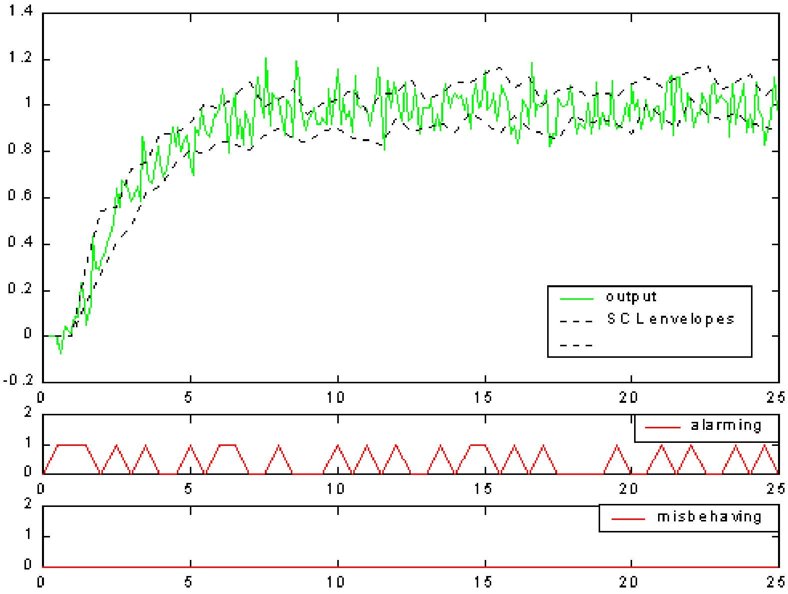


Figure 9. SCL detection strategy.

When a fault is present, the SCL strategy is able to discriminate between noise and the effect of the fault. For instance, the noisy signal displayed in Figure 9 becomes alarming several times but never misbehaving. The SCL strategy has been compared to pure open-loop and closed-loop strategies in the case of abrupt and drift faults and happens to be a valuable trade-off, in particular SCL results in lower detection times [14] [40][50].

3.2. The fault isolation module

Having detected one or more misbehaving variables, our system searches for the original possible cause(s) and elaborates a list of diagnosis candidates. Let us call *Causal System Description* (*CSD*) the formal description of the system given by its causal structure informed with delay propagation times and component labels as presented in section 2.3, *COMP* the set of physical components composing the physical system, and *OBS* the set of observations at some time point.

According to the logical theory of diagnosis, a diagnosis is a minimal set of components for which the invalidation of the normal behaviour assumption yields *(CSD, COMP, OBS)* consistent [15]. In the CA~EN diagnosis approach, the influences themselves are the elements of *COMP*. Faulty influences are turned back into their corresponding faulty components. From now on and given the equivalence, we will work indifferently with components or with influences.

Time aspects include delay times, as well as the dynamics introduced by the different operating modes captured by the influence activation conditions.

Diagnoses are computed from the collection of *conflict sets*, i.e. sets of components such that the observations indicate that at least one of the components in the set must be behaving abnormally.

The diagnosis process is initiated as soon as a variable is reported as misbehaving and the causal structure is used to retrieve the hypotheses underlying the propagation paths leading to the detected inconsistency. For this variable, say *X,* the conflict generation procedure traces backwards in the causal graph, following the intuition that the influences which may be at the origin of the misbehaviour of *X* are those related to the edges belonging to the paths reaching node *X* from the first upstream measured variable nodes. This set of influences, called the *ascendant influences* of *X*, is recorded as a conflict set. The discrepancies reported by misbehaving variables have each an associated conflict set.

Let us notice that knowing that some variable is not misbehaving may be informative for refining the diagnosis if some specific exoneration assumptions are made [11][50].

The diagnosis generation is based on generating the minimal hitting sets of the collection of conflicts. As new symptoms for a given fault may appear across time, it is important the diagnosis procedure be incremental. In CA~EN, we use a timed extension of Levy's algorithm [20] which is an incremental revised version of Reiter's [38]. Diagnoses are given as sets of faulty components labeled by their corresponding time of failure. Figures 3 and 4 are screens from the TIGER SHEBA system that illutrates the isolation procedure on the Frame 6 turbine.

4. Timed fault isolation

The standard approaches for continuous systems, either in the FDI community or in the DX community, have a static view of diagnosis and do not account for aspects referring to time in the system, such as delays. For example, the usual FDI fault signatures rely on symptoms that are based on residuals and it is assumed that all the symptoms are available simultaneously. This is not obviously the case in many situations. The same is true for the DX logical model-based diagnosis approach since diagnoses are generated from conflicts independently of their time of appearance (see [11]) for the relation between fault signatures and conflicts).

It is only recently that this issue has been approached in some pieces of work, like [10][9][34]. The CA~EN ideas were ahead of their time since the knowledge representation formalism already included the representation of timed delays and the fault isolation procedure delivered time stamped diagnosis. Our intuition was indeed that available time information such as order of symptom appearance or date of symptom might be determinant in some diagnosis situations. However, it was somewhat too early and this issue was immature in the diagnosis community at that time. Hence, although the basics were already there at the end of the TIGER project, it is only recently that the timed fault isolation algorithm, now able to deal with symptoms appearing in time, has been fully completed (a draft is submitted this year to the DX'06 Workshop [41]). These new results are reported in the next sections.

4.1 Case study for timed diagnosis

Through the following sections we use the example presented in [34] which is a two tanks system. It consists in continuously supplying water to two consume areas (s_1 and s_2 are the corresponding flows) from two cascaded reservoirs (y_1 and y_2 are the water levels in the respective reservoirs). The water transport between reservoirs is modeled as an open flow channel with a pump. τ_1 between the two reservoirs and τ_2 between the pump and reservoir 1 are the transport time delays. Variable u_{out1} represents the output flow from reservoir 1 and u_{pump} represents the flow through the pump. The system is shown in Figure 10 and its *CSD* is given in Figure 11. The "*meas*" index indicates measured variables and *dx(t)* stands for the variable $x(t+\delta t)$, where we recall that δt is the sampled time period. In the *CSD*, every influence is related to its underlying physical component. In Figure 11, C1 corresponds to the level sensor of Reservoir 1; C2 to the level sensor of Reservoir 2; C3 to the flow sensor of consume area 1; C4 to the flow sensor of consume area 2; C5 to the output pipe; C6 to the pump; C7 to the upper tank and C8 to the lower tank. One should notice that several influences may correspond to the same physical component as it is the case for C7 and C8.

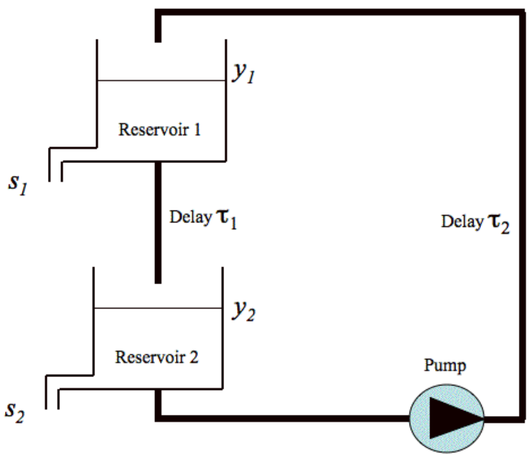


Figure 10. The water transport system example (from [34]).

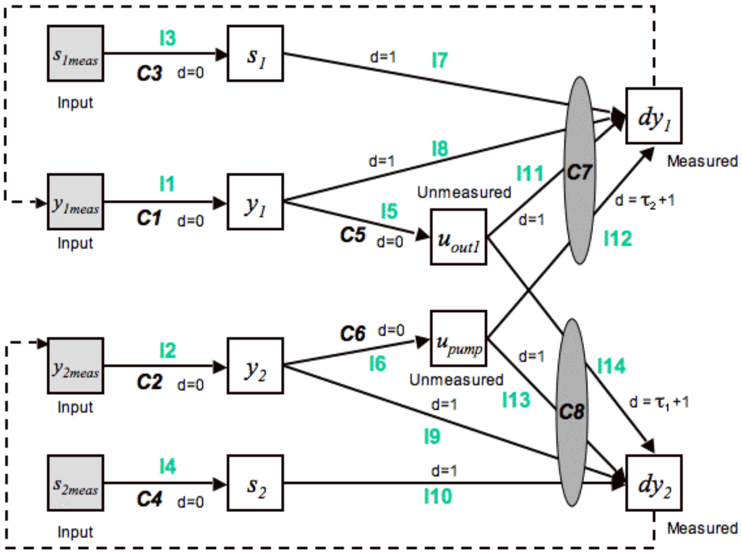


Figure 11. *CSD* of the water transport system example.

4.2. Accounting for time in conflict generation

Time information is explicitly represented in *CSD* by the delays associated to influences (cf. Figure 11). A time label called *failure time* is introduced to indicate when a component, say *C*, has started to fail. In other words, given the current time *t,* the failure time means that the component *C* must be faulty since at least time (t – *failure time*) and this is noted $C_{failure\ time}$. The elements of the conflicts that are generated when a variable misbehaves in *CSD* are hence each labelled by a failure time. These conflicts are called *time labeled conflicts.*
The following considerations are taken into account for reasoning about time in the causal model *CSD*, i.e., assigning the failure times to the components when generating conflicts and accounting for the activation conditions truth values:

1. The influence's delay represents the time needed by the effect variable to react to a variation of the cause variable. For example, the effect variable dy_1 needs a delay τ_{2+1} to be influenced by the cause variable u_{pump}.
2. The effect of a fault on a component associated to an influence propagates instantaneously to the effect variable, regardless of its related delay. For example a fault in C1 propagates instantaneously to the effect variable y_1.

3. The occurrence of a fault on a component associated to an influence can account for a downstream misbehaving variable at time *t* only if the influence was active at time *t – AccuDelay*, where *AccuDelay* is equal to the sum of the delays on the path going from the directly influenced variable to the misbehaving variable. For example a fault in C2 can account for a misbehavior of dy_1 if and only if influence I2 was active at time $t - ((\tau_2+1) + 0)$.
4. The failure time of a component is calculated by addition of delays from each one of the influences on the path going from the directly influenced variable to the misbehaving variable. If a component has two or more ways to reach the same variable, minimum sum delay is enough to explain that the variable has been influenced by the faulty component. This case is referred as the *One component-Two paths-One conflict* case.

4.3. Timed Diagnosis generation

The diagnosis generation is devised to be run at each sampling time, based on the generated conflict sets at that given instant. In Levy's diagnosis algorithm [20][21], a *leaf* on a Hitting-Set tree represents an associated node number identifier n_i and a set of edges $H(n_i)$ that describes the path from the leaf up to the root. At the end of the diagnosis procedure, the resulting diagnoses are given by the leaves of the generated HS-tree. For each new conflict, the open branches are extended and it is sufficient to prune the new leaves n_i for which there exists n_j, $n_j \neq n_i$, such that $H(n_i) \supseteq H(n_j)$. The diagnoses that explain the target system's misbehaviors are made up of the edges that go from each open leaf (i.e. that has not been closed) to the root. In this way Levy's algorithm gives a means to compute diagnoses incrementally, under the assumption of permanent faults.

In this section, we extend Levy's algorithm to provide an incremental diagnosis generation algorithm that deals with time in two cases: simultaneous symptom occurrence and symptoms occurring in time. The algorithm takes as input time labeled conflicts and outputs time labeled diagnoses.

The following principle, called *Maximum Failure Time Principle* is used to determine the diagnosis element failure times from the conflict element failure times:

Maximum Failure Time Principle. Let us consider that one component is involved in the misbehavior of two, or more, variables, i.e. it appears in several conflicts with different failure times (Figure 12 and the example below illustrate this case for C1). Then, this component is labeled with the same failure time in the generated diagnoses. This failure time is taken as the maximum failure time of the component in the conflicts. The reason for choosing the maximum among all the failure times is that it must simultaneously explain the misbehavior of all the misbehaving variables, i.e., simultaneously cover all conflicts.

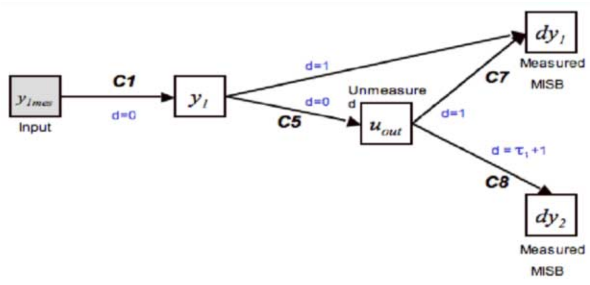


Figure 12. One component in two conflict sets.

For example, consider the case of two misbehaving variables MISB(dy_1) and MISB(dy_2) and the two corresponding conflicts:

Conf(dy_1) = {$C1_1$, $C2_{\tau_2+1}$, $C3_1$, $C5_1$, $C6_{\tau_2+1}$, $C7_0$}
Conf(dy_2) = {$C1_{\tau_1+1}$, $C2_1$, $C4_1$, $C5_{\tau_1+1}$, $C6_1$, $C8_0$}

Component *C1* has a failure time of 1 in the first conflict and $\tau_1 + 1$ in the second one. Then, the single component diagnosis based on C1 must have failure time $max[1, \tau_1 + 1] = \tau_1 + 1$. Hence the corresponding time labeled single component diagnosis will be $C1_{\tau_1+1}$.

4.3.1. Management of simultaneous symptoms

This section analyzes the case of multiple simultaneous symptoms occurrence. Without loss of generality, let us consider the case of two simultaneous misbehaving variables MISB(dy_1) and MISB(dy_2), which lead to the two time labeled conflicts Conf(dy_1) and Conf(dy_2) given above. The hitting-set tree corresponding to the management of these two conflicts by the algorithm is shown in Figure 13. The diagnoses D1 to D8 given below result from the open leaves *ol1* to *ol8* with the following indexes *{ol1=1, ol2=2, ol3=4, ol4=5, ol5=9, ol6=12, ol7=15, ol8=18}* in the hitting-set tree. Notice that the *Maximum Failure Time Principle* has been used for all the single component diagnoses D1 to D4.

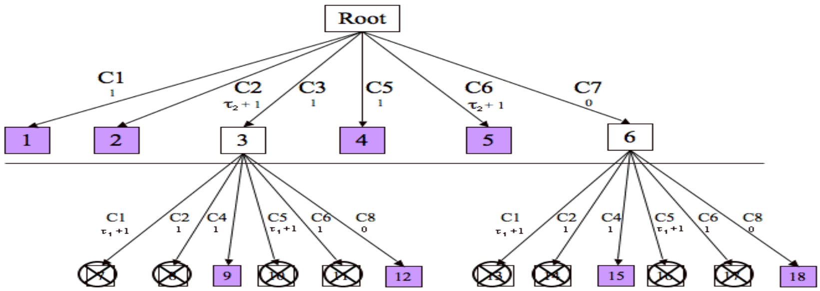


Figure 13. Timed diagnosis with simultaneous symptom occurrence.

D1 = $\{C1_{\tau_1+1}\}$
D2 = $\{C2_{\tau_2+1}\}$
D3 = $\{C5_{\tau_1+1}\}$
D4 = $\{C6_{\tau_2+1}\}$
D5 = $\{C3_1, C4_1\}$
D6 = $\{C3_1, C8_0\}$
D7 = $\{C7_0, C4_1\}$
D8 = $\{C7_0, C8_0\}$

4.3.2. Management of symptoms occurring in time

When misbehaving variables and hence the corresponding conflicts appear at different times, the dates of appearance can be advantageously used to discard some time based inconsistent diagnoses. In the diagnosis generation algorithm, this means that some leaves can be closed because of time-based inconsistency (cf. Figure 14).

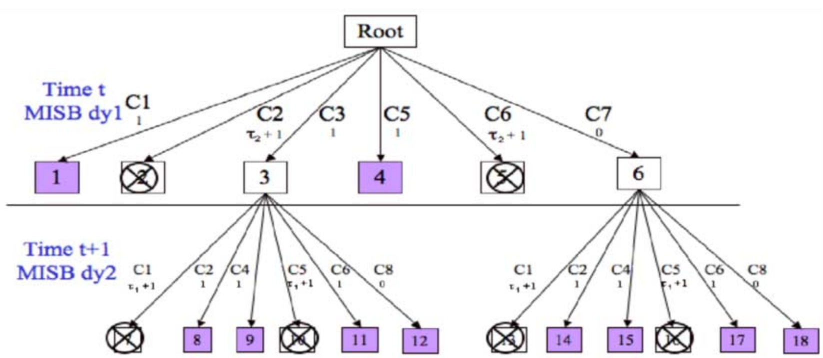


Figure 14. Diagnoses at different times *(δt=1, τ_1=3 and τ_2=5).*

The following principle, called *Minimal Required Time Principle* is used to check time based consistency. It is based on the fact that the failure times of a common component involved in two consecutive conflicts must account for the time interval elapsed between the symptoms occurrence.

Minimum Required Time Principle. Without loss of generality, let us assume two misbehaving variables appearing at times t and $t+n\delta t$, n being an integer, and a component C involved in the two corresponding conflicts with failure times f_1 and f_2, respectively. Then, C leads to a time based consistent diagnosis if and only if:

$$f_2 - f_1 \geq n\delta t$$

Checking the above condition in the diagnosis scenario of Figure 14, it happens that component C1 fulfills the requirements to be a diagnosis: $f_1 = 1$ and $f_2 = \tau_1+1$, hence $(3+1) - 1 = 3 \geq 1$. The same happens with C5. On the other hand C2 and C6 are not considered as single diagnoses because of time based inconsistency: $f_1 = \tau_2+1$ and $f_2 = 1$, hence $1 - (5+1) = -5 < 1$.
Our algorithm discards the diagnoses candidates that do not satisfy the *Minimum Required Time Principle*. For the remaining diagnosis candidates, we use the following to determine the failure times to be assigned to their elements:

Single component diagnosis candidates : the Maximum Failure Time Principle applies.

Multiple component diagnosis candidates : Let us assume that Ci_{f1} is involved in a conflict at date t and that Cj_{f2} is involved in a conflict at date $t + \delta t$, then the generated diagnosis must update the failure time of Ci to account for $n\delta t$. The generated diagnosis candidate at time $t+n\delta t$ is hence $\{Ci_{f1+n\delta t}, Cj_{f2}\}$. This is referred to as the *Updated Failure Time Principle.*

Figure 14 illustrates the procedure. For example leaf 2 has been closed from the application of the *Minimum Required Time Principle* to C2. The generated diagnoses with failure times updated according to the *Maximum Failure Time Principle* and the *Updated Failure Time Principle* are the following :

Diagnoses at time t:

$\{C1_1\}, \{C2_{\tau_2+1}\}, \{C3_1\}, \{C5_1\}, \{C6_{\tau_2+1}\}, \{C7_0\}$

Diagnoses at time $t + \delta t$:

$\{C1_{\tau_1+1}\}, \{C5_{\tau_1+1}\}, \{C3_{1+\delta t}, C2_1\}, \{C3_{1+\delta t}, C4_1\}, \{C3_{1+\delta t}, C6_1\}, \{C3_{1+\delta t}, C8_0\}$,
$\{C7_{0+\delta t}, C2_1\}, \{C7_{0+\delta t}, C4_1\}, \{C7_{0+\delta t}, C61\}, \{C7_{0+\delta t}, C8_0\}$

Compared to other works, the timed fault isolation algorithm presented above provides a solution to timed diagnosis for continuous systems within the model-based logical theory of diagnosis.

5. Other CA~EN based contributions

5.1. Diagnosability analysis

In relation with model-based diagnosis and CA~EN, a work dealing with the important issue of diagnosability and sensor placement was also started within the TIGER-SHEBA project. The requirement came from the end users stating that methods for analysing the diagnosability of a system and determining which sensors are needed to achieve the desired degree of diagnosability would be highly valued. This agreed with the commonly accepted principle that the requirements for maintenance and diagnosis should be considered at the earliest stages of design. Actually, diagnosability analysis turns out to be a prerequisite for several tasks such as instrumentation design, end-of-line testing, testing for diagnosis, etc. The problem was formulated as:

- Assessing the level of discriminability of a system, i.e. given a set of sensors, the number of faults which can be discriminated; and its degree of diagnosability, i.e. the discriminability level related to the total number of anticipated faults.
- Characterizing and determining the minimal additional sensors that guarantee a specified degree of diagnosability.

The first method that was proposed within the scope of TIGER-SHEBA [49] was based on a causal model similar to the one used by CA~EN. The main ideas behind the method were to use the causal model to analyse the analytical redundancies introduced by hypothesized sensors and build an Hypothetical Fault Signature Matrix (HFS Matrix). However, it did not take into account that a behavioral model including algebraic relations has generally several causal interpretations. Further work enhanced the method by exploring all the model equation causal interpretations, hence resulting in the completeness of the analytical redundancies associated to the availability of sensors and performs. From that a full diagnosability assessment can be performed [52][53].

5.2. Hybrid model-based state tracking

Subsequent developments integrated the CA~EN approach with a discrete automata based reasoning level to handle hybrid systems and perform state tracking [3][4][5].

CA~EN is already able to deal with hybrid systems:

- the formalism allows for representing multiple mode systems implemented by associating activation conditions to the causal influences of the *CSD,*
- the fault isolation mechanism accounts for the activation conditions truth values when tracing back the *CSD* for generating the conflicts associated to a misbehaving variable.

However, the work improved CA~EN along the following lines:

- the conditions defining mode transitions are not longer assumed to be all observable,
- component behaviors can be described at different levels of abstraction, i.e. by qualitative or quantitative relations,
- fault models can be explicitly represented, allowing fault identification,
- faults are not assumed to be permanent,
- the diagnosis generation combinatorial problem is solved by a mechanism that focuses the search on the most probable diagnoses first.

Following these goals, the research concentrated on combining techniques for continuous systems fault detection and isolation, already present in Ca-En into a high-level qualitative diagnosis system. To deal with a hybrid state space, the idea was to slice the continuous state space into a finite set of regions, called *configurations,* where the system is expected to behave in each mode. On the other hand, the bounded uncertainty over the continuous state delimitates geometrical regions that are easily positioned in the configuration grid. The diagnoses are then expressed in terms of the modes and configurations in which every component might behave across time, indicated by consistency with the observations. Based on this hybrid formalism, a new set of algorithms interleaving the search for the most probable diagnoses with consistency checking were defined and implemented in the KOALA software prototype [3][4][5].

6. Conclusions

This paper is in memory of Rob Milne. The scientific experience of the TIGER and TIGER SHEBA European projects is recalled through the presentation of CA~EN, a model-based diagnosis software that was devised during the projects and integrated with the commercial tool TIGER™. More importantly, several directions of more recent work that all build on CA~EN are presented, showing that the work initiated during these projects is still very much alive.

This paper reports about a piece of work and, more importantly, about a chapter of life for which I send special model-based thoughts to Rob who rests quietly at 8450 meters on the shoulder of the biggest mountain in the world. He is one with the mountain.

Rob Milne and DX community during DX'04, Carcassonne (France), June 2004

Acknowledgements

I would like to thank all the people that have been involved in the work reported in this paper: the partners of the TIGER and TIGER SHEBA european projects, with particular thanks to Charlie Nicol from IA who was very patient with our sometimes idealistic views of the research, Teresa Escobet from UPC who worked with me sometimes even late into the night, Gabriela Calderon from Girona University who, more recently, helped with the timed diagnosis work.

References

[1] J. Armengol, L. Travé-Massuyès, J. Vehi, J.Ll. de la Rosa, A survey on interval model simulators and their properties related to fault detection. *Annual Reviews in Control*, vol. 24, No. 1. ISSN: 1367-5788. Elsevier Science. Oxford, UK., 2000, pp. 31-39.

[2] D.D. Bobrow, Qualitative reasoning about physical systems, Artificial Intelligence, 24 (Special Volume), 1984.

[3] E. Bénazéra, L. Travé-Massuyès, P. Dague, State tracking of uncertain hybrid concurrent systems. DX02, 106-114, 2002.

[4] E. Bénazéra, Diagnosis and Reconfiguration based on Hybrid Concurrent Models - Application to Autonomous Satellites, PhD thesis of Université Paul Sabatier, Toulouse, France, 2003.

[5] E. Bénazéra, L. Travé-Massuyès, The consistency approach to the on-line prediction of hybrid system configurations, *IFAC Conference on Analysis and Design of Hybrid Systems (ADHS'03)*, Saint-Malo (France), 16-18 June 2003, 6p.

[6] K. Bousson, L. Travé-Massuyès, Fuzzy Causal Simulation in Process Engineering, *13th International Joint Conference on Artificial Intelligence*, *IJCAI'93*, Chambéry, France, 1993, pp. 1536-1541.
[7] K. Bousson, L. Travé-Massuyès, Putting more Numbers in the Qualitative Simulator CA~EN, *Int. Conf., on Intelligent Systems Engineering*, Hamburg-Harburg, Germany, pp. 62-69, 1994.
[8] K. Bousson, L. Zimmer, L. Travé-Massuyès, Causal Model-Based Diagnosis of Dynamic Systems, *5th Int. Workshop on Principles of Diagnosis*, New Paltz, (NY), USA, 1994, pp. 34-41.
[9] C. Combastel, S. Gentil and J.P. Rognon, Toward a better integration of residual generation and diagnostic decision, *IFAC SafeProcess'00*, Washington DC USA, 2003.
[10] M.O. Cordier, C. Dousson, Alarm driven monitoring based on chronicles, *IFAC SafeProcess'00*. Budapest, Hungary, 2000.
[11] M.O. Cordier, P. Dague, F. Levy, J. Montmain, M. Staroswiecki, L. Travé-Massuyès, Conflicts versus analytical redundancy relations. A comparative analysis of the model-based diagnosis approach from the artificial intelligence and automatic control perspectives, Rapport LAAS N° 02584, *IEEE Transactions on Systems, Man and Cybernetics*. Part B, Vol.34, N°5, pp.2163-2177, Octobre 2004.
[12] T. Escobet, J. Quevedo, S. Tornil, L. Travé-Massuyès, Integration of Dynamic Models in a Gas Turbine Supervision Control, *7th IEEE International Conference on Emerging Technologies and Factory Automation ETFA '99*. Barcelona, Oct. 1999.
[13] T. Escobet, J.P. Gouyon, J.A. Jimenez, J. Quevedo, L. Travé-Massuyès, S. Tornil, Model Descriptions, Rapport technique D310.9 du projet Européen Trial Applications TIGER SHEBA MBD n° 27548, Juin 1999, 115p.
[14] T. Escobet, L. Travé-Massuyès, S. Tornil, J. Quevedo, Fault detection of a gas turbine fuel actuator based on qualitative causal models, *European Control Conference (ECC'01),* Porto (Portugal), 4-7 Septembre 2001, 2741-2746.
[15] W. Hamscher, L. Console and J. de Kleer (eds.) "*Readings in Model-Based Diagnosis*", Morgan Kaufmann, San Mateo, CA, 1992.
[16] E. Hyvonen, *Constraint reasoning based on interval arithmetic: the tolerance propagation approach*, Artificial Intelligence 58, pp. 71-112, 1992.
[17] Y. Iwasaki, Simon H.A. (1986): Causality in device behavior. *Artificial Intelligence* 29(1-3), pp. 3-32.
[18] B.J. Kuipers (1986): Qualitative Simulation. *Artificial Intelligence* 29, pp. 289-338.
[19] B.J. Kuipers (1994): *Qualitative Reasoning - Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA.
[20] F. Levy, *Contribution à la realisation d'un raisonneur à profondeur variable*, PhD Dissertation Thesis, University of Paris Nord, Villetaneuse (France), 1989.
[21] F. Levy, ATMS for default theory, Internal Report LIPN n°91-8, Université Paris Nord, Villetaneuse (France), 1991.
[22] R. Milne, L. Travé-Massuyès, Real Time Model Based Diagnosis of Gas Turbines, *AIENG'93 Int. Conf.*, Toulouse, France, 1993.
[23] R. Milne, L. Travé-Massuyès, Application Oriented Qualitative Reasoning, *QR'93 Workshop*, Orcas Island, (WA), USA, 1993.
[24] R. Milne, L. Travé-Massuyès, Real-Time Model Based Diagnosis of Gas Turbines in *« Artificial Intelligence in the Petroleum Industry »*, edited by B. Braunschweig and R. Day Editions Technip, Paris, pp. 197-214, 1995.
[25] R. Milne, L.Travé-Massuyès, Model-based aspects of the TIGER gas turbine condition monitoring system, *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS'97,* Hull, UK, pp. 420-425, 1997.
[26] R. Milne, C. Nicol, M. Ghallab, L. Travé-Massuyès, K. Bousson, C. Dousson, J. Quevedo, J. Aguilar-Martin, A. Guash, TIGER : real-time situation assessment of dynamic systems, *Intelligent Systems Engineering Journal*, vol. 3, n° 3, pp. 103-124, 1994.
[27] R. Milne, C. Nicol, L. Travé-Massuyès, J. Quevedo, TIGER : Knowledge based Gas Turbine Condition Monitoring, *Expert Systems Conference ES'95*, Cambridge, UK, 1995, (Award of the Best Application Paper).
[28] R. Milne, C. Nicol, L. Travé-Massuyès, J. Quevedo, TIGER : Knowledge based Gas Turbine Condition Monitoring, *AI Communications Journal,* vol. 9, n° 3, pp. 92-108, 1996.
[29] R. Milne, L.Travé-Massuyès, J. Quevedo, TIGER : numeric and qualitative model-based diagnosis, *13th IFAC World Congress (IFAC'96),* San Francisco, USA, pp. 127-132, 1996.
[30] R. Milne, C. Nicol, L.Travé-Massuyès, Tiger with Model Based Diagnosis: Initial Deployment, In: *Applications an Innovations in Intelligent Systems VIII, Ann Macintosh and Mike Moulton and Frans Coenen (eds), pp. 181-196*, ISBN 1-85233-402-9, 2000, Springer-Verlag, London, Proceedings of ES 2000, Cambridge, England, December 2000. Best Paper Award from the Specialist Group on Knowledge-Based Systems and Applied Artificial Intelligence (SGES).

[31] R. Milne, C. Nilcol , L. Travé-Massuyès, TIGER with model-based diagnosis: initial deployment, Knowledge-Based Systems, vol. 14, n° 3-4, 2001, pp. 213-222.

[32] J. Pastor, L.Travé-Massuyès, J.F Demonet, B. Doyon, P. Celsis, BioCaen : A causal qualitative network for cerebral information propagation modeling, *11th Int. Workshop on « Qualitative Reasoning about Physical Systems »,* Cortona, Italy, 1997.

[33] J. Pastor, M. Lafon, L. Travé-Massuyès, J.F. Demonet, B. Doyon, P. Celsis, Information processing in large-scale cerebral networks : the causal connectivity approach, *Biological Cybernetics*, 82, pp. 49-59, 2000.

[34] Puig, V., J. Quevedo, T. Escobet and B. Pulido, On the integration of fault detection and isolation in model-based fault diagnosis, *15th International Workshop on Principles of Diagnosis*, Carcassonne, France, 2004.

[35] C. J. Price 1, L. Travé-Massuyès, R. Milne, L. Ironi, B. Bredeweg, M. H. Lee, P. Struss, N. Snooke, P. Lucas, M. Cavazza, Visions for Model-based Reasoning, *2nd MONET Workshop on Model-Based Systems during IJCAI'05*, Edinburgh (GB) , 30 Juillet - 1er Août 2005, 6p.

[36] C. J. Price 1, L. Travé-Massuyès, R. Milne, L. Ironi, B. Bredeweg, M. H. Lee, P. Struss, N. Snooke, P. Lucas, M. Cavazza, Qualitative Futures, *19th International Workshop on Qualitative Reasoning (QR-05)*, Graz (Autriche), 18-20 Mai 2005, 9p.

[37] J. Quevedo, L. Travé-Massuyès, T. Escobet, S. Tornil, TIGER-SHEBA: Diagnòstic basat en models de turbines de gas, 2nd Investigacio Technologica Duran Farell Award, Universitat Politècnica de Catalunya, also LAAS-CNRS Internal Report n°02188, 2002 (in Catalan).

[38] R. Reiter, A theory of diagnosis from first principles, *Artificial Intelligence* 32(1), p. 57-96, 1987.

[39] S. Spanache, T. Escobet , L. Travé-Massuyès, Sensor placement optimisation using genetic algorithms, *15th International Workshop on Principles of Diagnosis (DX'04)*, Carcassonne (France), 11-14 Juin 2004, pp. 179-184.

[40] S. Tornil, , T. Escobet, L. Travé-Massuyès, Robust fault detection using interval models, 1*2th European Control Conference ECC'03*, Cambridge, UK, 2003.

[41] L. Travé-Massuyès, G. Calderon, Timed fault diagnosis for continuous systems, Internal Report LAAS n°06143, Mars 2006, 7p. Submitted to the *17th International Workshop on Principles of Diagnosis (DX'06)*, Peñaranda de Duero, Burgos (Spain), 26-28 June 2006.

[42] L. Travé-Massuyès, R. Milne, Application-oriented qualitative reasoning, *The Knowledge Engineering Review*, vol. 10, n° 2, pp. 181-204, 1995.

[43] L. Travé-Massuyès, R. Milne, Diagnosis of Dynamic Systems based on Explicit and Implicit Behavioural Models: An application to Gas Turbines in Esprit Project TIGER,
1) *Fifth Scandinavian Conference on Artificial Intelligence SCAI'95*, Tronheim, Norway, 1995 (invited), pp. 169-183.
2) *IJCAI Workshop on Second Engineering Problems for Qualitative Reasoning*, Montréal, Québec, 1995, pp. 1-17.

[44] L. Travé-Massuyès, R. Milne, Diagnosis of Dynamic Systems based on Explicit and Implicit Behavioural Models : An application to Gas Turbines in Esprit Project TIGER, *Applied Artificial Intelligence Journal*, vol. 10, n° 3, pp. 257-277, 1996.

[45] L. Travé-Massuyès, R. Milne, TIGER™: Gas turbine condition monitoring using qualitative model-based diagnosis, *IEEE Expert Magazine "Intelligent Systems & Applications"*, vol. 12, n° 3, May-June 1997, pp. 22-31.

[46] L. Travé-Massuyès, R. Milne, TIGERTM: Gas turbine condition monitoring. Focus on the Ca-En qualitative model-based system, *Journées d'Etude sur les Logiciels pour le Traitement de l'Image, du Signal et l'Automatique (ELISA'97),* Nancy (France), 25-26 Mars 1997, 13p.

[47] L. Travé-Massuyès, R. Milne, Gaps between research and industry related to model-based and qualitative reasoning, In the proceedings of the *ECAI'98 workshop on "Model Based Systems and Qualitative Reasoning*", Brighton (UK), August 23-28, 1998.

[48] L.Travé-Massuyès, R. Pons, Causal ordering for multiple mode systems, *11th Int. Workshop on « Qualitative Reasoning about Physical Systems »,* Cortona, Italy, 1997.

[49] L. Travé-Massuyès, T . Escobet, R. Milne, Model-based diagnosability and sensor placement. Application to a frame 6 Gas turbine sub-system, *7th International Joint Conference on Artificial Intelligence (IJCAI'01)*, Seattle (USA), 4-10 Août 2001, pp.551-556 ; also in Proc. of *12th International Workshop on Principles of Diagnosis (DX'01),* Sansicario, Via Lattea (Italie), 7-9 Mars 2001, pp.205-212.

[50] L. Travé-Massuyès, T. Escobet, R. Pons, S. Tornil, The CA~EN diagnosis system and its automatic modelling method, *Computación i Sistemas Journal*, vol 5, n°2, pp. 128-143, 2001.

[51] L. Travé-Massuyès, T. Escobet, R. Pons, S. Tornil, The Ca-En diagnosis system and its automatic modelling method, In Proc. of *Workshop on Diagnosis, Qualitative Reasoning and Socio-Economic Systems*, C. Alonso and J.A. Ortega Editors, Valladolid, Spain, 4-6 July 2001, ISBN 84-95499-35-5.

[52] L. Travé-Massuyès, T. Escobet, S. Spanache, Diagnosability analysis based on component supported analytical redundancy relations, *5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS'2003),* Washington (USA), 9-11 Juin 2003, pp.897-902

[53] L. Travé-Massuyès, T. Escobet, S. Spanache, X. Olive, Diagnosability analysis based on component supported analytical redundancy relations, Janvier 2004, 12p, to appear in IEEE Transactions on Systems, Man and Cybernetics, Part A.

[54] L. Travé-Massuyès, L. Ironi, P. Dague, Mathematical Foundations of Qualitative Reasoning, AI Magazine, Special Issue on Qualitative Reasoning, Vol. 24, N°4, pp. 91-106, 2003.

[55] L. Travé-Massuyès, P. Dague**,** *Modèles et raisonnements qualitatifs*, Hermes, Traité IC2 Information, Commande, Communications, N°ISBN 2-7462-0744-3, 2003, 361p

[56] D. Weld, J. de Kleer (eds.), *Readings in Qualitative Reasoning About Physical Systems* (1990), Morgan Kaufmann, San Mateo, CA (USA).

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Adventures in Model-based Reasoning

Chris Price [1]
University of Wales Aberystwyth

Abstract. Model-based reasoning is a practical way of reasoning about real world systems based on the behaviour of the system being modelled. This paper considers the useful characteristics of models for model-based reasoning, discusses the state of industrial adoption of model-based reasoning techniques, with particular reference to a previous study by Travé-Massuyès and Milne of industrial take-up of this technology. Finally, the paper considers the prospects for model-based reasoning over the next twenty years.

Keywords. Model-based reasoning, qualitative reasoning

1. Introduction

Rob Milne was an enthusiast for applying interesting technologies, and one of the technologies that he championed was model-based reasoning (e.g. [1,2,3]). He invested his time as well as his enthusiasm in this technology, as a member of the steering group for the MONET European Network of Excellence in Model-based and Qualitative Reasoning [22]. The term *model-based reasoning* is an unfortunate and yet appropriate name for the work of using models of a system to reason about the behaviour of the system being modelled.

The name is unfortunate, because it can be argued that any representation of a system is a model. For example, a rule-based diagnostic expert system for a car contains a model of the process of diagnosing problems with the car. Executing the rule-based system is a way of using that model for diagnosis. Similarly, a payroll program might be said to contain a model of the process of paying company staff. Model-based reasoning is being used in a more specific way in this paper.

The name is appropriate, because model-based reasoning attempts to use models that reflect the structure of the domain in a much more direct way than the example models given above. The rule-based car diagnosis system mentioned above would contain heuristics that reflect the experience of a human car mechanic — "When the car won't start, and the lights aren't on, it is because of battery-related problems". Model-based reasoning corresponds to what good engineers do when their heuristics do not detect where the problem is — they use their knowledge of how the system works to trace which components are involved in the correct operation of the system, making predictions about what should be observed if those components are working correctly.

[1] Correspondence to: Chris Price, Department of Computer Science, University of Wales, Aberystwyth, SY23 3DB, U.K. Email: cjp@aber.ac.uk.

For diagnosis, the general idea is that the model predicts what should be happening in the observed system, and that discrepancies between the model of the system and the observed real world system can be used to detect problems and to decide what might be causing the problems.

The attributes of the models used in model-based reasoning vary from domain to domain, but the models used will typically be:

- Structural
- Component-based
- Compositional
- Executable
- Granular

This paper considers why these attributes are significant in the light of the qualities required of model-based applications. It then reflects on the level of practical success of model-based applications, and looks to the future prospects for the technology.

2. Desired Qualities of Model-based Applications

The model attributes listed in the previous section are not desirable of themselves, but because they contribute to certain qualities in the applications built using those models. The kinds of qualities looked for in model-based systems are:

Ease of construction / reusability. Within many domains, many different systems can be built from the same types of component. Process plants are made up of pipes, containers, sensors, pumps, heaters, coolers. Electrical systems are made up of wires, relays, fuses, lamps, motors, electronic control units. It is desirable to be able to produce a model of the whole system by joining these components. If this type of model can be used as the basis for a diagnostic system, then a diagnostic system for a new design can be produced from the design, rather than hand-crafted as a rule-based system.

Visibility / explainability. In the past, simulations for complex systems such as nuclear power plants have been coded as large Fortran programs. The simulation is given a set of numerical inputs, and produces a set of numerical results. Such simulators have a number of implicit assumptions that are not apparent to the user. In model-based reasoning, such assumptions can be made explicit [4] and viewed by the user. A trace of the reasoning done by the simulator can also be provided to the user as explanation for the results of the system [5].

Ability to alter models. Many simulations only address the situation of a correctly working system. For model-based applications such as calculation of the consequences of a component failure [6], it is necessary to be able to simulate the behaviour of the system when it contains a broken component.

Coverage of possibilities. In many real world applications, there is a good deal of uncertainty attached to the values of parameters in a model. In some applications, such as ecological ones, it is because of the complexity of the system being modelled. In others, such as engineering applications, it might be because of uncertainty in the design, or because of wear of components during use. This uncertainty means that numerical modelling with exact parameter values can give precise but inaccurate results. One solution to this is to use a technique such as Monte Carlo simulation to try to cover all possible results for the system. Where there is uncertainty over the values of many variables, it can be difficult to guarantee coverage by this method. The solution employed by most model-based reasoning systems is to use less precise values, and apply qualitative or semi-qualitative reasoning to produce results that cover all possibilities.

3. Attributes of Models for Model-based Reasoning

The extent to which the qualities discussed in the previous section are achieved in a model-based application will depend on the extent to which the attributes outlined in the introduction are met, and so this section will examine those attributes, and show how they contribute to the qualities needed by the applications.

3.1. Structural Models

Model-based reasoning tends to use models that reflect the structure of the system in some way. For many engineering systems, the models will reflect the physical structure of the system. The degree of faithfulness of the modelling to the structure will depend on the types of answers required from the model-based system. Struss [7] gives an interesting study of the degree of faithfulness needed to model different kinds of failures in a thyristor. Another example might be modelling the ecology of a stream - it would not be necessary to model each rock and boulder in order to reason about the general flow of water through the stream, only to model the topology of the stream and its environs in a less specific way.

This attribute shows the main difference between a heuristic rule-based diagnostic system and a model-based diagnostic system. For a rule-based diagnostic system, what is being modelled is the process of producing a diagnosis. In the model-based diagnostic system, what is being modelled is the domain, and the diagnostic work is performed by a general diagnostic problem solver using the model [8].

3.2. Component-based Models

When models are structural, it makes it easier for the user to build models for a domain, but the real advantages come from having some kind of component-based models that can be reused for different systems in the same domain. Then, a description of the structure of the system, for example a circuit diagram or a topological description of a stream, can be used to automatically form a model of the system. This greatly enhances the reusability of the model-based reasoning. Components in this sense do not need to be physical components, just as the structure of a system does not need to be a physical structure. For example, a chemical process plant might be best modelled in terms of the chain of processes that are carried out by the plant. In such a model, it would not be necessary to model flow through pipes and vessels, only the processes. In this case, the processes would be the components in the model. Such a model would not be of much use in detecting leaks or broken pumps, but might be appropriate for applications wanting to reason about the interactions between the processes.

3.3. Compositional Models

This attribute was originally phrased as a *no function in structure* rule [9]: components were not allowed to contain the behaviour of the overall system of which they were a part. This is perhaps better expressed as a need for compositionality, or for system independent components. The modelled behaviour of the component should be such that it works correctly in a variety of different settings, and if a component is removed or replaced within a model of the system, then the new system model will correctly reflect the behaviour of the new system configuration.

Compositionality is an important aspect of the reusability of components just discussed in the context of component-based models, especially for applications where one is investigating the behaviour of the system in the event of a component failure.

While useful components might be regarded as system independent, they are not context independent. An example of the kind of assumptions that are made on components might be that an electrical relay models the electrical aspects of a relay's behaviour. It cannot be used to predict overheating problems caused by its thermal generation properties, or to model its use as a paperweight.

3.4. Executable Models

Many models, for example in software engineering, have only descriptive properties — they describe some aspect of a system, but have no predictive properties. Model-based reasoning relies on the fact that models can be executed to discover facts about the behaviour of the system being modelled. This attribute enables a range of applications that will be discussed in the next section.

3.5. Granular Models: Choosing an Appropriate Level of Precision

Because predictions often need to be made under conditions of partial knowledge, qualitative reasoning [10] has often been an appropriate level of precision for model-based reasoning about real world systems. Engineers, due in part to their training, often incline towards the use of numerical models for all modelling tasks, but Coghill et al. [11] usefully point out that there is an interaction between precision and accuracy in selecting modelling granularities. Numerical modelling is more precise than qualitative reasoning, but is often less accurate because of that. By reducing spurious precision where it is not known, accuracy is increased. For example, predicting a current of 500mA in a motor is more precise than just predicting that there is current flowing through the motor, but if the actual current is 320mA, then it is less accurate.

In the early days of model-based reasoning, it was often assumed that all models used were likely to be qualitative. In their 1995 survey paper [12], Travé-Massuyès and Milne comment that

> "It is interesting to note that most people doing application-oriented work cite the need for the integration of qualitative and quantitative models",

and that trend has continued over the past decade, with both quantitative and semi-quantitative models being used for a variety of applications. A promising area is the use of qualitative models to interpret and to guide the use of quantitative models in appropriate situations.

4. Industrial Take-up of Model-based Systems

One of the most useful previous reviews of industrially-oriented work in this area was the previously mentioned survey done by Rob Milne working with Louise Travé-Massuyès [12]. They considered 40 different projects at different levels of technological development, and split them into major application areas. They commented

> "At this time there are very few real industrial QR (qualitative reasoning) systems... it is hoped that a similar summary in a few years will paint a very different picture."

This section will attempt to update both the classification used, and the level of development reported, to reflect today's position. Travé-Massuyès and Milne classify the projects into six major application areas:

- continuous processes
- engineering
- ecology
- electronic circuits
- business and commerce
- medicine

Continuous processes. This was split into two subareas (monitoring and diagnosis, and modelling and analysis), but ideally one would aim to reuse the same models or at least related models for each area. This has continued to be an area where a good deal of work has happened. Perhaps Travé-Massuyès and Milne's own work on diagnosing gas turbines is the best example of successful use of model-based reasoning in this area. The model-based aspect of TIGER has been developed considerably since being mentioned in the 1995 survey, and is now deployed and useful as part of the wider TIGER system. The addition of model-based monitoring to TIGER, based on the qualitative interval predictor CA-EN extends TIGER to detect faults that would otherwise have been beyond its capability [3]. The CHEM project (http://www.chem-dss.org) has also successfully used model-based reasoning along with a range of other techniques to successfully diagnose problems on chemical process plants. Rob Milne was one of the leaders of the BRIDGE task group, funded by the European Commission as part of the MONET2 Network of Excellence in Model-based and Qualitative Reasoning . Over the past four years, BRIDGE has achieved two things that will encourage the industrial take-up of model-based technologies in the area of continuous processes over the coming years. It brought together the two main groups of researchers looking to apply model-based techniques in this domain, and gave them a common way of talking about problems [13]. It also produced a collection of articles and tutorials to facilitate industrial adoption of model-based diagnostic technologies, and example solutions to reference problems (available online from [22]).

Engineering. Monitoring, diagnosis and engineering design are picked out in the 1995 survey as aspects of engineering where there were a number of applications coming through into industry. [14] recently detailed applications in the automotive industry and gave a long list of companies using this technology in earnest, including several products deployed in industry for some time. While automotive applications of model-based reasoning in engineering are still developing, this can be regarded as an area where the technology has come of age. Other areas of engineering have also seen similar successes. An exciting example in the space industry is provided by the work by NASA on DS-1 (Deep-Space I). DS-1 was the first spacecraft to use an integrated model-based approach to space craft autonomy, where model-based diagnosis was integrated with mission planning in order to respond to unforeseen events such as jammed thrusters [15,16]. DS-1 was launched in 1998, and successful tests on the model-based autonomy were conducted in May 1999.

Ecology. This was an area where there had been one or two interesting research examples in the 1995 survey. The research examples have multiplied since the 1995 survey [17,18,19,20,21], but have not for the most part moved beyond the research laboratory. Perhaps the most interesting development in this area has occurred over the past year, with the formation of the new NaturNET-REDIME project by the European Commission (http://www.naturnet.org/). This project brings together ecologists and model-based reasoning researchers. One of the main strands of the project is on developing a qualitative reasoning and modelling workbench that will allow users to express and develop their conceptual understanding of issues relevant to sustainable development.

Electronic circuits. It seemed strange that this was identified as an area separate from engineering in the 1995 survey. It may be that this was because it had a quite successful application, and the authors wished to highlight it. It has not been an area of significant growth in applications since that time.

Business and commerce. This has been the least successful of the outlined areas of industrial applications of model-based systems. Very early applications of model-based reasoning were reported in the 1995 survey, but there has been little work in this area since that time.

Medicine. Like ecology, industrial developments in this domain have lagged behind those in engineering, and for similar reasons. There had been significant investment in developing engineering applications of model-based reasoning, and the same kinds of investment have not occurred in ecology and medicine. There are signs of this changing — the MONET network has mapped out the areas where investigative work is happening in biomedical domains [22]. The 1995 survey mentioned early work on pacemakers with qualitative reasoning. Since then, qualitative reasoning has also been involved in the production of the first fully digital pacemaker. The patient treatment advice tool is based on qualitative models of the processes occurring in the patient, and has been successful where other technologies failed to provide a usable tool [23].

One area that was insignificant at the time of the 1995 survey has grown to greater prominence since then. The application of model-based reasoning in education has had several notable successes over the past decade. Work in the United States at Junior High School level has shown that pupils themselves can build simple models that explain real world phenomena, and that teachers can confidently build qualitative models with little support from ’techies’. Betty’s Brain [24] is a piece of software that allows students to provide rules for Betty’s Brain and see how Betty reasons about situations based on the information they place in her “brain’. This gives students valuable feedback on how the rules they provide for models would work out in practice. VModel [25] enables students to construct simple process-based models from scenarios provided by a teacher. Studies indicate that students have a much better understanding of modelling after using these tools than they had before.

5. Futures for Model-based Systems

The MONET Network of Excellence in Model-based and Qualitative Reasoning has been facilitating work in model-based reasoning for more than a decade. From 2001 to 2004, the Network was funded as *MONET2* by the European Commission, and as part of its funded work, it identified key application areas where model-based reasoning has the potential to make a significant difference over the next ten to twenty years.

5.1. The Science-bot: Automated Education

The science-bot is an automated teacher that has expertise in a specific area of science, and can assist learners in understanding that area by interacting with them. It can adjust the communicative interaction to be appropriate to the specific user, setting appropriate problems for the learner to solve, diagnosing any problems the learner has, and providing learning materials that meet the learner's individual needs.

5.2. The Virtual Vehicle: from Conception to Recycling

In some ways, this is the least ambitious of the four visions, as it is really just an extension of what is already happening in the automotive industry. The use of models is reducing the amount of real prototyping that needs to be done, and saving significant costs. That virtual prototyping needs to be integrated and reusable throughout the lifetime of the product. Model-based reasoning is making a major contribution to this effort.

5.3. Understanding and Managing Complex Natural Systems

Many natural systems cannot be adequately modelled numerically, because of incomplete information. The dynamics of such systems result from complex interacting mechanisms, and are very often affected by both endogenous and exogenous factors. This is another area where qualitative reasoning is beginning to make a difference. Qualitative models properly integrated with quantitative methods hold the prospect of providing understandable, explainable models of complex systems. Bellazzi et al. [26] give an example of a successful early application of a hybrid model-based method for system identification.

5.4. Robust Autonomous Problem Solvers in the Face of Uncertain Situations

This extrapolates from the early work done both by NASA [15,16] and in Europe [27] in remote diagnostics, to the need for autonomous systems to be able to reason about their state and act on that reasoning. Satellites, planetary landers and rovers are being sent to remote locations, and need to be able to plan how to carry out their missions under unforeseen circumstances, because as their circumstances become more remote, it is becoming harder to foresee all possible events that might occur. Models of the state of the system form an important basis for such reasoning, and the kinds of model-based reasoning techniques that have been developed will be appropriate for such situations.

While model-based reasoning has a key contribution to make to the achievement of the each of the four visions outlined, other technological developments will also be needed. The Science-bot, for example, will also need advances in analogical reasoning and user modelling in order to be successful. However, for all of the applications outlined, and explored in greater detail in [28], we consider that model-based reasoning is central to the efficient production of workable applications.

6. Conclusions

Model-based reasoning has progressed significantly in the decade since Travé-Massuyès and Milne published their survey in 1995 [12]. In the automotive industry and the process industries, applications that in 1995 were promising proof of concept studies in universities have been developed into full blown commercial applications. Other application domains, such as biomedical and ecological applications seem to be just moving towards convincing proof of concept studies at present. Educational applications seem to have come from nowhere in 1995 to being one of the most exciting areas for applying this technology, whereas business and commerce applications have been less successful than was envisaged in 1995.

The visions presented in section 5 show that there are plenty of new challenges to be met in the application of model-based reasoning to real world problems, but perhaps the biggest challenge, given that model-based reasoning will be with us for some time to come, is finding a better name for it.

Acknowledgements

The area of work discussed in this paper has been stimulated by the efforts of the MONET Network of Excellence. Throughout the first ten years of MONET, Rob Milne was a vital member of MONET's Steering Committee. The committee met together on average three or four times a year, and Rob was present at the majority of those meetings and others on specific topics such as BRIDGE, despite the need to travel around Europe to the meetings and his otherwise hectic schedule. His coffee breaks during the MONET meetings were often spent on the phone, catching up with the business that he had left behind in order to be at the meeting. One of my abiding memories of him will be of him on his mobile, usually simultaneously looking for the place where he could catch the sun while making his phone call. Those regular meetings shared with Rob left me with two distinct impressions of him.

Firstly, his good sense and the depth of his industrial and political experience. Without forcing his opinions on the rest of the Steering Committee, he provided us with advice and guidance about how to work with the European Commission, about how to achieve the aims of the Network, about how to help companies see the benefits of model-based technology.

Secondly, the way he made time for the more important things despite his busy schedule. It has been noted that no-one says at the end of their life "I wish I'd spent more

time at the office". While Rob clearly spent a lot of time at his work, he also made time to enjoy life. He was often the first in the water if a MONET meeting was held by the sea, or planning skiing or climbing trips if the meeting was conveniently situated.

A MONET project dinner was always more fun if Rob was present. At the last of those dinners in Brussels in March 2005, we were celebrating a successful completion to MONET2, and presentation of the achievements to the European Commission. Rob spoke of the IJCAI 2005 conference, an event in Edinburgh in Summer 2005 that he had been planning since 1995 — ten years of effort that were finally coming to fruition. He also spoke about the excitement of his forthcoming expedition to Everest. He covered the dangers of an Everest expedition, and the chances of not coming back. We did not think he would become part of those statistics. We will miss him.

References

[1] Robert Milne: Model-based reasoning: The applications gap. IEEE Expert 6(6): 5-7 (1991).

[2] Louise Travé-Massuyès, Robert Milne: Gas-Turbine Condition Monitoring Using Qualitative Model-Based Diagnosis. IEEE Expert 12(3): 22-31 (1997).

[3] R. Milne, C. Nicol and L. Travé-Massuyès, TIGER with model based diagnosis: initial deployment, Knowledge-Based Systems 14(3-4) pp213-22, 2001.

[4] K. Forbus, B. Falkenheiner, Compositional modelling: finding the right model for the job, Artificial Intelligence, vol 51, 95-143, 1991.

[5] J. E. Hunt and C. J. Price, Explaining Qualitative Diagnosis, Engineering Applications of Artificial Intelligence, pp161-169, vol 1(3), September 1988.

[6] C. J. Price and N. S. Taylor, Automated Multiple Failure FMEA, Reliability Engineering and System Safety 76(1), pp1-10, 2002.

[7] P. Struss, *Testing Physical Systems*, in Proceedings AAAI-94, pp251–256, AAAI Press, 1994.

[8] J. de Kleer and B. Williams, *Diagnosing Multiple Faults*, Artificial Intelligence 32, pp97–130, 1987.

[9] J. de Kleer and J. S. Brown, *A Qualitative Physics based on Confluences*, pp. 7–83 in [10].

[10] D. Bobrow (ed), *Qualitative Reasoning About Physical Systems*, North-Holland, 1985 (originally published as Artificial Intelligence vol. 24, 1984).

[11] R. Leitch, Q. Shen, G. Coghill and M. Chantler. Choosing the right model. In IEE Proceedings - Control Theory and Applications, volume 146, pages 435-449, 1999.

[12] L. Travé-Massuyès and R. Milne. Application-oriented qualitative reasoning. Knowledge Engineering Review, 10(2) pp181-204, 1995.

[13] G. Biswas, M-O. Cordier, J. Lunze, L. Travé-Massuyès, M. Staroswiecki. Diagnosis of complex systems: Bridging the methodologies of the FDI and DX communities (Introduction to special section), IEEE Trans. SMC, vol. 34, No. 5, pp2159-2162, 2004.

[14] P. Struss and C. J. Price, Model-based systems in the automotive industry, AI Magazine special issue on Qualitative Reasoning, 17-34, vol. 24(4), Winter 2003.

[15] B. C. Williams, P. P. Nayak, A Model-based Approach to Reactive Self-Configuring Systems, Proceedings of AAAI-96, Portland, Oregon, p. 971-978, August 1996.

[16] B. C. Williams, P. P. Nayak, A Reactive Planner for a Model-based Executive, Proceedings of IJCAI-97, Nagoya, p. 1178-1185, August 1997.

[17] P.. Salles, R. Muetzelfeldt, and H. Pain. Qualitative Models in Ecology and their Use in Intelligent Tutoring Systems, Procs 10th Intl Workshop on Qualitative Reasoning, pp216-224, 1996.

[18] U. Heller and P. Struss. Transformation of Qualitative Dynamic Models - Application in Hydro-Ecology, Procs 11th Intl Workshop on Qualitative Reasoning, pp83-92, 1997.

[19] P. Salles and B. Bredeweg. Building Qualitative Models in Ecology, Procs 11th Intl Workshop on Qualitative Reasoning, pp155-164, 1997.
[20] J. Keppens and Q. Shen Towards Compositional Modelling of Ecological Systems via Dynamic Flexible Constraint Satisfaction, Procs 14th Intl Workshop on Qualitative Reasoning, pp74-82, 2000.
[21] P. Salles, B. Bredeweg, S Araújo, Qualitative models on stream ecosystem recovery: exploratory studies, Procs 17th Intl Workshop on Qualitative Reasoning, pp155-162, 2003.
[22] MONET information and documents are available online at *http://monet.aber.ac.uk*.
[23] P. Lucas, R. Kuipers, F. Feith. A system for pacemaker treatment advice, Proceedings ECAI 2004, pp. 735–739, 2004.
[24] K. Leelawong, J. Davis, N. Vye, G. Biswas, et al. The Effects of Feedback in Supporting Learning by Teaching in a Teachable Agent Environment, in 5th Intl Conf on the Learning Sciences, pp245-252, 2002.
[25] K. D. Forbus, K. Carney, B. Sherin, L. Ureel, VModel: a visual qualitative modeling environment for middle school children, 16th Innovative Applications of AI Conference, 2004.
[26] R. Bellazzi, R. Guglielmann, L. Ironi, C. Patrini: A hybrid input-output approach to model metabolic systems: An application to intracellular thiamine kinetics, Journal of Biomedical Informatics, 34(4), 221-248, 2001.
[27] E. Bénazéra, Diagnosis and Reconfiguration based on Hybrid Concurrent Models - Application to Autonomous Satellites, PhD thesis of Université Paul Sabatier, Toulouse, France, 2003.
[28] C. J. Price, L. Travé-Massuyès, R. Milne, L. Ironi, B. Bredeweg, M. H. Lee, P. Struss, N. Snooke, P. Lucas, M. Cavazza. Qualitative Futures, Procs 19th Intl Workshop on Qualitative Reasoning, pp29-37, QR'05, 2005.

Combining Dimensional Analysis and Heuristics for Causal Ordering – In Memory of Dr Rob Milne –

Qiang Shen [a,1] and Taoxin Peng [b]
[a] *Department of Computer Science, University of Wales, Aberystwyth SY23 3DB*
[b] *School of Computing, Napier University, Edinburgh EH10 5DT*

Abstract. This paper is a refined version of the work that the authors presented at the 13th International Workshop on Qualitative Reasoning, jointly with the late Dr Rob Milne. It is dedicated to Rob in recognition of his significant contribution and support for the research described herein.

The paper presents a novel approach for generating causal dependencies between system variables, from an acausal description of the system behaviour, *and* for identifying the end causal impact, in terms of whether a change in the value of an influencing variable will lead to an increase or a decrease in the value of the influenced variables. This work is based on the use of the conventional method for dimensional analysis developed in classical physics, in conjunction with the exploitation of general heuristics. The utility of the work is demonstrated with its application to providing causal explanation for a benchmark problem that involves a dynamic feedback loop. The results reflect well the common-sense understanding of the causality in such a system that is otherwise difficult to capture using conventional causal ordering methods.

Keywords. Causal ordering, Dimensional analysis, Causal impact, Fault diagnosis, Personal memories

1. Introduction

Knowledge of causality is essential in handling many application problems. In performing fault diagnosis, for example, when one or more variables are observed to have an abnormal value, it is necessary to find out what internal variables (which are perhaps unmeasurable) could have been causing the observed abnormalities, what other variables might be affected, and what other observations would be spuriously correlated to the observations of these variables. An explicit expression of the causal relations amongst system variables will enable the diagnostic system to generate convincing explanations for its findings.

Having recognised the great application potential, significant work has been developed for deriving a causal ordering between variables in a physical system [5,6,8,11,12]. However, such work only addresses the issue of determining those variables that may

[1]Correspondence to: Q. Shen, Dept. of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK. Tel.: +44 1970 621825; Fax: +44 1970 628536; E-mail: qqs@aber.ac.uk

affect other variables, without identifying the directions of causal effects in terms of whether an increase in the value of an influencing variable will lead to an increase or a decrease in the value of the influenced variables. Also, important limitations remain when attempting to utilise the existing techniques to draw inferences on the causal relations among variables dynamic systems that typically involve feedback loops [3]. Therefore, a technique which may reduce such restrictions is clearly very desirable.

To better understand causal relations between system variables, it is useful to investigate what represents the fundamental features of physical quantities and their relationships. When describing the behaviour of a physical variable, the most distinct characteristic is its physical dimension. For this reason, dimensional analysis [2,4] has long served as a basis upon which to create quantitative system models and/or to perform model-based simulation in classical physics and control engineering. Recently, it has also been applied to developing a method for qualitative reasoning [1], though this method does not establish an explicit causal ordering among system variables. Inspired by this observation, this paper presents a novel approach for analysing causality in physical systems by means of dimensional analysis in conjunction with general heuristics. Supported with such first principles about the physical systems, the proposed technique can be applied both to produce a causal ordering amongst system variables and to provide an identification for the directions of causal effects on the changes of variable values.

The rest of this paper is structured as follows. Theoretical foundations are introduced in the next section, showing the basis upon which to develop the present work. The causal ordering algorithm based on the dimensional analysis is then described in section 3. An illustrative application of this method is given in section 4. The technical part of the paper is concluded and further work pointed out in section 5. Finally, a brief note on Dr Rob Milne's significant contribution to this and the relevant work is given in section 6.

2. Theoretical Foundations

2.1. Dimensional Analysis

Informally, dimensional analysis [2,4] is a method by which information about a physical variable can be deduced from that of certain other variables, given a dimensionally consistent relationship between them. The present work makes use of the following three most basic properties of dimensional analysis.

- *Principle of dimensional homogeneity.* Given that $y = \sum_i a_i f_i(x_i)$ represents a physical law governing the behaviour of a certain system, with y and x_i being the system variables or their temporal derivatives and a_i being the corresponding parameters, all the $a_i f_i(x_i)$ must have the same dimension as y.
- *Product theorem.* If the value of a variable can be derived from measurements of given base variables $u, v, w, ...$, that value can then be written in the form of $Cu^{\alpha}v^{\beta}w^{\gamma}$ ..., where $C, \alpha, \beta, \gamma, ...$ are constants.
- *Buckingham's* Π*-theorem.* Given physical quantities $u, v, w, ...$ such that $f(u, v, w, ...) = 0$ is a complete equation (that reflects the underlying physical laws characterising the physical situation), then its solution can be written in the form $F(\Pi_1, \Pi_2, ..., \Pi_{n-r}) = 0$, where n is the number of arguments of f and r is the number of basic dimensions needed to express the variables $u, v, w, ...$; for all i, Π_i is a dimensionless number.

A dimensionless product Π can be generally expressed as follows:

$$\Pi_i = y_i \times (x_1^{\alpha_{i1}} ... x_r^{\alpha_{ir}})$$

where $x_1, ..., x_r$ are termed the repeating variables, $y_1, .., y_r$ the performance variables and $\{\alpha_{ij} | 1 \leq i \leq n - r, 1 \leq j \leq r\}$ the exponents.

Based on the laws of motion, three units are generally regarded as fundamental, namely, *length* (L), *mass* (M) and *time* (T) [4]. These are termed *base units* in this paper. Any other physical unit is regarded as a *derived unit*, since it can be represented by a combination of these base units. That is, the mass system is herein adopted for the expression of physical dimensions. This is not the unique option, though. For instance, the study of electricity and magnetism has shown the value of including other dimensions as base units.

Given the use of the mass system, the dimensions of any physical variable can be represented in the general form of $M^\alpha L^\beta T^\gamma$. This representation of a variable is referred to as the dimensional representation of that variable. Following such a representation, for example, the physical quantity *force* can be dimensionally represented by MLT^{-2}, *pressure* by $ML^{-1}T^{-2}$ and *velocity* by LT^{-1}.

2.2. Π-*calculus*

To be self-contained, the Π-calculus [1] that is employed in this work is briefly summarised below. For simplicity, each dimensionless number, Π_i is called a *regime*, and a collection of regimes is called an *ensemble* (typically, in modelling physical systems, an ensemble can be seen as a distinct component). If x_k is a variable that occurs in both regimes Π_i and Π_j, then x_k will be referred to as a *contact variable* between these two dimensional products. The set of variables x_j, $1 \leq j \leq r$ that repeat in each regime is called the *basis* of an ensemble.

There are three levels of calculation, depending on the relationships among the dimensionless products:

1. *Intra-regime analysis*. This is within a regime and provides the following method for calculating partial derivatives:

$$\partial y_i / \partial x_j = -(\alpha_{ij} y_i) / x_j$$

2. *Inter-regime analysis*. This is across regimes and gives the following method for calculating partial derivatives:

$$[\partial y_i / \partial y_j]^{x_p} = (\alpha_{ip} / \alpha_{jp})(y_i / y_j)$$

   where x_p is a contact variable for regimes Π_i and Π_j.
3. *Inter-ensemble analysis*. This is across ensembles and extends the inter-regime analysis to calculate inter-ensemble partial derivatives.

When the sign of a partial derivative, $\partial y_i / \partial x_j$ (or $[\partial y_i / \partial y_j]^{x_p}$), is obtained, the causal effect between ∂y_i and ∂x_j can then be inferred. For instance, if $\partial y_i / \partial x_j > 0$, then an increase in x_j will lead to an increase in y_i; if $\partial y_i / \partial x_j < 0$, then an increase in x_j will lead to a decrease in y_i. However, the calculation of the inter-ensemble partials is not straightforward. It will need domain-specific knowledge about connections between components. This issue will be discussed further in the next section.

3. The Proposed Approach

3.1. Basic Notations

To start with, the dimensional representation of a variable x is hereafter denoted by $D(x)$. Therefore, for *force* f, *pressure* p and *velocity* v, the following holds: $D(f) = MLT^{-2}$, $D(p) = ML^{-1}T^{-2}$ and $D(v) = LT^{-1}$.

Definition 1. Two physical variables x_1 and x_2 are *equivalent* to each other if and only if they have the same dimensional representation, i.e. $D(x_1) = D(x_2)$.

This definition correctly imposes a unique dimensional representation of physical variables within any given system.

Definition 2. A system variable is regarded to be *exogenous* if it is controlled only by factors external to the system; other variables are called endogenous. An equation that indicates a variable to be exogenous is called an *exogenous equation*; other equations are termed endogenous equations.

This definition reflects the understanding that if a variable is exogenous, its derivative is conventionally treated as exogenous as well. Exogenous variables are determined during the system modelling process. They jointly set up the scope of the physical system being modelled, separating it from the rest of the world. Such variables play an important role in causal ordering (and, indeed, in any system modelling approaches [9]).

In systems modelling, variables are normally represented as a function of time. Hence, the base unit *time* (T) is generally regarded as more fundamental than the other two base units *length* (L) and *mass* (M). The following definition reflects this heuristic.

Definition 3. For a given endogenous variable, x, the number of different base units appearing in its dimensional representation, excluding the dimension *time* if it has negative power, is called the *degree of commitment* of the variable, denoted as $cd(x)$; the algebraic sum of the exponents of all the base units involved is called the *degree of factorisation* of the variable, denoted as $fd(x)$; and the algebraic sum of the exponents of the base units, excluding that of dimension *time* if it has negative power is called the *degree of factorisation excluding negative time* of the variable, denoted as $fd\_t(x)$.

For example, given the dimensional representation of force $D(f) = MLT^{-2}$, the degree of commitment of force $cd(f) = 2$, the degree of factorisation $fd(f) = 1 + 1 + (-2) = 0$ and the degree of factorisation excluding negative time $fd\_t(f) = 1 + 1 = 2$.

A functional relation between two variables which is represented by an equation is reversible (or symmetric as referred to in [6]), if (i) x is expressed as a function of y and (ii) y can also be expressed as a function of x. However, causal relations are irreversible because the observation that x causes y clearly does not imply that y causes x. The purpose of causal ordering is, after all, to find the causal relations among the variables in a given model, which will convert a set of reversible equations into a set of irreversible constraints amongst the system variables.

The question is, given a set of system variables, how to determine which variable may have a higher degree of freedom to change its state? A variable is of a higher degree of freedom if it is more independent of, or less dependent upon the states of other system variables. Intuitively, variables with a lower cd and/or a lower fd_t value seem to be of a higher freedom degree and hence more independent. For example, suppose that a given system model includes the following two variables, *force* f and *velocity* v, with $D(f) = MLT^{-2}$ and $D(v) = LT^{-1}$. Hence, $cd(f) = 2$, $fd\_t(f) = 2$ and $cd(v) = 1$,

$fd\_t(v) = 1$. Using the above heuristics, f is treated as more dependent than v. A natural deduction from this is that if they both appear in the same equation then *force* can be regarded as depending on *velocity* at any given time instance, but not vice versa. This agrees with common-sense understanding of the physical relation between these two physical quantities.

In addition, variables that have a dimension of $time$ with negative power seem to be more independent, and the higher the negative exponent of the time dimension that a variable has, the less reliant its current state is upon the current states of other variables. In dynamic systems, this implies that, given two variables of different amounts of power on the time dimension and both being involved within one equation with the same cd and fd_t, the variable with a more negative time exponent tends to change its state before the other variable. For example, the change of physical quantity *velocity* v ($D(v) = LT^{-1}$) causes the change of quantity *length* s ($D(s) = L$), given $s = vt$ (i.e. motion with a constant speed), where t stands for the absolute time. This again matches the intuitive understanding of classical mechanics.

Generally speaking, a system is composed of some components. The behaviour of a system is determined by the behaviour of its components together with the specifications of their inter-connections, with the behaviour of each component being generally expressed as a set of equations. In this paper, the inter-connections between system components are specified using *structural constraints* which are imposed by the topological or geometrical linkages among these components. Such constraints imply the causal relations between the variables used to describe the boundary conditions of the components. This knowledge cannot be obtained from dimensional analysis, but from the design knowledge of the system [1]. Therefore, some of the variables may be treated as different types such that a variable may be regarded as an endogenous variable when a component is dealt with on its own, or it may be treated as exogenous when coupling between system components is considered.

Definition 4. Given an ordinary equation or a structural constraint relating system variables $u, v, ...; x, y, ...$, which may be quantitative or qualitative and may include temporal derivatives, $[x, y, ...] = [u, v, ...]$ is named a symbolic causal equation, which signifies that the variables on the left hand side (LHS) causally depend on the variables on the right hand side (RHS) (if such a causal relationship between the system variables considered can indeed be established). The order of those variables appearing on the left or right side is arbitrary.

It is worth noting that, given an ordinary equation, it is not necessary to impose a restriction on the number of the variables appearing on either side of its corresponding symbolic causal equation. This differs from the conventional representation of the irreversible (or asymmetric) causal equation [6], in which a constraint is imposed such that there is only one variable that is allowed to appear on the LHS. Further, in some cases, one side can be empty if all the variables involved in the original equation are causally inter-dependent.

For convenience, two binary predicates *Beless* and *Inequation* are introduced. Formula *Beless*(x_1, x_2) states that x_1 is less independent than x_2. Formula *Inequation*(x_1, x_2) states that variables x_1 and x_2 appear in the same equation within a given system model.

Observation. Let x_1 and x_2 be two endogenous variables in a given equation:

$$\begin{aligned} Beless(x_1, x_2) \iff & \ cd(x_1) > cd(x_2) \\ & \vee (cd(x_1) = cd(x_2) \wedge fd\_t(x_1) > fd\_t(x_2)) \end{aligned}$$

$$\vee(cd(x_1) = cd(x_2) \wedge fd\_t(x_1) = fd\_t(x_2)$$
$$\wedge fd(x_1) > fd(x_2))$$

A joint use of predicates *Beless* and *Inequation* allows the following concept to be defined.

Definition 5. A variable, x is called the *most dependent* variable in a given equation if it satisfies:

$$\neg\exists y(Inequation(x, y) \wedge Beless(y, x))$$

In order to deal with dynamic systems, which always involve differential equations, a further notion has to be introduced, which is a weaker form of the differential causality rule used in [6].

Definition 6. *Conditional Differential Causality Rule*. In a differential equation, if one of the temporal derivatives is one of the most dependent variables, then this derivative is causally dependent on all the other variables that are not a most dependent derivative within that equation.

This rule is weaker because (i) it does not require differential equations to be represented in first order, canonical form, which is required by many existing causal ordering methods (e.g. [6,8]), and (ii) it allows more than one most dependent derivative to co-exist on the left hand side of a symbolic equation. A differential equation is said to be in canonical form if and only if there is only one derivative in the equation, and the derivative is the only term appearing on the left hand side of the equation [6]. In addition, for the present approach, a derivative does not necessarily have to appear on the left hand side unless it is also a most dependent variable.

3.2. Causal Ordering Algorithm

Having defined the basic notations for representing potential causal relationships amongst variables in a system model, an efficient algorithm is devised to rearrange the model into a set of symbolic irreversible causal equations. This results in the following *dimensional analysis based causal ordering algorithm*, where the identification of variables' dimensions are done by performing a simple pattern matching procedure.

1. For each component in the system, do
   - (a) List the n variables that appear in the set of original equations, which define the behavior of the component, and identify their dimensional representations.
   - (b) For each endogenous variable, calculate its *cd*, fd_t and *fd*; for each exogenous variable, set its *cd*, fd_t and *fd* to zero.
   - (c) For each endogenous equation, create two sets: *Left* and *Right*, set them to be empty initially, and sort the variables in the equation with respect to the truth value of the binary predicate *Beless* and the *conditional differential causality* rule as follows:
     - i. Choose any pair of variables, x and y in the set of variables in the equation, do
       - A. If $Beless(x, y)$ is true, put y into set *Right*. If x is not in set *Left* then move all variables from set *Left* to set *Right* and put x into set *Left*. If x is already in set *Left*, then keep set *Left* unchanged. If there are no unsorted variables left, then stop; otherwise, choose another unsorted variable, say z, consider variables x and z and go to (i.A);

B. If $Beless(y, x)$ is true, use the same procedure as that in (i.A) but reverse the positions of x and y;

C. If neither $Beless(x, y)$ nor $Beless(y, x)$ is true, both x and y are put into set *Left*. If there are no unsorted variables left, then stop; otherwise, choose another unsorted variable, say z, consider the pair of variables x and z and go to (i.A);

ii. For the two sets of variables, *Left* and *Right*, if both of them are not empty, rearrange their elements to form the corresponding symbolic causal equation by doing the following:

A. If there are no derivatives appearing in set *Left*, then put variables in set *Right* to the right hand side (RHS) of the symbolic equation and variables in set *Left* to the left hand side (LHS);

B. If there are derivatives appearing in set *Left*, then put such derivatives to LHS and any other variables to RHS.

iii. Consider the variables on the right hand side. If the number of these variables is greater than one, then substitute the original set of variables in that equation with these variables (in order to identify, if any, further detailed causal relations amongst such variables) and go to (i).

2. For every derivative variable dx in the system model, create a symbolic equation such that $[x] = [dx]$.
3. For any two components which are connected, generate the additional symbolic equations by directly putting the boundary variables of one component to the right hand side and those of the other to the left hand side, with respect to the causal implication indicated by the given structural constraints.

This algorithm produces a set of symbolic equations, which specifies the causal relations among all the system variables. It is composed of two parts, the first (steps 1 and 2) dealing with equations within a single component and the second (step 3) coupling any two components. The coupling method will be illustrated further with a particular example in the next section. This algorithm supports the well-known *integral causality* [6], which indicates that the state of a variable depends on its own derivative. This is obvious since the *fd* value of dx/dt is always less than that of x. The resulting symbolic causal equations can be interpreted as a causal graph, which is a directed graph with the nodes being system variables and the links being created by the following method:

For any pair of variables, x and y, create a directed link from x to y if they jointly appear in one original equation or one structural constraint, and x is in the RHS set and y in the LHS set of the corresponding symbolic equation.

Formally, a directed graph $G = (V, E)$ consists of a finite non-empty set V of elements called *nodes* and a set E of ordered pairs between the elements of V called *links*. For simplicity, the set of all links pointing away from (pointing to) v, $v \in V$, is denoted by $E_{out}(v)$ ($E_{in}(v)$), and the set of all nodes associated with $E_{out}(v)$ ($E_{in}(v)$), excluding v, is denoted by $V_{out}(v)$ ($V_{in}(v)$). A node, v, is called an end node if $|E_{out}(v)| = 0$ or $|E_{in}(v)| = 0$, where $|E|$ denotes the number of the elements in set E.

In describing the behavior of certain system components or their relationships, the defining sets of equations and/or structural constraints may include more than one most dependent variable. This algorithm puts all such most dependent variables or derivatives into the LHS of the resulting symbolic causal equations, rather than artificially choosing just one of them to be put in the LHS. Although this may sound conservative, this treatment has so far appeared to produce reasonable and intuitive results.

There are at least two reasons that support the present approach. First, some of the variables in a model may represent physical quantities that stand in a fully equivalent position. It is too difficult, if not impossible, to tell the causal relations among them. Second, should there be additional information available on the description of the system, this may be utilised to discriminate further causal relations without being forced to assume one of the variables to be the most dependent and later to retract such assumptions. This approach is, therefore, more flexible compared to the approach where only one variable is allowed to be the derived variable regarding any one equation, an approach that is typically employed in the existing causal ordering techniques.

Note that step (1.c.iii) is a recursive procedure, allowing the algorithm to maximise the exploitation of dimensional information embedded in the system variables. This is justified on the ground that for any pair of variables which are inter-related, with both involved in one equation, if they are of a different independent level, then one is dependent on the other. However, empirically, in some cases this treatment may generate certain causal relations that are too detailed to be necessary for the explanation of the system's behaviour under consideration.

Another point worth mentioning is that the above algorithm requires no explicit equations, quantitative or qualitative, to be actually given, but an implicit indication of there being a relation between the variables involved. This is due to the fact that the causal relationships are generated by analysing the dimensions of each variable appearing in the given relations alone.

3.3. *Algorithm Extension*

The above algorithm, as with other existing causal ordering methods, only returns a description of cause-effect relationships between system variables given a set of reversible equations and a set of structural constraints. No identification of the actual causal impact is provided, in terms of whether an increase in the value of an influencing variable may lead to an increase or a decrease in the value of the influenced variables. Fortunately, a useful technique [1], also based on the dimensional analysis, for deriving such identification from a given physical device has been developed. Yet, this technique relies on the construction of the so-called regimes in the Π-calculus which, in turn, relies upon the identification of those variables that form the basis upon which to describe the system component being modelled. Although some heuristics that may help the selection of such variables are given in [1], a more explicit method is provided here, which covers those heuristics as specific cases:

1. For any component, write the set of all variables in the component as S, set $S_r = \phi$, where r is the size of the basis;
2. Find a variable x in S, satisfying

$$\neg\exists y\ (y \in S \wedge Beless(x, y))$$

   (a) If there is only one such variable, then this variable is selected and is put into set S_r;
   (b) If there are more than one such variable, then choose any one of them and put it into the set S_r.

   (c) Set $S = S - S_r$

3. If $|S_r| = r$ then stop, the set S_r is the basis for the corresponding component; otherwise go to step 2.

Although temporal derivatives are allowed in the causal graph generated by the above causal ordering algorithm, only the relations among variables are considered in the Π-calculus. In order to integrate the Π-calculus into the causal ordering algorithm, the causal graph needs to be modified to remove the derivatives. For this purpose, the *transitivity* that causality possesses as a basic characteristic is exploited. Transitivity states that x causes y and y causes z implies that x causes z. Thus, two causal links: a link from a variable x to a derivative dy and a link from the derivative dy to a variable z can be replaced by a link from the variable x to z. With this transitional procedure, if z is itself also a derivative and there is a link from z to another variable u, then the path from x to u can be replaced by a single link from x to u, and so on. The graph resulting from the use of the transitivity shows the dependencies amongst system variables only, excluding all the temporal derivatives. Such graphs are hereafter named *derived causal graphs*.

Given $G = (V, E)$ being a causal graph including some derivatives as its nodes, a derived causal graph $G' = (V', E')$ can be obtained from G using the following procedure:

1. Set $V' = V$ and $E' = E$.
2. Select a derivative node dx in V', do
   (a) If dx is not an end node, set $E' = E' - \{E'_{out}(dx) \cup E'_{in}(dx)\} \cup \{u \rightarrow v | u \in V'_{in}(dx), v \in V'_{out}(dx)\}$
   (b) If dx is an end node, set $E' = E' - \{E'_{out}(dx) \cup E'_{in}(dx)\}$
3. Set $V' = V' - \{dx\}$. If there are no derivative nodes left, stop; otherwise go to step 2

A derived causal graph can then be analysed by means of the Π-calculus, in order to obtain a *causal graph with impact signs*. In such a graph, links are annotated with a causal influence impact sign, $+$ or $-$. The $+$ sign from variable x to variable y signifies that x causes y to change in a monotonically increasing way, and the $-$ sign between them indicates that x causes y to change in a monotonically decreasing way. Such information is clearly very useful for performing many qualitative reasoning tasks [1,7,8].

The change from an ordinary causal graph to the corresponding causal graph with impact signs may appear to lose some of the representational power of the original, since derivatives are eliminated. However, the derivation of the new graph does not necessarily destroy the existing one. Detailed causal links involving derivatives can be readily preserved by recording the original causal graph, although there is no information on the actual impact of how an influencing variable may cause the influenced variable to change there.

4. Results

The above extended causal ordering algorithm has been implemented and applied to a number of interesting system models. To demonstrate the essential ideas, this section presents the results of applying it to the commonly used pressure regulator model [7].

The function of the pressure regulator is to maintain a constant pressure at its output. This device consists of two components: a pipe with an orifice and a spring valve, as

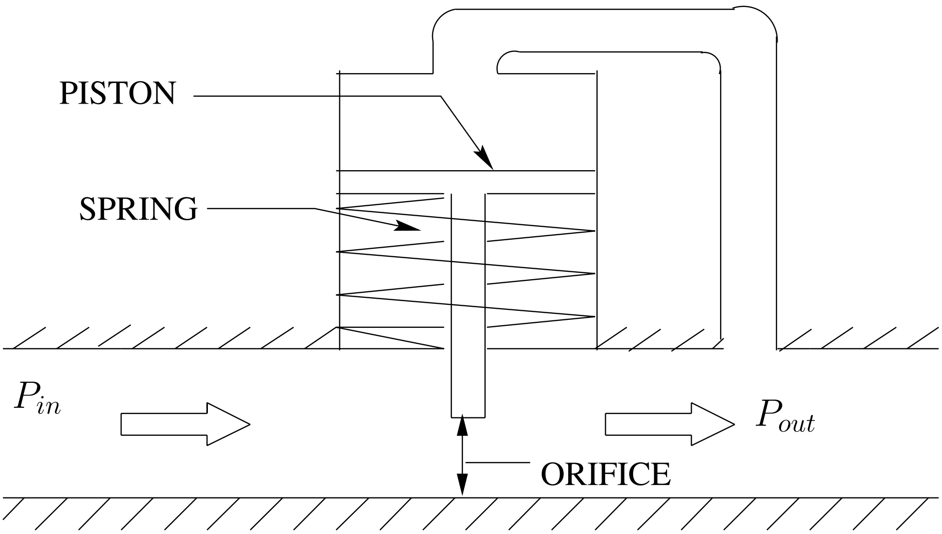


Figure 1. Pressure regulator

Table 1. Meaning of the variables

Variable	Variable Meaning
P_{out}	the outlet pressure
P_{in}	the inlet pressure
Q	the orifice flow rate
A_{open}	the orifice opening
ρ	the fluid density
x	the spring displacement
P	the pressure on the piston
k	the spring constant

shown in Figure 1. Each component is modelled individually first and the two components are then coupled to form the system model, subject to given structural constraints (see below). The meaning of the system variables is listed in Table 1.

The component of the pipe with an orifice can be modelled by the following equations:

$$\begin{array}{ll} P_{out} = c_1 Q & Q = (c_2 P_{in} + c_3 A_{open})/c_4 \rho \\ P_{in} \ = c_5 & \rho \ = c_6 \end{array}$$

where P_{in} and ρ are exogenous, and c_1 - c_6 are positive constants. The spring valve component can be modelled by:

$$-kdx = c_7 dP \qquad P = c_8$$

Table 2. Dimensional values

Variables	**Dimensions**	*cd*	*fd_t*	*fd*
P_{out}	$L^{-1}MT^{-2}$	2	0	-2
P_{in}	$L^{-1}MT^{-2}$	0	0	0
ρ	ML^{-3}	0	0	0
Q	L^3T^{-1}	1	3	2
A_{open}	L^2	1	2	2
x	L	1	1	1
P	$L^{-1}MT^{-2}$	0	0	0
k	MT^{-2}	0	0	0

In this model, c_7 and c_8 are positive constants and k is an exogenous variable whilst P is regarded to be exogenous with respect to the spring valve component, and so is the derivative dP. The dimensional values for each variable are worked out as shown in Table 2.

There are two structural constraints between the two components. One is the connection that transmits the outlet pressure in the pipe to the piston in the spring valve component. That is, the rate of change of the pressure on the piston, dP is determined by the outlet pressure in the pipe, P_{out}. The other constraint is that the motion of the piston affects the orifice opening; more specifically as the spring is compressed, the orifice reduces. This constraint indicates that the displacement x determines the opening A_{open}. Running the dimensional-analysis based causal ordering algorithm results in the following symbolic equations:

$$\begin{array}{llll} [P_{out}] &= [Q] & [Q] &= [P_{in}, \rho, A_{open}] \\ [A_{open}] &= [P_{in}, \rho] & [dx] &= [k, dP] \\ [dP] &= [P_{out}] & [A_{open}] &= [x] \\ [P] &= [dP] & [x] &= [dx] \end{array}$$

This leads to the causal graph as shown in Figure 2, including a feedback loop:

$$Q \rightarrow P_{out} \rightarrow dP \rightarrow dx \rightarrow x \rightarrow A_{open} \rightarrow Q$$

From this causal graph, a derived causal graph without derivatives involved can be generated (but is omitted here), according to the algorithm extension given in the previous section. For instance, the link from P_{out} to x comes from the path $P_{out} \rightarrow dP \rightarrow dx \rightarrow x$, while the link from P_{out} to P comes from the path $P_{out} \rightarrow dP \rightarrow P$.

In the pipe component, five variables and three dimensions are involved. It is easy to select Q_{in}, ρ and A_{open} as the basis by the algorithm extension. Similarly, P and k can be chosen as the basis for the spring component. Applying the Π-calculus to the derived causal graph, partial derivatives and hence the causal graph with impact signs for the pressure regulator can be obtained. The result is shown in Figure 3. To illustrate how the partials are calculated, consider the pipe component in which there are two regimes:

$$\Pi_1 = (Q\rho^{1/2})/(A_{open}P_{in}^{1/2}) \qquad \Pi_2 = P_{out}/P_{in}$$

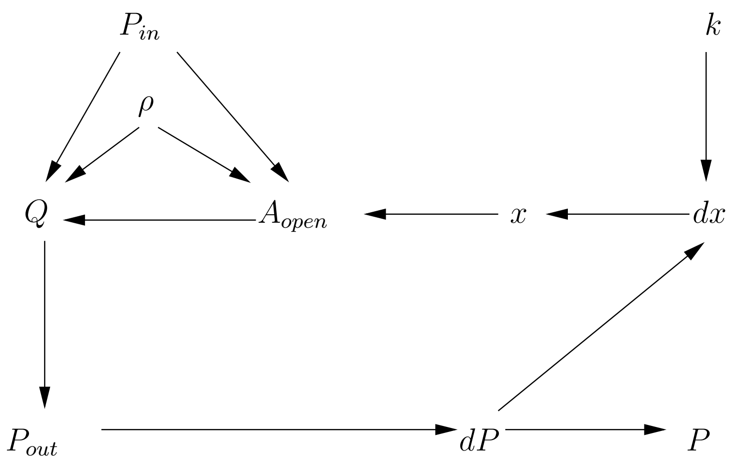


Figure 2. Causal graph for the pressure regulator

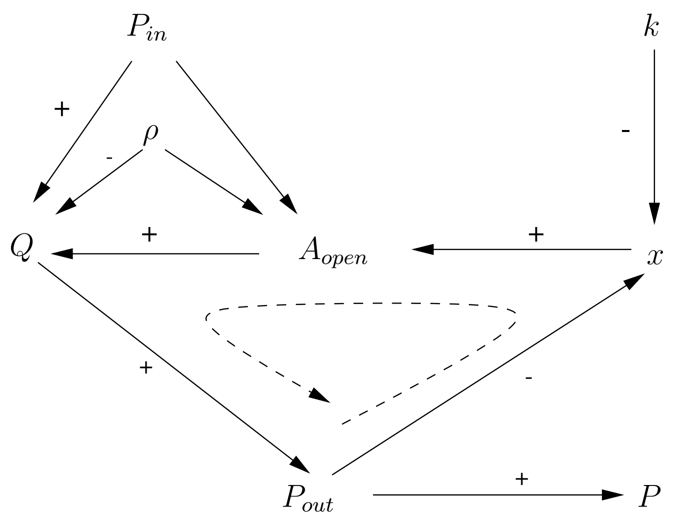


Figure 3. Causal graph with impact signs

The relative intra-regime partials are calculated as follows:

- From Π_1: $\partial Q/\partial P_{in} > 0$, $\partial Q/\partial A_{open} > 0$, $\partial Q/\partial \rho < 0$.
- From Π_2: $\partial P_{out}/\partial P_{in} > 0$.

Given P_{in} as a contact variable, the corresponding inter-regime partial is calculated by

$$[\partial P_{out}/\partial Q]^{P_{in}} = (\partial P_{out}/\partial P_{in})/(\partial Q/\partial P_{in}) > 0.$$

Finally, from the resulting causal graph, the behaviour of the pressure regulator can be explained as follows:

An increase in P_{in} leads to an increase in Q. This increase in Q leads to an increase in P_{out}. The increase in P_{out} leads to a decrease in the spring displacement x. This decrease in x leads to a decrease in A_{open} in the pipe orifice component. The decrease in A_{open} leads to the decrease in Q. Finally, this decrease in Q leads to a decrease in P_{out}.

This confirms that there is a (negative) feedback loop: an increase in P_{out} eventually leads to a decrease in P_{out}. This reflects the correct desired functionality of the system - to prevent the outlet pressure from deviating from a preset constant value.

It is interesting to compare the present technique with those proposed in [5,6,8,12], regarding the treatment of multi-component devices. In this work, if a device is composed of more than one component, each component is considered separately first and the structural constraints among the components are then taken into account while coupling pairs of the components. The pressure regulator example shows that this treatment is quite successful. However, using the existing methods, additional explicit equations must be given in order to represent the component connections or specific functionalities such as feedback loops. Finally, most of these approaches (including [5,6,8,12]) do not provide the impact signs of the actual causal effects, whilst the present work offers a mechanism for identifying such impacts.

5. Conclusions

This paper has presented a novel approach to generating a description of the causal relationships between system variables and, also, to identifying the actual impact of the causal effects by attaching a calculated positive or negative sign to each causal link, indicating whether a change in the value of an influencing variable will lead to an increase or a decrease in the value of influenced variables. The work rests its theoretical foundations on the conventional dimensional analysis developed in classical physics and the Π-calculus [1]. Experimental results have shown that the present approach retains the most appealing characteristics of the existing causal ordering techniques, being able to produce a causal explanation that reflects intuitive understanding of causal dependencies amongst the system variables.

However, further research remains to be done. This includes: a) investigating the possibility of integrating the present method with one or more other existing causal ordering algorithms, in order to maximise their benefits to generate better causal explanations; b) comparing the efficacy of the present approach with the most recent developments in causal modelling of dynamic systems (e.g. the approach based on Bond graphs [13]); and c) exploiting its potential in supporting model-based fault diagnosis.

6. Personal Memories

This research was jointly done with Dr Rob Milne. In fact, this article was based on a paper that Rob and we ourselves presented at the 13th International Workshop on Qualitative Reasoning [10]. The article is dedicated to Rob in recognition of his significant contribution to both the original ideas and the formulation of the technical methods presented in this paper.

Rob had been in collaboration with us since the mid-nineties, involving a good number of research and development projects, which ranged from UK EPSRC-funded, Rob's own company-sponsored, to undergraduate and MSc projects carried out within the University of Edinburgh. All of these projects had a clear focus on creating automated methods to generate intuitively meaningful and scientifically sound explanations for physical system (including process) behaviours. While being extremely busy as the Managing Director of Intelligent Applications Ltd (IAL), Rob always put in substantial effort in initiating and co-supervising these projects.

Rob was not only an entrepreneur and a world-leading researcher, but also an excellent educator – with his assistance and guidance, many of our project students have produced excellent results and graduated with distinction. Indeed, working with the real-world problems presented by Rob and finding solution methods through Rob's insightful suggestions, many of our students' project reports have led to externally peer-reviewed publications. Also, some of the prototypical software resulting from these projects has been revised and incorporated into their demonstration packages by IAL.

In particular, for the research reported herein (which was funded jointly by EPSRC and IAL), Rob had worked very closely with us. This paper summarises the work that has benefited considerably from many penetrating comments and constructive suggestions made by Rob, who was enthusiastically involved in monitoring and co-directing throughout the entire lifespan and beyond of the relevant research project. Starting from the very beginning when the grant proposal was constructed, Rob made many useful remarks on presenting the case for support to the EPSRC and offered a substantial amount of investment in the project. Without Rob and his company's support, the grant might not have been awarded in the first place.

As a world leader in developing and applying fault diagnostic techniques to physical systems in general and turbine machines in particular, Rob had considerable first-hand experience gained from working with industrial end-users. He encouraged us to recognise the importance of being able to produce convincing explanations to the user for any practical software intended for use to find faults in the real world. Such software should be capable of explaining why and how the found faults would have causally led to the observed symptoms that triggered the diagnostic process.

However, in the new generation of knowledge based systems for fault diagnosis, the popularly employed system models tend to be acausal. Although model-based reasoning systems typically have a certain embedded mechanism to produce logically sound explanations for the found faults, they generally lack the efficacy of conventional expert systems in terms of producing user-friendly cause-effect descriptions. Having recognised this, Rob strongly supported our work on deriving causal ordering from acausal system models. His exceptionally close involvement in this project made him an invaluable team member. The discussions and scientific debate taking place at our regular monthly meetings with Rob made the progress of this work possible. In addition, Rob provided real-world problem cases to help illustrate typical causal relationships between system variables. For many situations (e.g. real-time systems with feedback loops), this proved both intellectually challenging and scientifically interesting, and formed foundations for the creation of a number of novel ideas.

It is worth noting that, after the original paper on causal ordering was published (jointly authored by Rob and ourselves), Rob had suggested that we look into the possibility of integrating this promising work within the Tiger system (which is of course well-known around the world as Rob's major contribution to the research and application of automated fault diagnostic systems). A small number of proof of concept projects were carried out by our MSc students, co-supervised by Rob and his colleagues. Moreover, we have had many discussions in exploiting the possibility of proposing a much larger research project in extending and integrating this piece of work into Tiger. Tragically, Rob has left us, not being able to help us further in pursuing this research. It is sad that we have lost a long-time friend and collaborator. Nevertheless, we are determined to continue this line of investigation, following the advice and guidance provided to us by Rob. We hope a much improved version of causal ordering algorithm can be combined with an advanced fault diagnostic system like Tiger in the near future, realising a wish that Rob left with us.

References

[1] R. Bhaskar and A. Nigam. Qualitative physics using dimensional analysis. *Artif. Intell.*, 45(1-2):73–111, 1990.

[2] E. Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev.*, IV(4):345–376, 1914.

[3] P. Gawthrop and L. Smith. *Metamodelling: Bond graphs and dynamic systems*. Prentice Hall, 1996.

[4] H. Huntley. *Dimensional Analysis*. MacDonald & Co. (Publishers) Ltd., 1952.

[5] Y. Iwasaki and H.A. Simon. Causality in device behavior. *Artif. Intell.*, 29(1):3–32, 1986.

[6] Y. Iwasaki and H.A. Simon. Causality and model abstraction. *Artif. Intell.*, 67(1):143–194, 1994.

[7] J. De Kleer and J.S. Brown. A qualitative physics based on confluences. *Artif. Intell.*, 24(1-3):7–83, 1984.

[8] M. Lee and P. Compton. Context-dependent causal explanations. In *Proceedings of QR-94*, pages 176–186.

[9] R. Leitch, Q. Shen, G. Coghill, and M. Chantler. Choosing the right model. *Control Theory and Applications*, 146(5):435–449, 1999.

[10] Q. Shen, T. Peng, and R. Milne. Dimensional analysis based causal ordering. In *Proceedings of QR-99*, pages 193–202.

[11] J. Top and H. Akkermans. Computational and physical causality. In *Proceedings of IJCAI-91*, pages 1171–1176.

[12] L. Travé-Massuyés and R. Pons. Causal ordering for multiple mode systems. In *Proceedings of QR-97*, pages 203–214.

[13] O. Zaki, K. Brown, and D. Lane. Automatic generation of software agents for physical model from a description. In *Proceedings of UKCI-05*, pages 54–60.

I-Ex: Intelligent Extreme Expedition Support

Stephen POTTER[1], Austin TATE[1] and Rob MILNE[2]
[1]*Artificial Intelligence Applications Institute, University of Edinburgh, UK*
[2]*Intelligent Applications Ltd., UK*
`http://www.aiai.ed.ac.uk/project/ix/`

Abstract. The aim of the I-X research programme is to provide a general framework for performing *mixed-initiative synthesis tasks*, along with a set of tools that supports its use. This framework arises from and builds upon seminal work at the University of Edinburgh in the field of Artificial Intelligence planning. In this paper we describe the framework and tools, before describing the application of I-X to the task of planning and coordinating expeditions to remote locations – such as an attempt on Everest. We call this application *I-Ex*.

Keywords. Synthesis tasks, planning, coordination, intelligible messaging

1. Introduction

An expedition such as an attempt on Everest obviously requires a lot of coordination and planning, first during the preparatory stages, later during the expedition itself, and finally in concluding the effort. Where previously this would primarily be done 'manually', the rise of the internet and the World-Wide Web allows the introduction of greater amounts of computer support and wider access to information from a variety of sources.

The aim of the I-Ex project is to attempt to provide a means to integrate and structure the expedition activity with information from disparate sources and to allow access to intelligent technologies where appropriate. This computer support builds on the more general I-X architecture, which has been developed within the Artificial Intelligence Applications Institute at the University of Edinburgh. Simply put, the purpose of I-X is to provide intelligent computer support to people who are performing some task together. This task may be, say, designing a car, or performing an emergency evacuation, or coordinating an attempt on Everest. The I-X architecture supplies a framework that encourages a methodological approach to the task, based on cycles of issue-raising, handling and resolution. This is underpinned by the 'intelligible messaging' of issues, activities and other information among the agents in the system. This allows users to manipulate, transform and transmit information in context-sensitive ways that continually aim to move the process forward. I-X's foundations in AI planning technologies allow the invocation of automated planners that are able to suggest potential plans for achieving task sub-goals, while its open architecture facilitates the use of external computational services and information.

In this paper we first describe in general terms the I-X approach, and the technology and tools that implement it. Then we discuss the I-Ex project as an application of I-X, describing the necessary preparations for deployment along with the deployment itself, and the lessons learned from these. Finally we, draw some concluding remarks about the future of task-achieving architectures of this sort.

2. The I-X Approach

I-X is a research programme with a number of different aspects intended to allow humans and computer systems to cooperate in the creation or modification of some product or products such as documents, plans, designs or physical entities – that is, it supports *mixed-initiative synthesis tasks* [1].

The I-X approach involves the use of shared models for task-directed cooperation between agents which are jointly exploring (via some process, which may have been dynamically determined) a range of alternative options for the synthesis of an artifact (termed a *product*) such as a design or a plan. The <I-N-C-A> (Issues-Nodes-Constraints-Annotations) ontology [1] is used to represent a specific product as a set of constraints on the space of all possible products in an application domain. This ontology can be used to describe the product requirements or specification and the emerging description of the product itself. It can also describe the processes involved.

I-X provides a modular systems integration architecture that mirrors the underlying <I-N-C-A> ontology. It provides a “Model-Viewer-Controller” style of architecture. Plug-in components for Issue Handlers, Activity Performers, Constraint Managers, I/O Handlers and Process or Product Viewers allow for specific I-X systems to be tailored for specific tasks. The I-X approach draws on earlier work on Nonlin [2], O-Plan [3-7], Optimum-AIV [8,9], <I-N-OVA> [10,11] and the Enterprise Project [12,13] but seeks to make the framework generic and to clarify terminology, simplify the approach taken, and to increase the re-usability and applicability of the core ideas.

2.1. I-X Process Panels

I-X Process Panels (I-P^2) are used to support individual users who are carrying out processes and responding to events in a cooperative working environment – they are the primary means by which useful functionality is delivered to users (Figure 1). A panel supports the tracking of personal or group issues, the planning and execution of activities and the checking of constraints. It can be used to communicate with other panels and any other known services, agents and other co-operative working support tools to form a network of activity and process support in an organisation. This communication is achieved via a range of strategies which vary from simple direct internet ports, through custom name server and brokering systems, to comprehensive, secure, agent communications platforms such as the CoABS Grid [14] and KAoS [15].

I-X Process Panels and their predecessors, the Open Planning Process Panels (O-P^3) [7], have been used in a number of prototype and deployed application areas including:

- Air campaign planning process support [5];
- Non-combatant evacuation operations [6];
- US Army small unit operations for hostage rescue [16];
- Coalition and multi-national forces command and control [17,18];
- Search & rescue coordination [19,20];

- Help desks;
- Unmanned autonomous vehicle command and control;
- Cooperative working between e-Scientists [21].

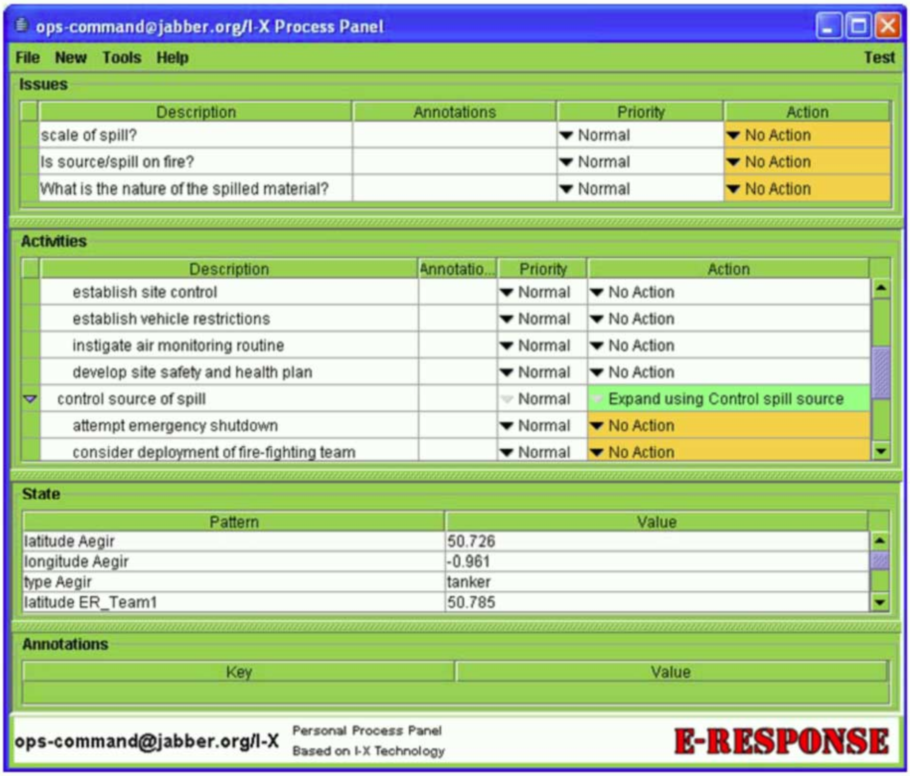


Figure 1. An I-X Process Panel (I-P^2), shown here providing an interface on to the response to a (simulated) marine oil spill, with *Issues* raised, the *Activities* currently in hand and the current *State*.

An I-X Process Panel can be seen, at its simplest, as an intelligent ‘to-do’ list for its user (the nature of the ‘intelligence’ of a panel will be the discussed below in section 2.2). However, and especially when used in conjunction with other users’ panels, it can become a workflow, reporting and messaging ‘catch all’, allowing the coordination of activity, and hence facilitating more successful and efficient collaborations. A panel corresponds to its user’s ‘view’ onto the current activity, through the presentation of the current items (from the user’s perspective) of each of the four sets of entities comprising the <I-N-C-A> model. The contents of these sets, along with the current context and state of the collaboration, are used to generate dynamically the support options the tool provides. For example, associated with a particular activity might be suggestions for performing it using known procedural expansions, for invoking an agent known to offer a corresponding capability, or for delegating the activity to some subordinate agent in the environment.

To summarise, an I-X Process Panel can accept requests to:

- Handle an issue;
- Perform an activity;

- Add a constraint;
- Note an annotation.

The panel allows its user to resolve issues and perform activities, and note the subsequent change in the state of the collaboration, through:

- Manual (user) activity;
- Invoking local computational capabilities (perform);
- Invoking external capabilities (invoke or query/answer);
- Rerouting or delegating to other panels or agents (pass);
- Planning and executing a composite of these strategies (plan or expand).

The panel receives "progress" or "completion" reports and other event-related messages and, where possible, interprets them to help the user:

- Understand the current status of issues, activities and constraints;
- Understand the current world state, especially status of process products;
- Control the interaction with other agents;
- Annotate the various elements of the model.

An I-X Process Panel can cope with partial knowledge and can operate even where little or no pre-built knowledge of the domain or knowledge of other panels and services is available – effectively becoming a simple to-do list and issue-tracking aid in that case.

The ease and freedom with which people use instant messaging tools in a variety of situations, and the realization that, in activity-oriented situations, many of the messages exchanged refer to items that could be described using the <I-N-C-A> ontology (such as particular issues, activities and various types of preferences and constraints), has led to the adoption of *intelligible messaging* for I-X. By intelligible messages, we mean messages that are structured, where appropriate, using <I-N-C-A> constructs; this provides the recipient (which may be a non-human agent) of any message with additional formal information about the content and encourages users to think about structuring their messages in a more formal, task-oriented manner, and allows access to the intelligence underpinning the I-X system. However, the users retain the ability to send 'normal' informal instant messages whenever this seems more appropriate – for example, when discussing things outside the scope of the activity. (An additional advantage of this approach is that by describing I-X Process Panels as providing instant messaging augmented with process, activity and task support, prospective users seem more readily able to grasp what the I-X approach can offer them.)

As I-X Process Panels have been developed and used in more cooperative and human-centric applications (such as in support of scientific meeting and group work [21]), intelligible messaging has come to influence the interfaces and become central to our approach. We have incorporated the use of a Jabber [22] communications strategy, which provides for Instant Messaging using XML content through third-party servers. This allows for simpler 'out of the box' deployments of the I-X Process Panels, and on a larger scale.

Running alongside and complementing the functionality offered by the process panels, the I-X tool suite provides a range of support and visualization tools (Figure 2).

Figure 2. The I-X tool suite.

2.2. I-Plan

The facilities available through the I-X Process Panels include a simple AI planner (I-Plan) used to provide context-sensitive options for the handing of issues, the performance of activities, and the satisfaction of constraints.

For any activity on the panel, an *Action* column shows its current status and, in a drop-down menu, the available options ("activity handlers") to perform the activity. The status of each activity is indicated by the colour shown in this column (as in Figure 1):

- White indicates that the item is currently not ready for execution (i.e., some temporal ordering, preconditions or other constraints might not be satisfied);
- Orange indicates that the action is ready to be performed and that all preconditions and constraints are met;
- Green indicates that the item is currently being performed;
- Blue indicates successful completion;
- Red indicates a failure for which failure recovery planning steps might be initiated.

The drop-down menu providing the set of options available to perform any item on the panel is dynamically generated and is context-sensitive, reflecting the local knowledge of the capabilities of other panels and available services. It also draws on the in-built planner and other plug-in modules providing specialized handlers. I-Plan can perform hierarchical partial-order composition of plans from a library of single-level plan schemas or *Standard Operating Procedures* (SOPs). This library can be augmented during the planning phase, either through a simple interface allowing the user to add specific ways to expand a given action (intended for users familiar with the application domain but not AI planning techniques) or with a more comprehensive graphical domain editor. Grammars and lexicons for the domain are built automatically during domain editing to assist the user.

The link to the underlying planner justifies our description of I-X as providing an *intelligent* system. The planner is able to match activity refinements (such as a decomposition of the activity into a number of simpler sub-activities) from the library and allow the user to access these by adding corresponding "expand" handlers to the action menu. At this point, instead of merely being able to tick-off an activity once done, users can exploit the knowledge in the library to break an activity down into sub-activities that, when all performed, accomplish the higher-level task. Of course, sub-activities can themselves be broken down further until a level of primitive actions is reached, at which point the library of procedures no longer contains any refinements that match the activities. One common way of handling an activity is through delegating the task to another (human or computer) agent in the system that is known to be capable of performing it. In this manner, the user collaboratively exploits the knowledge of other agents.

Future developments of I-Plan will provide more assistance with a "How do I do this?" option under the Action menu which will be able to account for other concurrent items on the panel, and account for mutual satisfaction of open variables and other constraints.

2.3. Other I-X Tools

The other tools in the I-X suite include messaging tools and various information viewers (e.g., map, 3D VRML and PDA interfaces) and editors, along with three specific tools: *I-DE*, *I-Q* and I-*Space*.

I-DE (I-X Domain Editor) allows the creation, maintenance and, ultimately, the publication of SOPs, generic approaches to archetypal activities.

I-Q (I-Query) is a generic I-X agent shell which, when embodied with the appropriate mechanisms, provides an agent with the ability to interact with a query service of some kind. It usually responds by adding facts or constraints into the current state of the panel. A typical application, for instance, might be the retrieval of information from some external source such as the Semantic Web.

I-Space is used to maintain organizational relationships with other agents in the environment. The nature of the relationship (for instance, supervisor-supervisee) will influence the nature of the activity-based interactions between these agents: the choices available to an agent will depend (amongst other things) both on its position in the organizational scheme of things and on its awareness of the capabilities and dynamic status (e.g. the current 'presence') of other agents. Exchange of agent and organization relationships with tools such as the KAoS Policy Administration Tool (KPAT) is possible [15].

2.4. I-X Message Formats

There are a number of messages that are used within the I-X Process Panels and that can be passed between panels and other services and agents.

- Issues, Activities, Constraints and Annotations;
- Current state information (as world state constraints);

- Plans (composites of Issues, Activities, Constraints and Annotations);
- Reports of progress or completion of nominated activities;
- Text-orientated ‘chat’ messages.

The first three relate to the core underlying ontology on which I-X is based. The other two message types provide status and other contextual information. Reports can be requested when the task of handling a certain activity is passed on to another agent; details of activity *progress* and *completion* (success/failure) are then reported to the original sender of the item. This provides a way to monitor activity progress and receive milestone reports.

2.5. State Information

Information about the current state of the environment can be passed to panels via *world state* constraints. These might come directly from sensors, or through some analysis or reporting system.

A specific type of state information we have found useful is the presence or status information maintained by instant messaging systems, so one can tell if another agent, panel or person is active and available for communications. The Jabber messaging protocol allows users to ‘subscribe’ to the presence of others; in this way, users can construct lists of their contacts, which the Jabber server dynamically augments with their current presence status.

Incoming completion reports and information about the current state sent as constraints can cause other activities to become executable when their pre-conditions or ordering constraints are satisfied.

I-X allows custom state information viewers to be added to expand or replace the default state viewer (a simple tabular representation) that is a constituent of a ‘standard’ I-P^2 panel. One example of a custom state viewer that has been found to be useful in a number of applications is the BBN OpenMap™ tool [23], which allows the information to be layered on a map (which might be topographical or something more conceptual, as the application demands). Changes to information in any viewer, or coming in via messages from outside of panels are synchronized.

2.6. The <I-N-C-A> Ontology

<I-N-C-A> (Issues-Nodes-Constraints-Annotations) is the basis of the ontology that underpins the I-X approach. It provides the framework for the representation used to describe processes and process products within I-X Process Panels, and gives a structure for the main types of activity-orientated I-X messages. <I-N-C-A> is designed to be a conceptual model that facilitates the expression and communication of task elements for both human and computer agents.

In <I-N-C-A>, both processes and process products are abstractly considered to be made up of a set of *issues*, *nodes* (corresponding to activities in a process, or parts of a physical product), *constraints* of various kinds (which describe relationships between nodes) and *annotations*, which provide rationale and other useful but less formal information. Before discussing in more detail each of the elements which comprise the <I-N-C-A> model, a brief discussion of the origins and development of the model will help give an understanding of its nature.

The forerunner of <I-N-C-A>, <I-N-OVA> [10], when first designed, was intended to act as a bridge to improve dialogue between a number of communities working on formal planning theories, practical planning systems and systems engineering process management methodologies. It was intended to reflect and support the then new work on automatic manipulation of plans, human communication about plans, principled and reliable acquisition of plan information, and formal reasoning about plans. It has since been utilized as the basis for a number of research efforts, practical applications and emerging international standards for plan and process representations. See [24] for a more detailed historical view of the relationships between <I-N-OVA>, the earlier work in AI on plan representations, and the results from the process and design communities and standards bodies.

2.6.1. Issues

Issues state unresolved questions about the task; these can concern unsatisfied objectives, questions raised as a result of analyses, etc. An issue can be thought of as implying potential further constraints to be added into the model in order to address the issue. Until recently, issues in I-X had a task or activity orientation to them, being mostly concerned with actionable items referring to the process underway – that is, with actions in the process space. Lately, however, we have adopted the gIBIS [25] approach of expressing issues as any of a number of specific types of question to be considered [26,27]. The types of questions advocated are:

- Deontic questions – what should we do?
- Instrumental questions – how should we do it?
- Criterial questions – what are the criteria?
- Meaning or conceptual questions – what does X mean?
- Factual questions – what is X? or, is X true?
- Background questions – what is the background to this project?
- Stakeholder questions – who are the stakeholders of this project?
- Miscellaneous questions – to act as a catch-all.

The first five of these are likely to be the most common in our task support environment. This is similar to the Questions-Options-Criteria approach [28], which was used for rationale capture for plans and plan schema libraries in our earlier work [29], and to the mapping approaches used in Compendium [30]. Compendium can in fact exchange its set of issues, activities and some types of constraints and annotations with an I-P^2 [21,31].

2.6.2. Nodes

The nodes in the specifications describe components that are to be included in the current design. Nodes can themselves be artifacts that can have their own structure with sub-nodes and other <I-N-C-A>-described refinements associated with them. When <I-N-C-A> is being used to describe processes, the nodes are usually the individual activities or their sub-activities. They are usually characterized by a *pattern* composed of an initial verb followed by any number of parameter objects, noun phrases, and qualifiers or filler words describing the activity. An example activity might be:

```
(transport package-1 from location-a to location-b)
```

Nodes can themselves be thought of as constraints that restrict the space within which an artifact may lie, with the sets of issues and constraints serving to restrict further the space of solution artifacts. Others have recognized the special nature of the inclusion of nodes (or activities) into a synthesized artifact (or plan) compared to all the other constraints that may be described. In the planning domain, Khambhampati and Srivastava differentiate Plan Modification Operators into *progressive refinements*, which can introduce new actions into the plan, and *non-progressive refinements*, which partition the search space with existing sets of actions in the plan [32]. They call the former genuine planning refinement operators, and think of the latter as providing the scheduling component.

2.6.3. Constraints

The constraints restrict the relationships between the nodes to describe only those artifacts within the design space that meet the requirements. The constraints may be split into *critical constraints* and *auxiliary constraints*; some constraint managers (solvers) can return ‘maybe’ answers to constraints of the latter type, indicating that the constraint being added to the model is satisfactory providing other critical constraints are imposed by other constraint managers. The ‘maybe’ answer is expressed as a disjunction of conjunctions of such critical or shared constraints. More details on the ‘yes/no/maybe’ constraint management approach used in I-X and the earlier O-Plan systems are available in [4].

The choices of which constraints are considered critical and which are considered auxiliary is itself a decision for an application of I-X, and specific decisions will need to be made about how to split the management of constraints within such an application. In other words, it is not pre-determined for all applications. A temporal activity-based planner would normally have object/variable constraints (equality and inequality of objects) and some temporal constraints (maybe just a simple *before* constraint) as the critical constraints. But, in a 3D design or a configuration application object/variable and other critical constraints (possibly spatial constraints) might be chosen. It depends on the nature of what is communicated between constraint managers in the application of the I-X architecture.

2.6.4. Annotations

The annotations add additional human-centric information or design and decision rationale to the information describing the artifact.

3. An Application of I-X: I-Ex

Rob Milne’s planned journey to the Himalayas, his eagerness to exploit fully the opportunities afforded by this trip and his interest in the I-X technology together provided us with a unique chance to apply I-X concepts to a new domain, namely that of coordinating an expedition to some remote place. The name given to this application was *I-Ex*, and it would test both the generality of the I-X and <I-N-C-A> concepts and the robustness of the I-X software under taxing conditions.

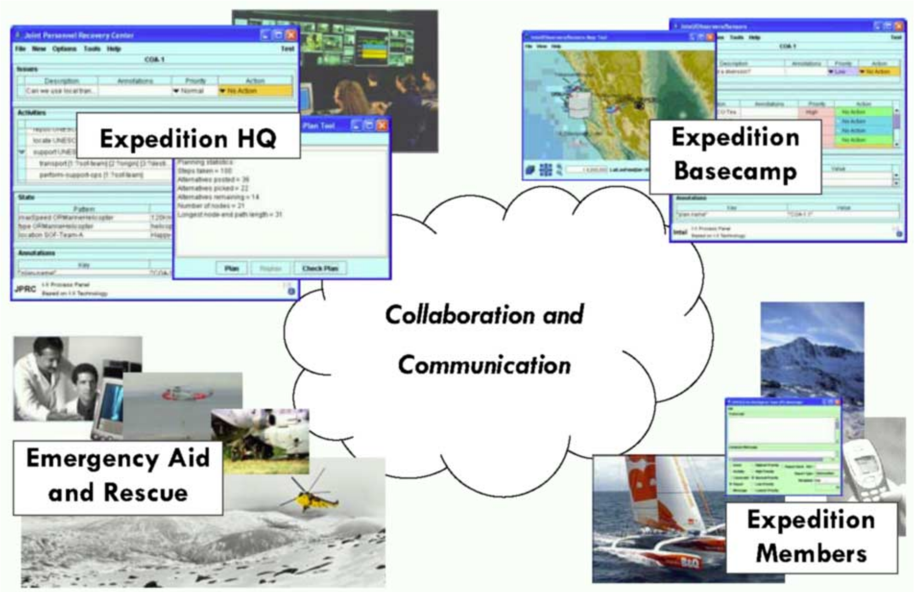


Figure 3. Overview of the I-Ex infrastructure.

Figure 3 gives an overview of the various components of I-Ex. The shared task is the expedition itself. In the event of, say, a significant deterioration in weather forecast over the coming hours, the I-X technology might be invoked to re-plan aspects of the climb. This might involve reasoning with available resources (for example, the number of oxygen bottles known to be stored at the South Col), temporal aspects (when will the bad weather pass over? how many hours of daylight are left?) and capabilities (available rescue services, if the situation worsens).

It was envisaged that there would be a number of different human agents involved:

- The mountaineer, carrying I-X on a lightweight Pocket PC/PDA, continually reports his progress and status to the base camp reporting system.
- Progress is monitored by an I-P^2 user at base camp (which has 'full' internet connectivity), and in turn supplies the mountaineer with next or revised objectives, latest weather updates, etc.
- An expedition support team, based, say, in the UK, and also using an I-P^2 is kept informed of progress, providing a full reporting system and updating the expedition web site accordingly.

In actuality, time and technology constraints meant that in this instance we would concentrate on the monitoring and communications aspect of the last two of these, effectively meaning that we would have an 'I-X-enabled' laptop at base camp that would communicate progress information to the 'support team' at Edinburgh.

3.1. Tailoring for I-Ex

Before deployment, it was necessary to tailor the general I-X system and tools for this particular application. These preparations concerned communications, task modelling, system 'packaging' and establishing external reporting mechanisms.

3.1.1. Communications in I-Ex

Jabber was chosen as the primary medium for I-X communications between base camp and the support team (with e-mail being used as a secondary channel for images and other non-I-X information). There were two principal reasons for this choice. First, Jabber allows asynchronous communications, in the sense that it is not necessary for the intended recipient of a message to be online when the message is sent: messages can be queued until the next time the recipient logs on. This becomes important when we cannot ensure that agents are online continuously (and where there is a time-difference between agents), and do not wish messages to be lost. However, there are other situations in which synchronous communications are desirable, such as time-critical tasks in which it is important that a particular activity begins almost immediately, and hence, candidates for performing it should exclude any off-line agents. In these cases I-X can prevent any messages being sent to off-line agents.

Secondly, the choice of Jabber meant that we were able to use a third-party server (namely jabber.org), already known from previous applications to be reasonably robust.

3.1.2. 'Packaging' I-Ex

In practice, 'the user' at base camp at any given time might be one of a number of individuals, all having different levels of computer skills and awareness of the I-Ex project objectives. In order to accommodate these different users, it was felt necessary to develop a basic 'reporting' version of I-P^2, which would be both easy to use and foolproof, as far as possible, in operation. The issues, state constraints and annotations would all be hidden in this reporting panel, leaving just the ascent plan in its current state. When appropriate, the user would simply select 'Done' from the action menu of the current activity; this would generate and send the appropriate reporting messages to the support team panel. In addition, the panel would automatically save the state of the current plan on closing (after having first generated a back-up file of the plan before the latest modifications) and reload it on subsequent start up.

While the aggregate effect of these changes was to limit the autonomy and flexibility of the base camp user and hide much of the underlying functionality, the lack of time and opportunity to familiarise the users meant that this seemed a small price to pay in return for robustness during the experiment.

3.1.3. Task Modelling for I-Ex

With Rob's input, we were able to develop an initial 'ascent plan'. Activities in this plan correspond to achieving particular targets, such as arrival at base camp, acclimatization at camp 1, through to reaching the summit and the subsequent descent to base camp. Some of these activities contain sub-activities; for instance, acclimatization at camp 1 involves the ascent to camp 1 and, after a suitable period, a return to base camp. Figure 4 shows part of this ascent plan (seen here in the minimal reporting panel).

The ascent plan is echoed in a 'monitoring plan', used to control the support team activity. This contains activities such as monitoring the arrival at base camp, which in turn consists of the sub-activities of awaiting a report of arrival at base camp and then updating the external progress reports on the web, to relate this information to a wider audience. Hence, every activity in the ascent plan has a counterpart in the monitoring plan, with the I-X reporting mechanism being used to automatically generate and send progress messages when the base camp user indicates that the activity is now 'Done'. Figure 5 shows the monitoring plan as seen in a standard I-P^2.

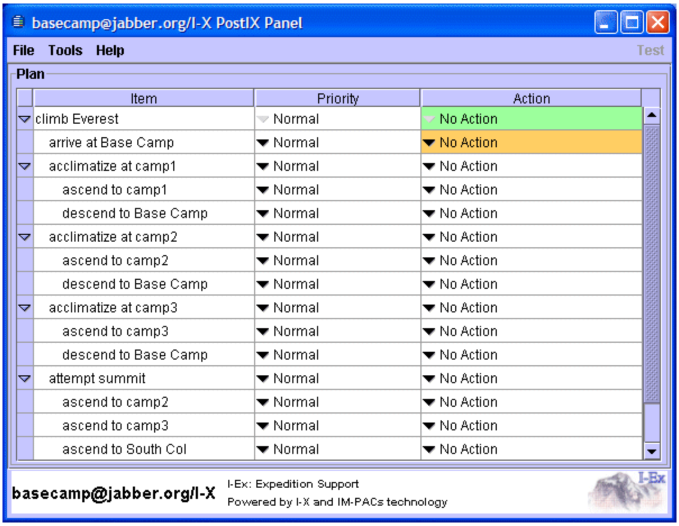


Figure 4. The base camp process panel, showing the initial activities in the ascent plan.

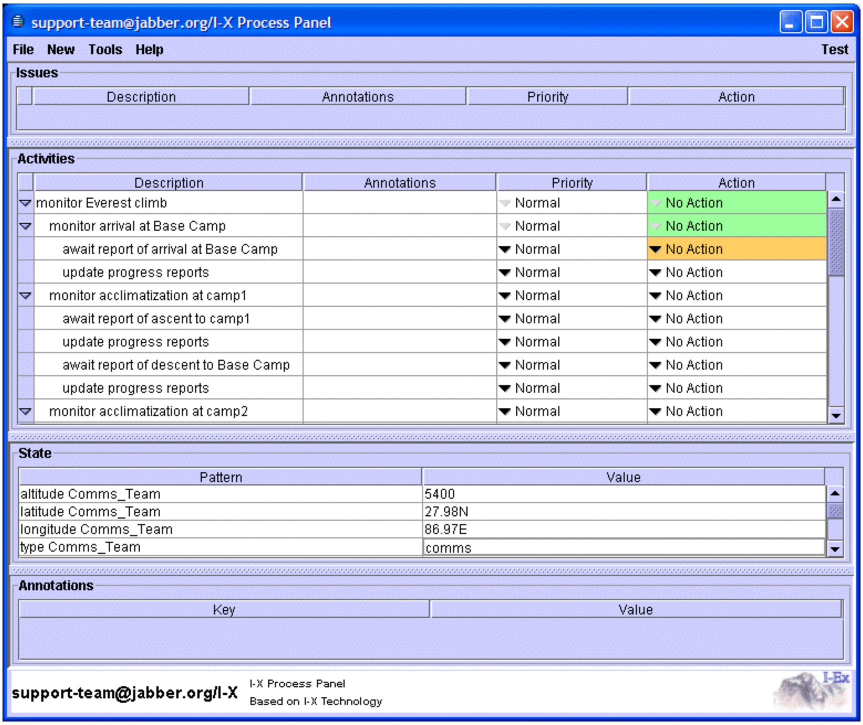


Figure 5. The support team process panel, showing the first steps in the monitoring plan.

3.1.4. External Reporting in I-Ex

A feature of many expeditions and journeys in the internet age is the expedition 'blog', which describes expedition progress and status, as well as photographs, more personal reflections, where appropriate, on the daily routine, surroundings, etc. The blog serves as a central point for broadcasting information intended to keep a general audience informed. Since reporting and monitoring were integral to the I-Ex project, it was suggested by Rob that the provision of a blog would form a natural extension to this, as well as providing him with a useful service. However, in addition to the 'normal' manual approach for updating the blog, we tried to introduce a 'semantic blogging' element, whereby I-X reports could, according to the message type, be automatically converted into the appropriate progress reports and posted on the web site, exploiting the meaning of the underlying <I-N-C-A> messages and the activity progress reporting mechanisms that are built-in to I-X. An autonomous I-X agent was developed for this task, but as it had not been adequately tested before the expedition team set off, it was not deployed. Instead, reports were manually converted into blog entries, and these were supplemented by conventional entries which were sent by e-mail. Figure 6 shows some typical blog entries that were written by Rob alongside a report generated from an I-X progress update.

3.2. I-Ex in Practice: Lessons Learned

The aims of the I-Ex deployment during the Everest expedition were somewhat limited in scope, being restricted to reporting tasks and not exercising the full potential of the I-X environment as a general task-achieving environment. However, within this scope, the deployed system was felt to be a success. The reporting aspects worked well, with the Jabber communications robust enough to enable transmissions when e-mail was unavailable, and the blog and associated web-site attracting much interest.

There are a number of more general lessons that we have been able to draw from the I-Ex application; these might be summarized in two points as follows:

- The importance of appropriate tailoring of the I-X tools: a mixed-initiative task is likely to involve humans of mixed abilities and backgrounds. It is important that the interfaces are modified accordingly to present an appropriate view onto the underlying model, and one which will enable the user to interact appropriately with the models and other participating agents.
- Reporting and monitoring: one of the key roles that technology such as I-X might come to play is in making explicit, in ways previously unimaginable, the current state of the task and its participants through the establishing and use of clear and, where appropriate, formal reporting structures, which may work on different levels: reporting between individual agents, reporting to groups of agents, broadcasting to the community of agents and possibly beyond.

4. Future Directions

Work on I-X and its applications in coalition command and control, search and rescue, help desks, etc. has indicated the value of adopting an 'augmented' style of the instant messaging paradigm which we call activity-orientated 'intelligible messaging'. This provides a platform for making AI planning technology available in an immediately usable form.

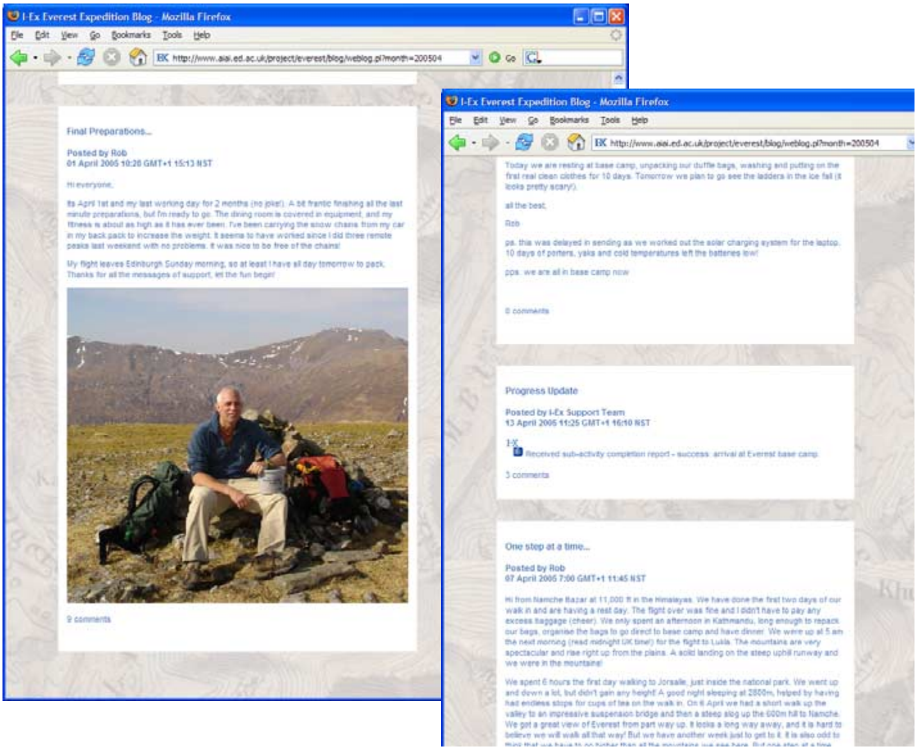


Figure 6. Some of Rob's blog entries. The *Progress Update* entry shown in the center of the right-hand browser was generated from the corresponding I-X activity completion report.

There are many opportunities for extending the initial approach and the technology that implements it. A more comprehensive I-Plan planner is to be incorporated as project work allows, and this will use the O-Plan 'repairing plans on-the-fly' repair technology [33] to recover from failures. The incorporation of more capable constraint managers and optimization algorithms to propose options is possible within the design.

More sophisticated and robust communications strategies are being studied, including logging of message traffic for quality control and audit purposes. We are particularly interested in the changes of process that might be triggered when the status of agents and panels alter (say moving from a synchronous on-line instant messaging mode of communication to an off-line store and forward messaging mode). Ways to describe panels and user/service capabilities in an organization, the roles they play and the authorities they have to act for one another are also being developed.

5. Conclusions

This paper has described the I-X approach to mixed-initiative synthesis tasks, whereby human and computer agents are assisted in their activity-oriented collaborations. The accompanying suite of tools provides human users with interactive visualizations of the current state of the collaboration from their particular perspectives, underpinned and supported with communications tools, planning aids and organization and domain modellers. As well as what might be thought of in conventional terms as synthesis tasks, the underpinning <I-N-C-A> ontology is general enough to allow the concepts to applied to more conceptual synthesis tasks such as mounting an expedition: the I-Ex application, though necessarily restricted in scope, is one example of this, with the communications and reporting aspects of the architecture firmly to the forefront.

More technical details on I-X and the current version of the I-P^2 software (available for demonstration purposes) are available at the project website [34].

6. Acknowledgements

Rob Milne is listed as an author of this paper, as we planned with him for a write up of the joint work on I-Ex. Rob assisted the I-X team in investigating the application of I-X technology to extreme expedition support and used a packaged version of the technology during his Everest expedition. He regularly reported to the Edinburgh team on his use of the system and provided valuable input throughout, and our intention was to jointly write-up the experiences of using I-Ex for academic publication.

This material is based on research within the I-X project and sponsored by a range of sources including the Defense Advanced Research Projects Agency (DARPA) and US Air Force Research Laboratory under agreement number F30602-03-2-0014. Thanks to our co-workers on the I-X project at AIAI (in particular Jeff Dalton, Jussi Stader and Gerhard Wickler) for supporting technologies and valuable discussions relating to this work.

The IM-PACs project is funded by Scottish Enterprise, the European Research Development Fund (ERDF) and AIAI at the University of Edinburgh. The project has received assistance from the Proof of Concept Fund managed by Scottish Enterprise. The Proof of Concept Funds supports the pre-commercialisation of leading-edge technologies emerging from Scotland's Universities, Research Institutes and National Health Service Trusts [Project Reference: 5-SEB-004].

7. References

[1] Tate, A. (2003) <I-N-C-A>: an Ontology for Mixed-initiative Synthesis Tasks, *Proceedings of the Workshop on Mixed-Initiative Intelligent Systems (MIIS)* at the International Joint Conference on Artificial Intelligence (IJCAI-03), Acapulco, Mexico, August 2003.

[2] Tate, A. (1977) Generating Project Networks, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-77)*, pp. 888-893, Cambridge, MA, USA, Morgan Kaufmann.

[3] Currie K. and Tate A. (1991) O-Plan: the Open Planning Architecture, *Artificial Intelligence*, 52, 49-86, 1991.

[4] Tate A. (1995) Integrating Constraint Management into an AI Planner, *Journal of Artificial Intelligence in Engineering*, 9(3), pp. 221-228, 1995.

[5] Tate, A., Dalton J. and Levine, J. (1998) Generation of Multiple Qualitatively Different Plan Options, *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems (AIPS-98)*, Pittsburgh PA, USA, June 1998.

[6] Tate, A., Dalton, J. and Levine, J. (2000) O-Plan: a Web-based AI Planning Agent, AAAI-2000 Intelligent Systems Demonstrator, in *Proceedings of the National Conference of the American Association of Artificial Intelligence (AAAI-2000)*, Austin, Texas, USA, August 2000.

[7] Levine, J., Tate, A. and Dalton, J. (2000) O-P3: Supporting the Planning Process using Open Planning Process Panels, *IEEE Intelligent Systems*, 15(6), November/December 2000.

[8] Aarup, M., Arentoft, M.M., Parrod, Y., Stokes, I., Vadon, H. and Stader, J. (1994) Optimum-AIV: A Knowledge-Based Planning and Scheduling System for Spacecraft AIV, in *Intelligent Scheduling* (eds. Zweben, M. and Fox, M.S.), pp. 451-469, Morgan Kaufmann.

[9] Tate, A. (1996a) Responsive Planning and Scheduling Using AI Planning Techniques - Optimum-AIV. In, "Trends & Controversies - AI Planning Systems in the Real World", *IEEE Expert: Intelligent Systems & their Applications*, 11(6), pp. 4-12, December 1996.

[10] Tate, A. (1996b) Representing Plans as a Set of Constraints – the <I-N-OVA> Model, *Proceedings of the Third International Conference on Artificial Intelligence Planning Systems (AIPS-96)*, Edinburgh, May 1996.

[11] Tate, A. (2000) <I-N-OVA> and <I-N-CA> - Representing Plans and other Synthesized Artifacts as a Set of Constraints, AAAI-2000 Workshop on Representational Issues for Real-World Planning Systems, *National Conference of the American Association of Artificial Intelligence (AAAI-2000)*, Austin, Texas, USA, August 2000.

[12] Fraser, J. and Tate, A. (1995) The Enterprise Tool Set – An Open Enterprise Architecture, in *Proceedings of the Workshop on Intelligent Manufacturing Systems, International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montreal, Canada, August 1995.

[13] Stader, J. (1996) Results of the Enterprise Project, in *Proceedings of Expert Systems '96, the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems*, Cambridge, UK, December 1996.

[14] Kahn, M.L. and Della Torre Cicalese, C. (2001) CoABS Grid Scalability Experiments, Autonomous Agents 2001 Conference, *2nd Int'l Workshop on Infrastructure for Agents, MAS, and Scalable MAS*, May 29, 2001.

[15] Bradshaw, J.M., Dutfield, S., Benoit, P. and Woolley, J.D. (1997) KAoS: Toward an industrial-strength generic agent architecture, In J. M. Bradshaw (Ed.), *Software Agents*, 1997, pp. 375-418, Cambridge, MA: AAAI Press/The MIT Press.

[16] Tate, A., Dalton, J., Jarvis, P. and Levine, J. (2000) Using AI Planning Technology for Army Small Unit Operations, Poster Paper in the *Proceedings of the Artificial Intelligence Planning and Scheduling Systems Conference (AIPS-2000)*, Breckenridge, Colorado, USA, April 2000.

[17] Allsopp, D.N., Beautement, P., Bradshaw, J.M., Durfee, E.H., Kirton, M., Knoblock, C.A., Suri, N. and Tate, A. and Thompson, C.W. (2002) Coalition Agents Experiment:Multi-Agent Co-operation in an International Coalition Setting, Special Issue on Knowledge Systems for Coalition Operations (KSCO), *IEEE Intelligent Systems*, 17(3), pp. 26-35, May/June 2002.

[18] Wark, S., Zschorn, A., Perugini, D., Tate, A., Beautement, P., Bradshaw, J.M. and Suri, N. (2003) Dynamic Agent Systems in the CoAX Binni 2002 Experiment, Special Session on Fusion by Distributed Cooperative Agents at the 6th International Conference on Information Fusion (Fusion 2003), Cairns, Australia, July, 2003.

[19] CoSAR-TS Team (2003) Coalition Search and Rescue Task Support Web Site. http://www.aiai.ed.ac.uk/project/cosar-ts/

[20] Siebra, C. and Tate, A. (2003) I-Rescue: A Coalition Based System to Support Disaster Relief Operations, *The Third International Association of Science and Technology for Development (IASTED) International Conference on Artificial Intelligence and Applications (AIA-2003)*, Benalmadena, Spain, September 2003.

[21] Buckingham Shum, S., De Roure, D., Eisenstadt, M., Shadbolt, N. and Tate, A. (2002) CoAKTinG: Collaborative Advanced Knowledge Technologies in the Grid, Proceedings of the Second Workshop on Advanced Collaborative Environments, Eleventh IEEE Int. Symposium on High Performance Distributed Computing (HPDC-11), July 24-26, 2002, Edinburgh, Scotland.

[22] Jabber (2003) Jabber XML Instant Messaging, http://www.jabber.org.

[23] BBN (2003) OpenMap™ Open Systems mapping Technology - http://openmap.bbn.com

[24] Tate, A. (1998) Roots of SPAR – Shared Planning and Activity Representation, *The Knowledge Engineering Review*, 13(1), 1998.

[25] Conklin, J. and Begeman, M.L. (1988) gIBIS: A Hypertext Tool for Exploratory Policy Discussion. *ACM Trans. Inf. Syst.* 6(4), 303-331, 1988.
[26] Selvin, A.M. (1999) Supporting Collaborative Analysis and Design with Hypertext Functionality, *Journal of Digital Information* 4, 1999.
[27] Conklin, J. (2005) *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. Wiley: Chichester.
[28] MacLean, A., Young, R. M., Bellotti, V., & Moran, T. (1991) Questions, Options, and Criteria: Elements of Design Space Analysis, *Human-Computer Interaction*, 6(3 & 4), 201-250, 1991.
[29] Polyak, S. and Tate A. (1998) Rationale in Planning: Causality, Dependencies and Decisions, *The Knowledge Engineering Review*, 13(3), September 1998, pp. 247-262.
[30] Buckingham Shum, S., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B. and Nuseibeh, B. (2006). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. In: A.H. Dutoit, R. McCall, I. Mistrik, and B. Paech (Eds.), *Rationale Management in Software Engineering* Springer-Verlag/Computer Science Editorial.
[31] Chen-Burger, Y. and Tate, A. (2003) Concept Mapping Between Compendium and I-X, Informatics Report Series, University of Edinburgh, EDI-INF-RR-0166, May 2003.
[32] Khambhampati, S. and Srivastava, B. (1996) Unifying Classical Planning Approaches, Arizona State University ASU CSE TR 96-006, July 1996.
[33] Drabble, B., Dalton, J. and Tate, A. (1997) Repairing Plans on the Fly, in *Proceedings of the NASA Workshop on Planning and Scheduling for Space, Oxnard CA, USA, October 1997*.
[34] I-X Project Website: http://www.aiai.ed.ac.uk/project/ix/.

Rob Milne: A Tribute to a Pioneering AI Scientist, Entrepreneur and Mountaineer
A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Applied Computational Humor and Prospects for Advertising

Oliviero Stock and Carlo Strapparava
ITC-irst, I-38050, Povo, Trento, Italy
{stock, strappa}@itc.it

Dedication to Rob Milne

Rob Milne was full of vital energy, as those who had the privilege of knowing him know well. Mountains would not stop him. In contributing to our field he was unique in his ability to put together research and intelligent and innovative applications. He did show this ability in his personal achievements and when he served the community. He was specifically capable of influencing others in promoting this vision and of launching new ideas that would catch attention and remain through time. Rob among things was also endowed with the ability to tell good jokes. I once heard him tell a very good one as he was delivering an invited talk precisely devoted to Artificial Intelligence research and applications. As it will become clear he was alluding to some characteristic behaviour of people in our Artificial Intelligence community in relation to technology transfer and ambition to material outcome of research. It was a great joke and for me it remained always bound to Rob. I would like to remember him by telling it now.

"A man of great compassion and faith once begged the Eternal. "You know how faithful I have always been. I have been Your servant throughout my life. I have five children to feed and I have no money. Please, please let me win the lottery!" Next day it is lottery time and nothing good for our man happens. So the following week he begs again: "Master of the Universe, You that can decide on all events, see the suffering of your humble servant. Please let me win the lottery". Also this time nothing good for our man happens. The following week he goes back to the Eternal and says: "My Lord, King of the Whole World, have I not done enough to celebrate Your glory? Is there anything I failed to do right in my life? Then why are You not helping me at the moment of need? Why don't You make me win the lottery?" A big thunder roared, the black clouds opened in the sky and a deep voice could be heard saying: "Meet me half way, buy a lottery ticket!"

Without having the ambition of building machines able to achieve this majestic level of humor, I believe something can be done also for reproducing humor capabilities on machines. Humor is an essential element of our life, without that our existence would be unbearable. There is a great potential for computational humor, and especially, I believe there is a real prospect for applications. The paper below talks about some work we have done in the direction of automatic production of humor, with a focus on the prospect of applications in the domain of advertising. I am convinced Rob would have liked to see it here.

- Oliviero Stock

Abstract. Humor is an important mechanism for communicating new ideas and change perspectives. On the cognitive side humor has two very important properties: it helps getting and keeping people's attention and it helps remembering. This qualities have made it an ideal tool for advertisement. Clearly, the world of advertisement has a great potential for the adoption of computational humor, if the latter is possible at all. Though humor has been called "AI-complete" some prototypes are able to produce expressions limited in humor typology but meant to work in unrestricted domains. In this paper some such work on computational humor is presented and the applied potential of computational humor for advertisement discussed. The followed approach is based on a number of resources with rather shallow internal representations, but with large coverage, and the design of some specialised reasoners.

Keywords. Computational Humor, Natural Language Processing, Creative Language

1. Introduction

Humor is one of the most interesting and puzzling aspects of human behavior. Despite the attention it has received in fields such as philosophy, linguistics, and psychology, there have been only few attempts to create computational models for humor recognition or production. Yet computational humor has the potential to change computers into extraordinarily creative and motivational tools.

Computer-human interaction needs to evolve beyond usability and productivity. There is a wide perception in the field that the future is in themes such as entertainment, fun, emotions, aesthetic pleasure, motivation, attention, engagement and so on. Humor is an essential element in communication: it is strictly related to the themes mentioned above, and probably humans cannot survive without it. While it is generally considered merely a way to induce amusement, humor provides an important way to influence the mental state of people to improve their activity. Even though humor is a very complex capability to reproduce, it is realistic to model some types of humor production and to aim at implementing this capability in computational systems. Let us now review a few elements that make humor so important from a cognitive point of view.

Humor and emotions. Humor is a powerful generator of emotions. As such, it has an impact on people's psychological state, directs their attention [1], influences the processes of memorization [2] and of decision-making [3], and creates desires and emotions. Actually, emotions are an extraordinary instrument for motivation and persuasion because those who are capable of transmitting and evoking them have the power to influence other people's opinions and behaviour. Humor, therefore, allows for conscious and constructive use of the affective states generated by it. Affective induction through verbal language is particularly interesting; and humor is one of the most effective ways of achieving it. Purposeful use of humorous techniques enables us to induce positive emotions and mood and to exploit their cognitive and behavioural effects. For example, the persuasive effect of humor and emotions is well known and widely employed in advertising. Advertisements have to be both short and meaningful, to be able to convey information and emotions at the same time.

Humor and beliefs. Humor acts not only upon emotions, but also on human beliefs. A joke plays on the beliefs and expectations of the hearer. By infringing on them, it causes surprise and then hilarity. Jesting with beliefs and opinions, humor induces irony and accustoms people not to take themselves too seriously. Sometimes simple wit can sweep away a negative outlook that places limits on people desires and abilities. Wit can help people overcome self-concern and pessimism that often prevents them from pursuing more ambitious goals and objectives.

Humor and creativity. Humor encourages creativity as well. The change of perspective caused by humorous situations induces new ways of interpreting the same event. By stripping away clichés and commonplaces, and stressing their inconsistency, people become more open to new ideas and points of view. Creativity redraws the space of possibilities and delivers unexpected solutions to problems. Actually, creative stimuli constitute one of the most effective impulses for human activity. Machines equipped with humorous capabilities will be able to play an active role in inducing users' emotions and beliefs, and in providing motivational support.

1.1. Humor in Advertising

From an application point of view we think the world of advertisement has a great potential for the adoption of computational humor. As things are now humanly produced humor appears widely used in advertisement.

Data confirm humor as a fundamental tool. In the UK, advertisements are based on humor in 33% of all cases and in 93% of those cases that show pertinence, according to a 1998 survey [4].

[5] have shown that perception of humor in promotional messages produce higher attention and in general a better recall than non humorous advertisement of the product category, of the specific brand and of the advertisement itself.

Investments are important, we are talking about a relevant economical sector. For instance in the UK alone in 2002 about 28,000 million euro were spent in advertisement (source Ac Nielsen and IPA). Media used for getting the message across to the public is quite conservative at present. For instance in Italy in 2004 the large Italian companies (the first 200 investors) aimed 74% of their investments towards TV commercials. For the total market TV commercials got about 56%, while radio 5%. So we can say that media based on some technology prevails, and also that within technology images have the lion's share. If we look at printed advertisement we observe that 21% advertisement investments went with newspapers, 14% with magazines and 2.4% with posters. So, it seems it is not just a matter of image but also that the market favors being bound to recent news.

Of course internet appears as a fundamental medium already now. Internet advertising was about 9.4 billion \$ (8,000 million euro) in 2004 according to Kagan Research LLC. And growth is very fast: Google advertisement revenues went from 0 to 3,400 million euro in five years according to Business Week.

The future will probably include two important factors: a) reduction in time to market and extension of possible occasions for advertisement; b) more attention to the wearing out of the message and for the need for planning variants and connected messages across time and space; c) contextual personalisation, on the basis of audience profile and perhaps information about the situation. Leaving alone questions of privacy as far as this paper is concerned (but of course advertisement and promotion can be for a good cause and for social values!), all three cases call for a strong role for computer-based intelligent technology for producing novel appropriate advertisements. We believe that computational humor will help producing those kind of messages that have been so successful in the “slow" or non personalised case before the advent of the ubiquitous computer as a medium.

We shall focus on *verbal humor*. In the following we give some background to the field, then report concrete results we had in the development of a limited but domain independent program for humorous acronyms. Following that we will present some resources and components for the extension to some kind of humorous expressions that can be used in automatic advertisements.

2. Background

While humor is relatively well studied in scientific fields such as linguistics [6] and psychology [7,8], to date there is only a limited number of research contributions made toward the construction of computational humor prototypes. A good review of the field can be found in [9]. Almost all the approaches try to deal in some way with “incongruity" at various levels of refinement [10,11,6]. Incongruity theory focuses on the element of surprise. It states that humor is created out of a conflict between what is expected and what actually occurs in the joke. Underlying incongruity is one of the obvious features of a large part of humor phenomena: ambiguity or double meaning.

One of the first attempts that deals with humor generation is the work described in [12], where a formal model of semantic and syntactic regularities was devised, underlying some types of puns (punning riddles). A punning riddle is a question-answer riddle that uses phonological ambiguity. The three main strategies used to create phonological ambiguity are syllable substitution, word substitution and metathesis. *Syllable substitution* is the strategy to confuse a syllable in a word with a similar or identical sounding word. An example of syllable substitution is shown in the following joke: “What do shortsighted ghosts wear? Spooktacles" [13].

Word substitution is the strategy to confuse an entire word with another similar- or identical-sounding word. An example of a joke with word substitution is : “How do you make gold soup? Put fourteen carrots in it" [13].

Metathesis is a strategy very different to syllable or word substitution. It uses reversal of sounds and words to suggest a similarity in meaning between two semantically distinct phrases. An example is “What is the difference between a torn flag and a postage stamp? One’s a tattered banner and the other’s a battered tanner." [12].

Punning riddles based on these three strategies are all suitable for computer generation. Ritchie and Binsted focussed on the word substitution based punning riddles, as lists of homophones (i.e. phonetically identical words) are already available.

The assumptions about the contents and the structure of the lexicon are as follows. The lexicon consists of a finite set of *lexemes* and of *lexical relations*. A lexeme is an abstract entity corresponding to the meaning of a word. If a word has two meanings, it has two corresponding lexemes. Every lexeme has a set of properties about the representation and the type of word. A lexical relation can be an explicit relation between two lexemes, like synonym or homophone, or a general inter-lexeme relation, applicable to more than one pair of lexemes.

In order to describe a punning riddle, two sorts of symbolic description have to be used: *schema* and *template*. A schema stipulates a set of relations witch must be hold between the lexemes used to build a joke. A template indicates the information necessary to turn a schema and lexemes into a piece of text. It contains fixed segments of text that are to be used and syntactic details of how lexemes have to be expressed.

In [12], this model was then exploited to implement a system called JAPE, able to automatically generate amusing puns.

In a recent work [14] automatic production of a funny and appropriate punchline at the end of short jokes is proposed. The authors present a model that describes the relationship between the connector (part of the set-up) and the disjunctor (the punchline). In particular they have implemented this model in a system which, given a joke set-up, can select the best disjunctor from a list of alternatives.

Another humor-generation project was HAHAcronym [15], whose goal was to develop a system able to automatically generate humorous versions of existing acronyms, or to produce a new funny acronym constrained to be a valid vocabulary word, starting with concepts provided by the user. The humorous effect was achieved mainly on the basis of incongruity. We will further discuss this system in more details in Section 3.

Humor recognition has received less attention. In [16] the application of text categorization techniques to humor recognition has been investigated. In particular the authors show that classification techniques are a viable approach for distinguishing between humorous and non-humorous text, through experiments performed on very large data sets. They restrict their investigation to the type of humor found in *one-liners*. A one-liner is a short sentence with comic effects and a peculiar linguistic structure: simple syntax, deliberate use of rhetoric devices (e.g. alliteration, rhyme), and frequent use of creative language constructions meant to attract the readers' attention. In fact, while longer jokes can have a relatively complex narrative structure, a one-liner must produce the humorous effect "in one shot", with very few words.

The humor-recognition problem is formulated as a traditional classification task, feeding positive (humorous) and negative (non humorous) examples to some automatic classifiers. The humorous data set consisted of a corpus of 16,000 one-liners collected from the Web using an automatic bootstrapping process. The non-humorous data were selected such that it is structurally and stylistically similar to the one-liners. In particular, four different corpora were selected, each composed by 16,000 sentences: (1) Reuters news titles [17]; (2) proverbs; (3) sentences picked from the British National Corpus (BNC)[18]; and (4) commonsense statements from the Open Mind Common Sense (OMCS) corpus [19]. The features taken into account were both content-based features, usually considered in traditional text categorization tasks, and humor-specific stylistic features, such as alliteration, presence of antonymy and adult slang. Regardless of the non-humorous data set playing the role of negative examples, the performance of the automatically learned humor-recognizer was always significantly better than apriori known baselines. Surprisingly, comparative experimental results showed that in fact it is more difficult to distinguish humor from regular text (e.g. BNC sentences) than from the other data sets.

Another work on humor recognition is the study reported in [20], focussing on a very restricted type of wordplays, namely the "Knock-Knock" jokes. The goal of the study was to evaluate to what extent wordplay can be automatically identified in "Knock-Knock" jokes, and if such jokes can be reliably identified from other non-humorous texts. The algorithm is based on automatically extracted structural patterns and on heuristics heavily based on the peculiar structure of this particular type of jokes. While the wordplay recognition gave satisfactory results, the identification of jokes containing such wordplays turned out to be significantly more difficult.

Finally [21] proposes a first attempt to recognize the humorous *intent* of short dialogs. According to the authors, computational recognition of humorous intent can be divided into two parts: recognition of a humorous text, and recognition of the intent to be humorous. The approach is based on detecting ambiguity both in the setup and in the punchline.

3. HAHAcronym

HAHAcronym was the first European project devoted to computational humor[1]. The main goal of HAHAcronym was the realization of an acronym ironic re-analyzer and generator as a proof of concept in a focalized but non restricted context. In the first case the system makes fun of existing acronyms, in the second case, starting from concepts provided by the user, it produces new acronyms, constrained to be words of the given language. And, of course, they have to be funny.

The realization of this system was proposed to the European Commission as a project that we would be able to develop in a short period of time (less than a year), that would be meaningful, well demonstrable, that could be evaluated along some pre-decided criteria, and that was conducive to a subsequent development in a direction of potential applicative interest. So for us it was essential that:

1. the work could have many components of a larger system, simplified for the current setting;
2. we could reuse and adapt existing relevant linguistic resources;
3. some simple strategies for humor effects could be experimented.

One of the purposes of the project was to show that using “standard" resources (with some extensions and modifications) and suitable linguistic theories of humor (i.e. developing specific algorithms that implement or elaborate theories), it is possible to implement a working prototype.

3.1. Resources

In order to realize the HAHAcronym prototype, we have refined existing resources and we have developed general tools useful for humorous systems. A fundamental tool is an incongruity detector/generator, that makes the system able to detect semantic mismatches between word meaning and sentence meaning (i.e. in our case the acronym and its context). For all tools, particular attention was put on reusability.

The starting point for us consisted in making use of some standard resources, such as WORDNET DOMAINS [22] (an extension of the well-known English WORDNET) and standard parsing techniques. The tools resulting from the adaptation will be reusable for other applications, and are portable straightforwardly to other languages (e.g. WORDNET DOMAINS is multilingual).

[1]EU project IST-2000-30039 (partners: ITC-irst and University of Twente), part of the Future Emerging Technologies section of the Fifth European Framework Program.

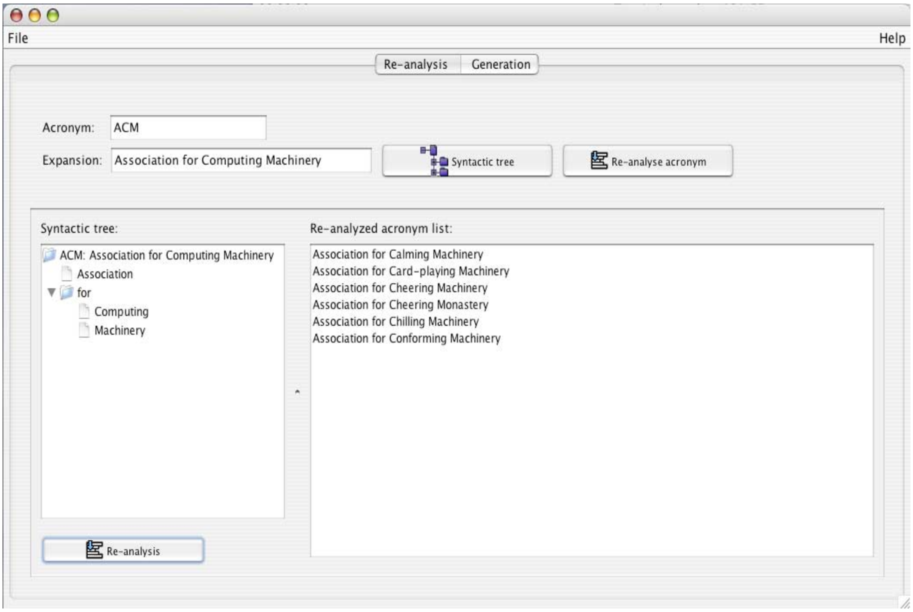


Figure 1. A screenshot of a reanalysis

Wordnet. WORDNET is a thesaurus for the English language inspired by psycholinguistics principles and developed at the Princeton University by George Miller [23]. Lemmata (about 130,000 for version 1.6) are organized in synonym classes (about 100,000 *synsets*). A synset contains all the words by means of which it is possible to express a particular meaning: for example the synset {`knight`, `horse`} describes the sense of "horse" as a chessman. The main relations present in WORDNET are *synonymy*, *antonymy*, *hyperonymy-hyponymy*, *meronymy-holonymy*, *entailment*, *troponymy*.

Wordnet Domains. Domains have been used both in linguistics (i.e. Semantic Fields) and in lexicography (i.e. Subject Field Codes) to mark technical usages of words. Although this is useful information for sense discrimination, in dictionaries it is typically used for a small portion of the lexicon. WORDNET DOMAINS is an attempt to extend the coverage of domain labels within an already existing lexical database, WORDNET. The synsets have been annotated with at least one domain label, selected from a set of about two hundred labels hierarchically organized.

We have organized about 250 domain labels in a hierarchy (exploiting Dewey Decimal Classification), where each level is made up of codes of the same degree of specificity: for example, the second level includes domain labels such as BOTANY, LINGUISTICS, HISTORY, SPORT and RELIGION, while at the third level we can find specialization such as AMERICAN_HISTORY, GRAMMAR, PHONETICS and TENNIS.

Opposition of semantic fields. On the basis of well recognized properties of humor accounted for in many theories (e.g. incongruity, semantic field opposition, apparent contradiction, absurdity) we have modelled an independent structure of domain opposition, such as RELIGION vs. TECHNOLOGY, SEX vs. RELIGION, etc... We exploit opposition as a basic resource for the incongruity generator.

Adjectives and Antonymy Relations. Adjectives play an important role in modifying and generating funny acronyms. WORDNET divides adjectives into two categories. *Descriptive adjectives* (e.g. `big`, `beautiful`, `interesting`, `possible`, `married`) constitute by far the largest category. The second category is called simply *relational adjectives* because they are related by derivation to nouns (i.e. `electrical` in `electrical engineering` is related to noun `electricity`). To relational adjectives, strictly dependent on noun meanings, it is often possible to apply similar strategies as those exploited for nouns. Their semantic organization, though, is entirely different from the one of the other major categories. In fact it is not clear what it would mean to say that one adjective "is a kind of" (ISA) some other adjective. The basic semantic relation among descriptive adjectives is antonymy. WORDNET proposes also that this kind of adjectives is organized in clusters of synsets associated by semantic similarity to a focal adjective. Figure 2 shows clusters of adjectives around the direct antonyms *fast/slow*.

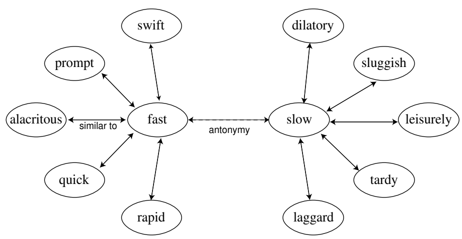


Figure 2. An example of adjective clusters linked by antonymy relation

Exploiting the hierarchy. It is possible to exploit the network of lexical and semantic relations built in WORDNET to make simple ontological reasoning. For example, if a noun or an adjective has a geographic location meaning, the pertaining country and continent can be inferred.

Rhymes. The HAHAcronym prototype takes into account word rhymes and the rhythm of the acronym expansion. To cope with this aspect we got and reorganized the CMU pronouncing dictionary (`http://www.speech.cs.cmu.edu/cgi-bin/cmudict`) with a suitable indexing. The CMU Pronouncing Dictionary is a machine-readable pronunciation dictionary for North American English that contains over 125,000 words and their transcriptions.

Parser, grammar and morphological analyzer. Word sequences that are at the basis of acronyms are subject to a well-defined grammar, simpler than a complete noun phrase grammar, but complex enough to require a nontrivial analyzer. We have decided to use a well established nondeterministic parsing technique. As far as the dictionary is concerned, we use the full WORDNET lexicon, integrated with an ad-hoc morphological analyzer. Also for the generation part we exploit the grammar as the source for syntactic constraints. All the components are implemented in Common Lisp augmented with nondeterministic constructs.

Other resources. An “a-semantic" or “slanting" dictionary is a collection of hyperbolic/attractive adjective/adverbs. This is a last resource, that some time can be useful in the generation of new acronyms. In fact a slanting writing refers to that type of writing that springs from our conscious or subconscious choice of words and images. We may load our description of a specific situation with vivid, connotative words and figures of speech. Some examples are: *abnormally*, *abstrusely*, *adorably*, *exceptionally*, *exorbitantly*, *exponentially*, *extraordinarily*, *voraciously*, *weirdly*, *wonderfully*. This resource is hand-made, using various dictionaries as information sources.

Other lexical resources are: a euphemism dictionary, a proper noun dictionary, lists of typical foreign words commonly used in the language with some strong connotation.

3.2. Architecture and Implementation

To get an ironic or “profaning" re-analysis of a given acronym, the system follows various steps and strategies. The main elements of the algorithm can be schematized as follows:

- acronym parsing and construction of a logical form
- choice of what to keep unchanged (typically the head of the highest ranking NP) and what to modify (e.g. the adjectives)
- look up for possible substitutions
- exploitation of semantic field oppositions
- granting phonological analogy: while keeping the constraint on the initial letters of the words, the overall rhyme and rhythm should be preserved (the modified acronym should sound similar to the original as much as possible)
- exploitation of WORDNET antonymy clustering for adjectives
- use of the slanting dictionary as a last resource

Figures 3 and 4 show a sketch of the system architecture.

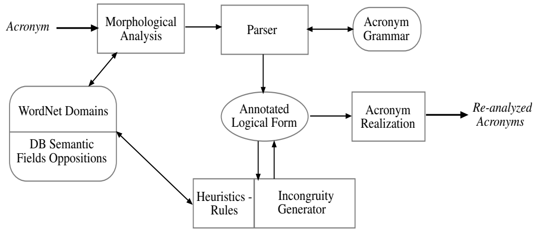


Figure 3. Acronyms Reanalysis

In our system, making fun of existing acronyms amounts to an ironical rewriting, desecrating them with some unexpectedly contrasting, but otherwise consistently sounding expansion.

As far as acronym generation is concerned, the problem is more complex. To make the task more attractive - and difficult - we constrain resulting acronyms to be words of the dictionary (APPLE is good, IBM is not). The system takes in input concepts (actually synsets, possibly resulting from some other process, for instance sentence interpretation) and some minimal structural indication, such as the semantic head. The primary strategy of the system is to consider words that are in ironic relation with the input concepts as potential acronyms. By definition acronyms have to satisfy constraints - to include the initial letters of lexical realization, granting that the sequence of initials satisfy the overall acronym syntax. Ironic reasoning comes mainly at the level of acronym choice and in the selection of the fillers of the *open slots* in the acronym.

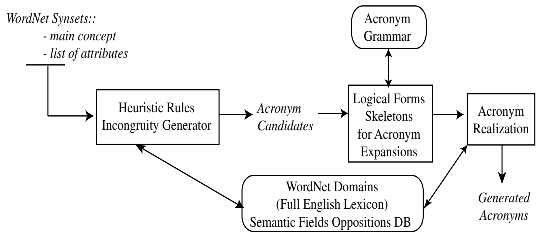


Figure 4. Acronyms Generation

For example, giving as input “fast" and “CPU", we get static, torpid, dormant. The complete synset for “CPU" is{ processor#3, CPU#1, central_processing_unit#1, mainframe#2}; so we can use a synonym of “CPU" in the acronym expansion. The same happens for “fast". Once we have an acronym proposal, a syntactic skeleton has to be filled to get a correct noun phrase. For example given in input “fast" and “CPU", the system selects TORPID and proposes as syntactic skeletons:

$$< adv >_T < adj >_O \texttt{ Rapid Processor } < prep > < adj >_I < noun >_D$$

or

$$< adj >_T < adj >_O \texttt{ Rapid Processor } < prep > < noun >_I < noun >_D$$

where “rapid" and “processor" are synonyms respectively of “fast" and “CPU" and the notation $<\texttt{Part\_of\_Speech}>_{Letter}$ means a word of that particular part_of_speech with $Letter$ as initial. Then the system fills this syntactic skeleton with strategies similar to those described for re-analysis.

The system is fully implemented in Common Lisp, exploiting CLOS and the Meta-Object protocol to import WORDNET DOMAINS and to implement the navigation/reasoning algorithms.

3.3. Examples

Here below some examples of acronym re-analysis are reported. As far as semantic field opposition is concerned we have slightly tuned the system towards the domains FOOD, RELIGION and SEX. We report the original acronym, the re-analysis and some comments about the strategies followed by the system.

ACM - Association for Computing Machinery
→ Association for Confusing Machinery
FBI - Federal Bureau of Investigation
→ Fantastic Bureau of Intimidation

The system keeps all the main heads and works on the adjectives and the PP head, preserving the rhyme and/or using the a-semantic dictionary.

`CRT` - Cathodic Ray Tube
→ `Catholic Ray Tube`
`ESA` - European Space Agency
→ `Epicurean Space Agency`
`PDA` - Personal Digital Assistant
→ `Penitential Demoniacal Assistant`
→ `Prenuptial Devotional Assistant`
`MIT` - Massachusetts Institute of Technology
→ `Mythical Institute of Theology`

Some re-analyses are RELIGION oriented. Note the rhymes.

As far as generation from scratch is concerned, a main concept and some attributes (in terms of synsets) are given as input to the system. Here below we report some examples of acronym generation.

Main concept: *processor* (in the sense of CPU);
Attribute: *fast*

OPEN - On-line Processor for Effervescent Net
PIQUE - Processor for Immobile Quick Uncertain Experimentation
TORPID - Traitorously Outstandingly Rusty Processor for Inadvertent Data_processing
UTMOST - Unsettled Transcendental Mainframe for Off-line Secured Tcp/ip

We note that the system tries to keep all the expansions of the acronym coherent in the same semantic field of the main concept (COMPUTER_SCIENCE). At the same time, whenever possible, it exploits some incongruity in the lexical choices.

3.4. Evaluation

Testing the humorous quality of texts is not an easy task. There have been relevant studies though, such as those by [24]. For HAHAcronym, a simpler case, an evaluation was conducted under the supervision of Salvatore Attardo at Youngstown University, Ohio. Both reanalysis and generation have been tested according to criteria of success stated in advance and in agreement with the European Commission, at the beginning of the project.

The participants in the evaluation were forty students. They were all native speakers of English. The students were not told that the acronyms had been computer generated.

No record was kept of which student had given which set of answers (the answers were strictly anonymous). No demographic data were collected. However, generally speaking, the group was homogeneous for age (traditional students, between the ages of 19 and 24) and mixed for gender and race.

The students were divided into two groups. The first group of twenty was presented the reanalysis and generation data. We tested about 80 reanalyzed and 80 generated acronyms (over twice as many as required by the agreement with the European Commission). Both the reanalysis module and the generation module were found to be successful according to the criteria spelled out in the assessment protocol.

The acronyms reanalysis module showed roughly 70% of acronyms having a score of 55 or higher (out of a possible 100 points), while the acronym generation module showed roughly 53% of acronyms having a score of 55 or higher. The thresholds for success established in the protocol were 60% and 45%, respectively.

One could think that a random selection of fillers could be often funny as well. A special run of the system was performed with lexical reasoning and heuristics disabled, while only the syntactical constraints were operational. If the syntactical rules had been disabled as well, the output would have been gibberish and it would be not fairly comparable with normal HAHAcronym production. This set of acronyms were presented to a different group of twenty students. The result was that less than 8% of the acronyms passed the 55 points score test, we conclude that the output of HAHAcronym is significantly better than random production of reanalysis.

Acronyms	*scored > 55*	*Success Thresholds*
Generation	52.87%	45%
Reanalysis	69.81%	60%
Random	7.69%	

Table 1. Evaluation Results

The successful experience with HAHAcronym encouraged us to investigate further scenarios in which computational humor techniques can be successfully applied. In our opinion one of the most promising is the automatic creation of funny advertisements, as we discuss in the next section.

4. Computational Humor and Advertisements

As we said in the introduction, humor is an important mechanism for communicating new ideas and change perspectives. On the cognitive side humor has two very important properties: it helps getting and keeping people's attention and it helps remembering.

Type and rhythm of humor may vary and the time involved in building the humorous effect may be different in different cases: some times there is a context - like in joke telling - that from the beginning let you expect for the humorous climax, which may occur after a long while; other times the effect is obtained in almost no time, with one perceptive act. This is the case of static visual humor, of funny posters or when some well established convention is reversed with an utterance. Many advertisement-oriented expressions have this property. The role of variation of a known expression seem to be of high importance and studies have also shown the positive impact on the audience of forms of incongruity in the resulting expressions [25,26].

As for memorization it is a common experience to connect in our memory some knowledge to a humorous remark or event. In a foreign language acquisition context it may happen that an involuntary funny situation was created because of so called "false friends" words that sound similar in two languages and may have the same origin but a very different meaning. The "false friends" acknowledgment is conducive to remembering the correct use of the word. Similarly, as shown experimentally (see the introduction), a good humorous expression has exceptionally good recall quality not only *per se* but also for product type and brand, especially if it somehow pertinent. For a large number of verbal expressions what it takes is the ability to perform *optimal innovation* [27] of existing material, with a humorous connotation. We must start from well known expressions that are firm points for the audience, and are connected to the concept or element we intend to promote. We should then be able to perform variations either in the external context, in case the material is ambiguous and the audience can be lured to a different interpretation; or, more often, within the expression itself, changing some material of the expression appropriately, while still preserving full recognisability of the original expression.

For instance a good advertising expression for a soft drink is “Thirst come, thirst served", obvious alteration of a known expression.

This process is for us an extension of what we have done within HAHAcronym. It takes of course more complex components, though. We are talking about natural language expressions, more difficult than acronyms for syntax, semantics, pragmatics. They involve world knowledge, common sense and so on.

In most fields of AI the difficulties of reasoning on deep world knowledge have been understood for a while. There is a clear problem in scaling up between toy experiments and meaningful large scale applications. The more so in an area such as humor, by many called “AI-complete", where good quality expressions require subtle understanding of situations and are normally the privilege of talented individuals. Our goal here is to produce general mechanisms limited to the humorous re-visitation of verbal expressions, but meant to work in unrestricted domains.

Our approach is based on a number of resources with rather shallow internal representations for most of the knowledge and on the design of some specialised reasoners. While we cannot say that for the time being the overall system is complete, we shall mention some of the resources and describe an important new component for the overall system. We consider the use of affective terms a critical aspect in communication and in particular for humor. Valence (positive or negative polarity) of a term and its intensity (the level of arousal it provides) are fundamental factors for persuasion and also for humorous communication. Making fun of biased expression or alluding to related “colored" concepts plays an important role for humorous revisitation of existing expressions.

4.1. Advanced Resources

As done for HAHAcronym, our approach relies on “standard", non humor-oriented resources (with some extensions and modifications) and on the development of specific algorithms for humor processing. Thus also for the generation of humorous advertisements we exploit resources such as pronunciation dictionary, lexical knowledge bases, starting from WORDNET, WORDNET DOMAINS; additionally we have introduced a specialized lexical repository for affective entries (see below), grammars, a repository of commonsensical sentences (OpenMind CommonSense)[19], database of proverbs, clichés, and idioms. Algorithms for making use of those resources include traditional components, like parser and morphological analyzer and specific reasoners.

Recently, we have developed an *affective similarity mechanism* [28], an important component for dealing with advertisement expressions.

All words can potentially convey affective meaning. Each of them, even those apparently neutral, can evoke pleasant or painful experiences. While some words have emotional meaning with respect to the individual story, for many others the affective power is part of the collective imagination (e.g. words “mum", “ghost", “war" etc.). Therefore, it is important to measure the affective power of a generic term. To this aim, we studied the use of words in textual productions, and in particular their co-occurrences with the words in which the affective meaning is explicit. We have to distinguish between words directly referring to emotional states (e.g. “fear", “cheerful") and those having only an indirect reference to emotion, depending on the context (e.g. words that indicate possible emotional causes such as “monster" or emotional responses such as “cry"). We call the former direct affective words and the latter indirect affective words. We developed an affective similarity mechanism that consists of (i) WORDNET-AFFECT [29], an affective lexical resource based on an extension of WordNet, in which the direct affective words and synsets are labeled with affective labels (in a similar way of WORDNET DOMAINS), and on (ii) a selection function (named affective weight) based on a semantic similarity mechanism automatically acquired in an unsupervised way from a large corpus of texts (100 millions of words), in order to individuate the indirect affective lexicon. Applied to a concept (e.g. a WORDNET synset) and an emotional category, this function returns a value representing the semantic affinity with that emotion. In this way it is possible to assign a value to the concept with respect to each emotional category, and eventually select the emotion with the highest value. Applied to a set of concepts that are semantically similar, this function selects subsets characterized by some given affective constraints (e.g. referring to a particular emotional category or valence). For example, if we give in input the verb “shoot" and an indication of negative valence, the system individuates the emotional category HORROR. Then, it extracts the target-noun “gun" and the causative evaluative adjective “frightening" and finally generates the noun phrase “frightening gun".

4.1.1. WORDNET-AFFECT *and the Emotional Categories*

A-Labels	*Valence*	*Examples of word senses*
JOY	positive	noun `joy#1`, adjective `elated#2`, verb `gladden#2`
LOVE	positive	noun `love#1`, adjective `loving#1`, verb `love#1`
APPREHENSION	negative	noun `apprehension#1`, adjective `apprehensive#3`
SADNESS	negative	noun `sadness#1`, adjective `unhappy#1`, verb `sadden#1`
SURPRISE	ambiguous	noun `surprise#1`, adjective `surprised#1`
APATHY	neutral	noun `apathy#1`, adjective `apathetic#1`,
NEGATIVE-FEAR	negative	noun `scare#2`, adjective `afraid#1`, verb `frighten#1`
POSITIVE-FEAR	positive	noun `frisson#1`
POSITIVE-EXPECTATION	positive	noun `anticipation#1`, verb `anticipate#1`

Table 2. Some of emotional categories in WORDNET-AFFECT and some corresponding word senses

WORDNET-AFFECT is an extension of WordNet database [23], including a subset of synsets suitable to represent affective concepts. Similarly to our method for domain labels, we assigned to a number of WordNet synsets one or more affective labels (*a-labels*). In particular, the affective concepts representing emotional state are individuated by synsets marked with the a-label EMOTION. There are also other a-labels for those concepts representing moods, situations eliciting emotions, or emotional responses. WORDNET-AFFECT is freely available for research purpose at `http://wndomains.itc.it`. See [29] for a complete description of the resource.

Recently, we extended WORDNET-AFFECT with a set of additional a-labels (i.e. the emotional categories), hierarchically organized, in order to specialize synsets with a-label EMOTION. In a second stage, we introduced some modifications, in order to distinguish synsets according to emotional valence. We defined four additional a-labels: POSITIVE, NEGATIVE, AMBIGUOUS, NEUTRAL. The first one corresponds to "positive emotions", defined as emotional states characterized by the presence of positive signals (or pleasure). It includes synsets such as `joy#1` or `enthusiasm#1`. Similarly the NEGATIVE a-label identifies "negative emotions" characterized by negative edonic signals (or pain), for example `anger#1` or `sadness#1`. Synsets representing affective states whose valence depends on semantic context (e.g. `surprise#1`) were marked with the tag AMBIGUOUS. Finally, synsets referring to mental states that are generally considered affective but are not characterized by valence, were marked with the tag NEUTRAL.

Another important property for affective lexicon concerning mainly adjectival interpretation is the stative/causative dimension [30]. An emotional adjective is said *causative* if it refers to some emotion that is caused by the entity represented by the modified noun (e.g. "amusing movie"). In a similar way, an emotional adjective is said *stative* if it refers to the emotion owned or felt by the subject denoted by the modified noun (e.g. "cheerful/happy boy").

4.1.2. Affective Semantic Similarity

A crucial issue is to have a mechanism for evaluating the similarity among generic terms and affective lexical concepts. To this aim we estimated term similarity from a large scale corpus.In particular we implemented a variation of Latent Semantic Analysis (LSA) in order to obtain a vector representation for words, texts and synsets.

In LSA [31], term co-occurrences in the documents of the corpus are captured by means of a dimensionality reduction operated by a Singular Value Decomposition (SVD) on the term-by-document matrix. For the experiments reported in this paper, we run the SVD operation on the British National Corpus[2].

The resulting LSA vectors can be exploited to estimate both term and document similarity. Regarding document similarity, Latent Semantic Indexing (LSI) is a technique that allows us to represent a document by means of a LSA vector. In particular, we used a variation of the *pseudo-document* methodology that takes into account also a *tf-idf* weighting schema (see [32] for more details). Each document can be represented in the LSA space by summing up the normalized LSA vectors of all the terms contained in it. Also a synset in WORDNET (and then an emotional category) can be represent in the LSA space, performing the pseudo-document technique on all the words contained in the synset. Thus it is possible to have a vectorial representation of each emotional category in the LSA space (i.e. the *emotional vectors*). With an appropriate metric (e.g. cosine), we can compute a similarity measure among terms and affective categories. We defined the *affective weight* as the similarity value between an emotional vector and an input term vector.

[2]The British National Corpus is a very large (over 100 million words) corpus of modern English, both spoken and written [18].

For example, the term “sex" shows high similarity with respect to the positive emotional category AMOROUSNESS, with the negative category MISOGYNY, and with the ambiguous valence tagged category AMBIGUOUS_EXPECTATION. The noun “gift" is highly related to the emotional categories: LOVE (with positive valence), COMPASSION (with negative valence), SURPRISE (with ambiguous valence), and INDIFFERENCE (with neutral valence).

Related Emotional Term	*Positive Emotional Category*	*Emotional Weight*
university	ENTHUSIASM	0.36
professor	SYMPATHY	0.56
scholarship	DEVOTION	0.72
achievement	ENCOURAGEMENT	0.76
	Negative Emotional Category	
university	DOWNHEARTEDNESS	0.33
professor	ANTIPATHY	0.46
study	ISOLATION	0.49
scholarship	MELANCHOLY	0.53
	Ambiguous Emotional Category	
university	AMBIGUOUS-HOPE	0.25
career	EARNESTNESS	0.59
rector	REVERENCE	0.57
scholar	REVERENCE	0.67
	Neutral Emotional Category	
university	WITHDRAWAL	0.12
faculty	APATHY	0.13
admission	WITHDRAWAL	0.31
academic	DISTANCE	0.35

Table 3. Some terms related to “university" through some emotional categories

The affective weight function can be used in order to select the emotional categories that can best express or evoke valenced emotional states with respect to input term. Moreover, it allows us to individuate a set of terms that are semantically similar to the input term and that share with it the same affective constraints (e.g. emotional categories with the same value of valence).

For example, given the noun *university* as input-term, it is possible to ask the system for related terms that have a positive affective valence, possibly focussing only to some specific emotional categories (e.g. SYMPATHY). On the other hand given two terms, it is possible to check whether they are semantically related, and with respect to which emotional category. Table 3 shows a portion of affective lexicon related to “university" with some emotional categories grouped by valence.

4.2. Use of the mechanism for humor production

Our approach can be synthesized in the following points:

- We tend to start from existing material, for instance well known popular expressions, and produce some “optimal innovation" characterized by irony, or some other humorous connotation.
- The variation is based on reasoning on various resources, such as those described above. While there is no deep representation of meaning and pragmatics at the complex expression level, there is more detailed representation at the individual component level, permitting some type of reasoning (an example is the affective lexical resources indicated above where lexical reasoning is strongly involved). For the more complex expressions we adopt more opaque but dynamic and robust mechanisms, based on learning and similarity, such as the one mentioned above. The potential is particularly strong in dealing with new material, such as reacting to novel expressions in a dialog or, as we are working at now, making fun of fresh headlines.
- The architecture of the system accommodates a large number of resources and revision mechanisms. The coordination of the various modules is based on a number of specific humor heuristic mechanisms inspired by the General Theory of Verbal Humor [6].

In general the results are meant to be in the form of a set of proposed expressions. The choice is left to the human: few things worse than poor humor can be experienced, so it is better not to risk... We are working, though, on post-processing based on mechanisms for automatic evaluation of humor as an evolution of initial work on humor recognition.

5. Conclusions

We have illustrated some initial prototypes for computational humor and discussed some prospects for the field of advertising. Although mainly we focus on humorous (mostly ironic) re-visitation of existing verbal expressions, we think this accounts for an important portion of the complex universe of humor. But also more complex creative expressions such as some of the Marx Brothers’ (see [33]) are not unrealistic. From an applied AI point of view, we believe that an environment for proposing solutions to advertising professionals can be a practical development of this work, for the moment leaving the last word to the human professional.

In the future, the potential of fully automatic humor production (and retrieval) will find a big opportunity if advertisements are to be linked to an evolving context, such as incoming news, or changing of location of the audience. More than that: if advertisements will be personalized, also humor will need to follow, given the success it has had in the “pre-technological" world of advertisement. Humor is appreciated differently by different individuals. Personality studies regarding this specific theme give important indications [8]. So computational humor generators will need to take into account the personality of the receiver and, if possible, his mood and situation to produce well received expressions. There will not be sufficient creative human beings - and anyway they would cost too much - for producing all the necessary quantity of humorous expressions in real time.

References

[1] S. Kitayama and P. Niedenthal. *The Heart's Eye: Emotional influences in Perception and Attention*. Academic Press, 1994.

[2] D. Kahneman. *Attention and Effort*. Englewood Cliffs: Prentice-Hall, 1973.

[3] A. Isen. Positive affect and decision making. In M. Lewis and J. Haviland, editors, *Handbook of Emotion*. Guilford, New York, 1993.

[4] M. Toncar. The use of humour in television advertising: revisiting the US-UK comparison. *International Journal of Advertising*, 20(4):521–540, 2001.

[5] S. Perry, S. Jenzowsky, C. King, H. Yi, J. Hester, and J. Gartenschlaeger. Humorous programs as a vehicle of humorous commercials. *Journal of Communication*, 47(1):20–39, 1997.

[6] S. Attardo. *Linguistic Theories of Humor*. Mouton de Gruyter, Berlin, 1994.

[7] S. Freud. *Der Witz und Seine Beziehung zum Unbewussten*. Deutike, Leipzig and Vienna, 1905.

[8] W. Ruch. Computers with a personality? lessons to be learned from studies of the psychology of humor. In *[34]*, 2002.

[9] G. Ritchie. Current directions in computational humour. *Artificial Intelligence Review*, 16(2):119–135, 2001.

[10] A. Koestler. *The act of creation*. Hutchinson, London, 1964.

[11] V. Raskin. *Semantic Mechanisms of Humor*. Dordrecht/Boston/Lancaster, 1985.

[12] K. Binsted and G. Ritchie. Computational rules for punning riddles. *Humor*, 10(1), 1997.

[13] K. Webb, editor. *The Crack-a-Joke Book*. Puffin, London, 1978.

[14] J. Stark, K. Binsted, and B. Bergen. Disjunctor selection for one-line jokes. In *Proc. of Intelligent Technologies for Interactive Entertainment (INTETAIN '05)*, Madonna di Campiglio, Italy, December 2005.

[15] O. Stock and C. Strapparava. Getting serious about the development of computational humor. In *Proceedings of 18th International Joint Conference on Artificial Intelligence (IJCAI03)*, Acapulco, Mexico, 2003.

[16] R. Mihalcea and C. Strapparava. Laughter abounds in the mouths of computers: Investigations in automatic humor recognition. In *Proc. of Intelligent Technologies for Interactive Entertainment (INTETAIN '05)*, Madonna di Campiglio, Italy, December 2005.

[17] D. Lewis, Y. Yang, T. Rose, and F. Li. RCV1: A new benchmark collection for text categorization research. *The Journal of Machine Learning Research*, 5:361–397, December 2004.

[18] BNC-Consortium. British national corpus., 2000.

[19] P. Singh. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access.*, Palo Alto, CA, 2002.

[20] J. Taylor and L. Mazlack. Computationally recognizing wordplay in jokes. In *Proceedings of CogSci 2004*, Chicago, August 2004.

[21] J. Taylor and L. Mazlack. Toward computational recognition of humorous intent. In *Proc. of the 27th Annual Conference of the Cognitive Science Society (COGSCI 05)*, Stresa, Italy, July 2005.

[22] B. Magnini, C. Strapparava, G. Pezzulo, and A. Gliozzo. The role of domain information in word sense disambiguation. *Journal of Natural Language Engineering*, 8(4):359–373, 2002.

[23] C. Fellbaum. *WordNet. An Electronic Lexical Database*. The MIT Press, 1998.

[24] W. Ruch. Special issue: Measurement approaches to the sense of humor. *Humor*, 9(3/4), 1996.

[25] M. Pricken. *Creative Advertising*. Thames & Hudson, 2002.

[26] A. Testa. *La parola immaginata*. Pratiche Editrice, 2000.

[27] R. Giora. Optimal innovation and pleasure. In *[34]*, 2002.

[28] A. Valitutti, C. Strapparava, and O. Stock. Lexical resources and semantic similarity for affective evaluative expressions generation. In *Proc. of the First International Conference on*

Affective Computing and Intelligent Interaction (ACII05), Beijing, China, October 2005.

[29] C. Strapparava and A. Valitutti. WordNet-Affect: an affective extension of WordNet. In *Proc. of 4^{th} International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1083 – 1086, Lisbon, May 2004.

[30] A. Goy. Lexical semantics of emotional adjectives. In S. Feist, S. Fix, J. Hay, and J. Moore, editors, *: Proceedings of Student Conference in Linguistics. MIT Working Papers in Linguistics 37*. MIT Press, 2000.

[31] T. K. Landauer, P. W. Foltz, and D. Laham. Introduction to latent semantic analysis. *Discourse Processes*, 25:259–284, 1998.

[32] A. Gliozzo and C. Strapparava. Domains kernels for text categorization. In *Proc. of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, June 2005.

[33] O. Stock. Password Swordfish: Verbal humor in the interface. In J. Hulstijn and A. Nijholt, editors, *Proc. of International Workshop on Computational Humour (TWLT 12)*, University of Twente, Enschede, Netherlands, 1996.

[34] O. Stock, C. Strapparava, and A. Nijholt, editors. *Proceedings of the The April Fools Day Workshop on Computational Humour (TWLT20)*, Trento, 2002.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Making Music with AI: Some examples

Ramón LOPEZ DE MANTARAS
IIIA-Artificial Intelligence Research Institute
CSIC-Spanish Scientific Research Council
Campus UAB 08193 Bellaterra

Abstract. The field of music raises very interesting challenges to computer science and in particular to Artificial Intelligence. Indeed, as we will see, computational models of music need to take into account important elements of advanced human problem solving capabilities such as knowledge representation, reasoning, and learning. In this paper I describe examples of computer programs capable of carrying out musical activities and describe some creative aspects of musical such programs.

Keywords. Artificial Intelligence, Computational Models of Music

Dedication to Rob Milne

Rob was a man that liked big challenges, especially in AI and in mountaineering, and therefore he was very enthusiastic about the challenge of replicating creativity by means of artificial intelligence techniques. In several occasions I had very long and stimulating discussions with him regarding artificial creativity in general and AI applications to music in particular, and he was well aware of the main developments in the area. This paper is dedicated to him, a truly creative man.

Introduction

Music is a very challenging application area for AI because, as we will see in this survey of a set of representative applications, it requires complex knowledge representation, reasoning, and learning. The survey is organized in three subsections. The first is devoted to compositional systems, the second describes improvisation systems, and the third is devoted to systems capable of generating expressive performances. It is unanimously accepted among researchers on AI and music that these three activities involve extensive creative processing. Therefore, although creativity is not the main focus of this paper, I believe that the computational systems described in this paper are valuable examples of artificially creative behaviour.

The books by Boden [1,2], Dartnall [3], Partridge & Rowe [4], and Bentley & Corne [5]; as well as the papers by Rowe & Partridge [6], and Buchanan [7] are very interesting sources of information regarding artificial intelligence approaches to creativity. Besides, for further information on AI and music I recommend the books edited by Balaban et al. [8] and by Miranda [9], and the book by Cope [10].

1. Composing music

Hiller and Isaacson's [11] work, on the ILLIAC computer, is the best known pioneering work in computer music. Their chief result is the "Illiac Suite", a string quartet composed following the "generate and test" problem solving approach. The program generated notes pseudo-randomly by means of Markov chains. The generated notes were next tested by means of heuristic compositional rules of classical harmony and counterpoint. Only the notes satisfying the rules were kept. If none of the generated notes satisfied the rules, a simple backtracking procedure was used to erase the entire composition up to that point, and a new cycle was started again. The goals of Hiller and Isaacson excluded anything related to expressiveness and emotional content. In an interview (see [11], p. 21), Hiller and Isaacson said that, before addressing the expressiveness issue, simpler problems needed to be handled first. We believe that this was a very correct observation in the fifties. After this seminal work, many other researchers based their computer compositions on Markov probability transitions but also with rather limited success judging from the standpoint of melodic quality. Indeed, methods relying too heavily on markovian processes are not informed enough to produce high quality music consistently.

However, not all the early work on composition relies on probabilistic approaches. A good example is the work of Moorer [13] on tonal melody generation. Moorer's program generated simple melodies, along with the underlying harmonic progressions, with simple internal repetition patterns of notes. This approach relies on simulating human composition processes using heuristic techniques rather than on Markovian probability chains. Levitt [14] also avoided the use of probabilities in the composition process. He argues that: "randomness tends to obscure rather than reveal the musical constraints needed to represent simple musical structures". His work is based on constraint-based descriptions of musical styles. He developed a description language that allows expressing musically meaningful transformations of inputs, such as chord progressions and melodic lines, through a series of constraint relationships that he calls "style templates". He applied this approach to describe a traditional jazz walking bass player simulation as well as a two-handed ragtime piano simulation.

The early systems by Hiller-Isaacson and Moore were both based also on heuristic approaches. However, possibly the most genuine example of early use of AI techniques is the work of Rader [15]. Rader used rule-based AI programming in his musical round (a circle canon such as "Frère Jacques") generator. The generation of the melody and the harmony were based on rules describing how notes or chords may be put together. The most interesting AI component of this system are the *applicability rules*, determining the applicability of the melody and chords generation rules, and the *weighting rules* indicating the likelihood of application of an applicable rule by means of a weight. We can already appreciate the use of metaknowledge in this early work.

AI pioneers such as Herbert Simon or Marvin Minsky also published works relevant to computer music. Simon and Sumner [16] describe a formal pattern language for music, as well as a pattern induction method, to discover patterns more or less implicit in musical works. One example of pattern that can be discovered is "the opening section is in C Major, it is followed by a section in dominant and then a return to the original key". Although the program was not completed, it is worth noticing that it was one of the firsts in dealing with the important issue of music modeling, a subject that has been, and still is, widely studied. For example, the use of models based on generative grammars has been, and continues to be, an important and very useful approach in music modeling (Lerdahl and Jackendoff [17])

Marvin Minsky in his well known paper *Music, Mind, and Meaning* [18] addresses the important question of how music impresses our minds. He applies his concepts of agent and its role in a society of agents as a possible approach to shed light on that question. For example, he hints that one agent might do nothing more than noticing that the music has a particular rhythm. Other agents might perceive small musical patterns such as repetitions of a pitch; differences such as the same sequence of notes played one fifth higher, etc. His approach also accounts for more complex relations within a musical piece by means of higher order agents capable of recognizing large sections of music. It is important to clarify that in that paper Minsky does not try to convince the reader about the question of the validity of his approach, he just hints at its plausibility.

Among the compositional systems there is a large number dealing with the problem of automatic harmonization using several AI techniques. One of the earliest works is that of Rothgeb [19]. He wrote a SNOBOL program to solve the problem of harmonizing the unfigured bass (given a sequence of bass notes infer the chords and voice leadings that accompany those bass notes) by means of a set of rules such as "If the bass of a triad descends a semitone, then the next bass note has a sixth". The main goal of Rothgeb was not the automatic harmonization itself but to test the computational soundness of two bass harmonization theories from the eighteenth century.

One of the most complete works on harmonization is that of Ebcioglu [20]. He developed an expert system, CHORAL, to harmonize chorales in the style of J.S. Bach. CHORAL is given a melody and produces the corresponding harmonization using heuristic rules and constraints. The system was implemented using a logic programming language designed by the author. An important aspect of this work is the use of sets of logical primitives to represent the different viewpoints of the music (chords view, time-slice view, melodic view, etc.). This was done to tackle the problem of representing large amounts of complex musical knowledge.

MUSACT [21] uses Neural Networks to learn a model of musical harmony. It was designed to capture musical intuitions of harmonic qualities. For example, one of the qualities of a dominant chord is to create in the listener the expectancy that the tonic chord is about to be heard. The greater the expectancy, the greater the feeling of consonance of the tonic chord. Composers may choose to satisfy or violate these expectancies to varying degree. MUSACT is capable of learning such qualities and generate graded expectancies in a given harmonic context.

In HARMONET [22], the harmonization problem is approached using a combination of neural networks and constraint satisfaction techniques. The neural network learns what is known as harmonic functionality of the chords (chords can play the function of tonic, dominant, subdominant, etc) and constraints are used to fill the inner voices of the chords. The work on HARMONET was extended in the MELONET system [23, 24]. MELONET uses a neural network to learn and reproduce higher-level structure in melodic sequences. Given a melody, the system invents a baroque-style harmonization and variation of any chorale voice. According to the authors, HARMONET and MELONET together form a powerful music-composition system that generates variations whose quality is similar to those of an experienced human organist.

Pachet and Roy [25] also used constraint satisfaction techniques for harmonization. These techniques exploit the fact that both the melody and the harmonization knowledge impose constraints on the possible chords. Efficiency is however a problem with purely constraint satisfaction approaches.

Sabater et al. [26], approach the problem of harmonization using a combination of rules and case-based reasoning. This approach is based on the observation that purely rule-based harmonization usually fails because in general the rules don't make the music, it is the music that makes the rules. Then, instead of relying only on a set of imperfect rules, why not making use of the source of the rules, that is the compositions themselves? Case-based reasoning allows the use of examples of already harmonized compositions as cases for new harmonizations. The system harmonizes a given melody by first looking for similar, already harmonized, cases, when this fails, it looks for applicable general rules of harmony. If no rule is applicable, the system fails and backtracks to the previous decision point. The experiments have shown that the combination of rules and cases results in much fewer failures in finding an appropriate harmonization than using either technique alone. Another advantage of the case-based approach is that each newly correctly harmonized piece can be memorized and made available as a new example to harmonize other melodies; that is, a learning by experience process takes place. Indeed, the more examples the system has, the less often the system needs to resort to the rules and therefore it fails less. MUSE [27] is also a learning system that extends an initially small set of voice leading constraints by learning a set of rules of voice doubling and voice leading. It learns by reordering the rules agenda and by chunking the rules that satisfy the set of voice leading constraints. MUSE successfully learned some of the standard rules of voice leading included in traditional books of tonal music.

Certainly the best-known work on computer composition using AI is David Cope's EMI project [28, 29]. This work focuses on the emulation of styles of various composers. It has successfully composed music in the styles of Cope, Mozart, Palestrina, Albinoni, Brahms, Debussy, Bach, Rachmaninoff, Chopin, Stravinsky, and Bartok. It works by searching for recurrent patterns in several (at least two) works of a given composer. The discovered patterns are called signatures. Since signatures are location dependent, EMI uses one of the composer's works as a guide to fix them to their appropriate locations when composing a new piece. To compose the musical motives between signatures, EMI uses a compositional rule analyzer to discover the constraints used by the composer in his works. This analyzer counts musical events such as voice leading directions; use of repeated notes, etc. and represents them as a statistical model of the analyzed works. The program follows this model to compose the motives to be inserted in the empty spaces between signatures. To properly insert them, EMI has to deal with problems such as: linking initial and concluding parts of the signatures to the surrounding motives avoiding stylistic anomalies, maintaining voice motions, maintaining notes within a range, etc. Proper insertion is achieved by means of an Augmented Transition Network [30]. The results, although not perfect, are quite consistent with the style of the composer.

2. Synthesizing expressive performances

One of the main limitations of computer-generated music has been its lack of expressiveness, that is, lack of gesture. Gesture is what musicians call the nuances of performance that are unique (in the sense of conveying the "personal touch" of the musician) and subtly interpretive or, in other words, creative.

One of the first attempts to address expressiveness in music performances is that of Johnson [31]. She developed an expert system to determine the tempo and the articulation to be applied when playing Bach's fugues from "The Well-Tempered Clavier". The rules were obtained from two expert human performers. The output gives the base tempo value and a list of performance instructions on notes duration and articulation that should be followed by a human player. The results very much coincide with the instructions given in well known commented editions of "The Well-Tempered Clavier". The main limitation of this system is its lack of generality because it only works well for fugues written on a 4/4 meter. For different meters, the rules should be different. Another obvious consequence of this lack of generality is that the rules are only applicable to Bach fugues.

The work of Bresin, Friberg, Fryden, and Sundberg at KTH [32, 33, 34, 35] is one of the best known long term efforts on performance systems. Their current "Director Musices" system incorporates rules for tempo, dynamic, and articulation transformations constrained to MIDI. These rules are inferred both from theoretical musical knowledge and experimentally by training using, in particular, the so-called analysis-by-synthesis approach. The rules are divided in three main classes: *Differentiation rules*, which enhance the differences between scale tones; *Grouping rules*, which show what tones belong together; and *Ensemble rules*, that synchronize the various voices in an ensemble.

Canazza et al [36] developed a system to analyze how the musician's expressive intentions are reflected in the performance. The analysis reveals two different expressive dimensions: one related to the energy (dynamics) and the other one related to the kinetics (rubato) of the piece. The authors also developed a program for generating expressive performances according to these two dimensions.

The work of Dannenberg and Derenyi [37] is also a good example of articulation transformations using manually constructed rules. They developed a trumpet synthesizer that combines a physical model with a performance model. The goal of the performance model is to generate control information for the physical model by means of a collection of rules manually extracted from the analysis of a collection of controlled recordings of human performance.

Another approach taken for performing tempo and dynamics transformation is the use of neural network techniques. Bresin [38] describes a system that combines symbolic decision rules with neural networks is implemented for simulating the style of real piano performers. The outputs of the neural networks express time and loudness deviations. These neural networks extend the standard feed-forward network trained with the back propagation algorithm with feedback connections from the output neurons to the input neurons.

We can see that, except for the work done by the group at KTH that considers three expressive parameters, the other systems are limited to two such as rubato and dynamics, or rubato and articulation. This limitation has to do with the use of rules. Indeed, the main problem with the rule-based approaches is that it is very difficult to find rules general enough to capture the variety present in different performances of the same piece by the same musician and even the variety within a single performance [39]. Furthermore, the different expressive resources interact with each other. That is, the rules for dynamics alone change when rubato is also taken into account. Obviously, due to this interdependency, the more expressive resources one tries to model, the more difficult is finding the appropriate rules.

We have developed a system called SaxEx [40] SaxEx is a computer program capable of synthesizing high quality expressive tenor sax solo performances of jazz ballads based on cases representing human solo performances. Previous rule-based approaches to that problem could not deal with more than two expressive parameters (such as dynamics and rubato) because it is too difficult to find rules general enough to capture the variety present in expressive performances. Besides, the different expressive parameters interact with each other making it even more difficult to find appropriate rules taking into account these interactions.

With CBR, we have shown that it is possible to deal with the five most important expressive parameters: dynamics, rubato, vibrato, articulation, and attack of the notes. To do so, SaxEx uses a case memory containing examples of human performances, analyzed by means of spectral modeling techniques and background musical knowledge. The score of the piece to be performed is also provided to the system. The heart of the method is to analyze each input note determining (by means of the background musical knowledge) its role in the musical phrase it belongs to, identify and retrieve (from the case-base of human performances) notes with similar roles, and finally, transform the input note so that its expressive properties (dynamics, rubato, vibrato, articulation, and attack) match those of the most similar retrieved note. Each note in the case base is annotated with its role in the musical phrase it belong to as well as with its expressive values. Furthermore, cases do not contain just information on each single note but they include contextual knowledge at the phrase level. Therefore, cases in this system have a complex object-centered representation.

Although limited to monophonic performances, the results are very convincing and demonstrate that CBR is a very powerful methodology to directly use the knowledge of a human performer that is implicit in her playing examples rather than trying to make this knowledge explicit by means of rules. Some audio results can be listened to at www.iiia.csic.es/arcos/noos/Demos/Aff-Example.html. More recent papers by Arcos and Lopez de Mantaras [41] and by Lopez de Mantaras and Arcos [42], describe this system in great detail.

Based on the work on SaxEx, we have developed TempoExpress [43], a case-based reasoning system for applying musically acceptable tempo transformations to monophonic audio recordings of musical performances. TempoExpress has a rich description of the musical expressivity of the performances, that includes not only timing deviations of performed score notes, but also represents more rigorous kinds of expressivity such as note ornamentation, consolidation, and fragmentation. Within the tempo transformation process, the expressivity of the performance is adjusted in such a way that the result sounds natural for the new tempo. A case base of previously performed melodies is used to infer the appropriate expressivity. The problem of changing the tempo of a musical performance is not as trivial as it may seem because it involves a lot of musical knowledge and creative thinking. Indeed, when a musician performs a musical piece at different tempos the performances are not just time-scaled versions of each other (as if the same performance were played back at diferent speeds). Together with the changes of tempo, variations in musical expression are made (see for instance the work of Desain and Honing [44]). Such variations do not only affect the timing of the notes, but can also involve for example the addition or deletion of ornamentations, or the consolidation/fragmentation of notes. Apart from the tempo, other domain specific factors seem to play an important role in the way a melody is performed, such as meter, and phrase structure. Tempo transformation is one of the audio post-processing tasks manually done in audio-labs. Automatizing this process may, therefore, be of industrial interest.

Other applications of CBR to expressive performance are those of Suzuki et al. [45], and those of Tobudic and Widmer [46, 47]. Suzuki et al. [45], also use example cases of expressive performances to generate multiple performances of a given piece with varying musical expression, however they deal only with two expressive parameters. Tobudic and Widmer [46] apply instance-based learning (IBL) also to the problem of generating expressive performances. The IBL approach is used to complement a note-level rule-based model with some predictive capability at the higher level of musical phrasing. More concretely, the IBL component recognizes performance patterns, of a concert pianist, at the phrase level and learns how to apply them to new pieces by analogy. The approach produced some interesting results but, as the authors recognize, was not very convincing due to the limitation of using an attribute-value representation for the phrases. Such simple representation cannot take into account relevant structural information of the piece, both at the sub-phrase level and at the inter-phrasal level. In a subsequent paper, Tobudic and Widmer [47], succeeded in partly overcoming this limitations by using a relational phrase representation.

The possibility for a computer to play expressively is a fundamental component of the so-called "hyper-instruments". These are instruments designed to augment an instrument sound with such idiosyncratic nuances as to give it human expressiveness and a rich, live sound. To make an hyper-instrument, take a traditional instrument, like for example a cello, and connect it to a computer through electronic sensors in the neck and in the bow, equip also with sensors the hand that holds the bow and program the computer with a system similar to SaxEx that allows to analyze the way the human interprets the piece, based on the score, on musical knowledge and on the readings of the sensors. The results of such analysis allows the hyper-instrument to play an active role altering aspects such as timbre, tone, rhythm and phrasing as well as generating an accompanying voice. In other words, you have got an instrument that can be its own intelligent accompanist. Tod Machover, from MIT's Media Lab, developed a hypercello [48] and the great cello player Yo-Yo Ma premiered, playing the hypercello, a piece, composed by Tod Machover, called "Begin Again Again..." at the Tanglewood Festival several years ago.

3. Improvising music

Music improvisation is a very complex creative process that has also been computationally modeled. It is often referred to as “composition on the fly”. Because of the hard real time constraints involved, music improvisation it is creatively speaking more complex than composition (where musicians have the time to revise and improve their work) and since it obviously requires expressiveness too it is perhaps the most complex of the three music activities addressed in this paper. An early work on computer improvisation is the Flavours Band system of Fry [49]. Flavours Band is a procedural language, embedded in LISP, for specifying jazz and popular music styles. Its procedural representation allows the generation of scores in a pre-specified style by making changes to a score specification given as input. It allows combining random functions and musical constraints (chords, modes, etc.) to generate improvisational variations. The most remarkable result of Flavours Band was an interesting arrangement of the bass line, and an improvised solo, of John Coltrane’s composition *Giant Steps*.

GenJam [50] builds a model of a jazz musician learning to improvise by means of a genetic algorithm. A human listener plays the role of fitness function by rating the offspring improvisations. Papadopoulos and Wiggins [51] also used a genetic algorithm to improvise jazz melodies on a given chord progression. Contrarily to GenJam, the program includes a fitness function that automatically evaluates the quality of the offspring improvisations rating eight different aspects of the improvised melody such as the melodic contour, notes duration, intervallic distances between notes, etc.

Franklin [52] uses recurrent neural networks to learn how to improvise jazz solos from transcriptions of solo improvisations by saxophonist Sonny Rollins. A reinforcement learning algorithm is used to refine the behavior of the neural network. The reward function rates the system solos in terms of jazz harmony criteria and according to Rollins style.

The lack of interactivity, with a human improviser, of the above approaches has been criticized [53] on the grounds that they remove the musician from the physical and spontaneous creation of a melody. Although it is true that the most fundamental characteristic of improvisation is the spontaneous, real-time creation of a melody, it is also true that interactivity was not intended in these approaches and nevertheless they could generate very interesting improvisations. Thom [53] with her Band-out-of-a-Box (BoB) system addresses the problem of real-time interactive improvisation between BoB and a human player. In other words, BoB is a "music companion" for real-time improvisation. Thom's approach follows Johnson-Laird's [54] psychological theory of jazz improvisation. This theory opposes the view that improvising consists of rearranging and transforming pre-memorized "licks" under the constraints of a harmony. Instead he proposes a stochastic model based on a greedy search over a constrained space of possible notes to play at a given point in time. The very important contribution of Thom is that her system learns these constraints, and therefore the stochastic model, from the human player by means of an unsupervised probabilistic clustering algorithm. The learned model is used to abstract solos into user-specific playing modes. The parameters of that learned model are then incorporated into a stochastic process that generates the solos in response to four bar solos of the human improviser. BoB has been very successfully evaluated by testing its real-time solo tradings in two different styles, that of saxophonist Charlie Parker, and that of violinist Stephane Grapelli.

Another remarkable interactive improvisation system was developed by Dannenberg [55]. The difference with Thom's approach is that in Dannenberg's system, music generation is mainly driven by the composer's goals rather than the performer's goals. Wessel's [56] interactive improvisation system is closer to Thom's in that it also emphasizes the accompaniment and enhancement of live improvisations.

A very recent and very remarkable interactive musical system is that of Pachet [57]. His system, Continuator, is based on extended multilayer Markov models to learn to interactively play with a user in the users' style and therefore it allows to carry musical dialogues between a human and the system.

4. Apparently or really creative

The described computational approaches to composing, performing, and improvising music are not just successful examples of AI applications to music. In my opinion are also valid examples of artificially creative systems because composing, performing, and improvising music are, undoubtedly, highly creative activities. Margaret Boden pointed out that even if an artificially intelligent computer would be as creative as Bach or Einstein, for many it would be just apparently creative but not really creative. I fully agree with Margaret Boden in the two main reasons for such rejection. These reasons are: the lack of intentionality and our reluctance to give a place in our society to artificially intelligent agents. The lack of intentionality is a direct consequence of Searle's Chinese room argument, which states that computer programs can only perform syntactic manipulation of symbols but are unable to give them any semantics. This critic is based on an erroneous concept of what a computer program is. Indeed, a computer program does not only manipulate symbols but also triggers a chain of cause-effect relations inside the computer hardware and this fact is relevant for intentionality since it is generally admitted that intentionality can be explained in terms of causal relations. However, it is also true that existing computer programs lack too many relevant causal connections to exhibit intentionality but perhaps future, possibly anthropomorphic, embodied artificial intelligences, that is agents equipped not only with sophisticated software but also with different types of advanced sensors allowing to interact with the environment, may have enough causal connections to have intentionality.

Regarding social rejection, the reasons why we are so reluctant to accept that non human agents can be creative is that they do not have a natural place in our society of human beings and a decision to accept them would have important social implications. It is therefore much simpler to say that they appear to be intelligent, creative, etc. instead of saying that they are. In a word, it is a moral but not a scientific issue. A third reason for denying creativity to computer programs is that they are not conscious of their accomplishments. However I agree with many AI scientists in thinking that the lack of consciousness is not a fundamental reason to deny the potential for creativity or even the potential for intelligence. After all, computers would not be the first example of unconscious creators, evolution is the first example as Stephen Jay Gould [58] brilliantly points out: *If creation demands a visionary creator, then how does blind evolution manage to build such splendid new things as ourselves?*

References

[1] M. Boden, *The Creative Mind: Myths and Mechanisms*, Basic Books 1991.
[2] M. Boden (Ed.); *Dimensions of Creativity*, MIT Press 1994.
[3] T. Dartnall (Ed.); *Artificial Intelligence and Creativity*, Kluwer Academic Pub. 1994.
[4] D. Partridge, and J. Rowe; *Computers and Creativity*, Intellect Books, 1994.
[5] P.J. Bentley and D.W: Corne (eds.), *Creative Evolutionary Systems*, Morgan Kaufmann 2002.
[6] J. Rowe, D. Partridge, Creativity: A survey of AI approaches, *Artificial Intelligence Review* 7 (1993), 43-70.
[7] B.G. Buchanan, Creativity at the Metalevel: AAAI-2000 Presidential Address, *AI Magazine* 22:3 (2001), 13-28.
[8] M. Balaban, K. Ebcioglu, and O. Laske (Eds.); *Understanding Music with AI*, MIT Press1992.
[9] E. R. Miranda (Ed.); Readings in Music and Artificial Intelligence, Harwood Academic Pub. 2000.
[10] D. Cope, *The Algorithmic Composer*, A-R Editions, Computer Music and Audio Digital Series, Volume 16, 2000.
[11] L. Hiller, and L. Isaacson. *Musical composition with a high-speed digital computer* (1958). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. *Machine Models of Music*. 9-21. Cambridge, Mass.: The MIT Press, 1993.
[12] S.M. Schwanauer, and D.A. Levitt, D.A., ed. *Machine Models of Music*. Cambridge, Mass.: The MIT Press, 1993.
[13] J.A. Moorer. *Music and computer composition* (1972). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. *Machine Models of Music*. 167-186. Cambridge, Mass.: The MIT Press. 1993.
[14] D.A. Levitt. *A Representation for Musical Dialects*. In *Machine Models of Music*, ed. Schwanauer, S.M., and Levitt, D.A., 455-469. Cambridge, Mass.: The MIT Press. 1993.
[15] G.M. Rader. *A method for composing simple traditional music by computer* (1974). Reprinted in Schwanauer, S.M., and Levitt D.A., ed.. *Machine Models of Music*. 243-260. Cambridge, Mass.: The MIT Press. 1993.
[16] H.A. Simon, and R.K. Sumner. *Patterns in Music* (1968). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. *Machine Models of Music*. 83-110. Cambridge, Mass.: The MIT Press. 1993.
[17] F. Lerdahl, and R. Jackendoff. *A Generative Theory of Tonal Music*, MIT Press 1983.
[18] M. Minsky. *Music, Mind, and Meaning.* (1981). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. *Machine Models of Music*. 327-354. Cambridge, Mass.: The MIT Press. 1993.

[19] J. Rothgeb. *Simulating musical skills by digital computer* (1969). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. *Machine Models of Music*. 157-164. Cambridge, Mass.: The MIT Press. 1993.

[20] K. Ebcioglu. *An expert system for harmonizing four-part chorales*. In *Machine Models of Music*, ed. Schwanauer, S.M., and Levitt, D.A., 385-401. Cambridge, Mass.: The MIT Press. 1993.

[21] J. Bharucha. *MUSACT: A connectionist model of musical harmony*. In *Machine Models of Music*, ed. Schwanauer, S.M., and Levitt, D.A., 497-509. Cambridge, Mass.: The MIT Press. 1993.

[22] J. Feulner. *Neural Networks that learn and reproduce various styles of harmonization.* In Proceedings of the 1993 International Computer Music Conference. San Francisco: International Computer Music Association. 1993.

[23] D. Hörnel, and P. Degenhardt. *A neural organist improvising Baroque-style melodic variations*. In Proceedings of the 1997 International Computer Music Conference, 430-433. San Francisco: International Computer Music Association. 1997.

[24] D. Hörnel, and W. Menzel. Learning musical strucure and style with Neural Networks. *Journal on New Music Research*, 22:4 (1998), 44-62.

[25] F. Pachet, and P. Roy. *Formulating constraint satisfaction problems on part-whole relations: The case of automatic harmonization*. In ECAI'98 Workshop on Constraint Techniques for Artistic Applications. Brighton, UK. 1998.

[26] J. Sabater, J.L. Arcos, and R. Lopez de Mantaras. *Using Rules to Support Case-Based Reasoning for Harmonizing Melodies*. In AAAI Spring Symposium on Multimodal Reasoning, pp. 147-151. Menlo Park, Calif.: American Association for Artificial Intelligence. 1998.

[27] S.M. Schwanauer. *A learning machine for tonal composition*. In *Machine Models of Music*, ed. Schwanauer, S.M., and Levitt, D.A., 511-532. Cambridge, Mass.: The MIT Press. 1993.

[28] D. Cope. *Experiments in Music Intelligence*. A-R Editions 1996.

[29] D. Cope. *Pattern Matching as an engine for the computer simulation of musical style*. In Proceedings of the 1990 International Computer Music Conference. San Francisco, Calif.: International Computer Music Association. 1990.

[30] W. Woods. Transition network grammars for natural language analysis. *Communications of the ACM* 13:10 (1970), 591-606.

[31] M.L. Johnson. *An expert system for the articulation of Bach fugue melodies*. In *Readings in Computer Generated Music*, ed. D.L. Baggi, 41-51. Los Alamitos, Calif.: IEEE Press. 1992.

[32] R. Bresin. *Articulation rules for automatic music performance*. In Proceedings of the 2001 International Computer Music Conference 2001. San Francisco, Calif.: International Computer Music Association. 2001.

[33] A. Friberg. *A quantitative rule system for musical performance*. PhD dissertation, KTH, Stockholm. 1995.

[34] A. Friberg, R. Bresin, L. Fryden, and J. Sunberg. Musical punctuation on the microlevel: automatic identification and performance of small melodic units. *Journal of New Music Research* 27:3 (1998), 271-292.

[35] A. Friberg, J. Sunberg, and L. Fryden. Music From Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. *Journal on New Music Research*, 29:3 (2000), 199-210.

[36] S. Canazza, G. De Poli, A. Roda, and A. Vidolin. *Analysis and synthesis of expressive intention in a clarinet performance*. In Proceedings of the 1997 International Computer Music Conference, 113-120. San Francisco, Calif.: International Computer Music Association. 1997.

[37] R.B. Dannenberg, and I. Derenyi. Combining instrument and performance models for high quality music synthesis, *Journal of New Music Research* 27:3 (1998), 211-238.

[38] R. Bresin. Artificial neural networks based models for automatic performance of musical scores, *Journal of New Music Research* 27:3 (1998), 239-270.

[39] R.A. Kendall, and E.C. Carterette. The communication of musical expression. *Music Perception* 8:2 (1990), 129.

[40] J.L. Arcos, R. Lopez de Mantaras, and X. Serra; Saxex: A Case-Based Reasoning System for Generating Expressive Musical Performances. *Journal of New Music Research* 27:3 (1998), 194-210.

[41] J.L. Arcos, and R. Lopez de Mantaras; An Interactive Case-Based Reasoning Approach for Generating Expressive Music. *Applied Intelligence* 14:1 (2001), 115-129.

[42] R. Lopez de Mantaras, and J.L. Arcos; AI and Music: From Composition to Expressive Performance. *AI Magazine* 23:3 (2002), 43-57.

[43] M. Grachten, J.L. Arcos, and R. Lopez de Mantaras; *TempoExpress, a CBR Approach to Musical Tempo Transformations*. In Proceedings of the 7th European Conference on Case-Based Reasoning (Eds. P. Funk and P. A. Gonzalez Calero), Lecture Notes in Artificial Intelligence 3155 (2004), 601-615.

[44] P. Desain and H. Honing. *Tempo curves considered harmful*. In "Time in contemporary musical thought" J. D. Kramer (ed.), Contemporary Music Review. 7:2, 1993.

[45] T. Suzuki, T. Tokunaga, and H. Tanaka; *A Case-Based Approach to the Generation of Musical Expression*. In Proceedings of the 16th International Joint Conference on Artificial Intelligence, Morgan Kaufmann (1999), 642-648.

[46] A. Tobudic, and G. Widmer; *Playing Mozart Phrase by Phrase*. In Proceedings of the 5th International Conference on Case-Based Reasoning (Eds. K.D. Ashley and D.G. Bridge), Lecture Notes in Artificial Intelligence 3155 (2003), 552-566.
[47] A. Tobudic, and G. Widmer; *Case-Based Relational Learning of Expressive Phrasing in Classical Music*. In Proceedings of the 7th European Conference on Case-Based Reasoning (Eds. P. Funk and P. A. Gonzalez Calero), Lecture Notes in Artificial Intelligence 3155 (2004), 419-433.
[48] T. Machover, *Hyperinstruments: A Progress Report*, MIT Media Lab Internal Research Report, 1992.
[49] C. Fry. *Flavors Band: A language for specifying musical style* (1984). Reprinted in Schwanauer, S.M., and Levitt, D.A., ed. 1993. *Machine Models of Music*. 427-451. Cambridge, Mass.: The MIT Press.
[50] J.A. Biles. GenJam: *A genetic algorithm for generating Jazz solos*. In Proceedings of the 1994 International Computer Music Conference. San Francisco, Calif.: International Computer Music Association. 1994.
[51] G. Papadopoulos, and G. Wiggins. *A genetic algorithm for the generation of Jazz melodies*. In Proceedings of the SteP'98 Conference, Finland. 1998
[52] J.A. Franklin. *Multi-phase learning for jazz improvisation and interaction*. In Proceedings of the Eight Biennial Symposium on Art and Technology. 2001.
[53] B. Thom. *BoB*: *An improvisational music companion*. PhD dissertation, School of Computer Science, Carnegie-Mellon University. 2001.
[54] P.N. Johnson-Laird. *Jazz improvisation: A theory at the computational level.* In *Representing Musical Structure*, ed. P. Howell, R. West, and I. Cross. London, UK and San Diego, Calif.: Academic Press. 1991.
[55] R.B. Dannenberg. Software design for interactive multimedia performance. *Interface* 22:3 (1993), 213-218.
[56] D. Wessel, M. Wright, and S.A. Kahn. *Preparation for improvised performance in collaboration with a Khyal singer*. In Proceedings of the 1998 International Computer Music Conference. San Francisco, Calif.: International Computer Music Association. 1998.
[57] François Pachet: Beyond the Cybernetic Jam Fantasy: The Continuator. *IEEE Computer Graphics and Applications* 24:1 (2004), 31-35.
[58] S. J. Gould; Creating the Creators, *Discover Magazine*, October (1996), 42-54.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Intelligent Applications Ltd: A History 1986 to 2005

Charlie NICOL and Linsay TURBERT

Abstract. This paper is in memory of Rob Milne, the founder and managing director of Intelligent Applications. Rob passed away on 5th June 2005 whilst attempting to climb Mount Everest. http://www.aiai.ed.ac.uk/project/everest/.

1. Introduction

Intelligent Applications was formed by Rob in 1985.

Figure 1. Rob founding Intelligent Applications.

After obtaining a Degree in Artificial Intelligence (AI) from Massachusetts Institute of Technology, Rob Milne decided to read for his Doctorate at Edinburgh University, which had acquired a world-wide reputation for excellence in this field.

Rob was impressed by Silicon Glen and the large number of high technology companies. He felt strongly that there was tremendous potential for development of Artificial Intelligence based applications in the area and was surprised how little was being done to exploit this. Rob recognised significant opportunities for Scotland as the base for a high technology company. He also very much enjoyed his time in Edinburgh and while there, met his wife, Val. He resolved that he would establish his own business at some time to exploit these opportunities and that it would be in the Edinburgh area.

The next few years were spent with the US Military in Washington where Rob was fast becoming one of the world authorities on AI. During that time he worked with many government agencies and with the top members of the AI community. This provided him with considerable experience in the applications of AI. He identified a need for a company to specialise in the application of AI techniques to the near term problems of industry in the field of machine condition monitoring and fault diagnosis.

To assist in the establishment of the business, Rob was aided by Peter Bennett, a Business Consultant with extensive commercial experience who became a Director and shareholder of the company. Initial start-up funding was provided by two investment trusts managed by Baillie Gifford & Co., who specialise in early stage financing of high technology companies.

INTELLIGENT APPLICATIONS LIMITED

Specialists in the application of Artificial Intelligence

Figure 2. Intelligent Applications logo.

Objectives:
To produce products using Artificial Intelligence technology and sell these mainly through third party vendors.

1985 to 1987 In the early years, the company focused on product definition, development and early sales. With the first three products essentially complete, sales started by the end of the second year. Initial market feedback and beta testing resulted in further enhancements to the products. Rob identified many opportunities to exploit the new technologies that existed in AI so as to help solve industrial problems. Of the three products developed in the first years, SYNERGIST was developed and was the only product sold world-wide to help automate the expensive process of manually testing printed circuit boards. ANNIE was designed to exploit the increasing usage of AI in process control by providing a low cost entry level system for process monitoring. VIOLET was designed to capture the emerging demand for better and automatic data interpretation for machinery condition monitoring.

1988 Completed development of Amethyst, derived from the original Violet product to meet the special market need of IRD Mechanalysis Inc. The company performed its first projects based on the Annie product. At this time, the original Baillie Gifford venture capital investors decided to close their fund and offered to sell their interest to Rob and Peter. That transaction was completed leaving and majority of the company owned between Rob and Peter with small shareholdings allocated to Charlie Nicol and Jon Aylett.

1989 A number of turnkey applications were undertaken for various Blue Chip clients including a very large system for British Steel. During that period, Intelligent Applications developed more working expert system based applications than any other UK company. In addition, the Amethyst product started to show considerable market success.

1990 Project development had given the company a wealth of experience in its field and insight into the practical problems for industry in adopting the benefits of Expert System solutions to its diagnostic and maintenance problems. The achievement of profitability from project work had also been difficult.

As a result, the company shifted its emphasis from project work to development of new and existing products. The Amethyst product was further enhanced to build upon its early successes. Two additional new products were brought to market and considerable research was conducted into further product directions.

The company has therefore progressed from early stage product development, through large-scale project development, to a market driven, product based company.

During its fifth year, the company received a prestigious SMART Award from the DTI in order to develop a new product based on its electronics testing experience. Subsequently, the company was further honoured by being awarded a Queens Award for Technology, the first expert systems company to do so.

1991 In the sixth year, the company started to benefit from the groundwork that had been laid. Amethyst continued to sell strongly, the new Test-It product established an initial user base with strong repeat orders from Compaq, and the development of new products received considerable support. The company received a second successive SMART Award, this time to support a product for building energy analysis.

The company received funding support from the DTI to lead a seven company collaborative project to begin the initial development of a machine tool product. At the end of the year, the company was awarded a European ESPRIT project. IA was also taking the lead of a £2.8 million project involving six companies over three years to develop a product for gas turbines. Despite the severe U.K. recession, the company made a respectable profit, which was ahead of its original projections.

1992 to 1995 Intelligent Applications undertook the initial development of the TIGER gas turbine monitoring system, in collaboration with others assisted by substantial EU funding assistance.

1995 to 2001 Intelligent Applications made a deal with John Brown Engineering for resale of TIGER. The company also made sales of its Cased Based Reasoning (CBR) technology and skills. Intelligent Applications made the strategic decision to invest heavily in TIGER as its major product and to minimise work in other areas.

2001 to 2005 Intelligent Applications was acquired by Sermatech in January 2001 (part of Teleflex group). The company continued to develop and sell TIGER, but diversified in 2003 to provide web based applications for Teleflex Corporate.

Figure 3. Rob on the Sermatech Exhibition Stand.

1.1. Awards

- Queen's Award for Technology 1991
- DTI Manufacturing Intelligence Competition (1989, 1990, 1991, 1992)
- LEEL 'Small Firms Access to Europe Competition Export Award'
- SMART Award (1990, 1991, 1992)

1.2. The Staff

The personnel strategy at Intelligent Applications had always been to have a small number of highly qualified technical people who are competent, not only in the software development, but also in the engineering application domains.

A hallmark of the company's approach was that the staff were all experienced software developers, and so were able to become effective very quickly in the particular application area in which they were involved. By having a deep understanding of the application problem as well as the tools available to solve it, they were able to produce effective solutions.

1.3. Location

Intelligent Applications was located in modern premises in Livingston, just west of Edinburgh, Scotland. This is the heart of Scotland's Silicon Glen, and within ready access to a wide range of software expertise and personnel.

Figure 4. Intelligent Applications Offices in Livingston.

1.4. Background

Intelligent Applications specialised in the intelligent interpretation of data. This intelligent interpretation took two forms, addressing two different market sectors. For the manufacturing, heavy engineering and process industries, Intelligent Applications specialised in intelligent monitoring and diagnostic systems for industrial equipment and electronic systems. Systems in this area had been either, on-line automated diagnostic and monitoring systems, or off-line intelligent diagnostic assistant systems.

In the financial sector, the intelligent interpretation of data took the form of data mining services. Intelligent Applications worked in a number of areas and were successful in areas as diverse as predicting customer retention problems related to mortgages to predicting paper defect problems for paper mills. The major areas of activity of the company were

Vibration based condition monitoring	Electronic fault diagnosis
Gas turbine fault diagnosis	Building energy management
Machine tool fault diagnosis	Case base reasoning

1.5. Other Experience

Intelligent Applications built more working diagnostic systems using knowledge based and artificial intelligence technology than any other UK company. This gave them a considerable base of applications experience for example:

- Intelligent gas chromatography analysis for nuclear power stations.
- Continuous diagnostics of the steel making process.
- Automated ladder logic diagnostics for programmable logic controllers.
- Automated diagnostic system for double-ball bar analysis of machine tools.
- Intelligent monitoring and diagnostics of building air conditioning, and environmental control systems.

2. Products From Intelligent Applications

2.1. TIGER™

Tiger Gas Turbine Condition Monitoring and Diagnostics: Intelligent Applications was the leader of a major European ESPRIT Project TIGER, to develop advanced condition monitoring and diagnosis systems for gas turbines. Project partners were; Exxon Chemical, Dassault Aviation, John Brown Engineering, CNRS-LAAS and UPC In Barcelona. The project aimed to develop a sophisticated monitoring and diagnosis tool for a range of gas turbine systems. An initial version of the system produced significant cost benefit to Exxon Chemical.

The TIGER project was successful and became a product. . . .

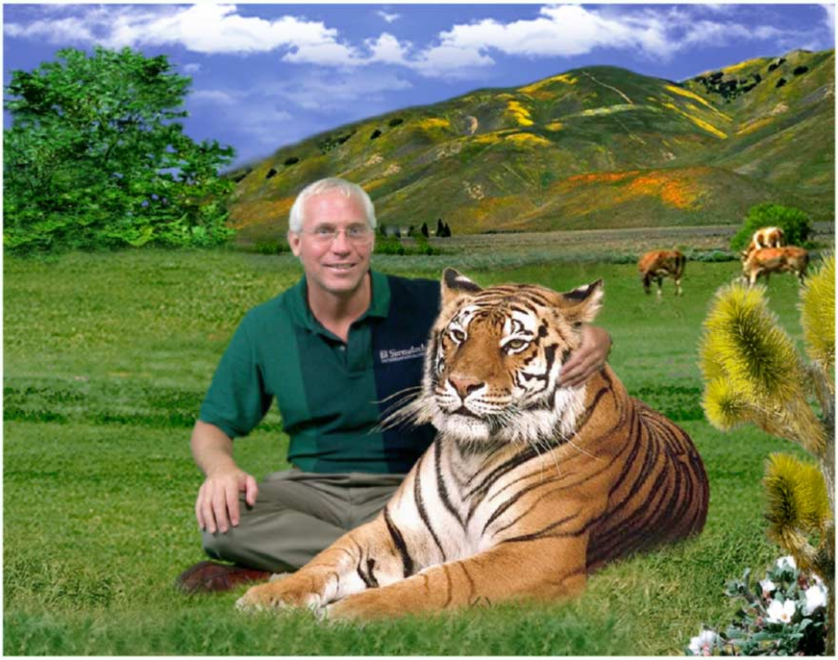


Figure 5. Rob with a tiger – used as his screen background for many years.

TIGER was and still is the most advanced gas turbine condition monitoring system available. It combines real-time data acquisition and knowledge based system techniques with a unique capability for temporal management and dynamic systems monitoring, and also a unique module for qualitative prediction and model based diagnosis.

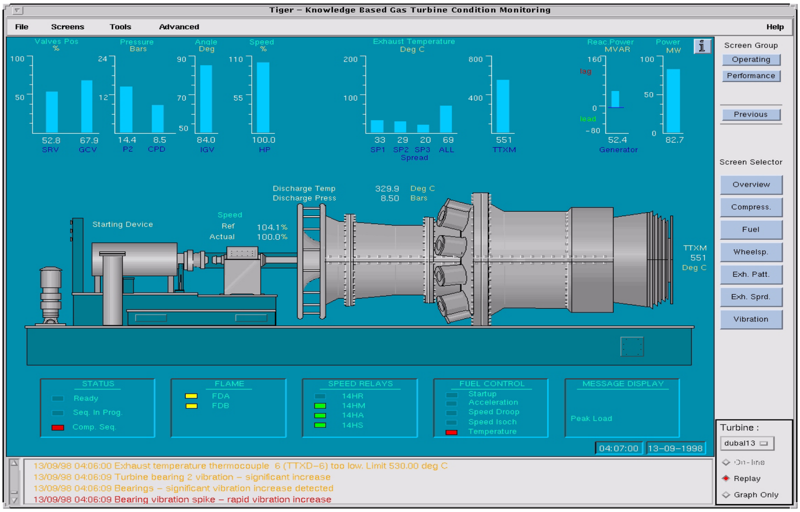


Figure 6. TIGER software - overview of a GE Frame 6 gas turbine.

TIGER collects data at once per second intervals from the gas turbine controller and performs a complete interpretation and diagnosis every second.

TIGER is invaluable in spotting problems at a very early stage, as well as providing a rapid diagnosis in the case of a trip or major problem.

TIGER is a very comprehensive and powerful product, with an advanced user interface, diagnostic links to manuals, support for data archiving and trending, in addition to the presentation of diagnostic conclusions.

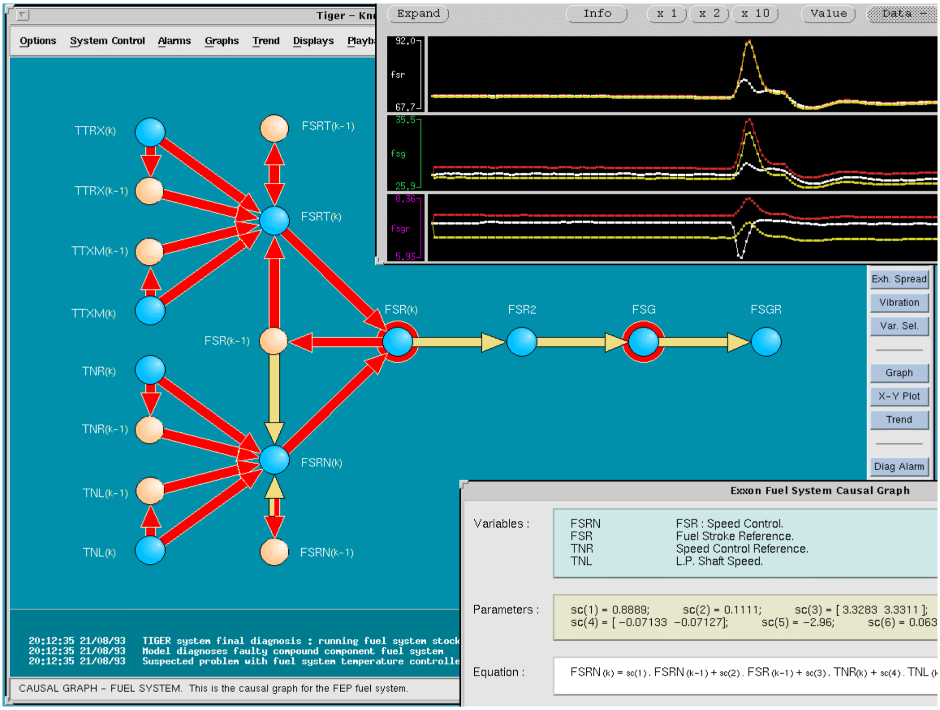


Figure 7. TIGER software - fault detection using CAEN qualitative models.

It is a major innovation in the gas turbine monitoring area, and has received a number of awards including the European Information Technology Prize from the European Commission, and a SMART award from the UK's Department of Trade & Industry.

The TIGER product was trademarked by Intelligent Applications in the UK and throughout Europe. In addition, the core algorithms were patented by Rob and C. Nicol.

2.2. Amethyst

The Amethyst product performed intelligent diagnostics of data collected with hand-held data collectors to support vibration based condition monitoring. Amethyst was sold worldwide by IRD Mechanalysis in the United States and sold many hundreds of copies over the years.

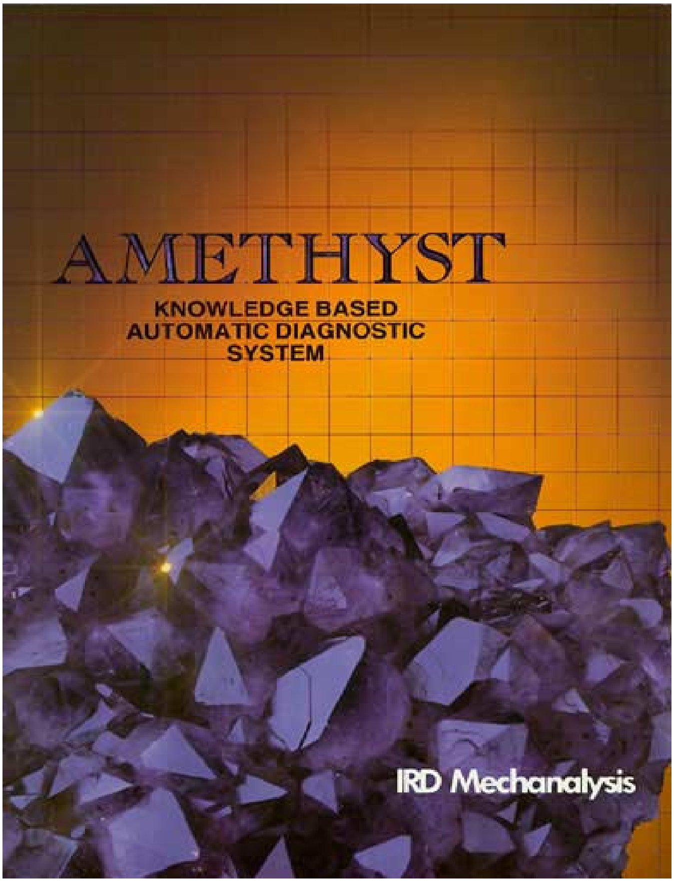


Figure 8. Amethyst vibration based condition monitoring brochure cover.

Amethyst was packaged as a fully automated solution such that when an engineer collected vibration based condition-monitoring data using a data collector, Amethyst would automatically scan all the new data, perform a diagnosis and generate reports for the engineer.

Because of the success of Amethyst, Intelligent Applications received the prestigious Queens Award for Technology in 1991. The development of Amethyst gave Intelligent Applications considerable experience in diagnostic of vibration data for rotating machinery. French, Italian and Spanish versions were available.

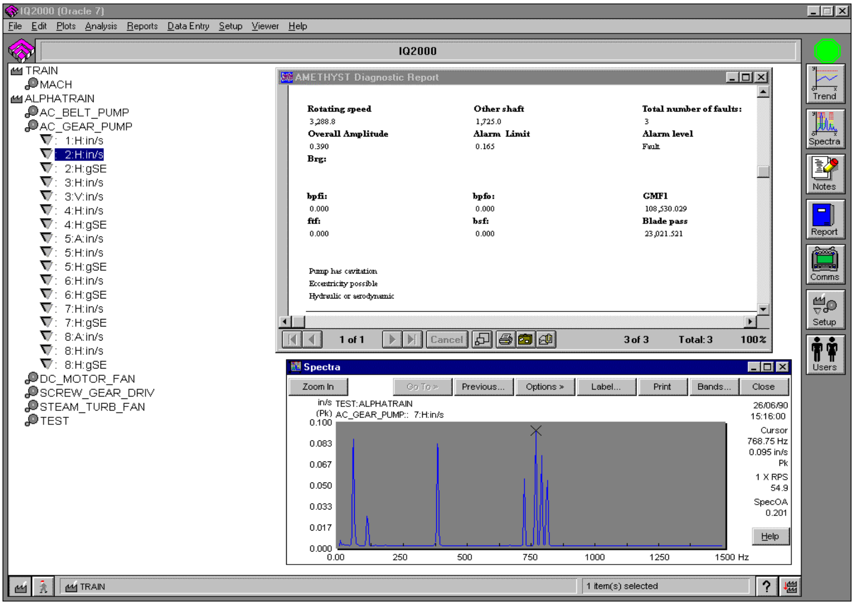


Figure 9. Amethyst diagnosing vibration problems.

2.3. TEST-IT

In the area of off-line intelligent diagnostic assistants, Intelligent Applications developed a specialised product known as Test-It to enable the rapid development of intelligent troubleshooting assistance.

After many years of developing intelligent diagnostic systems, Intelligent Applications developed Test-It to be conceptually straightforward to understand and yet able to represent the powerful range of troubleshooting steps a technician might require. A substantial Test-It application was built, to assist with troubleshooting the electronics that support the command consoles for a submarine.

The Test-It system incorporated a combination of rule based expert system technology with case based reasoning technology. It was ideal for developing intelligent diagnostic assistance to guide technicians through the troubleshooting of any large complex electrical, electronic or electro-mechanical system.

Intelligent Applications received a prestigious 1990 SMART Award (Small Firms Merit Award for Technology), and a follow on 1991 SMART Stage 2 Award to develop a new product for manual diagnosis of electronic circuit boards. This easy to use product not only allowed low skilled people to test systems, but also provided extensive fault reporting and history management. Users included Apricot Computers, Compaq Computers, Delco, Hewlett Packard and OKI Printers.

Intelligent Applications also built specialised diagnostic systems integrated with automatic test equipment for companies such as British Airways.

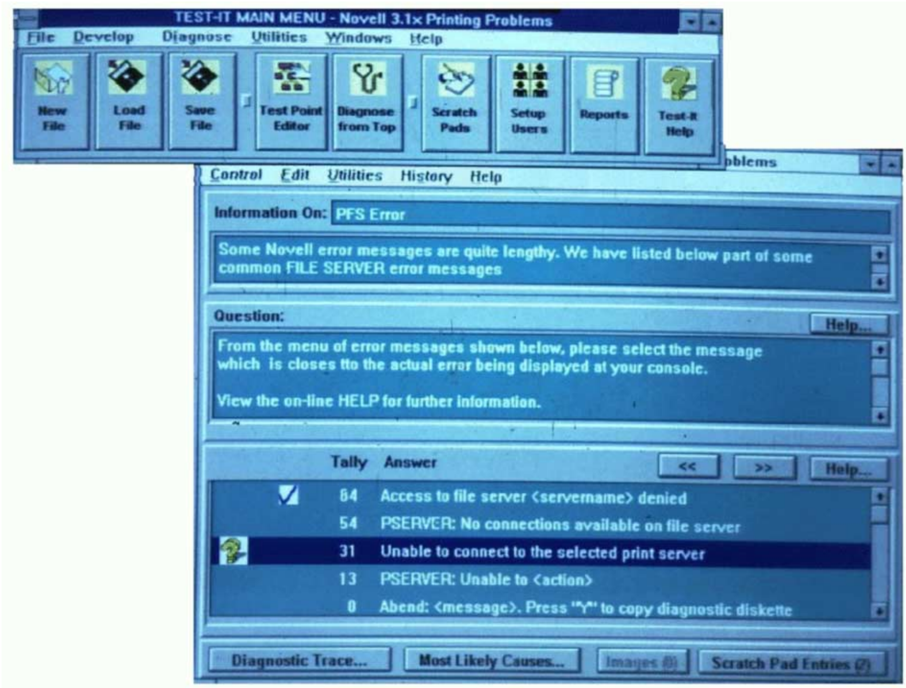


Figure 10. Test-it troubleshooting assistance for printers.

2.4. Other Products

Annie: The Analogue Interface Expert, which performed mechanical health-monitoring, control, and alarm monitoring in engineering applications. In addition to automatic monitoring, Annie allowed data logging, trend analysis of logged data and analysis of real or simulated data.

Annie coupled a PC-based or VAX VMS based expert system shell with a data acquisition system, via either a serial interface to a process control system, or directly to transducers via an IBM PC A/D conversion system.

The user built a knowledge base (in the expert system) that was specific to the particular application. The rules in the knowledge base used Annie software to interface between the expert system and the data acquisition system. The rules would interpret the data, based on the knowledge of an expert.

The Annie product was usually sold with consultancy for the provision of a data acquisition interface and rule base development. Annie proved successful in providing the kernel of such application development projects.

Violet: A vibration-based, machine health monitoring system. Any machine with critical high speed rotating parts, such as helicopters, mills, pumps or shafts, could be monitored by this product. Violet was a software tool which helped to determine the machinery's condition by analysing the vibration produced by the machinery as it operated. Vibration data was captured by a variety of industry standard systems and analysed immediately, for continuous automatic monitoring, and later, for trend analysis.

Violet coupled a PC-based expert system shell with a data acquisition system via spectrum analysis. Transducers or accelerometers on the machinery connected to a spectrum analyser which produced a spectrum of the vibration signatures. Violet could function with data from tape recordings, hand-held data collectors or on-line monitoring systems.

The user placed the knowledge necessary to interpret the spectrum data into a Violet knowledge base. This knowledge was in the form of rules that interpret the machine's health based on the frequencies and amplitudes of the vibration data. Violet was the basis of the *Amethyst* product.

Synergist: assisted with the repair of the electronic printed circuit boards. It was determined from market feedback however, that the *Synergist* product was too large, too complicated and too powerful. Responding to this market feedback, Intelligent Applications re-built the entire product concept which resulted in the *Test-It* product.

2.5. Projects Carried Out By Intelligent Applications

Machine Tool Diagnosis (FMS MAINT Project): Intelligent Applications led a UK consortium investigating ways of using expert systems to improve the fault diagnosis of machine tools. The project included major users at British Aerospace and Rolls-Royce. A fault diagnosis expert system was installed at these factories. The project continued with 7 companies from Finland and 6 from the UK supported under the EUREKA program.

Polar-Check Measurement Simulator: The Polar-Check Measurement Simulator was a product that helped the user understand how some of the most common faults occurred from analysing the path of a machine tool trying to contour a perfect circle. This easy to use Windows based software allowed the user to enter the error magnitude giving immediate graphical feedback. The software allowed the user to read in actual double ball bar measurement files and adjust the simulator areas to match them. This helped the user to visualise the underlying error affecting product quality. Simulation results could be printed out as reports, including the simulation parameter values and the resulting graphs.

Polar-Check Measurement Analyser: The Polar-Check Measurement Analyser offered low cost diagnostics for machine tool quick checks. It was aimed at identifying the presence of the most common machine tool errors and defining their magnitude. This historical trending of stored measurements also helped the user to detect and react to desirable trends before they cause defective products. This easy to use Windows based software quickly guided the user to the needed measurements and analysis and produced a wide range of reports both from individual measurements and of the error history of the machine tool.

Building Energy Management: Intelligent Applications received a 1991 SMART Award and a follow on SMART Stage 2 to develop a new product, DITTO, to identify energy wastage in building and heating systems. The product automatically compared the heating profile of a room with the desired profile to identify faults in the system and energy wastage. This work was preceded by a system to diagnose alarms in building energy control systems.

Case Based Reasoning: An important new technology was Case Based Reasoning (CBR). This was used to build problem-solving systems based on past solutions. Rather than explicitly develop a set of rules, the system stored past examples and could intelligently retrieve the most similar past solution to a difficult problem. Intelligent Applications distributed a world leading CBR product and offered consulting services to assist customers in developing successful systems. Customers included; Coopers (GB), National Power, UWCC, Dundee Institute of Technology, Tampere Institute of Technology and NIKEL.

Data Mining: This area involved applying learning techniques to find hidden patterns in databases. For example, to identify which customers always pay on time and which always pay late. Intelligent Applications undertook consultancy to use the CBR data mining tools for specific, confidential, customer requirements.

Customer Retention: Intelligent Applications used data mining to try and identify problems in customer retention for a major financial institution. Intelligent Applications were trying to predict which customers might change their mortgage to another financial institution in the near future. Prior to this work, the financial institution had no way of predicting this information. Using the techniques of knowledge guided induction, Intelligent Applications were able to predict which customers were likely to leave their financial institution with an accuracy of over 90%.

Caledonian Paper: Intelligent Applications had applied data mining in a diverse range of areas. A major success was predicting defects in paper for a major paper mill. Based on historical information of which reels of paper have had problems or degradations of quality. Intelligent Applications were able to build a predictive model that detected deviations in the paper making process that later lead to defects in the final paper product. This system was installed on-line for continuous monitoring at a major paper mill in Scotland and was used to detect faults at an early stage while there is still time to correct them.

Competitive bidding: Intelligent Applications help predict what the competitors price would be in a competitive price bidding situation for a major utility, giving information about the national grid and past performances. Resulting predictions were much more accurate than the existing human experts, and the company continued to refine this work.

Knowledge Based Vehicle Scheduling: Working with Edinburgh University's AIAI, Intelligent Applications implemented a Vehicle Scheduling System for a Scottish based company. The system kept track of the statutory drivers hours and rest period, and helped to then schedule the delivery of orders throughout Scotland, whist maintaining the necessary constraints on the vehicles available, the driving time, and the drivers regulations

CORECT: The CORECT project involved five UK companies in developing a product to help support the configuration of complex modular products. The sales person specified the general requirements and a proposed solution. A knowledge based system then evaluated the proposed solution and made sure it is consistent and would work properly.

Fix-Novell Printing: Fix-Novell Printing offered low cost intelligent diagnostics for Novell Networks. It was aimed at helping the majority of network users and administrators to quickly resolve many of the printing problems that occurred on Novell LANs.

British Steel Ravenscraig: Intelligent Applications built a very large real-time expert system to monitor the Steel Making Process and identify faults which affected the overall process. It was one of the largest real-time expert system projects projects ever developed in the United Kingdom.

British Steel Scunthorpe: Intelligent Applications built an on line expert system to monitor the waste gas recovery from the Steel Making Process so that this gas can be used as fuel. The system was capable of real-time monitoring the gas collection system and rapidly detecting faults so that they could be repaired very quickly. The system received an award in the 1990 DTI Manufacturing Intelligence competition.

Exxon Chemical: Intelligent Applications built an on-line continuous monitoring system of the vibration characteristics of the main pumps and compressors on the plant. The system rapidly identified problems so that they can be rectified more quickly. A second system was installed that provided on-line monitoring and diagnosis of gas turbine faults, evolving into the TIGER system. This work was also supported by the Offshore Supplies Office.

South of Scotland Electricity Board, Torness Nuclear Power Plant: Intelligent Applications built an on-line continuous monitoring expert system which looked at the mix of gases within the reactor vessel based on gas chromatography analysis to rapidly identify any problems occurring.

Expert Bearing Analysis: Working in co-operation with Diagnostic Instruments, Intelligent Applications embedded an expert system into a small battery powered spectrum analyser. This allowed the engineer to get a complete and detailed analysis of the bearing condition on the spot. This was the first time an expert system had been embedded into a battery-powered instrument and represented a major step forward for the condition monitoring industry.

3. Intelligent Applications Limited Awards

Figure 11. Award trophy wall and table at Intelligent Applications.

3.1. Queens Award for Technology

Queens Award for Technology – 1991 Intelligent Applications Limited of Livingston, Scotland has become the first company to receive a Queens Award for Technology for an expert system based application.

Figure 12. Rob receiving the Queens Award for Technology – 1991.

Intelligent Applications gains the Award for the development of 'Amethyst', an expert system for the automatic diagnosis of rotating machinery faults. It is used in vibration based condition monitoring and achieves in a few minutes what previously took 4-8 hours. The vibrations are collected using a hand-held vibration data collector, the output of which is analysed on an IBM PC compatible computer to indicate precise fault conditions enabling lower skilled staff to be employed and eliminates unnecessary strip down and repair leading to increased overall plant efficiency.

Amethyst was developed in conjunction with IRD Mechanalysis (UK) Ltd in Chester, UK and is sold by them worldwide. It has now sold over 300 copies, making it the most widely used expert system application in engineering.

This is the first time that an expert system based product received a Queens Award for Technology.

Figure 13. Rob with the Queens Award for Technology – 1991.

3.2. ITEA'96 Awards

The European Information Technology Prize – 1996 The President of the European Commission, Jacques Santer, has announced that the new TIGER product developed by Livingston based Intelligent Applications Limited is one of the Winners of the *European Information Technology Prize, 1996*.

Figure14. Rob receiving the European Information Technology Prize – 1996.

The TIGER system automatically detects problems or reductions in efficiency of a gas turbine, such as used in the 'dash for gas' power stations. TIGER uses an innovative combination of advanced artificial intelligence technologies to capture the knowledge and experience of the best engineers and check every second whether the gas turbine is okay or problems are developing. Any reduction in efficiency leads to higher fuel consumption and an increase in pollution. Although gas turbines are most well known as aircraft engines, Intelligent Applications is starting with ground based industrial gas turbines used in the power generation, chemical and process industries since the regulations of the aviation industry mean that it will take some time to get established for aircraft.

Figure 15. The TIGER booth at The European Information Technology Prizegiving – 1996.

3.3. SMART Awards

Small Firms Merit Award for Research and Technology – 1997 The Scottish Office, Industry Department have announced that Livingston based hi-technology software company Intelligent Applications have won one of the SMART awards in their 1997 competition.

Figure 16. Rob receiving the Small Firms Merit Award for Research and Technology – 1997.

Intelligent Applications' recently launched TIGER knowledge based gas turbine condition monitoring system and is already revolutionising the automatic monitoring of gas turbines worldwide by using state of the art artificial intelligence techniques. This SMART award will allow them to increase their technological lead in the marketplace by developing a model based monitoring system using a combination of traditional and artificial intelligence modelling and model based diagnosis techniques. The resulting system will provide more sensitive early detection of problems, helping to operate modern gas turbines with greater efficiency, hence reducing fuel consumption and air pollution.

Small Firms Merit Award for Research and Technology – 1992 Intelligent Applications Ltd has been awarded a Stage II Small Firms Award for Research and Technology (SMART) for continuing their work with their DITTO product.

Figure 17. Rob receiving the Small Firms Merit Award for Research and Technology – 1992.

Small Firms Merit Award for Research and Technology – 1991 SMART award awarded to Intelligent Applications for their feasibility study to apply expert systems to energy consumption management. The resulting product – DITTO – was able to analyse the difference between the desired internal temperature and the actual internal temperature profile, whilst also spotting any mechanical and people-related problems. The innovation in this system was to overcome the problems experienced in companies trying to manage their energy use with mechanical methods by applying artificial intelligence techniques to energy analysis for the first time. The anticipated product was an expert system product which was fully automatic, on-line, and very easy to use. It continuously monitored energy consumption with no human intervention.

Small Firms Merit Award for Research and Technology – 1990 Intelligent Applications Ltd is one of this years winners of the Small Firms Award for Research and Technology (SMART) for their expert system based software diagnostic tool for fault location on printed circuit boards and electronic assemblies. This should fill a market demand for this type of test equipment for PCB manufacturers, users, and also within the electronics repair industry. This is the first DTI SMART award for expert systems.

Figure 18. Rob receiving the Small Firms Merit Award for Research and Technology – 1990

3.4. Manufacturing Intelligence Awards

DTI - Manufacturing Intelligence Award – 1992 Certificate of Merit awarded to Exxon Chemical at Mossmorran in Fife and Livingston based Intelligent Applications Ltd. The award is based on the considerable cost benefit of an automatic diagnostic expert system for the gas turbine at the Fife plant. Exxon and Intelligent Applications have developed an expert system that monitors how well the critical gas turbine is working and provides an early warning of problems. An expert system is a computer program that captures the expertise of a human expert; in this case the plant machinery experts.

DTI - Manufacturing Intelligence Award – 1992 Certificate of Merit awarded to Glaxochem Ltd and Intelligent Applications Ltd for Amethyst Ring, an expert system which diagnoses machinery problems from vibration data. Vibration generated in a running machine has long been recognised as a prime indicator of its condition but the analysis of this vibration to provide the diagnostics is complex and Amethyst Ring automates this process.

DTI - Manufacturing Intelligence Award – 1990 Certificate of Merit in recognition of achievement of their particular systems were awarded to British Steel Technical and Intelligent Applications Ltd for their exemplary use of expert system techniques to improve gas collection in the Basic Oxygen Steel making process.

DTI - Manufacturing Intelligence Award – 1989 Awarded to Intelligent Applications Ltd for their expert system, Amethyst, which automates the interpretation of rotating machinery vibration data in order to facilitate condition based maintenance. The Certificate of Merit was awarded in recognition of this small company's success with this product in the export market.

Figure 19. Rob Receiving the DTI - Manufacturing Intelligence Award – 1989.

4. Press Release

Intelligent Applications Acquired by Sermatech International, USA for their TIGER™ Gas Turbine Condition Monitoring Software Livingston, Scotland. 22 May 2001

Intelligent Applications Ltd, the Livingston based high technology software firm have announced that they have been acquired by Sermatech International, USA. Sermatech have acquired Intelligent Applications (IA) to facilitate bringing IA's TIGER™ Gas Turbine Condition Monitoring Software System to the US and other key global power generation markets.

TIGER enables industrial gas turbine operators, primarily power generation companies, to understand the condition of their units in real time; detect potential problems at the earliest possible stage and analyse what happened when problems occur.

Sermatech International is a leading supplier of products and services that include turbo machinery life cycle support services; protective coatings; manufactured turbine components for use in power generation; industrial and aerospace applications. They are part of the Teleflex Group, listed on the New York Stock Exchange. The acquisition of IA is part of their product portfolio expansion into gas turbine services, including monitoring and diagnostics.

This will provide great benefit to users since TIGER's knowledge-based analysis provides a unique window into the operation and maintenance needs of a gas turbine. The system has consistently benefited users by pinpointing previously unidentified critical faults at an early stage, then alerting the operator to plan for the needed maintenance actions. *"It's been our experience that virtually all of TIGER's users become members of what we call the 'TIGER 150K Club', which means that TIGER has regularly provided benefits and savings of at least $150,000 in one year",* Dr. Milne summarised.

Through the data it analyses, TIGER enables users to maximise turbine efficiency and availability by:

- Reducing unit start-up times and overall maintenance costs
- Optimising the timing and length of outages
- Optimising the timing of turbine wash cycles
- Minimising operational and subsystem malfunctions
- Identifying needed maintenance
- Maximising turbine output in relation to NOx

TIGER provides a comprehensive view of a turbine's operation through continuous sampling and analysis of over 600 data inputs that are updated once per second. TIGER's "knowledge" system is based on more than 750 turbine operation incidents recorded over the equivalent of 25 years of turbine experience, and continues to grow as the number of TIGER users increases.

Intelligent Applications Ltd, located in Livingston, West Lothian, is an industry leader in developing intelligent software systems technology for gas turbine condition monitoring. The company specialises in online fault detection and monitoring software utilising artificial-intelligence-based technologies. IA has developed the greatest number of successful systems of this kind in the United Kingdom, and has been recognised with the Queen's Award for Technology for its innovations. For more information about TIGER and IA, visit the company's World Wide Web site at www.intapp.co.uk.

Sermatech Technical Services, headquartered in Limerick, PA (USA), is a leading provider of gas turbine life cycle support services; its capabilities range from protective coating services and specialised rotor and component repairs and upgrades to long term maintenance arrangements; field services and outage management. Sermatech Technical Services is part of Sermatech International, a leading supplier of products and services that include turbo machinery life cycle support services, protective coatings and manufactured turbine components for use in power generation, industrial and aerospace applications. Additional information about Sermatech International can be obtained from the company's World Wide Web site at www.sermatech.com.

Announcing the multi-million pound deal, Dr. Robert Milne, Managing Director of Intelligent Applications said, "*There is tremendous synergy with the two companies. Sermatech are well established in the gas turbine market place, not only in the US, but worldwide. Our TIGER product is the ideal system to enable them to expand their business into monitoring and diagnostic services. At the same time, it opens the rapidly growing US marketplace to Intelligent Applications*."

Intelligent Applications will continue to operate as an autonomous software development company for TIGER under the leadership of Dr. Robert Milne as Managing Director and Dr. Charlie Nicol as Technical Director. The company is expected to grow considerably from its current staff of 10.

5. The End

In August 2005, Teleflex, the controlling corporation of Intelligent Applications continued its exit from the gas turbine market by moving some staff from Intelligent Applications into its corporate IT division, selling off the intellectual property rights belonging to Intelligent Applications and making the remaining staff redundant. The company is now dormant.

6. Epilog

The TIGER software developed by Intelligent Applications is alive and well with Turbine Services Limited in Glasgow http://www.turbineserviceslimited.com/.

Rob's legacy continues. . .

Figure 20. Rob at his desk at Intelligent Applications, Livingston.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Rob Milne: The British Airways Years

Rick MAGALDI
AI Manager for Engineering, British Airways AI Unit, Retired

Abstract. A particular period of Rob Milne's professional contact with Industry that many of his academic colleagues and friends may not be familiar with is that during which he provided consultancy to British Airways. Rob became involved in a number of projects carried out by the British Airways Artificial Intelligence Unit from the late 1980's through to the late 1990's. The following is an account of some of the more important project areas that Rob and his colleagues from Intelligent Applications participated in, as well as highlighting the important contribution that he made to many of the business decisions required of the Artificial Intelligence Unit in developing knowledge-based solutions for the airline, and to the industry in general.

Keywords. Automated diagnostics, real-time systems, expert systems, knowledge management, case-based reasoning, neural networks

Introduction

Rob Milne and I became acquainted during the mid 1980's due to our mutual interests in applying artificial intelligence techniques to problems in the field of engineering, and in particular the area of equipment failure diagnostics. We subsequently worked together on a variety of initiatives, during which time we travelled the world together, attending countless business meetings and conferences.

A particularly interesting period of our various collaborations was during the time I worked for British Airways. During the late 1980's, British Airways formed an Artificial Intelligence Unit under the leadership of both Keith Rapley (Information Management Research) and Louis Busutil, the unit being tasked with introducing the then perceived benefits of knowledge-based systems to as many areas of the airline as possible. My role in this enterprise was to investigate and develop applications for the engineering division of the airline.

From the outset, it became particularly important to establish what expertise could be called upon from outside of the airline to support potential applications delivery. Market analyses quickly established that the company Intelligent Applications founded by Rob Milne was likely to be of significant help in providing the necessary specialist experience and expertise that would be required in developing diagnostics systems. This was particularly so on the development of a knowledge-based application called Caddis (**C**omputer **A**ided **D**iagnostics and **D**ecision **I**nformation **S**ystem) [1].

1. Avionic Equipment Diagnostics and Expert Systems

Caddis was intended to provide an on-line real time diagnostic aid for use by avionics technicians in the British Airways automated test equipment (ATE) workshops at London Heathrow. It was designed for use with the Collins ITS-700 ATE rig and directed towards aiding diagnostics on the EFIP-701 flight instrument processor [2]. The EFIP-701 being a symbol generator, processing and displaying a variety of flight information to cockpit displays such as weather radar, aircraft altitude and position.

Rob Milne and Chris Nelson of Intelligent Applications provided the real-time system design and knowledge, as well as assisting Brett Trudgeon in the design and construction of the knowledge-based components and user interfaces. The system development and testing took place at both Livingston and Heathrow. The Caddis diagnostic aid was implemented in Crystal, and consisted of a backward chaining inference engine, rule-base and Annie, a software package supplied by Intelligent Applications. Annie consisted of a set of functions, which allowed communications between Caddis and the ITS-700 test rig, and between Caddis and a Hewlett Packard digital oscilloscope.

This was a fairly ambitious project, considering that many of the major suppliers of avionic test equipment at that time had nothing similar in their development pipeline, and were only just becoming aware of the potential of expert systems to augment the functionality of their products. Caddis was eventually installed in the Avionics workshops at Heathrow and was subject to intensive testing over a period of several months, during which time Chris Nelson worked closely with British Airways avionic workshop ATE technicians to de-bug the system and have it accepted as an in-service tool.

This was eventually achieved, and the system proved to be successful in generating useful diagnostics. However, the relocation of avionic workshops away from Heathrow to a new site in South Wales led to the system having to be decommissioned and re-installed at the new location. This was achieved after some months delay, and was supported by Intelligent Applications for the remaining period of the support contract. Lack of continuing sponsorship by the new owners combined with a new generation of manufacturer-supplied ATE's eventually led to decline in its usage.

However, valuable lessons were learned during this project regarding the importance of delivering cost-beneficial solutions within acceptable timescales, the often-unforeseen difficulties of integrating several technologies and the need to carefully consider maintenance overheads and mid-life improvements within knowledge-based system life cycles.

2. Engine Condition Monitoring and Fault Diagnostics

During the latter period of Caddis development, Rob and I had various discussions with British Airways propulsion engineers about real-time condition monitoring for jet engines. We spent some time working on a funding proposal advocating the benefits of artificial intelligence support to take snapshots or for continuous monitoring of jet engine operating parameters during flight. The benefits of this would be that aircraft performance and propulsion engineers could gain deeper insights into engine performance by these means, as distinct from those gained from data analyses by using conventional algorithmic and statistical methods.

The outcome from these discussions was rather disappointing; British Airways engineers, although interested in the approach, felt that the project was too complex to be undertaken in-house and that it was a programme that was better undertaken by a manufacturer such as Rolls-Royce. It was realised that funding would be very difficult to justify in the then current financial climate, even for a feasibility study. Rob, keen to continue exploring the approach, switched his focus to ground-based gas turbine applications. This resulted in the very successful TIGER programme and eventual industrial success of this system.

3. The Application of Neural Networks

Later events within British Airways led to a softening of attitude towards new technologies such as artificial intelligence due to the realisation that other airlines had started to carry out their own application developments, and varying degrees of success were being reported. Additionally, manufacturers were beginning to take an interest and were beginning to consider the use of artificial intelligence tools and applications for product development and support. Consequently, several studies were initiated in-house and jointly with Rolls-Royce by Dr Simon Cumming of British Airways AI Unit. Simon ran a very comprehensive programme of investigative work funded by the British Airways propulsion department, which investigated new means of providing better prediction of incipient faults in gas turbines [3].

Much of the work carried out focussed on the application of neural network techniques (supervised and unsupervised learning) to handle data gathered from flight data recording (FDR) and condition monitoring systems. These were COMPASS in the case of Rolls-Royce, and GEM in the case of General Electric engines. The COMPASS system utilised data collected by FDR in-flight, and downloaded from a data-recording cassette after landing. The GEM system relayed snapshot data, which was transmitted to a ground station during flight, then by telex to a ground-based monitoring system.

Both systems monitored performance parameter values and deviations from these calculated using a thermodynamic model. However, these systems could not predict faults, but relied on the experience of the engineer concerned to interpret the data. This could be extremely time consuming and tedious given the amount that had to be evaluated. It was very difficult to spot and interpret trends in a consistent manner, a fact that had operational as well as safety implications. This provided the motivating factor in investigating and developing intelligent decision support methods and tools. It was envisioned that these would assist in interpretation, remember previous faults and focus on interesting parts of the data having analysed all of the collected data.

The condition monitoring programme initiated by Simon Cumming shared many similarities with Rob Milne's earlier proposal, but varied in the technological basis of the approaches, Rob's being model-based, and those of Simon utilising neural computing techniques. The basic elements of Simon's work utilised modelling and prediction, classification and context-based approaches. Some of this work was jointly investigated and supported by Dr Peter Cowley of Rolls-Royce Applied Sciences

Laboratories, Dr David Willshaw of the Centre for Neural Systems, University of Edinburgh and Dr John Pilkington of Scientific Computers.

Rob Milne never directly participated in any aspect of Simon's work but each maintained a keen interest in progress being made within their respective areas of interest and expertise, and Rob always regretted that we never had the go-ahead for his earlier proposals.

4. Knowledge Management

By the mid 1990's, British Airways had begun to re-develop its business strategy to meet the requirement of becoming a more cost-effective, revenue generating organisation. In doing so there was a gradual realisation that its core assets were the people that it employed – yet it had to reduce its workforce considerably. To counter the effects of such organisational shifts the British Airways AI Unit put forward a number of proposals to assist in the re-engineering of the business, which focussed on data mining and knowledge management [4] [5].

Implicit in these initiatives was the idea that artificial intelligence specialists could be more effectively employed within the business. Rather than concentrating on narrow-based artificial intelligence applications, they should be deployed to help expand the then existing knowledge sharing infrastructure based on electronic mail and groupware. This shift of focus led to an increasing interest in data mining and case-based reasoning.

5. Case-Based Reasoning

Applications utilising case-based reasoning in British Airways had been investigated in the early 1990's and a number of technology demonstrators had been funded [6] [7]. These had been built using the ReMind software development tool marketed by Cognitive Systems. Most of the applications, including one for aircraft systems diagnostics called CaseLine, demonstrated the usefulness of the approach, but the technology was often brittle under real-world operational conditions. This unfortunately led a number of business sponsors to lose interest and withdraw funding. I believe that Rob spent some time negotiating a VAR agreement with Cognitive Systems, but for a number of reasons – not least the demise of Cognitive Systems, decided to reduce his interest in case-based solutions to engineering diagnostics, and concentrate on model-based approaches. However, Rob and I continued to discuss possibilities of case-based help-desk systems in certain areas such as avionics fault tracking.

To this end we spent some time talking to Rockwell Collins about case-based reasoning and customer support, as well as briefing them of our experiences with ATE support and the Caddis diagnostics tool. This eventually led to our visiting the main Rockwell Collins factory based in Cedar Rapids USA. Rob always loved flying, and being a consultant for British Airways, he was placed in Business Class. I was travelling in First Class and was anxious that Rob be seated with me for the journey so we could discuss our strategy for the meeting with Rockwell Collins staff. I enquired of the Cabin Services Director if an upgrade was possible, and he said he would personally organise it. Rob was eventually located and ushered into First Class.

I shall always remember the look on Rob's face; it was an unashamed picture of glee, rather like that of a small child viewing a Christmas tree adorned with lights and gifts. (This was Rob's first experience of long distance first class travel). Nevertheless, he quickly settled in like the professional that he was and proceeded to spend the whole flight playing with the adjustable seats, cabin entertainment system, drinking champagne and revelling in the excellent food. We never did get to discuss business; Rob was far too occupied with enjoying himself!

Having enjoyed a productive business meeting, our return to London was almost a mirror image of the outbound journey, we were greeted at the cabin door by the very same Cabin Services Director, who immediately recognised Rob and promptly escorted him to First Class and gave him a seat. Rob was extremely impressed, and thereafter regarded British Airways as the best airline in the world. Rob enjoyed many years of flying with the airline, and invariably I managed to get him his upgrade. The smile and contentment on his face always made the effort worthwhile.

Despite the withering of interest in case-based solutions in most other parts of the airline, the engineering division continued to maintain interest in the approach, particularly as a tool to support diagnostics. Experience gained with prototyping an earlier system (CaseLine) led to interest from the Concorde technical support group in extending this work to develop a case-based demonstrator for fault investigation on the Olympus powerplant. Funding was obtained and a system called Ariadne was constructed using the ReMind case-base development tool and Visual Basic [8]. This was a successful system, which was encouraging for future developments.

Other case-based developments in British Airways Engineering during this period were being supported by the Airbus A320 CFM-56 propulsion engineers. This work was being managed by SNECMA, a premier French aero engine manufacturer, and *Acknosoft* the Paris-based case-based reasoning software company, which had developed a system called Cassiopee, which formed part of the broad-based INRECA

Programme funded by the European Commission. My only connection with this project was as a past member of the INRECA technical review team. I believe that this programme continued for several years, and that the developed system now forms part of the product support package for the CFM-56 engine.

The AI Unit continued independently with its own programme of work for support to Concorde and continued to monitor the case-based system development tools market with a view to re-engineering and extending the existing Ariadne system. In 1997, a new case-based reasoning tool produced by Atlantis Aerospace of Toronto Canada called SpotLight (now handled by CaseBank Technologies) became available, and this was subsequently adopted as the most suitable tool for future developments. An alliance with a respected aerospace company, already involved in the deployment of case-based maintenance support systems was seen within British Airways as a desirable and secure way forward in gaining future funding approval.

Spotlight was eventually used to re-engineer the Ariadne system to provide an operational diagnostics system, supporting Concorde route operations across the Atlantic. Again, Rob took a keen interest in this development and we frequently discussed the technical details and potential of such systems for equipment diagnostics in a range of other application areas.

Rob always loved Concorde, and on several occasions, I was able to take him on visits to the Concorde maintenance hanger. He considered Concorde to be a symbol of excellence, a perfect combination of both form and function. Concorde was never really the outstanding commercial success that was originally dreamed of, except in its final years. However, as Rob would have agreed, we occasionally need to prove that we can achieve something outside of the norm, to do something exceptional and damn the accountants!

Rob never managed to fly on Concorde, but I managed to have British Airways to fund a model of the aircraft, suitably mounted, and awarded as a prize in recognition of his work on TIGER. This award took pride of place alongside his Queen's Award for Technology. It seems poignant that Concorde is now retired and that Rob is no longer with us either!

6. Conclusions

I left British Airways in 2002 and only saw Rob Milne occasionally afterwards, although we often communicated by telephone and email. I believe that British Airways continued to play host to the British Computer Society's Specialist Group on Artificial Intelligence, of which Rob was the Treasurer. Their conference committee met annually at the British Airways HQ at Waterside, which was managed by Dr Simon Cumming who continued professional links with Rob in my absence.

My lasting impression of Rob is one of a wise and generous person who gave inspiration and encouragement to everyone he encountered. I believe that he helped steer many important decisions on the direction and future development of AI developments within British Airways Engineering. Of equal importance was the assistance he gave to the continuing dialogue we had with major equipment suppliers such as Boeing, Rolls-Royce and Rockwell Collins, as well as the Industrial community in general.

His support enabled many important decisions regarding research directions and equipment purchases by others. He was very generous in providing support to various groups within the airline and aircraft manufacturing industry. One example being his support to Allied Signal and CaseBank Technologies in their various business decisions on the merits of model-based vs. case-based technology solutions for product maintenance support.

Rob was able to give an informed objective evaluation, sharing his experiences with TIGER, freely supplying documents and performance metrics. The fact that Rob was willing to put aside his own commercial interests and to give much of his own valuable time and expertise spoke volumes of Rob's professionalism and kindness. I believe that this was a recurring theme in many other areas of Rob's social and professional life.

I always felt privileged to be both friend and colleague over the many years we worked together. Rob always seemed to move effortlessly between the many goals that he set, the relationships, and various interests that he enjoyed. Much of his enthusiasm infected those around him in equal measure and provided much needed inspiration to face difficult problems. He will be greatly missed!

References

[1] B. Trudgeon, Caddis Technical Documentation, Internal British Airways documents (1993).
[2] R.V. Magaldi, Knowledge Based Systems and Avionic Equipment Failure Diagnostics, Proceedings Royal Aeronautical Society London, Artificial Intelligence Techniques for Improving Aircraft Maintenance Efficiency, Hamilton Place London (1991). ISBN 0-903409-84-4.
[3] S. N. Cumming, D.J. Barnes, R.V. Magaldi, S. Mott, Use of Neural Network Methods for Aircraft Engine Condition Monitoring, paper submitted to: IEEE Aerospace Conference, Neural Networks for Aerospace Applications, Aspen Colorado, USA (1997).
[4] J. Barr, R.V. Magaldi, Corporate Knowledge Management for the Millennium, EWCBR-96, Third European Workshop on Case-Based Reasoning, Lausanne Switzerland (1996). Published in Advances in Case-Based Reasoning, Eds I. Smith & B. Faultings, Springer Verlag (1996). ISBN 3-540-61955-0
[5] R.V. Magaldi, A Case-Based Approach to the Management and Deployment of Knowledge Assets, American Association of Artificial Intelligence, Workshop: Exploring Synergies of Knowledge Management and Case-Based Reasoning (1999), AAAI Press.
[6] R.V. Magaldi, CBR for Troubleshooting Aircraft On the Flightline, Institution of Electrical Engineers, Computing and Control Division Colloquium on Case-Based Reasoning: Prospects for Applications, Professional Group C4 (Artificial Intelligence), IEE Savoy Place London (1994), Digest No: 1994/057.
[7] R.V. Magaldi, Maintaining Aeroplanes in Time-Constrained Operational Situations using Case-Based Reasoning, EWCBR-94, Second European Workshop on Case-Based Reasoning, Chantilly France (1994).
[8] R.V. Magaldi, A Proposal for the use of Case-Based Diagnostics Support to Minimise Technical Delays on Concorde Route Operations, Internal British Airways document. (2001).

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Rob and the BCS SGAI

ANN MACINTOSH
International Teledemocracy Centre
Napier University, 10 Colinton Road, Edinburgh, EH01 5DT, Scotland

Abstract. This essay describes the impressive organisational and communication skills and tenacity displayed by Rob during his voluntary work the British Computer Society's Specialist Group on Artificial Intelligence - SGAI. The chapter centres around Rob's work and enthusiasm for the Group and its associated annual conference. It explores Rob's interest in show-casing applications of AI techniques to support industry and finance and, I hope, helps to explain Rob's passion to go to Cambridge every winter (or rather almost every winter, there was one when he chose California instead and none us could quite understand why) for the past *n* years – the exact number now lost in the depths of time.

Keywords. BCS SGAI AI Conference

1. Introduction

Rob and I became involved in the workings of the BCS SGAI, or rather the SGES as it was called then, roughly about the same time, some fifteen years ago and that is when I first met Rob. It is interesting to reflect how our contributions to the Group switched backwards and forwards over those 15 years. First he would be Chair of the Conference (1991 and 1992) then I would be Chair (1993 and 1994). First I would be Treasurer of the Group, then he would be treasurer of the Group. In fact you could say it was like that to the very end – first he was Chair of the Local Organising Committee of IJCAI and then, sadly in June 2005, I took his place as Chairman.

So, in this Chapter, I will relate some of the history of his involvement in the Specialist Group. However, you cannot work with someone for over 15 years without becoming a friend as well as a colleague, so the final part of the chapter is devoted to Rob the friend, and look at the times when my son, Robin, and his son, Alex, ran in cross country races. Rob provided encouragement and advice to all those who ran in these toughly competitive races on those bitterly cold and snowy days in a Scottish winter.

2. The BCS SGAI

In 1980 Donald Michie became the founding member of the British Computer Society's Specialist Group on Expert Systems, the SGES. It is perhaps worth reflecting on the fact that both Donald and Rob had been at the University of Edinburgh, the former as the prestigious head of an international research group, the other starting out as a PhD student in AI. In the end both becoming world known for their respective work in this academic discipline, yet both retaining very strong links with the Specialist Group.

The BCS acts as an umbrella for groups working in specific technical subjects concerned with Information Systems. The BCS Specialist Group on Expert Systems provides support and advice to researchers, developers and users of AI technology of all kinds. Among the many activities the Group is responsible for is an annual series of international conferences on AI and a regular journal "Expert Update".

The SGES, was in the 1980s a very popular grouping within the BCS and Rob and I were asked to be members of its Committee at about the same time. Rob because of his well-respected application-oriented work in model-based reasoning and myself because of my then work with AIAI at Edinburgh. Committee meetings were held in London on a quarterly basis and Rob and I would regularly fly down and back in the day to contribute to the Group. In the early 1990s we began to see the membership figures fall and at the same time other specialist grouping appeared, devoted to areas such as neural networks and genetic algorithms. We often argued that our Group was not just "Expert Systems" but encompassed all AI subject areas and many times we argued together that the Group needed to change its name so as to appear more relevant to its potential membership base. Rob was a strong advocate for a change of name and tenaciously argued his point. After many light-hearted debates and discussions on possible new names, in 2000 we eventually changed the name of the Group to the SGAI.

I was Treasurer of the Group for 3 years, looking after its various bank accounts, administering membership subscriptions and preparing the annual accounts for the BCS. It was a time-communing role and one where, naturally, exact attention to detail was needed. I record this fact, because after 3 years I felt I had to resign the position as it was taking up too much of my time (although I did move on to become Vice Chair of the Group). Rob took over as Treasurer and remained in that role for the next 6 years. Through all the ups and downs of his business life he continued to devote time to his role as treasurer. During this time, and probably because of his entrepreneurial skills, he realised that the group was "missing out" on thousands of pounds of VAT refunds. For the next 3 years he argued and debated the main points of the case with both the VAT-man and the BCS. No way was Rob going to give in, and eventually recovered a considerable sum of the money owed to the Group. Needless to say we have had no further trouble with the VAT-man!

3. The Annual AI Conference

The SGAI annual international conference is where Rob's organisational and communication skills really excelled themselves. The Specialist Group has held a regular conference since 1981, although neither Rob nor myself played any part in these very early years. The Conference, held in the UK, provides a venue to debate the latest AI techniques. The refereed papers published in the conference proceeding are on research in a range of sub-disciplines such as natural language processing, knowledge engineering and evolutionary computing. The authors come from all over the world.

Given Rob's business background, it is perhaps not surprising that he argued the case for a special section of the conference to be devoted to showcasing the latest innovations in the application of AI to manufacturing industry, commerce, medicine and government. In 1990 the first proceedings of the papers from the Applications

Stream of the conference were published under the title “Business Benefits of Expert Systems”.

This 1990 conference provided the opportunity for both Rob and I to deliver presentations on our respective subject areas. My paper considered the successful training paths that had been developed at AIAI to train knowledge engineers. It focused on the continuing education programme of the University of Edinburgh delivered to companies such as ICL, Standard Life and Hitachi through a 10 week study programme in Knowledge Based Systems. Rob’s considered the business benefits of applying AI techniques [1].

Back in the 1990s he was already exploring the major issues which would have an important bearing on whether a company could gain the financial and business benefits from expert systems. He clearly recognised that the issue of business benefits was the single most critical restriction on expert system applications. I present here a very telling quote from that 1990 paper:

> *In summary, there are very considerable benefits from using expert systems, these benefit are diverse. In general, however, it is not possible to put a cost quantification on these benefits. At the same time, the number one problem stopping the development of expert systems application is a lack of a cost justification. You can see that it is going to be a long hard battle”* [1]

It is important to realise that not only did he recognise this fact, he also practised what he preached and ensured that his own company carefully considered the business benefits of each new application they developed. This fact is perhaps self-evident given Rob’s unique academic and entrepreneurial success in applying AI techniques, although, as Rob says above, the real financial success is indeed always a long hard battle, and so it proved to be with Rob and Intelligent Applications Ltd.

In both 1991 and 1992 Rob was the Conference Chair [2] [3]. This was a role in which he excelled and also thoroughly enjoyed; especially chairing the conference dinner and ensuring that we all had sufficient drinks for the numerous toasts of the evening! The social side of these conferences with Rob could fill whole other chapter and I’m sure each one of the contributors to this book has their own favourite story to relate – but perhaps it’s enough to say that in Rob’s company there was never a dull minute, and now these happy days live on in our memories.

Not content with a special parallel stream within the conference dedicated to the innovative application of AI, Rob, argued for, and once again achieved, the backing for a financial prize for the best refereed application paper. This was introduced the year that Rob was the Programme Chair for the Application Stream in 1994 [4]. Although I supported Rob’s concept of a ‘prize’, I must admit I was worried that Rob would always win it. My concerns were not justified in 1993 when the first prize went to the authors Linda Young and Stanley Cauvain of the Four Milling and Baking Research Association – although I did suspect that Rob encouraged Linda to submit a paper on her interesting use of expert systems to support bread-making. Then came 1995, when it was my turn to take over from Rob and be the Programme Chair for the Applications Stream and with the support of my Programme Committee select the best refereed paper – of course the author was Rob! Being the Programme Chair I was very concerned that this looked liked an “inside job”, and to make matters even worse, Rob was again the Conference Chair. I freely admit I did everything in my power to suggest that perhaps Rob’s paper was not the best – but all to no avail, the independent committee was adamant that his paper was by far the best and therefore rightly deserved the "application prize". So all my early concerns were not groundless and I had just chosen the wrong year to be worried! Rob was really delighted to win and the presentation he gave of his paper at the conference was really exceptional and left none of the delegates in doubt that this was a worth winner. I recall noting in the preface of the proceedings:

> *1995 has been yet another "application" success for the conference with this volume containing over 20 papers chosen from some 60 submitted to the Applications Stream.* [5]

Rob's paper was on, his now very famous and successful, Tiger system. The paper was entitled "TIGER: Knowledge Based Gas Turbine Condition Monitoring" and was co-authored by Charlie Nicol from Intelligent Applications and research partners from LAAS-CNRS in France and the Polytechnic University of Catalonia in Spain. Over the succeeding years Rob and Tiger were going to become synonymous, few of us could forget the pictures and screen savers Rob accumulated all showing the "tiger". Even in that early 1995 paper Rob was able to write:

> *The TIGER condition monitoring system has been in continuous use at the Exxon Chemical Fife Ethylene Plant for over two years. It has been instrumental in identifying the underlying problems for a number of situations, one of which resulted in considerable cost benefit to Exxon.* [6]

It is interesting to note that Rob references an earlier paper he wrote on "Online Diagnostic Expert System for Gas Turbines" published in the 1992 proceedings of the Expert Systems Conference.

The TIGER system was part funded by the Commission of the European Communities and I recall at the time it was a flagship project for the Commission, one they could really say had benefited European industry. In 1997, when Rob and I shared the Chairing the of the Application stream, myself as Chair, Rob as Deputy Chair and it was in that role that Rob invited Patrick Corsi from the European Commission, to be our Keynote Application speaker. Incidentally, Patrick was the Project Officer responsible for the TIGER project. In November 2005, while in Brussels, I happened to meet Patrick again. We reminisced about the years of ESPRIT funded projects and he still remembered the success TIGER achieved through the hard work of Rob.

I remained Programme Chair from 1995 to 1997 and then it was time for the role to swing back to Rob, so in 1998 he was again took over the Chair. At the same time he was also devoting considerable amounts of his energy to the European Conference – ECAI - and I remember we both went to Budapest for the 12th European Conference in 1997. Rob was on the Program Committee and we were both representing the SGAI at ECCAI (European Co-ordinating Committee on Artificial Intelligence) annual meeting. Of course Rob, naturally, went on to become the President of ECCAI.

His last major achievement for the SGAI and for the whole UK AI community was to ensure that *the* major international AI conference – IJCAI - finally returned to the UK for the first time in over 30 years. Rob undoubtedly would have been proud of the success of the 2005 conference in Edinburgh.

Rob was definitely missed at the 2005 SGAI annual conference in Cambridge, but in order to ensure that future generations of AI researchers and entrepreneurs continue to follow in his footsteps, the BCS Executive Committee has donated an Annual Prize to the conference for the best applications paper. Also, thanks to the sponsorship of Alan Montgomery, the Conference will also present "The Rob Milne Memorial Award" (see Figure 1) to the best-refereed paper.

Figure 1: The Rob Milne Memorial Award

The author's names will be etched on the glass trophy and the award itself will be on permanent display on the BCS Headquarters in London. There is also a smaller version of the award which the winners are able to keep. Rob would have liked these glass awards shaped as they are as mountains, the larger like a Munro and the smaller just a Corbett!

4. A Personal Note

I could not complete this essay without mentioning the other sides of Rob I got to know over those fifteen years.

We both have sons of a similar age, Robin, my son is a year younger than Alex, Rob's son. They both had an interest in cross-country running and here Rob and my life crossed yet again. Alex as was a member of George Heriot's School Cross County team and Robin as a member of the City of Edinburgh club team, often ran in the same races. I recall the many times I would shout out to encourage Alex running round a bend in what seemed like a blizzard while Rob was enthusiastically cheering Robin on. Although unlike me, a stationary observer, Rob was usually running round the course with them!

Then the school cross-country came to an end and both sons went on to universities. I will always remember sitting next to Rob on a plane flying to an SGAI committee meeting, and Rob, the then highly successful businessman, turning to me and saying “I don’t want to spoil Alex so how much allowance do you think I should give him for University, what are you going to give Robin? ”. I only hope Alex was pleased with my answer!

Then Robin tool up rock climbing so Rob was my natural source of intelligence on the subject although I have to admit I didn’t fully understand all the advice he gave out, such as what cam was best for what climbs and types of rocks.

Rob has been a close colleague over the past decade and half, and a true friend. His genial and easy manner often belied his toughness and determination both as a businessman and as a mountaineer.

References

[1] Milne, R; “Where is the Business Benefit?”; Invited paper presented at British Computer Society - Expert Systems '90 Conference, London, September 1990.

[2] Graham, I. M., and Milne, R. (eds.). Research and Development in Expert Systems XIII, Proceedings of Expert Systems 91 the 11th Annual Conference of the BCS Specialist Group on Expert Systems, London, September, 1991.

[3] Bramer, M., and Milne, R. (eds.). Research and Development in Expert Systems XIV, Proceedings of Expert Systems 92 the 12th Annual Conference of the BCS Specialist Group on Expert Systems, 1992.

[4] Milne, R, and Macintosh A. (eds.). Applications and Innovations in Expert Systems II, Proceedings of Expert Systems 94 the 14th Annual Conference of the BCS Specialist Group on Expert Systems, Cambridge, December, 1994.

[5] Macintosh A. and Cooper C., (eds.). Applications and Innovations in Expert Systems III, Proceedings of Expert Systems 95 the 15th Annual Conference of the BCS Specialist Group on Expert Systems, Cambridge, December, 1995.

[6] Milne, R., Nicol, C., Trave-Massuyes, L., and Quevedo, J.; “TIGER: Knowledge-Based Gas Turbine Condition Monitoring” in Macintosh A. and Cooper C., (eds.). Applications and Innovations in Expert Systems III, Proceedings of Expert Systems 95 the 15th Annual Conference of the BCS Specialist Group on Expert Systems, Cambridge, December, 1995.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Rob Milne's contribution to the Young Software Engineer of the Year Awards

Ian RITCHIE
Chairman of Judges, Young Software of the Year Awards

Abstract: Notes on the contribution made by Rob Milne to the Young Software Engineer of the Year Awards, now renamed 'The Rob Milne Memorial Award'.

Introduction

Rob Milne's enthusiasm was infectious. He brought his natural American sense of optimism and zest for life to everything that he did, and his dynamic contribution to any project was always certain to raise the game for all involved.

At first, it may seem a little puzzling that this young scientist who hailed from Montana and was putting in his time, post-PhD, with the US Army as Chief Scientist of the AI Center at the Pentagon, should chose to return to Scotland to start his business in the mid-80s. Part of the reason was that he had completed his PhD at Edinburgh University and very much enjoyed living in the city. He had also identified a potential market to apply his technology for the condition monitoring of turbines, widely used in the oil exploration industry that at that time was blossoming off the coast of Scotland. But probably far more relevant to Rob was the opportunity most weekends to participate along with an enthusiastic community of other mountaineers in ice climbing adventures among Scotland's many mountains.

I remember asking Rob why he had chosen to move from Washington DC back to Scotland. His reply was that Washington was far too hot and humid and that he found the summers unbearable. He much preferred the cold and ice of Scottish winters.

1. Judging the Young Software Engineer of the Year Award

Apart from meeting up as fellow technology entrepreneurs from time to time, my main experience of Rob was as a fellow judge in the annual Young Software Engineer of the Year awards. Every year since the awards were instituted in 1990, the judges gathered at the end of June to assess the project reports submitted by final year computer science students from Scotland's Universities. Rob was one of the judges in that first year and continued as a judge for every year since.

Rob was a relatively unusual individual who was both a successful businessman and also one who stayed academically active and very much in touch with the research areas of his discipline. This made him an ideal judge and his expertise in identifying excellence in software engineering was always impeccable.

The judging sessions were usually held in the library of the old mansion which stands in the middle of the grounds of Edinburgh Zoo, and were punctuated by the cries of the nearby wildlife collection. Reading all of these theses might seem a daunting task, but it was always made into a pleasure partly because of the enthusiasm that Rob showed for the job. He showed up one year with a bunch of highlighter pens and, since then, we have always made a special note of the funny, fascinating or puzzling things that people put into their thesis. It was quite typical of Rob that he saw the fun side of everything that he did.

After the judging session, the judges would have lunch together and this was always an opportunity to catch up with Rob's latest climbing adventure. Rob seemed somehow to organise his visits to prospective customers, or attendances at technical conferences, so that they would coincide with the chance to climb yet another interesting mountain.

In recognition of Rob and his huge contribution to the Scottish Young Software Engineer of the Year, it was decided to rename the award 'The Rob Milne Memorial Award'.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Rob Milne the Mountaineer

Rab Anderson
Scottish Mountaineering Club

I'm not really quite sure when I met Rob for the first time, or indeed how. There is no doubt he would have remembered though. What I do recall at the time was hearing about an American student at Edinburgh University who was gaining quite a reputation as a bold ice-climber, particularly on Ben Nevis where he was efficiently ticking his way through a number of the established climbs, some of them solo, without a rope.

There was no doubt that Rob took to the Scottish Winter climbing scene like a duck to water. But then again, if you knew him, then he would, wouldn't he? Although it did take me a few years of climbing with him before I fully understood why. Somehow we ended up climbing together, perhaps more by design on Rob's part than by fault. Anyway, in January 1980 we made the first ascent of a hard and bold ice climb on Creag an Dubh Loch in the Cairngorms, following the line of a summer climb called the Pink Elephant. The name White Elephant seems to have stuck for this, its winter counterpart, a great climb which has perhaps only been repeated once or twice to date. Rob and I had climbed together pretty much every year since then and as well as establishing a number of significant hard new routes together, we also repeated many climbs and shared many good days out on the hill. In March of 2005 shortly before heading off to Everest, Rob proudly announced the fact that we had climbed one hundred winter routes together. I didn't know he had been counting but knowing him he no doubt he had some efficient way of storing the information and his computer had announced it to him!

They say it takes two to tango and climbing is no different really. Unless of course you are into soloing without a rope, which is something that some of us who climb have done and continue to do on an occasional basis. The thing about soloing is that once you are more than 10 metres above the ground and you happen to fall off, then there is a good chance that you will die! Once you have accepted this, then it doesn't really make any difference whether you are 20 metres above the ground or 1,000 metres. The human brain is quite an impressive thing really. Not only does it compute this fact and convince you to do it but once you are doing it, it keeps you in touch with what you are doing. Now there are a number of reasons why people solo; the impetuousness of youth and the feeling of invulnerability; one's state of mind aided perhaps by the break up of a relationship; mind over matter; escapism and the complete feeling of freedom it provides. There is also the desire to go climbing and the difficulty in finding a suitably motivated partner. The one thing about soloing though, is that few do it on a regular basis, which is hardly surprising given that at some point the consequences of failure are realised. Anyway, I digress, so back to Rob's soloing. People do things and sometimes you don't analyse why, until an event. Such is the case here and following Rob's death I found myself asking why did he solo back then? It was certainly something I did not associate with his measured approach. Well, at the time he was young, he was bold, he was into his climbing and he was in a new country that could be said to be the birthplace of ice-climbing. Not rocket science then, for anyone who knew

Rob would figure out that, he was evidently having trouble finding like-minded partners and his enthusiasm got the better of him, so he did it himself.

Finding climbing partners can be difficult and finding regular partners even more so but if you want to climb, then find them at some point you must. Rob duly did and soon built up a good list of contacts. If you take the one hundred winter routes that Rob had counted we had climbed together and you divide this by the twenty or so years that we had been climbing, then you come up with a relatively unimpressive five routes a season. That is until you are aware we did not just climb with each other all of the time, for we both did routes with others. Neither does this account for the days when we tried to climb but the fickle Scottish weather got the better of us, so we either returned empty handed, or went hillwalking instead. What is impressive though, is the length of time we climbed together, for the difficulties encountered in winter climbing are not conducive to long lasting partnerships. It is perhaps also impressive that we remained enthusiastic and motivated through all the external influences brought about by the responsibilities of work and family life. One of the main contributing factors in all of this of course was the fact that we got on together and as well as rewarding, our days on the hill were enjoyable.

Due to the harshness of the Scottish winter environment and the nature of the Scottish weather I was a bit more selective about when I went out. Not so Rob, and as well as his keenness, I was always amazed at how he managed to turn up climbing partners at short notice. Rob was organised, so he used to let me know all his business commitments well in advance. I'm sure he knew that I never wrote them down, I'm not that organised. He would then contact me early in the week before to set the weekend up. Again, I'm not that organised and although I had lots of objectives up my sleeve I also knew that the weather could not be guaranteed, so I usually left him hanging until closer to the weekend. There were many occasions when I decided, even as late as the Friday evening, that I didn't think it was worth going out. Rob never got pissed off with this, he just accepted it and got on with things, which was one of his traits. He invariably turned up partners and went out. The weather was always okay, or so he said!

In his early years in Scotland, Rob joined The Jacobites', an Edinburgh climbing club, no doubt in his efforts to find partners. He wrote two articles for this Club's annual journal, both repeated in 2005 following his death. The first was "On Top of America" and described his ascent of Mount McKinley in Alaska with Brian Sprunt, an Aberdonian climber who was subsequently killed on the North Face of The Matterhorn. The second was "Up and Down the Ben" and described his ascent of the hard Ben Nevis ice climb Galactic Hitchhiker with one companion and his rapid descent of an even harder climb called Pointless with another. On this enforced descent, his companion for the day took a big fall and pulled the pair of them off into the snow below, down which they slid for some distance. Both were remarkably unscathed. To quote Rob "I believe I hit my head so was not harmed". Following this incident, and by way of illustrating Rob's determined nature, he went back in the Spring after the snows had melted to find the camera that he had lost during the fall. He painstakingly traversed back and forwards up the slopes beneath the route and actually found the camera, together with pieces of equipment that others had dropped. However, like the name of the climb he had been on it was a pointless exercise since both the camera and the film had been ruined.

Rob was kind of the stabilising influence in our climbing partnership, laid back is perhaps a better word. I was more focused about specific objectives and quite uptight with it all too. I don't think Rob would have minded if I said that I was the stronger climber, particularly since I trained more on indoor walls and spent all season rock climbing. Rob was mountain fit though, and by that I mean his body was more suited to long winter approaches and ploughing through deep snow, particularly since he did so much hillwalking all the year round. To some extent Rob just wanted to go and climb and have a good time. I generally played my cards close to my chest and kept my objectives to myself. Its not that I didn't trust Rob, I just knew of his enthusiasm to share what he had been up to and as a result it was easier to keep quiet. The backstop is that it's also easier to fail on something when nobody knows what you wanted to do in the first place! We would often end up going climbing somewhere "For a Look", without me really telling Rob what the objective was. There were times when we would arrive at a cliff and I could see him looking around at the unclimbed bits to see what he might find himself on. Needless to say this meant there were a few occasions when he found himself standing at the foot of some hard looking objective with little time to psych up for it, however, he took it all in his stride, as usual.

Interestingly, Rob accepted this approach and over the years it certainly worked out because we were a strong partnership and we did a lot of good routes. This secrecy might seem strange to some but when looking for new climbs it doesn't pay to let people know what you are up to. It also doesn't pay to fail on routes and have that information leak out. An example of this is the year I had been in France on an early season rock climbing trip and had pretty much written the rest of the Scottish winter season off. I returned to have the usual set-up call from an enthusiastic Rob, when he announced that the winter was not over. I wasn't that keen having felt warm rock beneath my fingertips but Rob most definitely was, as ever. He sensed my lack of motivation, so he mentioned that he had been out with two of the Ayrshire hot-shots and that conditions were good. They had in fact failed on an attempt to climb Raven's Edge in Glen Coe for the first time. Now, I'm still not sure to this day whether Rob dangled this little piece of information in front of me on purpose, in the hope that I might be tempted to go for this route with him. It worked though, because unbeknown to Rob this was a route very high on my agenda and one I had been waiting for years to come into suitable winter condition. That weekend one of the best winter climbs in Scotland was duly dispatched by the pair of us. This route was also another occasion where Rob's mountain legs were put to good use. The descent slopes from the top of the route are avalanche prone and in the gathering darkness we could not see them properly, so we decided to plough our way to the top of the mountain. Rob stoically took on the lion's share of breaking trail, all the way to the summit and back down. This was not for the only time either.

There were many other moments of support that he provided. Not my big sharp stick approach but his subtle, more laid back approach. One example that springs to mind is the day we had wandered across to the Shelter Stone in the Cairngorms to make the second ascent of Postern. It was a repeat ascent so Rob knew the objective that day. However, we were later than we should have been and the first ascent party had climbed the route over two days, bivouacking just below the top. We were after the route in the day and had no sleeping gear. In winter climbing you deal with what's in your face and it wasn't until we were standing beneath the route, geared up and ready to go, that I realised it was a bit late in the day. In reality we should have turned back and when I asked Rob whether we should go for it he simply drawled "Sure" and said we should just take it easy and see how we got on with it. Well we did it, but only just, topping out in the gathering darkness into a full-blown Cairngorm blizzard. I had an epic at the top but that was my fault and that's another story, but it certainly gave Rob something to talk about, for it's not every day that your leader has to abandon one of their two ice-axes half way up a hard pitch with the pick facing out!

Another example of Rob's nature was around Christmas one year when he accompanied my wife, Chris, and I into the Northern Corries of Cairn Gorm, or perhaps more correctly I should say, we accompanied him. We had been out the night before and had overdone it a touch, so the first thing was that Rob had to drive. To be honest if the weather forecast had not been so good, I think we would have pulled out. Anyway, we went and the second thing was that the drive, followed by an early morning approach walk in sub-zero temperatures, had our heads reeling and stomachs churning. So much so that when we got to the cliff I had to say to Rob that I wasn't up for anything hard and that whatever we did he would probably have to lead. Anyway, we found a suitable established climb and an undeterred Rob set off. Luckily the second pitch was easier and I managed to get it together enough to lead this. I was pleased at keeping the expense of the night before down but after the exertions of the first pitch Chris was not so fortunate. Now, if I had been in Rob's shoes I would have been pretty pissed off at being lumbered with two people who were obviously not fit to be on the hill and were more than likely going to spoil my day out, especially with one of them emptying the content of their stomach all over the ledge I was standing on. Not so Rob, for in his usual philosophical way he simply drawled "Better off out than in". Needless to say, the eventual result was a good day and as a bonus I spotted an unclimbed crackline, which was duly stored away in my mind and then climbed by Rob and I a few years later.

When you establish new routes you can name them. In the early days Rob named a route in the Cuillin of Skye after his soon to be wife, Val. We climbed a new route on the Lost Valley Buttress in Glen Coe and Rob came up with the name Tiger Feet, since this tied in nicely with the neighbouring climb called Sabre Tooth and with his interest in tigers and the name of one of his business products. As already stated, I am less subtle and on the neighbouring Lost Valley Minor Buttress we did a climb beside Grannies Groove and I called it Old Fart's Corner. Rob just shrugged his shoulders at this.

Rob liked technology, which is hardly surprising given his background. From early mobile phones and altimeters, to handheld GPS units and the latest NASA map downloads and so on. We all accept mobile phones nowadays but I remember when they were a novelty and an occasion when Rob proudly pulled this amazing little phone out of his pocket. It was the kind with an aerial that you had to pull out with your teeth. Well it looked cool doing it that way and when you were gloved-up and sitting on a ledge in the middle of a snowed-up cliff in the Cairngorms you didn't have an option. Anyway, like kids, we sat there on a ledge half way up a climb phoning people up and giving them the "guess where we are" and "you should see how good the climbing conditions are" routines. I also recall the Anemometer he got for Christmas one year, which he proudly produced from his pocket when we were out hillwalking. It had a scale that went to 70mph and when he held it up, apart from nearly having the thing torn out of his hand, the readout went straight to 70mph and stuck there. I dryly shouted "Yeah Rob it's windy!". To which he philosophically replied that perhaps he ought to get one with a scale that went up to one hundred.

I can remember the heap of a car that Rob had in his early student days. I don't recall what make it was but the colour was rust. It was a temperamental machine and had to be push-started now and again, which is not a lot of fun when you have just come off the hill in winter and one leg does not want to stagger in front of the other. You also damn nearly froze to death in it because the heater was useless. It was perhaps a typical student car, but certainly not a winter climbers' car.

Rob went to the States and then he was back. The military gave him three years at Edinburgh University and he gave them three back and now he was here again and he wanted to climb, so we did, with a big route on the Triple Buttress of Coire Mhic Fhearchair in April 1986. A climb we called The West Buttress Direttisima, which as far as I am aware has only seen one other ascent. He eventually set up in business and he got the ultimate driving machine, a BMW 5 series with electric windows and a trip computer. Although it might seem strange now, both were novelties at the time. The cars that followed were big SAAB's with bum warmer seats. Luxury when you come off the hill frozen. I'm glad that Rob had decided to make amends for those early trips in his rusted ice box. The cars were perhaps one of the only outward signs that Rob was doing well. The other sign was that he was away on business a lot. However, this never seemed a disadvantage to Rob, because he always had the knack of squeezing in a bit of climbing wherever he went. In fact, there were years when he probably did more climbing abroad than I did, as was proved by the tan he always wore.

Rob was into what might be called the grander mountain environment and as a result we never really climbed together in Scotland on rock in the summer. He had been born in Montana but he moved to Colorado and that's when he started to climb, so he quickly progressed to the snowy mountains where he found he was most at home. In 1975 he made the first ascent of the North-East Ridge of Mount Vancouver in the Yukon and a few years later he visited the Kitchatna Spires in Alaska. In the same year that I first climbed with him, 1980, he climbed the first of the summits in his eventual quest to climb the highest points on each of the seven continents, The Seven Summits. This was Denali in Alaska, or as some know it, Mount McKinley. In the Alps he climbed the Eiger North Face. When he was in back in the States for those few years, he went on an all star trip to the Karakoram in the Himalaya, where they pulled-off the first ascent of the brilliant Lukpilla Brakk spire. This trip was also when Rob managed to acquire a full Gore-Tex shell suit from the US Military, which he proudly wore around the Scottish hills for many years. The only problem was that it was camouflaged, luckily not arctic, however, mottled green black and brown are not exactly the best for photos either!

In the summer, or when there was no snow, or the weather was poor he was always off hillwalking and was known for squeezing the most out of his days out, both in terms of his time and his legs. He did some mammoth link-ups in his quest to tick the lists of Scottish hills. He had climbed all of the hills in Scotland over 3000 feet (914m). These are known collectively as The Munros and there are 284 of them. He had also climbed all of the 511 associated Munro Tops, which many do not do. Rob was also close to completing The Corbetts, the collective name for the list of hills in Scotland between 2500 feet (762m) and 3000 feet (914m), of which there are 219. He had also discovered the Border Hills in the South of Scotland, the Donalds, of which there are 89 over 2000 feet (610m). This latter group of hills was close enough to work so that he could even bag one early in the morning with a late start at the office, such was his enthusiasm.

Rob was keen to complete The Corbetts and The Donalds in the same year that he completed the Seven Summits, of which Everest was the last.

One of the last times that I saw Rob, he was drunk. Not just on alcohol, although there had been a lot of that, but on success. It was the end of the 2005 Scottish International Winter Climbing Meet at Glenmore Lodge in the Cairngorms and they were having party. It had been a great week and a lot of climbing had been done. Rob was one of the host climbers. In fact, Rob had acted as a host on a number of the Mountaineering Council of Scotland and British Mountaineering Council, International Winter Climbing meets. I am sure that some of the International climbers must have found it a little strange to find that one of their hosts was actually a smooth shaved American, rather than a bearded and kilted Scotsman. It would not have taken them long to realise though, that Rob was no ordinary American, for he was a local with a great knowledge and understanding of the Scottish Hills and a temperament ideally suited to the task in hand.

Rob was a member of the Alpine Club, but more importantly of the Scottish Mountaineering Club, the SMC. I am pretty sure that as with the Alpine Club, Rob attended every SMC dinner when he was in this country, which is quite a feat for I have been to about four in twenty five years. Rob became involved in the Club's affairs, taking on the role of Convener of the Publications Sub Committee, as well as a trustee of the Scottish Mountaineering Trust. The SMC is the major publisher of climbing and hillwalking guidebooks covering Scotland and it never seemed out of place that an American should be involved in the affairs of club such as the SMC, which is steeped in history and tradition. It was a position that ideally combined his business talents with his climbing skills and in-depth knowledge of the Scottish Hills. Rob also became the Editor of the SMC's Hillwalkers' Guidebooks and was one of those behind the latest, successful book, the North-West Highlands, as well as a CD Rom to The Corbetts, the Second Edition of The Corbetts book and the publishing of e-books through the SMC's website.

Rob certainly played an important part in the Scottish climbing scene and he will be missed by many.

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Robert Milne – Mountain Writing & Scottish Mountaineering Publications Work

Ken Crocket Ph.D., *Webmaster*, Scottish Mountaineering Club [www.smc.org.uk], *Director*, Scottish Mountaineering Trust (Publications) Ltd. November 2005

Abstract. This article takes a slightly unusual approach to Rob Milne's writing as it is inextricably bound up with the subject matter – exploratory mountaineering. The latter is peculiarly intensive and demanding, especially in winter, and as a result has often induced exciting writing from its practitioners. For those not involved in this aspect of mountaineering therefore, a little background is often helpful. This cannot be an exhaustive list; Rob may have penned articles for local magazines or journals which are rarely indexed. Article excerpts are in italics.

Keywords. Scotland, mountaineering, winter climbing, publishing, writing

Beginnings

Rob was an extremely competent all-round mountaineer; in *stricto senso* he could move safely over and up the mountains while walking or engaged in technical rock or ice climbing. Big hills, small hills, all were fun to Rob. Like many enthusiasts he began expanding his interest into writing and production work on guidebooks.

Rob mainly wrote about his mountaineering exploits in the pages of the blue-stockinged Scottish Mountaineering Club Journal. This highly respected journal has been published since 1891, beginning with several issues per year and now an eagerly-awaited annual publication. In its pages are recorded new climbs in Scotland, as well as exploits of SMC members throughout the world and articles, with an understandable Scottish bias, from just about anyone who can write competently. Rob could write competently.

'*Epitome*' [1] is an account of the second ascent of a winter climb in Lochnagar in the company of Brian Sprunt. *'The Aberdonians were hosting both the Polish climbers and their Edinburgh rivals. Edinburgh were huddled in one dark corner whispering of new route possibilities while Aberdeen were in a different corner muttering about the same routes. In the middle, the Poles put us all to shame with their one arm pull-ups.'*

As with many other climbers, it was winter days on the Scottish mountains which brought out the words from Rob. His sense of humour was evident from the start. A sense of humour is a pre-requisite for the often grim winter days which have to be endured so as to experience the odd, good day.

'Deadman's Groove' [2] describes a first winter ascent on The Cobbler, that wonderfully gnarly hill looking down on a sleepy Arrochar at the head of Loch Long. In this article Rob was climbing with Edinburgh friend Rab Anderson. Rob pokes a bit of fun at himself, *'As usual, I had no idea of what the climb might be, where it was, or how hard it might be. It wasn't that it was a secret; I just didn't know anything about routes on most crags. (Mine is not to reason why…)'*

One gets the impression here that Rob was relaxing on his climbs, happy for someone else to do the preparatory work but always throwing himself (in a relaxed manner of course) into the physical effort. The then Editor of the SMCJ (KVC) commissioned a special cartoon for this article. It shows a climber standing on the frozen belay ledge, holding the ropes while up above, almost out of sight, the worried leader toils away. Rob was to enjoy many good days with his friend Rab.

'The Crack' [3] saw Rob back in the Cairngorms with Rab Anderson, climbing a hard mixed winter route. Mixed means that the climb throws all kinds of defences at the probing mountaineer – rock, ice, snow, frozen turf (if one is lucky). Rob is becoming more precise in his writing here, just as his climbing footwork had to become more precise. Even so, a wee fall is taken before the day is won. Rob is also beginning to show signs of a desire to help organise the grading system for climbs in Scotland, as harder standards are shown to be confined by an out-of-date grading system. It's the 1992 Ferrari meeting the 1926 Silverstone track syndrome.

Widening Horizons

'Climbing Is All About Having Fun' [4] In this tale, written jointly with Louise Travé-Massuyès, a petite Frenchwoman from Toulouse, there is the story of a big wall ascent in Yosemite Valley – the North-West face on Half Dome. To add to the fun, a night is spent on the face in a rain storm, the first rain in the Valley for three weeks. But Rob is laid back about it all – *'Mid morning, it started to rain harder. I shouted over: "This is a good sign, it will stop soon." In her French accent she graciously replied: "Bullsheeet, and I'm frozen." '* It was May, and their second night hanging on the wall was a frosty one. Despite the discomfort the climbing went well and they found themselves on the summit. Louise finished the article – *'The landscape was cratered with wind carved sandy bowls. Clouds drifted in and partly covered the summit. This gave the strange atmosphere of being somewhere else . . . exactly where we wanted to be – out of the world.*

We didn't have an epic, far from it. The climb had been fun, and we had had fun. And climbing is all about having fun.'

'First Winter Ascent of Raven's Edge' [5] In this article, one of the few Rob wrote for the glossies, he described multiple attempts on this route on the Buachaille in Glen Coe, climaxing with the final successful ascent with Rab Anderson. He also wrote about the nature of new routeing...*'The challenge for those wanting to seek new routes is not just finding a small vertical unexplored world, but finding something which someone else would want to climb afterwards. The pleasure comes from creating a classic, not just a new route...On the north face of the Buachaille, one of those lines*

was waiting - Raven's Edge.' The route is graded VII, 7, which indicates its level of difficulty, as at that time grades stopped at VIII.

'Mixed Company' [6] was written for the US market. It began *'Modern Scottish mixed climbing is like a good single-malt whiskey (sic): you take it without ice.'* The article was well illustrated by Rab's photos. This was a time when few UK-based climbers wrote for the American magazines.

Editorials

By now, Rob had been building up his climbing experience, layer on layer. But, as if to show that he was still firmly grounded, in 1998 he became the General Editor of District Guides for the SMC. District guides are the hillwalking guides for Scotland, taking in such classic areas as the Central Highlands, The Islands, the Cairngorms and so on. It's a post requiring a good general knowledge of the Scottish mountains, as well as the ability to handle the sometimes awkward characters who write the guides.

'A Day On The Hill' [7] is an account of the first winter ascent of Stage Fright VI,7, Coire an Lochain, Cairngorms, with Graeme Ettle, one of the leading lights in Scottish winter climbing. A hard mixed route in a busy arena, Rob has conversations with adjacent climbing teams, interspersed with either looking after Graeme's ropes, or leading himself. *'Fond memories of the hill start with good climbs and good mates. We had had a superb day. But of course it was superb – it was a day on the hill.'*

'Reluctant Persistence' [8] sees Rob and Rab back on The Cobbler, slowly and reluctantly, but persistently, grinding their way up a hard mixed route. Aching calf muscles, fear fighting focus. *'If we complete this route, that will be 20 winters since Rab and I first did a new route together, a long partnership that has seen the rise of modern hard mixed climbing.'* The team win their way, and make the first winter ascent of Gangway (VI,7). The writing here has matured, with Rob pausing now and then to take stock of his inner feelings on the climb, exploring the dialogue between mind and matter. (It was also my final Journal as Editor, but the beginning of a happy period of working with Rob on other publications – KVC.)

'The Corbetts and Other Scottish Hills' [9] Rob jointly edited this guide with Hamish Brown. The Corbetts are the 219 Scottish hills whose heights lie between 2,500 ft and 3,000 ft (762m – 914.4m). They may be lower than the Munros, but their challenge is, if anything, even greater, as they are usually climbed in isolation. (Rob finished, or compleated his Munros in 1997, and was very close to finishing the Corbetts.)

Rob had much to do in the way of communication and organisation for this guide, but on top of that I had invited him to be my fellow Editor on a parallel project, 'The Corbetts and Other Scottish Hills' [10]

Both Rob and myself were interested in technology, and marrying technology with guidebook production made it even more fun. Partly due to the company who produced the CD-ROM for us, and partly due to Rob's efficiency, this publication was a model of smooth production. By this time, e-mail was widely used, making the task even easier.

Not all publishing work is visible of course, and for some years both of us had been members of the SMC Publications Sub-Committee. On this we met with other club members and thrashed out policies and styles for new guidebooks and other publications. Ethics came into it too, as new climbing practices came up against old rocks or icy walls. Rob and myself were both firm advocates of the use of GPS units on the mountains, as superb backup for safe navigation and as potential publishing tools. We enjoyed teasing the uninitiated, but gently, as we tried to bring them into the fold.

Rob was persuaded to take up the post of Convenor of the Publications Sub-Committee in 2002, with the stepping down of his predecessor. Despite his initial reluctance, being already the possessor of a very busy life, he quickly picked up the reins and became a popular and highly effective office-bearer. There were many changes taking place in our publishing sphere, and Rob smoothed the way. He would have been my fellow Editor on a new edition of another CD-ROM – 'The Munros', the list of 284 Scottish mountains of over 3,000 ft. Additionally, we had just begun the planning for a new book of selected Scottish hillwalks, with Rob as a fellow author. But, like many other potential publications, routes, and other projects I'm sure, it was not to be.

References

[1] SMCJ 32, 1983, pp 123-128
[2] SMCJ 34, 1991, pp 578-582
[3] SMCJ 35, 1992, pp17-20
[4] SMCJ 35, 1994, pp 432-439
[5] Climber, Jan 1997
[6] Climbing, Feb 1994
[7] SMCJ 36, 1998, pp 469-474, one illustration
[8] SMCJ 38, 2000, pp 277-281
[9] Scottish Mountaineering Trust, 2002, 2nd Edition, ISBN 0907521711, 282 pp
[10] Scottish Mountaineering Club, 2002, CD-ROM. ISBN: 0 907521 72 X

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Experience of Robert William Milne

COMPANY: Intelligent Applications Limited
1 Michaelson Square
Livingston
West Lothian
Scotland UK
EH54 7DP

JOB TITLE: Managing Director

TELEPHONE: +44 (0)1506 47 20 47

FAX: +44 (0)1506 47 22 82

EMAIL: rmilne@bcs.org.uk

POSITIONS HELD

Company Founder, Managing Director
Intelligent Applications Ltd *1986 to 2005*

The company's primary focus is the application of expert systems technology. He has primary responsibility for identifying areas where expert systems technology can provide industrial benefit, in conceptualising a solution, directing the needed software development, and ensuring that a successful application is delivered. He is also responsible for corporate activities from strategic planning, to sales and marketing to day-to-day management. He developed the idea for the company's key TIGER product and has lead its development and commercialisation.

Chief Scientist
Dept of the Army, Artificial Intelligence Centre *January 1985 to January 1986*
The Pentagon, Washington D.C.

Responsible for the founding of the Army AI Centre, including initial training of the centre's personnel, establishment of an applications development program, and briefing the senior officers in the US Army on the prospects for AI. Provided consultancy to many Army AI activities.

Assistant Prof. of Electrical Engineering
Dept of Electrical Engineering *March 1982 to January 1985*
Air Force Institute of Technology
Wright-Patterson AFB, Ohio 45433

Initially taught Masters degree courses in software engineering. Founded the WPAFB AI program, including developed a Masters degree level AI program, short courses in AI, research projects and consultancy to many Army and Air Force AI activities.

EDUCATION

Ph.D. in Artificial Intelligence from the University of Edinburgh, 1983.
B.Sc. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, 1978.

BORN

13 July 1956, Libby, Montana, USA: US Citizen and UK Citizen

PUBLICATIONS

TIGER with Model Based Diagnosis: Initial Deployment
Knowledge Based Systems Journal. Vol. 14, pp. 213–222. Publishers: Elsevier Science B.V. (2001).
Co-author: C. Nicol and L. Trave-Massuyes

Fault Diagnosis for Gas Turbines Based on The Control System
Engineering Applications of Artificial Intelligence. Vol. 13, pp. 477–484. Published by Elsevier Science Ltd (2000)
Co-authors: A. Guasch and J. Quevedo

Dimensional analysis based causal ordering
Proceedings of the 13th International Workshop on Qualitative Reasoning, pp. 193–202, (1999).
Co-authors: Q. Shen and T. Peng

Predicting Paper Making Defects On-Line Using Data Mining
Knowledge-Based Systems Journal. Vol. 11, pp. 331–338. Publishers: Elsevier Science B.V. (1998).
Co-authors: M. Drummond and P. Renoux

Knowledge Based Systems For Condition Monitoring
Handbook of Condition Monitoring – Techniques and Methodology. Chapter 22, pp. 521–540. Editor: A. Davies. Publishers: Chapman & Hall (1998).

A fuzzy expert system for vibration cause identification in rotating machines
Proceedings of the 6th International Conference on Fuzzy Systems, pp. 555–560, (1997).
Co-authors: C. Siu and Q. Shen

A fuzzy rule- and case-based expert system for turbomachinery diagnosis
Proceedings of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, pp. 556–563, (1997).
Co-authors: C. Siu and Q. Shen

Predicting Paper Making Defects On-Line Using Data Mining
Applications and Innovations in Expert Systems V. Proceedings of ES97 Conference, pp. 59–68. Editors: A. Macintosh and R. Milne. Publishers: SGES Publications (1997).
Co-authors: M. Drummond and P. Renoux

Predicting Paper Making Defects On-Line Using Data Mining
Studies in Informatics and Control Journal. Vol. 6, N° 4, (1997).
Co-authors: M. Drummond and P. Renoux

Gas Turbine Condition Monitoring Using Qualitative Model Based Diagnosis
IEEE Expert magazine, "Intelligent Systems & Their Applications". Vol. 12, N° 3, pp. 21–31. Publishers: IEEE Computer Society (1997).
Co-author: L. Trave-Massuyes

Knowledge Based Systems For Maintenance Management
Handbook of Condition Monitoring and Maintenance Management in Industries. pp. 377–393. Publishers: Elsevier Advanced Technology (1996).

TIGER: Knowledge Based Gas Turbine Condition Monitoring
BCS-SGES Newsletter. Autumn, N° 34, pp. 16–31. (1996).
Co-authors: C. Nicol and J. Quevedo

TIGER: Knowledge Based Gas Turbine Condition Monitoring
AI Communications Journal. Vol. 9, N° 3, pp. 1–17. Publishers: IOS Press (1996).
Co-authors: C. Nicol and J. Quevedo

Continuous Expert Diagnosis: Is The Future So Far Away?
Condition Monitor International Newsletter. N° 114, pp. 5–9. Publishers: Elsevier Advanced Technology (1996).

Diagnosis of Dynamic Systems Based on Explicit and Implicit Behavioural Models: An Application To Gas Turbines In ESPRIT Project TIGER
Applied Artificial Intelligence. Vol. 10, N° 3, pp. 257–277. Publishers: Taylor & Francis (1996).
Co-author: L. Trave-Massuyes

Increasing The Availability of FMS With Multi-Media Supported Diagnostic Expert Systems
Journal of Intelligent Manufacturing. N° 7, pp. 399–404. Publishers: Chapman & Hall (1996).
Co-author: S. Torvinen

Integrated Analytical and Fuzzy Simulation for Monitoring and Supervision of Machinery
International Journal of Manufacturing System Design. ISSN: 0218-3382, 2 (3), pp. 177–192, (1995).
Co-author: M. Ulieru

TIGER: Knowledge Based Gas Turbine Condition Monitoring
Applications and Innovations In Expert Systems III. Proceedings of ES95 Conference. pp. 23–43. Editors: A. Macintosh and C. Cooper. Publishers: SGES Publications (1995).
Co-authors: C. Nicol, L. Trave-Massuyes and J. Quevedo

Continuous Expert Diagnosis: Is The Future So Far Away?
Modern Power Systems Journal. Vol. 15, N° 10, pp. 19–21, ISSN 0260-7840. Publishers: Wilmington Publishing Ltd (1995).

Real-Time Model Based Diagnosis of Gas Turbines
Artificial Intelligence In The Petroleum Industry 'Symbolic and Computational Applications'. Chapter 6, pp. 197–215. Editors: Bertrand Braunschweig and Ron Day. Publishers: Editions Technip (1995).
Co-author: L. Trave-Massuyes

Application Oriented Qualitative Reasoning
The Knowledge Engineering Review. Issue 10:2, pp. 181–204. Publishers: Cambridge University Press (1995).
Co-author: L. Trave-Massuyes

Real-Time Situation Assessment of Dynamic Systems
Intelligent Systems Engineering Journal. Autumn, pp. 103–124 (1994).
Co-authors: C. Nicol, M. Ghallab, L. Trave-Massuyes, K. Bousson, C. Dousson, J. Quevedo, J. Aguilar and A. Guasch

Building Successful Applications: The Wrong Way and The Right Way
KI-94 Conference, Anwendungen der Kunstlichen Intelligenz. pp. 24–36. Publishers: Springer-Verlag (1994).

On-Line Diagnostic Expert System For Gas Turbines
Profitable Condition Monitoring (Keynote Address). pp. 47–54. Editor: B.K.N. Rao. Publishers: Kluwer Academic (1992).

Amethyst: The Development Experience
Industrial Applications of Knowledge-Based Diagnosis. pp. 115–141. Editors: G. Guida and A. Stefanini. Publishers: Elsevier Science (1991).

Petroleum Applications: How Do We Realise The Potential?
Revue De l' Institut Francais Du Petrole. Vol. 47, N° 3. (1992).

The Role of Expert Systems In Condition Monitoring
Profitable Condition Monitoring. Editor: B.K.N. Rao. Publishers: Kluwer Academic. (1992)

Expert Systems: Opportunities In The Mineral Industry
Minerals Engineering Journal. Vol. 5, N° 10–12, pp. 1307–1323. Publishers: Pergamon Press (1992).
Co-author: R. Bearman

Book Review: Expert Systems – A Manager's Guide by Karl Wigg
Reviewed by Dr. Milne for The Knowledge Engineering Review. Vol. 6, N° 4. (1991).

Model Based Reasoning: The Applications Gap
IEEE Expert magazine (1991).

Why Are Expert Systems Hard To Define
BCS-SGES Newsletter, Spring (1991).

Why Are Expert Systems Hard To Define (Short Version)
DTI Manufacturing Intelligence Newsletter, Spring (1991).

Research & Development – Improving the Maintenance of Machine Tools
Department of Trade and Industry Managing Maintenance Booklet (1991).

Amethyst: Vibration Based Condition Monitoring
The Handbook of Expert Systems Applications In Manufacturing. Editor: Jessica Keyes, Rex Maus. Publishers: McGraw Hill Inc. (1990).

An Expert System for PLC Diagnostics
Personal Computers in Industrial Control, M.C.U Conferences. pp. 139–143. (1990).

Case Studies in Condition Monitoring
Knowledge-Based Systems for Industrial Control, IEE Control Engineering Series 44. J. McGhee, M.J. Grimble and P. Mowforth, Editors, Peter Peregrinus Ltd Publications, pp. 255–266, 1990.

Amethyst: Automatic Diagnosis of Rotating Machinery Faults
Operational Expert System Applications in Europe. pp. 244–258. Editor: Gian Piero Zarri. Publishers: Pergamon Press (1990).

Amethyst: An Expert System for the Diagnosis of Rotating Machinery
Comadem'90 Conference, 'Condition Monitoring and Diagnostic Engineering Management', pp. 287–292 (1990).

Intelligent Data Interpretation
Practical Experiences In Building Expert Systems, pp. 103–106. Editor: Max Bramer. Publishers: John Wiley & Sons Publications (1990).

Expert Systems On-line
Systems & Control Encyclopaedia Supplementary Vol. 1, pp. 243–251. Editor: Madan G. Singh. Publishers: Pergamon Press (1990).

Diagnostic Strategies
Systems & Control Encyclopaedia Supplementary Vol. 1, pp. 155–160. Editor: Madan G. Singh. Publishers: Pergamon Press (1990).

Monitoring Process Control Systems
Advances in Engineering Software, Vol. 12, N° 3, pp. 129–132. Publishers: Computational Mechanics Publications (1990).

Cathie: Expert Interpretation of Gas Chromatography
Journal of Chromatography. pp. 341–347. Publishers: Elsevier Science Publishers B.V. (1989).

Artificial Intelligence for Vibration Based Health Monitoring
Journal of Condition Monitoring, Ruth Higginbotham, Vol. 2, N° 3, pp. 213–218. Editor: BHRA, The Fluid Engineering Centre (1989).

Resolving Lexical Ambiguity in a Deterministic Parser
Lexical Ambiguity Resolution. Editor: Steven L. Small. Publishers: Morgan Kaufmann Publishers, Inc. (1988).

Synergist: A Schematic Capture and Fault Diagnosis System
Intelligent CAD Systems 1: Theoretical and Methodological Aspects, Publishers: Springer-Verlag, Eurographics Seminar Book (1987).

Artificial Intelligence for On-line Diagnosis
IEE Proceedings, Vol. 134, N° 4 (1987).

Strategies for Diagnosis
IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-17, N° 3. (1987).

Artificial Intelligence Applied to Condition Health Monitoring
Chartered Mechanical Engineer. (1986).

Resolving Lexical Ambiguity in a Deterministic Parser
Computational Linguistics Journal, Vol. 12, N° 1 (1986).

A Few Problems with Expert Systems
Expert Systems in Government Symposium, pp. 66–67. Editor: Kamal Karna. Publishers: IEEE Computer Society (1985).

Information Management
Expert Systems, Expert Systems in Government Symposium, pp. 414–416. Editor: Kamal Karna. Publishers: IEEE Computer Society (1985).

Diagnosing Faults Through Responsibility
ACM Annual Conference, Association for Computing Machinery. pp. 88–91 (1985).

Fault Diagnosis Through Responsibility
Proceedings of the Ninth International Joint Conference on Artificial Intelligence, pp. 424–427 (1985).

Reasoning about Structure, Behaviour and Function
SIGART Newsletter. N^{o} 93, pp. 4–59. Editor: B. Chandrasekaran and R. Milne (1985).

The Theory of Responsibilities
SIGART Newsletter. N^{o} 93, pp. 25–30 (1985).

Artificial Intelligence
Simulation and Modelling Bulletin, Vol. 3 (1983).

Predicting Garden Path Sentences
Journal of Cognitive Science, Vol. 6, pp. 349–373 (1982).

PRESENTATIONS: (Most Invited)

Model-Based Diagnosability and Sensor Placement Application to a Frame 6 Gas Turbine Sub-System
IJCAI-2001 Conference, Seattle, Washington (USA), 04–10 August 2001.
Co-author: L. Trave-Massuyes and T. Escobet

TIGER: With Model Based Diagnosis: Initial Deployment
ES2000 Conference, Cambridge (UK), 11–13 December 2000.
Co-author: C. Nicol and L. Trave-Massuyes

TIGER: Continuous Diagnosis of Gas Turbines
ECAI-2000 ‘14th European Conference on AI’ (Berlin), 20–25 August 2000.
Co-author: C. Nicol

TIGER: Knowledge Based Gas Turbine Condition Monitoring
IEE Colloquium on ‘AI Based Applications for the Power Industry’, London (UK), 1 June 1999.
Co-author: L. Trave-Massuyes and J. Quevedo

TIGER: Intelligent Condition Monitoring and Diagnosis
Powergen'98 Conference
Co-author: C. Nicol, I. McKay and J. Hibbert

Gaps Between Research & Industry Related To Model Based and Qualitative Reasoning
ECAI 98 '13th European Conference on AI', Brighton (UK), 22–28 August 1998.
Co-author: L. Trave-Massuyes

Predicting Paper Making Defects On-Line Using Data Mining
ES97 Conference, Cambridge (UK), 15–17 December 1997.
Co-author: M. Drummond and P. Renoux

TIGER: Intelligent Continuous Monitoring of Gas Turbines
Model Based Reasoning Colloquium, London (UK), 17 November 1997.
Co-author: C. Nicol

Model Based Aspects of The TIGER Gas Turbine Condition Monitoring System
SafeProcess'97 Conference, Hull (UK), 26–28 August 1997.
Co-author: L. Trave-Massuyes

Using Data Mining to Learn To Be Able To Predict Paper Making Defects On-Line
Unicom Seminar 'CBR & Data Mining: Putting The Technology To Use', London (UK), 18–19 March 1997.
Co-author: C. Nelson

FMS MAINT: Maintenance For The Future
The Maine Event Conference, London (UK), 14 November 1996.
Co-author: P. Malinen

FMS MAINT: Maintenance For The Future
IEE Colloquium on AI in Manufacturing, London (UK), 11 February 1997.

Will This Paper Break! Discovering How To Predict Process Problems
Unicom Seminar 'Data Mining and Data Warehouse'96: Interaction of Information and Decision Technologies', London (UK), 4–6 November 1996.

TIGER: Applying Hybrid Technology for Industrial Monitoring
ECAI'96 '12th European Conference on AI', Budapest (Hungary), 11–16 August 1996.
Co-author: L. Trave-Massuyes and J. Quevedo

FMS MAINT: Maintenance For The Future
Euromaintenance'96 Conference, Copenhagen, 21–23 May 1996.
Co-author: P. Malinen

TIGER: Knowledge Based Gas Turbine Condition Monitoring
ES95 Conference, Cambridge (UK), 11–13 December 1995.
Co-author: L. Trave-Massuyes and J. Quevedo

Joint Presentation to Industry with L. Trave-Massuyes
UNICOM, London (UK), November 1995.

Joint Presentation to Industry with L. Trave-Massuyes
SITEF, October 1995, Toulouse (France).

CBR For Diagnosis: A Case Study for Electrical Machinery Diagnostics
ICCBR'95 Conference, Portugal, 23–26 October 1995.
Co-author: C. Nelson

Diagnosis of Dynamic Systems Based On Explicit and Implicit Behavioural Models: An Application To Gas Turbines in ESPRIT Project TIGER
IJCAI'95 Qualitative Reasoning Workshop, Montreal (Canada), 20 August 1995.
Co-author: L. Trave-Massuyes

Diagnosis of Dynamic Systems Based On Explicit and Implicit Behavioural Models: An Application to Gas Turbines in ESPRIT Project TIGER
SCAI'95 Conference, Trondheim (Norway), 29–31 May 1995.
Co-author: L. Trave-Massuyes

TIGER: Knowledge Based Gas Turbine Condition Monitoring
Powergen'95 Conference, Amsterdam (Netherlands), 16–18 May 1995.
Co-author: C. Nicol, D. Hunt, R. Fisher and C. Cloughley

TIGER: Numeric & Qualitative Model Based Diagnosis
IEE Colloquium on 'Qualitative & Quantative Modelling Methods for Fault Diagnosis', London (UK), 24 April 1995.
Co-author: L. Trave-Massuyes

Electrical Drive Diagnosis Using Case Based Reasoning
IEE Colloquium on 'CBR: Prospects for Applications', London (UK), 7 March 1995.
Co-author: C. Nelson

Knowledge Guided Data Mining
IEE Colloquium on 'Knowledge Discovery in Databases', London (UK), 1–2 February 1995.
Co-author: C. Nelson

TIGER: A Model Based and Rule Based Real-Time Tool For Manufacturing and Diagnosis of Dynamic Systems
IEE Colloquium on 'Real-Time Knowledge Based Systems', London (UK), 31 January 1995.
Co-author: C. Nicol

The Secrets of Successful Applications
SPICIS'94 Conference, Singapore, 14–17 November 1994.

Automatic Diagnostics Development Based on a Programmable Logic Controller
DX'94 Conference, New York (USA), 17–19 October 1994.
Co-author: T. Guasch

Building Successful Applications: The Wrong Way and The Right Way
KI-94 Conference, Saarbruecken (Germany), 23 September 1994.

Application Oriented Qualitative Reasoning
QR'93 Conference, Washington (USA), 16–20 May 1994.
Co-author: L. Trave-Massuyes

Real-Time Model Based Diagnosis of Gas Turbines
Institute of Measurement and Control, Fault Diagnosis Conference, London (UK), 26 January 1994.
Co-author: L. Trave-Massuyes

Real-Time Model Based Diagnosis of Gas Turbines
ES93 Conference, Cambridge (UK), 13–15 December 1993.
Co-author: L. Trave-Massuyes

Machine Tool Breakdown Diagnosis
IEE Colloquium on Expert Systems in Machine Tool Based Diagnosis and Maintenance, London (UK), 29 September 1993.

Machine Tools Breakdown Diagnosis
IMECHE'93 Conference on AI in Automation, Birmingham (UK), 20 September 1993.

Real-Time Model Based Diagnosis of Gas Turbines
Eurocaipep'93 Conference, Aberdeen (UK), 20–22 September 1993.
Co-author: L. Trave-Massuyes

Real-Time Model Based Diagnosis of Gas Turbines
Fourth International Workshop on Principles of Diagnosis, Wales (UK), 6–8 September 1993.
Co-author: L. Trave-Massuyes

Real-Time Model Based Diagnosis of Gas Turbines
AI ENG'93, Toulouse (France), 29 June – 1 July 1993.
Co-author: L. Trave-Massuyes

On-Line Diagnostic Expert System For Gas Turbines
TOOLDIAG'93, Toulouse (France), 5–7 April 1993.

On-Line Diagnostic Expert System For Gas Turbines
ES92 Conference, Cambridge (UK), 15–17 December 1992.

An Overview of Expert Systems in Manufacturing
Midlands Manufacturing Intelligence Club meeting, Leicester (UK), 25 June 1992.

An Overview of Expert Systems in Manufacturing
Belfast AI Group, Belfast (Ireland), 2 July 1992.

Why Some AI Applications Are Successful and Most Fail
ECAI'92 Symposium, Vienna (Austria), 7 August 1992.

Expert Reasoning With Numeric Histories
SICICA'92 Symposium, Malaga (Spain), 20–22 May 1992.
Co-authors: M. Drummond and E. Bain

Amethyst
IEE Colloquium, London (UK), 25 February 1992.

Predicting Faults With Real-Time Diagnosis
30th IEEE Conference on Decision and Control, Brighton (UK), 11–13 December 1991.
Co-authors: M. Drummond and E. Bain

Amethyst: Automatic Diagnosis of Rotating Machinery Faults
European Conference on Industrial Applications of Knowledge Based Diagnosis, Milan (Italy), 17–18 October 1991.

Petroleum Applications: How Do We Realise The Potential
European Conference on Artificial Intelligence in Petroleum, France, 16 October 1991.

Integration: The Key To Second Generation Applications
Safeprocess'91 Conference, Baden-Baden (Germany), 10–13 September 1991.

Using Expert Systems To Assist With Machine Tools
ES91 Conference, London (UK), September 1991.

Portable Bearing Diagnostics Using Enveloping & Expert Systems
Comadem'91 Conference, Southampton (UK), 17 July 1991.
Co-authors: J. Aylett, S. McMahon and T. Scott

Real-Time Means Planning Ahead To Look Back
AI in Engineering Conference, Oxford (UK), 3–4 July 1991.
Co-authors: M. Drummond and E. Bain

Second Generation Expert Systems: The Applications Gap
Avignon'91 Conference, Avignon (France), 29 May 1991.

A Diagnostic Expert System Embedded In A Portable Vibration Analysis Instrument.
IEE Colloquium, London (UK), 13 May 1991.
Co-authors: J. Aylett, S. McMahon and T. Scott

Expert System Applications In Manufacturing
British Computer Society Evening Session, Aberdeen (UK), 13 March 1991.

Expert System Applications In Manufacturing
British Computer Society Evening Session, Edinburgh (UK), 16 January 1991.

The Future For AI
The Petroleum Science and Technology Institute with the AIAI Seminar, Edinburgh (UK), 30 November 1990.

Expert Systems In Condition Monitoring
Maintenance Energy Environment Technology Association, Maintenance Conference, Dublin (Ireland), 29 November 1990.

An Expert System For PLC Diagnostics
Personal Computers in Industrial Control, Stevenage (UK), 22 November 1990.
Co-author:

An Overview of AI in the UK
Scottish Expert System Group Inaugural Meeting, (Special Interest Group of the British Computer Society), Glasgow (UK), 17 October 1990.

Why Do Most Applications Fail?
The Midlands Expert System Group Inaugural Meeting, (Special Interest Group of the British Computer Society), Birmingham (UK), 16 October 1990

Where Is The Business Benefit?
Expert Systems'90 Conference, London (UK), 19 September 1990.

An Expert System For Vibration Diagnostics
IMECHE Gearbox Measuring Vibration Conference, London (UK), April 1990.

Speech Recognition Using a Deterministic Parser
National Aerospace and Electronics Conference, May 1984.
Co-author: R. Routh

A Spoken English Recognition Expert System
CRISIS Conference, Colorado Springs, Colorado (USA), January 1984.
Co-author: R. Routh

TUTORIALS

Building Successful Applications

IJCAI'92, Chambery (France).
SPICIS'94 Conference, Singapore, 14–17 November 1994.
ECAI'94 '11th European Conference on AI', Amsterdam (The Netherlands), 8–12 August 1994.
AAAI'95, Washington D.C. (USA).
SCAI'95 Conference, Trondheim (Norway), 29–31 May 1995.
ECAI'96 '12th European Conference on AI', Budapest (Hungary), 11–16 August 1996. (Amsterdam, Washington D.C. and Chambery with L. Trave-Massuyes)

Applied Qualitative Reasoning joint tutorials with L. Trave-Massuyes

IJCAI'93, Chambéry (France), August 1993.
AAAI'93, Washington DC (U.S.A), July 1993.
AI for Applications, Orlando (U.S.A), March 1993.
Expert Systems Conference, Cambridge (U.K), December 1992.
ECAI'92, Vienna (Austria), August 1992.

Building Case Based Reasoning Systems

ES Conference, Cambridge (UK).

MEMBER OF CONFERENCE ORGANISING COMMITTEES

Expert Systems, British Computer Society Specialist Group on Expert Systems

London & Cambridge (UK) – 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003
(Conference Chairman, 1992, 1886)

IJCAI – International Joint Conference on Artificial Intelligence

IJCAI'99 – 16th International Joint Conference – Stockholm (Sweden) (Programme Committee)
IJCAI'01 – 17th International Joint Conference – Seattle (Programme Committee)
IJCAI'05 – 19th International Joint Conference – Edinburgh (Local Arrangements Chair)

ECAI – European Coordinating Committee for Artificial Intelligence

ECAI'98 – 12th European Conference – Brighton (UK) (Programme Committee)
ECAI'00 – 14th European Conference – Berlin (Germany) (Programme Committee)
ECAI'02 – 15th European Conference – Lyon (France) (Conference Committee)

DX99 Workshop – Tenth International Workshop on Principles of Diagnosis

Loch Awe, Scotland (UK) – 1999 (Co-organiser)

QUARDET'93 – III IMACS Workshop

Barcelona (Spain) – 1993

Sixth International Conference on Industrial & Engineering Applications of AI and Expert Systems
Edinburgh (UK) – 1993

TOOLDIAG'93 Conference. International Conference on Fault Diagnosis
Toulouse (France) – 1993

Ninth IEEE Conference on Artificial Intelligence for Applications
IEEE Computer Society, Florida (USA) – 1993

Fourth International Conference on Profitable Condition Monitoring
BHR Group, Stratford-upon-Avon (UK) – 1992

ECAI'92 Conference – 10th European Conference on Artificial Intelligence
Vienna (Austria) – 1992, Programme committee

IMACS International Workshop on Decision Support Systems and Qualitative Reasoning
Toulouse (France) – 1991 and Barcelona (Spain) – 1993

Workshop on Computer Software Structures Integrating AI/KBS Systems in Process Control
Bergen (Norway) – 1991

IEEE Computer Society Expert Systems & Their Applications
Avignon (France) – 1987, 1988, 1989, 1990, 1991

IEE Vacation School on Knowledge-Based Systems for Process Control
Glasgow (UK) – 1990

Editorial Board for Journal of Intelligent Robotic Systems: 1989–2005

Editorial Board for the Journal of AI in Engineering: 1989–2005

Editorial Board, AI Magazine for the AAAI: 2000–2005

Editorial Board, AI Comms for IOS Press: 2000–2005

Editorial Board, Journal of Applied Artificial Intelligence: 1995–2005

PROFESSIONAL ACTIVITIES

President, European Coordinating Committee for Artificial Intelligence (ECCAI) (2000–2004).
Director, Scottish Software Federation: the industry body for IT and software companies in Scotland (1997–2000).
Director, ScotlandIS: Successor to the Scottish Software Federation (2000–2002).

National Committee Member, British Computer Society Specialist Group on Expert Systems since 1987.
Treasurer, British Computer Society Specialist Group on Expert Systems since 1996.
Member, American Association of Artificial Intelligence. Since 1982.
Member, British Computer Society, Specialist Group on Expert Systems.
Member, A.I.S.B., Society for Artificial Intelligence and Simulation of Behaviour.
Past Chairman, Artificial Intelligence Program, Air Force Institute of Technology.
Past Chairman, Dayton SIGART (Special Interest Group in AI of the ACM).
Past Member, Association for Computational Linguistics.
Past Member, Association for Computing Machinery.
Nominated for National Treasurer, ACM SIGART.
Chartered Engineer, European Engineer.
Consultant to the European Commission for Expert Systems.

FELLOWSHIPS OF ACADEMIC SOCIETIES

Fellow of British Computer Society.
Fellow of European Coordinating Committee for Artificial Intelligence (elected 2000).
Fellow of Royal Society of Edinburgh (elected 2003).

PAST CONSULTING ACTIVITIES

Consultant to the US Army Research Office in AI.
Consultant to the US Air Force Avionics Laboratory in AI.
Consultant to the US Air Force Rome Air Development Centre in AI.
Consultant to the US Defence Advanced Research Projects Agency in the Military Applications of AI.
Consultant to Schlumberger Electronics (UK).

DIRECTORY LISTINGS

Who's Who In The World	Men of Achievement
Dictionary of International Biography	Third Edition of Five Hundred Leaders of Influence
5000 Personalities of The World	International Who's Who of Contemporary Achievement

A. Bundy and S. Wilson (Eds.)
IOS Press, 2006

Author Index