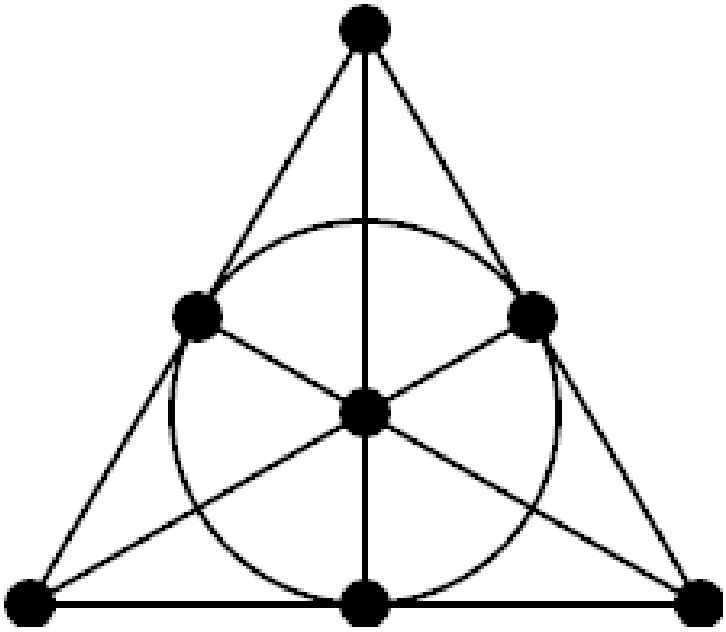
NEWSLETTER

BNVKI
AIABN

February 2005
Vol. 22, No. 1
ISSN 1566-8266



Finite Geometries, part 1

Benelearn 2005

Go at the Frontiers of AI

News from the Belgium-Netherlands Association for Artificial Intelligence

Three out of Three Ain't Bad

Editor-in-chief

In the previous issue of this Newsletter in the minutes of the last general assembly of the BNVKI-AIABN there was mention of a short debate whether the BNAIC should strive for being a high-quality conference or a gathering place for the Dutch-Belgian AI community. The main consensus was that both aims are not disjoint. Of course, to implement the conditions for this, a third aim should be added: organizers must find a pleasant environment for hosting the conference.

We are pleased to announce that the BNAIC 2005 will take place on October 17 and 18, 2005 in the beautiful city of Brussels, in the palace of the Koninklijke Vlaamse Akademie van België voor Wetenschappen en Kunsten. This palace not only is a very nice old building (see below), but also has a very nice interior, as evidenced by the praising words by our chairman at page 4 of this issue. So we feel that the organizers have already achieved one of the goals for a successful BNAIC 2005. In order to achieve also a second goal we evoke all AI researchers in Belgium and The Netherlands to send in high-quality papers. The Call for Papers is on page 22 of this issue. More information about the BNAIC 2005, such as a Call for Registration, will appear in the next issue of the Newsletter, but please reserve the 17th and 18th of October already in your agenda. I hope to see you all in Brussels then, and I'm confident that all three goals may be fulfilled: three out of three ain't bad!

BNAIC'05 venue: exterior of the palace of the Koninklijke Vlaamse Akademie van België voor Wetenschappen en Kunsten.

In this issue the reader will find the first part of a contribution by Henk Visser on Finite Geometries. This is an intriguing research subject for already about a century, one of the pioneers being the Italian mathematician Gino Fano (everyone knows the figure at the front cover being named after him). As evidenced by Henk Visser's article new insights into this "old" area can still be gained, assisted by the clever use of computer programs.

Further, Werner Uwents reports on a successful Benelearn 2005 conference, organized by the Human Media Interaction Group of the University of Twente. Erik van der Werf reports on a workshop in Maastricht, in collocation with his Ph.D. defence, showing the state-of-the-art in Go programs. Finally, Marie-Francine Moens reports on the 17th Annual Conference on Legal Knowledge and Information Systems, held recently in Berlin.

Finally I want to inform you that starting from the June issue we plan to have special-topic issues, among others reporting on the research of the most important groups in selected research areas. The next issue will contain more information, as well as a call to our readers for contributions.

http://como.vub.ac.be/bnaic2005

TABLE OF CONTENTS

The photographs in this issue are by courtesy of Karl Tuyls (pages 2 and 4) and Martijn van Otterlo (pages 7 and 8).

Front cover: Fano's famous "projective plane" (see Henk Visser's contribution on pages 4-7 of this issue).

BNVKI-Board News

Han La Poutré

On Friday 11 March, a delegation of the BNVKI board visited the upcoming BNAIC organisors, in Brussels: Ann Nowé, Bernard Manderick, Karl Tuyls, and Katja Verbeeck. For the two Dutch board members, Cees Witteveen and myself, this meant a pleasant train trip to a beautiful city we do not visit so often. For the Belgian member, Marc Denecker, this was a more common visit.

The BNAIC organisors guided us towards the place where the BNAIC would take place. We passed the royal palace in Brussels, and ended up at a second palace: the palace of the Koninklijke Vlaamse Akademie van België voor Wetenschappen en Kunsten (KVAB). This palace originated from the beginning of the nineteenth century, and was actually built as the palace for the Dutch crown prince Willem. We had a sightseeing walk through the palace, where we were accompanied by an employee of the Akademie, who told us all about halls, Indonesian wooden floors, crystal lights, acoustics, and expensive toilets with majestic mirrors. All this was actually built for crown prince Willem. However, according to our guide, just when it was finished, the Belgian became "stout" (nasty/bold). Which meant: the independence of Belgium from the Netherlands. Our Belgium guide obviously liked to talk about the palace, and we liked to hear about it.

Well, to make a long story short: the next BNAIC will take place in October, in this impressive palace in Brussels. I'm sure that BNAIC attendees will like it a lot and probably also want to stay in Brussels the weekend before. And yes: you can find the first call for papers for the BNAIC elsewhere in this Newsletter, as well as some nice pictures. This will be an interesting and memorable BNAIC…. again.

BNAIC'05 venue: interior of the palace of the Koninklijke Vlaamse Akademie van België voor Wetenschappen en Kunsten.

Finite Geometries, Part 1

Henk Visser
Haarlem

The method of transposition, that is, solving problems after transposing them to another domain, has proved to be useful in both finite projective and finite affine geometry.[1] These are "pure" line geometries: points are seen as parts of lines, and there are no other geometrical figures than points and lines. An overview will be given in section one. But we can also define "circle geometries" by postulating that points are to be seen as parts of circles and then the obvious question is whether the method of transposition can also be fruitfully used in this field of research. This is the subject of section two. In section three the question is addressed whether the results can be generalized for "curve geometries" in general.

FINITE LINE GEOMETRIES

The main idea of "transpositional tricks" in line geometries can easily be demonstrated by the problem of finding models for axioms such as

1. for each two distinct points, there is exactly one line containing them both
2. for each two distinct lines, there is exactly one point contained by both
3. not all points are on the same line
4. there exists at least one line
5. every line contains exactly 3 points[2]

For someone who is not familiar with this kind of problem: the "classical" model is Fano's famous "projective plane" (Figure 1).

Presumably it was found by Fano step by step, and by trial and error, starting with three points on a line.

But the problem can also be systematically solved by representing seven points on a circle and then choosing a triangle, representing a line containing three points, in such a way that it has exactly one point in common with each of its rotations over $2k\pi/7$ (k = 1, 2, 3, …, 6) around the centre of the circle (Figure 1).

[1] See Henk Visser (2001), "Transpositional tricks", *BNVKI Newsletter*, Vol. 18, No. 2, reprinted in the *ALP Newsletter,* (14) 3, available at http://www.cwi.nl/projects/alp/newsletter/aug01/nav/transpositional/transpositional.html

[2] There is no need for transpositions if it is required that every line contains exactly two points: just imagine a triangle.

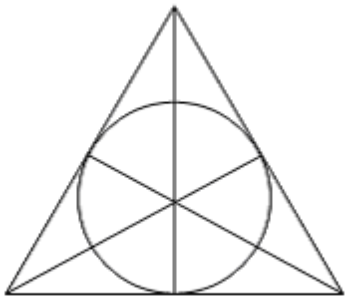


Fig. 1

This problem may in its turn be "transposed" to the problem of writing 7 as a sum of three numbers in such a way that each number below 7 appears exactly once as a partial sum, including cyclical combinations, in order to achieve that there are no more overlappings than in one single point:

$7 = 1 + 2 + 4$

$1 = 1$
$2 = 2$
$3 = 1 + 2$
$4 = 4$
$5 = 4 + 1$ (cyclical)
$6 = 2 + 4$

The resulting model can be pictured in different ways. First of all, Figure 2 can be "completed" by adding the other six triangles, preferably in different colors.

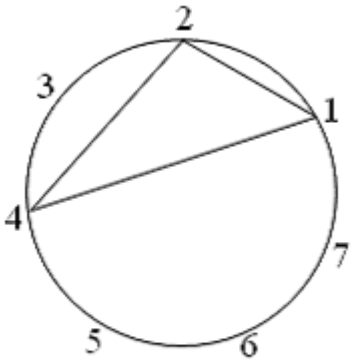


Fig. 2

There is also the following "numerical" representation:

124
235
346
457
561
672
713

The cyclical character of the model is conspicuous. Notice that it has been lost in Fano's representation. Therefore I prefer the following representation in which each point possesses two "locations" (Figure 3):

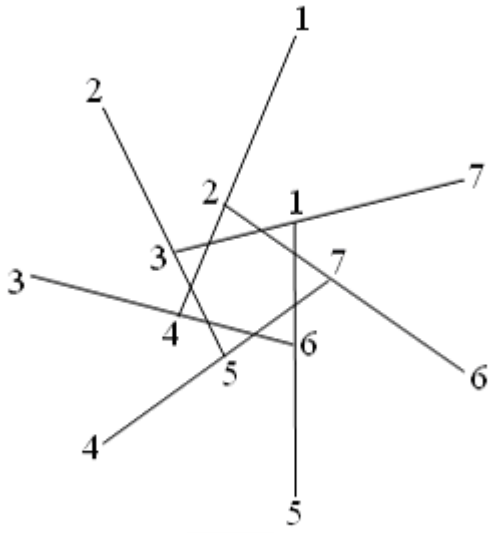


Fig. 3

By varying axiom 5, for example postulating that every line contains exactly 4 points, we get models for the resulting geometry by the following partitions of 13:

$13 = 1 + 2 + 6 + 4$
$13 = 1 + 3 + 2 + 7$

There is no need to pursue the matter further. In this way the so-called finite *projective* "line geometries" may acquire models, though it can be proved that not every variation of axiom 5 leads to a solution. For example, there is no projective line geometry in which every line contains exactly 7 points.[3] The corresponding number of points, to wit 43, has no partitions with the required property.

Similar considerations hold for so-called finite *affine* line geometries. Here the above axiom 2 is replaced by the following axiom:

2*. through a point not on a given line there is exactly one line which does not meet the given line

[3] This result is due to my colleague Dr. Donkers, whose java-program solved the partition problem for projective line geometries up to 10 points on every line.

It is easy to make a model for such a geometry if it is postulated that every line contains exactly 2 points: just imagine a tetraeder. But the problem of finding a model when every line contains exactly 3 points is difficult. Therefore it seems practical to shift to a domain of numbers and to solve the corresponding combinatorial problem, possibly with the following result:[4]

123 246 349 478 569
145 258 357
167 279 368
189

However, there is little "structure" in the thus acquired solution, so let us turn to a representation with a circle, this time with center 1 and eight points 2, 3, … 9 on it. Now it is possible to draw three "lines", that is, triples of points, that do not meet each other (Figure 4):

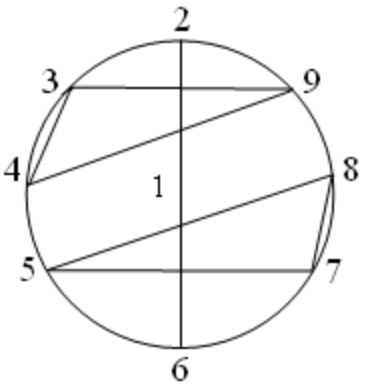


Fig. 4

Starting with the line 126, we see that there are six points that are not "on" it, but through each of these points there is exactly one line which does not meet the given line. This has been done in such a way that only two more lines, 349 and 578, are needed. Now we rotate the three lines over $\pi/4$ and we get the lines 137, 452 and 689. In the same way, we get the lines 148, 563, 792 and 159, 674, 823. In sum:

126 349 578
137 452 689
148 563 792
159 674 823

It is important that for each two distinct points, there is exactly one line containing them both. This is reflected in the property of the partition

8 = 1 + 5 + 2

that each number under 8 except 4 appears exactly once as a partial sum, including cyclical combinations:

1 = 1
2 = 2
3 = 2 + 1
5 = 5
6 = 1 + 5
7 = 5 + 2

There is also a perspicuous representation in which each point has two locations (Figure 5):

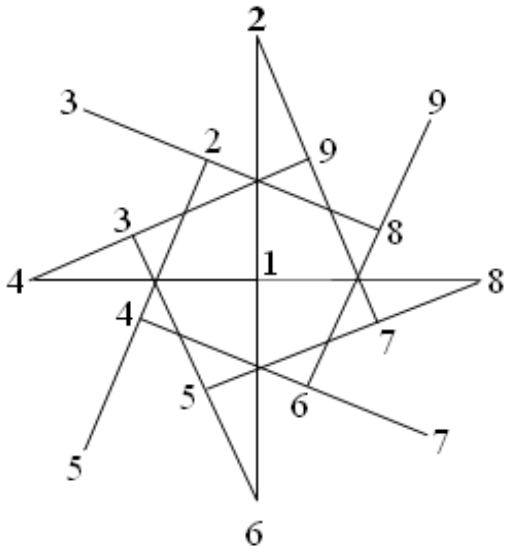


Fig. 5

For "higher order" finite affine line geometries it is useful to look for partitions of the number that precedes the number of points. For example, if every line should contain exactly 4 points, there is the following nice partition of 15:

15 = 1 + 2 + 4 + 8

Notice that there are no partial sums for 5 and 10. As a result, the corresponding geometrical picture with two "generating" lines is as follows (Figure 6):

[4] John Wesley Young, *Lectures on Fundamental Concepts of Algebra and Geometry*. New York: The Macmillan Company, 1923, p. 42.

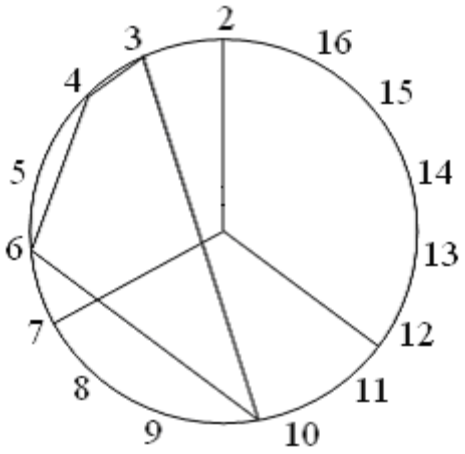


Fig. 6

If every line should contain exactly 5 points, there are two partitions of 24:

$$24 = 1 + 2 + 8 + 9 + 4$$
$$24 = 1 + 3 + 5 + 2 + 13$$

However, there is no model of the affine line geometry in which every line contains exactly 6 points.[5]

The fact that some finite projective or affine geometries "do not exist" is no reason to regard the whole system as insignificant, on the contrary: there are "enough" models and it is a challenge to delineate and explain the exceptions. But even when there would have been "few" models, their explanation might have been a serious question. Let us now see if this occurs with finite circle geometries.

Part 2 will be published in the next issue of the *BNVKI Newsletter*.

Benelearn 2005

February 17-18, 2005
University of Twente

Report by Werner Uwents
Katholieke Universiteit Leuven

This year, the Benelearn conference was held on February 17 and 18 at the University of Twente in Enschede, and organized by the Human Media Interaction Group. Major themes were computational learning theory, evolutionary approaches and data mining. Special attention was also given to some applications of machine learning. The conference attracted some 35 participants.

Martijn van Otterlo shows the history of Benelearn.

DAY 1

Invited speaker on the first day was Kristian Kersting. He discussed the integration of probabilistic reasoning with first order logic and machine learning. In particular, he showed how Bayesian networks, hidden Markov models and Markov decision processes can be upgraded to a logical setting.

Following this first invited lecture, two talks about statistics and theory were scheduled. Unfortunately, one of them had to be canceled. In the other talk, Jan Poland and Marcus Hutter discussed asymptotic behaviour for discrete minimum description length. Using the Hellinger distance, they showed that MDL has good asymptotic properties concerning the convergence to the true distribution.

After lunch, three contributions on learning optimal behaviour were presented. Frans Oliehoek, Matthijs Spaan and Nikos Vlassis addressed the problem of how to play optimally against a fixed opponent in a two-player card game with partial information. They modeled this problem as a partially observable Markov decision process and solved it by dynamic programming.

In another contribution of Jan Poland and Marcus Hutter, they specify a master algorithm for 'active experts problems'. This means that the actions of the master, making a decision based on the recommendations of some experts, may influence the behaviour of the adversary. This algorithm is combined with a universal expert class to derive some theoretic results.

Funlade Sunmola and Jeremy Wyatt studied reinforcement learning in an environment with

[5] It is again Dr. Donkers who also wrote a program for the partition problem in finite affine line geometries.

inherent determinism, in this case from known process templates. The templates are specified by a process language to which there is an underlying context-free grammar. This way, they are able to extend the standard optimistic model selection algorithm to incorporate a process filter.

The final session of this first day was about data mining approaches. In the first talk, Werner Uwents and Hendrik Blockeel proposed an extension for neural networks to process relational data. Some issues in training these networks and results for some experiments were discussed.

Jeroen De Knijf and Ad Feelders presented a way to incorporate constraints in frequent tree mining. This is very similar to the use of constraints in frequent itemset mining, for which it has shown to improve performance substantially. Considering only monotone constraints, an opportunistic pruning method can be used which leads to a considerable speedup compared to the basic algorithm.

To improve spectral clustering methods, Igor Fischer and Jan Poland investigated the effect of amplifying the block structure of the affinity matrix. Experiments confirm the advantage of this amplification, especially for very non-compact clusters.

In the last talk, by Ida Sprinkhuizen-Kuyper, Evgueni Smirnov and Georgi Nalbantov, a proof was given that reliability information results in information gain. A combination of version spaces and support vector machines was used to illustrate this result.

On Thursday evening, there was a chance to meet colleagues at an informal drink. This was followed by an excellent conference dinner at the local faculty club.

DAY 2

The second day started with an invited lecture by Samy Bengio. He talked about the challenges of multichannel sequence processing, which plays an important role in applications involving multiple information sources. Two novel approaches were discussed: asynchronous hidden Markov models, for processing multiple asynchronous streams, and layered hidden Markov models, for decomposing large and complex multistream problems into layered architectures.

The invited speaker was followed by a session about evolutionary approaches. Marco Wiering, Filippo Mignogna and Bernard Maassen tackled the problem of forest fire control using evolving neural networks. Multiple cooperating agents are learned to stop a forest fire in a simulator. The generation and assignment of subgoals is done with evolving multi-layer perceptrons.

Coevolution, another topic in evolutionary learning, was discussed by Edwin de Jong. This method can be used to chose the tests for evaluating candidate solutions adaptively. More specifically, a new algorithm was introduced that guarantees monotonicity in maximizing the utility of a candidate solution.

The afternoon of this second day was reserved for some applications of machine learning. Michal Sindlar and Marco Wiering developed a method to analyze facial expressions. They use multi-layer perceptrons, trained on data from the mouth and eyes regions.

Andrea Malossini, Enrico Blanzieri and Raymond Ng tried to assess the reliablity of support vector machines for data analysis of microarrays. It turns out that systematically mislabeling examples can dramatically decrease the accuracy of the classifier.

Finally, Rutger Rienks, Ronald Poppe and Mannes Poel tried out some methods for speaker prediction based on head orientation. A naive Bayes classifier, neural networks and humans were compared. Apparently, the different classifiers do not generalize well over different meetings, possibly influenced by human specific gaze behaviour.

This concluded a successful 14th Benelearn conference. Altogether, it provided an excellent opportunity for Belgian and Dutch machine-learning researchers to meet their colleagues and be informed about current research activities in Belgium and the Netherlands.

Conference Homepage:
http://hmi.ewi.utwente.nl/conference/Benelearn2005

The three M's: Marco Wiering, Martijn van Otterlo, and Mannes Poel. Marco was the organiser of Benelearn 2002, the other two the organisers of Benelearn 2005.

Workshop Go at the Frontiers of AI

January 26, 2005
Universiteit Maastricht

Erik van der Werf
Maastricht

Since the founding years of Artificial Intelligence (AI) computer games have been used as a testing ground for AI algorithms. After several decades of research, AI algorithms have now reached world-champion level in various games. Go is the main exception. Go is by far the most complex popular board game in the class of two-player perfect-information games. It is played by an estimated 25 to 50 million players, in many countries around the world, and has received significant attention from AI research. However, unlike for games such as Chess, Checkers, Draughts, and Othello, there are still no Go programs that can challenge a strong human player and as such it has become the 'task par excellence' for AI game research.

On Wednesday January 26, 2005, a symposium was organised, in collocation with the Ph.D. defence of Erik van der Werf, to bring together AI researchers working on games to exchange ideas, present their latest findings, and push forward the state-of-the-art in computer Go. Five speakers each gave a 30-minutes lecture in front of a large international audience of researchers, Ph.D. students, master students and Dutch Go enthusiasts (who had picked up the news of this event through the website of the Dutch Go association and a radio interview with the author by VPRO's Noorderlicht).

The first speaker was Jacques Pitrat of the Université Pierre et Marie Curie, Paris, France. His lecture, titled *Games, rules and meta rules*, discussed new ideas on how to tackle difficult problems at a more abstract 'meta' level compared to the traditional tree-search approach. Several interesting examples were shown where abstraction and meta-reasoning could be used to solve problems that would be extremely difficult to tackle by traditional tree search. Although it is not directly apparent whether the examples given by professor Pitrat would not have been solvable using more up-to-date search techniques, such as PN* or DF-PN, the lecture gave rise to a lively discussion and initiated several interesting ideas for future research.

The second speaker of the day was Tristan Cazenave of the Université Paris 8, Paris, France. His lecture, titled *Solving open life and death problems*, presented research on developing effective means to determine life and death in so-called open positions, which are the most common case in real games. Cazenave presented his algorithm for generalized widening in detail and compared it to several related algorithms. Experiments indicate that generalized widening combined with alpha-beta and enhancements works quite well. However, there still remains much work to be done, especially since there are several alternative techniques, such as DF-PN, for which it is not yet clear how their performance compares.

After a 15 minutes break the symposium continued with the third speaker, Alex Meijer of the TU Delft, Delft, The Netherlands. He gave a lecture with the double-title *Formalizing the Go Player's Mind or Multi-Objective Game Theory*. The main emphasis of this lecture was on extending traditional game theory to make it more practically applicable for achieving multiple dependent as well as independent sub-goals. Although the new theories are quite promising the question of how to automatically identify the right goals, and possibly establish an order of importance for these goals remains a fruitful area of future research.

The fourth speaker of the symposium was Bruno Bouzy of the Université Rene Descartes, Paris, France. Bruno Bouzy is the author of the Go program INDIGO, which in recent computer Go tournaments has progressed from a mediocre player to a serious contender in the sub-top. In his lecture, titled *Toward a statistical approach of computer Go*, an overview was presented of the new Monte-Carlo techniques used by INDIGO. These techniques are the main reason for the recent successes and most notably the Bronze medal in 19×19 Go at the 2004 computer Olympiad in Ramat-Gan, Israel. Several improvements of the traditional Monte-Carlo approach were discussed, including the use of pseudo-random games, and pre-computing statistics for the opening. Future work will explore the use of learning techniques as well as further exploitation of parallelism.

The last speaker of the day was Erik van der Werf of the Universiteit Maastricht, Maastricht, The Netherlands. His lecture, titled *Inside Magog*, presented several ideas and techniques used by the Go program MAGOG. Compared to other top Go programs MAGOG is quite young. Most of the strong Go programs have started over a decade ago. Yet, despite of this the program has already been quite successful, most notably by the bronze medal obtained in the 9×9 Go tournament of the 2004 Computer Olympiad in Ramat-Gan, Israel. The presentation revealed valuable insights on how to combine search and knowledge, the evaluation

function, quiescence search, and the use of realisation probabilities. Further, it underlined the importance of the move predictor, which is used to search with a bonus, and debunked the myth of MAGOG being a pure search-based program.

The last event of the day was a lively panel discussion with the audience. It started with the question whether computer Go could reach world-champion level before 2025. This date was selected, because the author believed that under ideal circumstances progress of one stone per year could be possible, thus leading to 9 dan level around 2025. However there was a general consensus among the members of the panel that such ideal circumstances are quite unlikely and the question was quickly rephrased to the question in which year computers would reach world-champion level. In the following discussion there remained to be a huge difference between the pessimistic and the more optimistic estimates, which can be summarized as: probably not before 2040 and maybe not even in this century. With such estimates it is clear that computer Go will remain to be one of the most challenging areas of AI game research for many years to come.

Three Ph.D. Thesis Summaries

Techniques for Improving the Efficiency of Inductive Logic Programming in the Context of Data Mining

Ph.D. Thesis by Jan Struyf
Katholieke Universiteit Leuven

Summary

DATA MINING

Companies and other organizations collect vast amounts of data in the hope to obtain from these data the knowledge that would help them improve the processes that have generated the data. A pharmaceutical company could for example collect various properties about the chemical compounds they use in drugs. Such data contains *implicit* knowledge about the compounds, which can, once extracted, be of great value to the company. *Data mining* systems are able to automatically extract such useful knowledge from the available data. The extracted knowledge can be represented as rules or more general models. For example, a data mining system could extract a rule predicting that a chemical compound is active with regard to a given disease if its molecular structure includes particular features. Having such knowledge can significantly reduce the development cost of new drugs.

INDUCTIVE LOGIC PROGRAMMING

Attribute-value data mining (AVDM) systems are frequently used in practice because they are efficient and easy to use. In AVDM, the data is represented in one single table where each instance is described by a fixed number of attributes. Many practical databases however require a relational representation with several tables. For example, to store the molecular structure of chemical compounds, one would use three tables: 'compound', 'atom', and 'bond'. To be able to mine relational data with AVDM, the different tables have to be summarized into one table, which may result in a loss of important information. *Relational* data mining (RDM) systems do not have this disadvantage as they can process relational data in its original format.

Most research concerning RDM has been situated in the field known as *inductive logic programming* (ILP), in which mining systems are developed that use first order predicate logic to represent the relational input data. ILP systems are more expressive, but also considerably slower than most AVDM systems. There are two main reasons for this: the hypothesis space searched for suitable models is larger in ILP than in AVDM and evaluating a single candidate model is computationally more expensive as well.

CONTRIBUTION

The thesis describes and evaluates a range of techniques that make ILP systems *faster* and more *scalable*. More specifically, it considers (1) query optimization techniques, (2) techniques for optimizing data representation and access, and (3) techniques for optimizing particular mining algorithms. Although these categories are rather orthogonal, similar ideas can be applied in each one of them.

Most ILP systems build models by searching a large space of candidate models. For example, to construct a suitable rule, an ILP system searches a large space of candidate rules, e.g., using a greedy search strategy. For each rule encountered during the search, a heuristic such as accuracy is computed. Rules with a high heuristic value are refined further in subsequent search steps, while clearly sub-optimal rules are discarded. To compute the accuracy of a given candidate rule, the system has to test the condition of the rule for each instance in the database. In ILP, the condition of a rule is a conjunction of predicate calls (a *query*). Answering such queries is known to be NP-complete, i.e., the runtime of all exact query execution algorithms

grows exponentially with the rule length. For long rules, which are for example typical for data sets storing molecules, using straightforward query execution algorithms quickly becomes computationally infeasible.

The thesis therefore evaluates several optimization techniques that speed up query execution. In a first chapter on this topic, four static query transformations are discussed: the *theta*, *cut*, *once* and *smartcall* transformation. Query transformations replace a given query by a new one that can be executed faster. E.g., by removing redundant components from the query and by avoiding unnecessary backtracking over independent components. For each transformation, both the reduction in query execution time and the transformation overhead are analyzed theoretically as well as experimentally. The thesis also investigates how the transformations can be best combined. An experimental analysis on artificial data sets with different complexity parameters shows that the combination *theta+smart+once* has the best overall performance.

A second chapter is devoted to a *reordering* transformation. This transformation places the components of a query, based on their selectivity, in the most efficient order. The chapter describes how the parameters of the transformation can be estimated, discusses a greedy approximation that works faster for long queries, and investigates how the reordering transformation can be combined with the four transformations discussed previously. An experimental evaluation on biochemical and other data sets shows that combining the transformations can result in a speedup of several orders of magnitude.

A third chapter discusses *multi-query* optimization techniques. Such techniques optimize the execution of a set of queries sharing common sub-expressions. The chapter describes how these techniques can be combined with the query transformations discussed before. This is non-trivial because the transformations may have a negative effect on the size of the common sub-expressions. In particular, the so-called query-pack multi-query technique is extended so that it can efficiently include transformed queries.

During model generation, ILP systems typically generate redundant candidate models, e.g., models that are logically equivalent to previously generated ones or models that can be proved to be sub-optimal apriori. Equivalence and subsumption tests can be used to remove such redundant candidates. However, these tests are computationally expensive and must be repeated for each candidate. The thesis therefore proposes a number of algorithms that speed up subsumption and equivalence tests by combining methods similar to the multi-query optimization techniques with hashing. An experimental evaluation shows that the proposed methods can significantly reduce the runtime of ILP pattern mining systems.

In some applications, the available data does not fit in main memory. It is then important to have (1) a space efficient representation of the data, and (2) algorithms that optimize data access on disk. The thesis proposes a new framework for representing data that has both properties. Specific to mining applications is the notion of 'instances' (e.g., the molecules, etc.). The ILP system will process each instance in turn while evaluating candidate models. A possible optimization therefore is to create a data block on disk for each instance including all relevant data. The blocks can then be processed one by one: load one block from disk, execute the queries on the block, remove the block from memory and proceed with the next one. A drawback is that such a representation leads to redundant storage if the blocks are not disjoint. The framework therefore extends this mechanism to include auxiliary blocks storing data that is common to several examples. It also provides methods to efficiently process data sets that store complex objects (e.g., molecules) and state based data sets (e.g., data sets occurring in agent technology). An experimental evaluation shows that the framework allows a compact representation of the data that can be mined efficiently by ILP systems.

Finally, a technique is proposed that reduces the computational overhead of cross-validation. N-fold cross-validation is a mining technique that is frequently used to estimate the predictive performance of a model. It is however computationally expensive because the mining algorithm has to be executed N times. The thesis proposes an efficient cross-validation algorithm for decision trees that is inspired on the multi-query optimization techniques discussed before and avoids repeated construction of common parts of the cross-validation trees. For some data sets, the method significantly reduces the overhead of cross-validation. It can also be extended to rule set cross-validation and to bagging.

CONCLUSION

Knowledge becomes increasingly important in the information age. Data mining techniques can extract such knowledge automatically from given data. Data sets with a complex, relational structure require the expressivity offered by ILP systems. This expressivity comes however at the price of an increased computational cost, resulting in some

cases in excessive runtimes. Jan Struyf therefore investigates in his Ph.D. thesis a number of techniques that are able to reduce the runtime of ILP systems. The work includes query optimization techniques, techniques optimizing data access and a case study of a technique optimizing a data mining algorithm. The experimental evaluation shows that the proposed techniques significantly reduce the execution time of ILP systems so that it becomes practically acceptable, even for complex biochemical mining tasks.

The thesis can be downloaded at:
http://www.cs.kuleuven.ac.be/publicaties/doctoraten/cw/CW2004_12.abs.html

This research was funded by the Fund for Scientific Research (FWO) of Flanders.

Methodological Aspects of Designing Induction-based Applications

Ph.D. thesis by Floor Verdenius
Universiteit van Amsterdam

SIKS Dissertation Series No. 2005-1.
ISBN 90-6754-825-1

Summary

This thesis considers methodological support for industrial application of machine learning techniques. Machine learning (ML) research has delivered a large number of promising techniques for tasks such as classification and numerical prediction. Since the mid-eighties, application of ML techniques has gained much interest, both from academic researchers and from industrial workers. ML techniques that build models from exemplary data are considered *ready for real-world application,* at least from the technical point of view.

The models that result from ML application can be incorporated within knowledge based systems. In this mode, ML techniques perform knowledge acquisition and knowledge maintenance for specific inferences. Alternatively, to obtain the same knowledge models, human expertise is elicited by knowledge engineers. When the models are also automatically updated, enabling the system to function in a changing environment, the system becomes adaptive. Another mode of ML application is as data analysis tool. The goal then is to detect patterns in data. The resulting models are exploited for exploring regularities in a domain and, after validation of the models, also for human problem solving. This application has become known under the headings of Data Mining and Knowledge Discovery in Databases.
As the research focus to date has mainly focused on technological issues, the need arises to rationally integrate the resulting techniques in real world applications, and to select, given a specific task, the best technique to perform that task. This thesis starts with a discussion of properties of process models that support the application of machine learning techniques. Numerous process models are available that support the development of information and knowledge systems in software engineering and knowledge engineering. Typical process models consist of an activity structure and tools that support analysis and design steps.

A number of ML design and analysis process models are proposed in literature. The existing process models focus on the mere application of learning techniques in an exploratory data mining context, sometimes related to one specific technique group, such as neural networks. All these approaches start from the assumption that the final application consists of learning techniques. In real-world situations however, functional (and financial) considerations dominate. Unfortunately, there is no process model available that considers the application of learning techniques in the design process of a complex embedded system, where the deployment of learning techniques has to be considered against alternatives.

In spite of a growing number of reported learning applications in literature, the current application practice is unknown at the beginning of this work. To identify bottlenecks in the application practice and to determine the application requirements of industrial practitioners, a survey of ML applications is organized in 1994. The survey addresses industrial companies, research companies and research institutes in the Netherlands.

The survey results show that ML techniques are often exploratory, and driven by technological availability. The application practice is ad hoc, and makes hardly any use of existing ML tools. Moreover, only a fraction of the available techniques is actually used in practice. Respondents indicate a need for pragmatic support in technique selection and efficient production of working solutions. To update our 1994 results with more recent data, the survey is complemented with recent results of a poll on the KDNuggets Internet site. These results suggest that, though the scale of application has changed (especially due to the success of data mining and knowledge discovery over the last decade), the main results of the 1994

survey still hold, except that tool usage has increased substantially.

Three main questions arise from the results of the survey:

- **Overall**: *Can a process model be formulated that both considers the design issues for machine learning modules as well as the integration of the machine learning module in the embedding system?*
- **Top Down**: *How to decompose a complex functional requirement so that the potential use of ML techniques is recognized and they are properly integrated in the ultimate design?*
- **Bottom Up**: *How to select, as a part of the design process, an appropriate ML technique for realizing the learning behavior as specified in the design?*

Two cases illustrate these issues. The Product Treatment Support System (PTSS) is a planning aid for warehouse and quality managers that store and treat agricultural produce over a long period of time. The system helps them to design batch specific treatment condition recipes (temperature, relative humidity, and gas conditions). In the system, statistical estimation, decision trees, neural networks and constraint satisfaction modules co-operate to match the performance of human experts in the domain.

In the wastewater domain two systems are discussed that represent different ways of incorporating knowledge and ML techniques. The Super Charger Scheduler combines a case-based reasoning approach with an indexing mechanism based on self organizing maps. The system, offering a new, previously non-existent functionality, could not exploit existing knowledge. Moreover, a fast response time and reliable output is more important than fully comprehensible results. The Water Quality Simulation Module on the other hand assembles complex simulation models from partial sub-models on the basis of a high-level model specification. Sub-models, with strict pre- and post-conditions, have to be handled properly. In the assembling process, knowledge on the actual configuration of the plant and on the functionality of the sub-models is used. For this domain, comprehensibility is a crucial factor for acceptance by domain experts. Moreover, the body of available knowledge for this system has been substantial.

The survey results and the experiences with these two systems illustrate the need for an explicit activity to support the mastering of the two signaled design problems: handling available data and knowledge sources properly, and identifying, from the functional perspective, the opportunities for applying ML techniques. The Method for Designing Induction-based Applications (MEDIA) makes the required functionality foremost in designing machine learning applications.

In the MEDIA model the design process of complex systems is subdivided in three 'levels'. At the *application level,* the system functionality is decomposed. Potential inductive components are identified based on available knowledge and data sources. These inductive components are further designed and implemented in the next two levels. For the design of the knowledge intensive modules, MEDIA relies on existing tools and approaches that have been developed in the knowledge engineering framework, e.g., within CommonKADS. At the *Data Analysis level,* selection of appropriate techniques is supported. At the *Technique level,* tuning of the technique design and technique parameters optimizes the performance of the selected learning technique. Technique optimization guidelines are normally provided by the developer of specific techniques, and not in the MEDIA framework.

The task decomposition that is assumed in the application layer of the MEDIA model can be put into practice with the approach that is developed in Chapter 6. ML components are combined with knowledge-based components dependent on the available data and knowledge sources in the domain. The cost and the potential quality of the components guide the decomposition of the high level task. At any step in the decomposition, the basic consideration is whether it is worthwhile to build the component as a learning module, as a knowledge-based module, or to decompose it further, making use of available sources. With the PTSS system, the method is illustrated to derive a solution that balances cost and quality. The resulting decomposition also has a better cost-quality ratio than a pure ML solution or a pure knowledge-based solution. The approach makes use of estimated cost and quality figures for both knowledge-based and machine learning components. Although an outline of estimators is sketched as an extension of existing software engineering process models, further work is required in this area.

The second tool that is developed within the MEDIA framework supports technique selection, to be used at the data analysis level. One of the outcomes of the survey is that in practice techniques are not selected on 'technical' criteria, but on the basis of availability and other non-functional arguments. The general framework of *guarded technique selection* implies a fundamental shift in

how technique selection is conceived. Guarded technique selection (GTS) assumes that techniques (or groups of functionally closely related techniques) are 'supervised' by dedicated software components, the *guards*. A guard checks whether a data set matches with the type of patterns that is suited for its master technique. The result is a *degree of suitability*. Collecting all degrees of suitability gives a selection of techniques that are 'most suitable' for constructing a model from the data set.

This concept is illustrated by constructing the guard for *orthogonal techniques*. Orthogonal techniques are those learning techniques that construct concept descriptions based on class boundaries orthogonal to one attribute axis. This group of techniques includes some of the more commonly applied ML techniques such as univariate decision tree learners. By analyzing the types of class boundaries, and matching these with typical patterns in the entropy behavior of data sets, it is found that opportunities to apply orthogonal techniques can be identified without exhaustive comparative analysis of the data set with all available ML techniques. Within the assumptions that underlie the used prototypes, the identification of orthogonal concept types functions substantially better than the meta-learning alternative of MetaL. Combined with other analytical instruments, prototype matching can help to rationalize the process of technique selection in the near future.

With the MEDIA model, the rational decomposition approach and the guarded technique selection, this thesis delivers a set of coherent tools to support the industrial application of machine learning techniques. The MEDIA model helps practitioners to structure the development process. It changes the commonly applied explorative approach to a more goal-oriented approach, and at the same time offers the opportunity to integrate, in the design phase, knowledge-based and learning components. Finally, this work leads to a more comprehensible selection of ML techniques for a task. The proposed guarded technique selection provides a sound theoretical foundation for technique selection where it formerly was based on more arbitrary considerations. Moreover the maintenance and re-engineering of learning components is facilitated by this approach.

Using Generative Probabilistic Models for Multimedia Retrieval

Ph.D. thesis by Thijs Westerveld
Universiteit Twente

Enschede: Neslia Paniculata CTIT series, ISSN 1381-3617; No. 04-67 Thesis Enschede.
ISBN 90-75296-13-4

Summary

This thesis discusses information retrieval from multimedia archives, focusing on documents containing visual material. We investigate search and retrieval in collections of images and video, where video is defined as a sequence of still images. No assumptions are made with respect to the content of the documents; we concentrate on retrieval from generic, heterogeneous multimedia collections. In this research area a user's query typically consists of one or more example images and the implicit request is: "Find images similar to this one." In addition the query may contain a textual description of the information need. The research presented here addresses three issues within this area.

First, we show how generative probabilistic models can be applied to multimedia retrieval. For each document in the collection a probabilistic model is built. For each of these models we then compute the probability that the query is generated from the model and the documents corresponding to the models with the highest probability are shown to the user. The assumption is that these are the most relevant documents, i.e., those with characteristics corresponding to the query characteristics. Visual information is modelled using Gaussian mixture models and information derived from language (e.g., the speech of the video soundtrack) is modelled using statistical language models.

The second issue addressed is the parallel between the use of generative probabilistic models for multimedia retrieval and comparable models for text. This thesis describes how the techniques developed for language relate to the multimedia techniques presented here and how these parallels can be leveraged.

Third, this thesis studies evaluation. We tested different model variants using a number of collections including the test collections of TRECVID, the international workshop series for benchmarking video retrieval. On average, language-based approaches outperform approaches based on visual information. However, for some

queries visual information is important. A combination of both modalities gives the best results when searching a heterogeneous multimedia collection.

De Connectie

New AI Magazine *De Connectie*

Ella Keijzer
Universiteit Utrecht

De Connectie is a new nationwide Dutch magazine on AI for students and other interested people. *De Connectie* is spread amongst all students studying AI at the VU, UvA, RuG, UU and UM. The magazine is published four times a year and its aim is to provide an overview of the broad researchfield of AI. It offers background stories on AI-related subjects, interviews with AI-related people and it keeps you informed on the latest developments in AI research in the Netherlands and perhaps even on international research. Every issue will have an interesting theme to it, which will be elaborated in various ways. There will be students writing on their internship, interviews with experts, articles on applications of AI in business and there will be 'popular-scientific' articles by researchers. *De Connectie* has started cooperating with the *BNVKI Newsletter* and we will try to contribute to every newsletter. Would you like to contribute to *De Connectie* with a column, an interview or an advertisement, or would you like to subscribe to *De Connectie*? Look for all the information at www.connectie.org or send an e-mail to redactie@connectie.org.

How to Assess a Ph.D. Thesis?

Jaap van den Herik
IKAT, Maastricht

Students expect to find deep ideas in Ph.D. theses. Once they have browsed through some theses they are at a loss. Does this thesis contain deep ideas? Does it contain any relevant idea or is it a genuine thesis of which many are published before? Writing a thesis is always a performance. If M.Sc. students are looking for material for their M.Sc. thesis and they dive into the text of a recently published Ph.D. thesis it frequently happens that they state: "is this all?" The true answer is: "yes, this is all, but for you it will be a fantastic job to achieve a comparable performance." Writing a Ph.D. thesis is not easy. Most Ph.D. students experience the truth of this statement as soon as they have started their Ph.D. research consisting of four stages. The *development of ideas* is the second stage, mostly overestimated in time by the fresh student. *Reading the literature* is the first stage preceding the stage of developing ideas, since it will be the basis for further research. However, the bulk of the work is in the third stage *experimentation* (or building a model) and in the fourth stage *writing*. The essence of an idea should pass these four stages successfully. If this happens the author can be proud and the proudness may be even greater if the author succeeds in letting the essence blink in the title of the thesis.

This announcement contains a list of five theses ranging from language via engineering to games. A really broad spectrum of AI research. The editors would like to congratulate the researchers with the substantiation of their ideas. The work is done, the real progress may start and the ideas may be disseminated. The *BNVK Newsletter* wishes you much success in reaching the next milestone.

Next to the five Ph.D. announcements, we publish three full abstracts of previously published Ph.D. theses. They provide a good insight into the ideas of the authors, Jan Struyf (*Techniques for Improving Efficiency of Inductive Logic Programming in the Context of Data Mining*), Floor Verdenius (*Methodological Aspects of Designing Induction-based Applications*) and Thijs Westerveld (*Using Generative Probabilistic Models for Multimedia Retrieval*). Moreover, there is a report by Erik van der Werf on a workshop titled *Go at the Frontiers of AI* preceding his Ph.D. defence on January 27, 2005. Our conclusion is a research question: can an AI program find the essence of a Ph.D. thesis?

Franc Grootjen (January 26, 2005). *A Pragmatic Approach to the Conceptualisation of Language*. Radboud Universiteit Nijmegen. Promotores: Prof. dr. ir. Th. P. van der Weide and Prof. C.H.A. Koster.

Zoran Stojanović (February 22, 2005). *A Method for Component-Based and Service-Oriented Software Systems Engineering*. TU Delft. Promotor: Prof. dr. H.G. Sol. Adjunct promotor: Dr. A.N.W. Dahanayake.

Nirvana Meratnia (February 23, 2005). *Towards Database Support for Moving Object Data*. Universiteit Twente. Promotor: Prof. dr. P.M.G. Apers. Assistant-promotor: Dr. Ir. R.A. de By.

Gabriel Infante-Lopez (April 6, 2005). *Two-Level Probabilistic Grammars for Natural Language Parsing*. Universiteit van Amsterdam. Promotores: Prof. dr. M. de Rijke and Prof. dr. R. Scha.

Pieter Spronck (May 20, 2005). *Adaptive Game AI*. Universiteit Maastricht. Promotores: Prof. dr. H.J. van den Herik and Prof. dr. E.O. Postma.

SECTION KNOWLEDGE SYSTEMS IN LAW AND COMPUTER SCIENCE

Section Editor
Marie-Francine Moens

Abstracts of the 17th Annual Conference on Legal Knowledge and Information Systems

December 8-10, 2004
Berlin, Germany

DIALECTICAL ARGUMENTS AND CASE COMPARISON

Bram Roth and Bart Verheij
Department of Artificial Intelligence, University of Groningen

Abstract. The basis of legal case-based reasoning is the doctrine of *stare decisis*: decisions in new cases should follow decisions in similar old cases. This paper takes as a starting point the 'case comparison' interpretation of the *stare decisis* doctrine. In this interpretation one establishes by case comparison which previously decided cases are sufficiently similar to a new case, after which the old conclusions are adopted in the new case. The paper shows how one can formally account for case comparison in terms of the dialectical arguments that cases give rise to. An innovation over previous work is that dialectical arguments are now formally defined, yielding a more transparent formal treatment of case comparison.

CONSTRUCTION OF A LEGAL ONTOLOGY FROM A EUROPEAN COMMUNITY LEGISLATIVE TEXT

Sylvie Despres and Sylvie Szulman
CRIP5 and LIPN, Université Paris, Paris

Abstract. The paper describes the construction of an application ontology around the concept of employee in a European community legislation text. Firstly, the terminae method is presented. Then it is shown how the text contributes to the building of this ontology. Afterward the use of a top level ontology as DOLCE and of a legal ontology as lri-core are detailed. The essential points of this study are: the application of the terminae method to the legal domain and its alignment to core ontologies.

INTERNET, PORTAL TO JUSTICE?

Tom van Engers, Radboud Winkels, Alexander Boer, and Emile de Maat
Leibniz Center for Law, University of Amsterdam

Abstract In the Netherlands people with low income are entitled to receive legal aid. In almost every big city in the Netherlands centres for delivering this legal aid are present. Currently these centres are reformed into a Legal Services Counter (*Het Juridisch Loket*). This reform process has several aims, including improving accessibility, improving the quality and standardisation of legal aid throughout the country, and improving efficiency and consequently diminishing operational costs. To achieve this, the organisation will be turned into a service oriented one and new technologies, like call centre technology and Internet will be used. The Leibniz Center for Law has developed an Internet based knowledge system that should support both future employees of the legal services counters as well as their clients. In the paper we present our experiences in designing and implementing this system in a changing organisation and the impact of the organisational change on the design team.

LETSUM, AN AUTOMATIC LEGAL TEXT SUMMARIZING SYSTEM

Atefeh Farzindar and Guy Lapalme
RALI, Université de Montréal / Quebec, Canada

Abstract. The paper presents work on the development of a new methodology for automatic summarization of justice decision. We describe LetSum (Legal text Summarizer), a prototype system, which determines the thematic structure of a judgment in four themes *Introduction*, *Context*, *Juridical Analysis* and *Conclusion*. Then it identifies the relevant sentences for each theme. We discuss the evaluation of produced summaries with statistical methods and also human evaluation based on jurist judgment. The results so far indicate good performance of the system when compared with other summarization technologies.

AGATHA: Automation of the Construction of Theories in Case Law Domains

Alison Chorley and Trevor Bench-Capon
Department of Computer Science, The University of Liverpool, Liverpool, UK

Abstract. Some recent accounts of reasoning with legal cases view reasoning with cases as theory construction. In the paper AGATHA (ArGument Agent for THeory Automation) system is described, which will automatically generate theories intended to explain a body of case law by following a process inspired by the style of argumentation found in case-based reasoning systems. Thus AGATHA behaves like a case-based reasoner, but has as its end product a theory, which can be examined, critiqued or input to our other tool, CATE, for refinement or to generate executable code.

XML Retrieval Models for Legislation

Marie-Francine Moens
Interdisciplinary Centre for Law and Information Technology (ICRI), *Katholieke Universiteit Leuven, Belgium*

Abstract. Legislation contains text-rich documents and is increasingly marked with XML tags. The XML markup can – among other uses – be exploited to more precisely answer free information queries. The paper reports on different XML retrieval models explicitly designed for the retrieval of legislation and which are based on the vector space model and the probabilistic language model. In addition search data structures are discussed for legislative databases that support these retrieval models. It is shown that the models provide more advanced access to the content of statutes.

Modelling Contracts Using RuleML

Guido Governatori and Antonino Rotolo
School of ITEE, The University of Queensland, Brisbane, Australia and CIRSFID and Law Faculty, University of Bologna, Bologna, Italy

Abstract. The paper presents an approach for the specification and implementation of e-contracts for Web monitoring. This is done in the setting of *RuleML*. We argue that monitoring contract execution requires also a logical account of deontic concepts and of violations. Accordingly, *RuleML* is extended to cover these aspects.

Models of "*Novelle*" and Normative Grammar

*Andrea Bolioli**, Luca Dini**, Pietro Mercatali*, and Francesco Romano**
** ITTIG-CNR, Florence, Italy and**CELI S.r.l., Turín, Italy*

Abstract. The paper presents a method for mining information within legal texts, in particular in regards to corpora of statutes. Text mining, or more in general Information Extraction, can provide valuable help to people involved in research about the linguistic structure of statutes, and, as a side effect can be the seed for a new generation of applications for validation and conversion in the legislative domain.

Sentence Classification Experiments for Legal Text Summarisation

Ben Hachey and Claire Grover
School of Informatics, University of Edinburgh

Abstract. The paper describes experiments in building a classifier which determines the rhetorical status of sentences. The research is part of a text summarisation project for the legal domain which uses a newly compiled and annotated corpus of judgments of the UK House of Lords. Rhetorical role classification is an initial step which provides input to the sentence selection component of the system. We report results from experiments with four classifiers from the Weka package (C4.5, naïve Bayes, Winnow and SVMs). We also report results using maximum entropy models both in a standard classification framework and in a sequence labelling framework. The SVM classifier and the maximum entropy sequence tagger yield the most promising results.

Reinterpreting Arguments in Dialogue: an Application to Evidential Reasoning

Floris Bex and Henry Prakken
Institute of Information and Computing Sciences, Utrecht, University, The Netherlands and Faculty of Law, University of Groningen, The Netherlands

Abstract. The paper presents a formalisation of two typical legal dialogue moves in a formal dialogue game for argumentation. The moves concern two ways of reinterpreting a general rule used in an argument, viz. by 'unpacking' and 'refining' the rule. The moves can be made not only by the user but also by the attacker of the rule, in order to reveal new ways to attack it. The new dialogue

game is illustrated with examples from legal evidential reasoning, in which these types of moves are particularly common.

A CASE STUDY ON AUTOMATED NORM EXTRACTION

T.M. van Engers, R. van Gog, and K. Sayah
Leibniz Center for Law, University of Amsterdam, Dutch Tax and Customs Administration/Centre for Process and Product development and University of Utrecht

Abstract. Within the (E-)Power research program a new approach for supporting the chain of processes from the creation of legal texts to the implementation of normative (juridical) information systems has been developed. According to this approach creating formal knowledge models starts with the analysis of the legal text. This process executed by knowledge analysts is very time consuming. Within the Power program an automated concept and norm extraction tool and a model generation tool using linguistic techniques have been developed to improve modelling productivity. The paper describes how this tool can be used to translate legal texts into a formal model (in our case the UML/OCL standard). This is done by giving some examples of specific legal constructs and discussing the modelling process.

"MURDERED BY PERSONS UNKNOWN" – SPECULATIVE REASONING IN LAW AND LOGIC

J. Keppens and B. Schafer
University of Wales Aberystwyth, Department of Computer Science, Aberystwyth, Ceredigion and University of Edinburgh, School of Law, Joseph Bell Centre, Edinburgh

Abstract. The paper discusses some features of evidentiary reasoning in law, more specifically the introduction of "alternative suspects" by defence solicitors, within the framework of a first-principle-(or model-)based decision support system for crime investigation. The work introduces some observations from legal reasoning and legal doctrine to argue that the existing approach may not be able to capture some important distinctions and concepts of legal reasoning. The paper then introduces some ideas from dynamic logic and update semantics, in particular the concept of "pegs", as good candidates to add the necessary expressive power to model-based reasoning systems.

A CONTENT MANAGEMENT SYSTEM BASED ON AN EVENT-BASED MODEL OF VERSION MANAGEMENT INFORMATION IN LEGISLATION

Alexander Boer, Radboud Winkels, Tom van Engers, and Emile de Maat
Leibniz Center for Law, University of Amsterdam

Abstract. Understanding electronic legislation requires a comprehension of its lifecycle and the events and corresponding transformations that make up this lifecycle. A metadata description attached to the legislative resource represents a time point in this lifecycle. The paper advocates an event-based model of version management for legislation. The research group currently develops a Content Management System (CMS) based on this model. The paper discusses the event-based model, sketches the infrastructure of the CMS, and relates it to timestamping in ^{META}Lex documents.

DIRECT: ONTOLOGY-BASED DISCOVERY OF RESPONSIBILITY AND CAUSALITY IN LEGAL CASE DESCRIPTIONS

Joost Breuker and Rinke Hoekstra
Leibniz Center for Law, University of Amsterdam

Abstract. The paper describes DIRECT, a system for automatic discovery of responsibility and causal relations in legal case descriptions based on LRI-Core, a core ontology that covers the main concepts that are common to all legal domains. These domains have a predominant common-sense character – the law is still for the people – and typical legal concepts such as norm, role responsibility, contract etc. have a grounding in abstract common-sense conceptualizations. The paper presents a number of design principles that follow from the common-sense stance in developing the LRI-core: the most important being cognitive plausibility. Furthermore, an approach is presented to enable the automatic analysis of cases described in terms of the ontology. Such analysis will determine causal chains in situation descriptions on the basis of which responsibility attribution can take place.

THE SEKT LEGAL USE CASE COMPONENTS: ONTOLOGY AND ARCHITECTURE

V.R. Benjamins, J. Contreras, M. Blázquez, L. Rodrigo, P. Casanovas, and M. Poblet
International Software Components, Madrid and Law and Technology Institute, Autonomous University of Barcelona

Abstract. SEKT stands for Semantically Enabled Knowledge Technologies. Using previous and recently accomplished work on judicial and transnational lawyering prototypes (IURISERVICE and NETCASE projects), we define an ontology of professional legal knowledge (OPLK) as a regular base for a multilayered architecture. The main idea is to build an iFAQ to convey practical legal knowledge from more experienced judges to younger ones in their first appointment. This must be considered a preliminary or first approach. The ontology is still under development.

ANNOUNCEMENTS

Symposium Machine Learning for Commercial Game AI

May 19, 2005
IKAT, Maastricht

For decades, artificial intelligence in analytical games, such as Chess and Go, has been the subject of successful academic research. Only recently this research has been extended to encompass artificial intelligence in commercial games, such as strategy games and roleplaying games.

On May 19, 2005, the Institute for Knowledge and Agent Technology (IKAT) of the Universiteit Maastricht, in collaboration with SIKS, organises the symposium *Machine Learning for Commercial Game AI*. The symposium aims at presenting state-of-the-art research in artificial intelligence for commercial games. The focus is on the use of machine-learning techniques to improve the behaviour of agents in commercial games, in particular the agents' ability to deal intelligently with human player tactics.

Keynote speakers at the symposium are prof. Jonathan Schaeffer, of the GAMES group of the University of Alberta, Canada, and dr. David W. Aha, of the Naval Research Lab, Washington, DC.

To participate, please register with Joke Hellemons, email: hellemon@cs.unimaas.nl. Registration is free of charge.

For further details, please see http://www.cs.unimaas.nl/ikat/symposium_aicg/.

Call for Papers Workshop at the Annual Meeting of the Association of Computational Linguistics (ACL 2005)

June 29, 2005
Ann Arbor, Michigan

As experience with machine learning for solving natural-language processing tasks accumulates in the field, practitioners are finding that feature engineering is as critical as the choice of machine-learning algorithm, if not more so. Feature design, feature selection, and feature impact (through ablation studies and the like) significantly affect the performance of systems and deserve greater attention. In the wake of the shift away from knowledge engineering and of the successes of data-driven and statistical methods, researchers in the field are likely to make further progress by incorporating additional, sometimes familiar, sources of knowledge as features. Although some experience in the area of feature engineering is to be found in the theoretical machine-learning community, the particular demands of natural-language processing leave much to be discovered.

This workshop aims to bring together practitioners of NLP, machine learning, information extraction, speech processing, and related fields with the intention of sharing experimental evidence for successful approaches to feature engineering, including feature design and feature selection. We welcome papers that address these goals. We also seek to distill best practices and to discover new sources of knowledge and features previously untapped.

The workshop will include an invited talk by Andrew McCallum of the University of Massachusetts at Amherst.

SUBMISSION

Submitted papers should be prepared in PDF format (all fonts included) or Microsoft Word .doc format and not longer than 8 pages following the ACL style. More detailed information about the format of submissions can be found here: http://www.aclweb.org/acl2005/index.php?stylefiles

The language of the workshop is English. Submissions should be sent as an attachment to the following email address: ringger@microsoft.com. All accepted papers will be presented in oral sessions of the workshop and collected in the printed proceedings.

Submissions are invited on all aspects of feature engineering for machine learning in NLP. Topics may include, but are not necessarily limited to: Novel methods for discovering or inducing features, such as mining the web for closed classes, useful for indicator features; Comparative studies of different feature selection algorithms for NLP tasks.; Interactive tools that help researchers to identify ambiguous cases that could be disambiguated by the addition of features.; Error analysis of various aspects of feature induction, selection, representation; Issues with representation, e.g., strategies for handling hierarchical representations, including decomposing to atomic features or by employing statistical relational learning; Techniques used in fields outside NLP that prove useful in NLP; The impact of feature selection and feature design on such practical considerations as training time, experimental design, domain independence, and evaluation; Analysis of feature engineering and its interaction with specific machine learning methods commonly used in NLP; Combining classifiers that employ diverse types of features; Studies of methods for defining a feature set, for example by iteratively expanding a base feature set; Issues with representing and combining real-valued and categorical features for NLP tasks.

IMPORTANT DATES

Paper submission deadline:	April 20, 2005; Noon, PST (GMT-8)
Notification of acceptance:	May 10, 2005
Camera-ready copy:	May 17, 2005

ORGANIZATION

Chair and contact person:
Eric Ringger
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
ringger@microsoft.com

http://research.microsoft.com/~ringger/FeatureEngineeringWorkshop/

Call for Tutorials
WI'05 and IAT'05

September 19-22, 2005
Compiègne, France

The 2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence (WI'05) and Intelligent Agent Technology (IAT'05) takes place on September 19-22, 2005, Compiègne University of Technology, Compiègne, France.

HOMEPAGES

Web Intelligence (WI'05):
http://www.comp.hkbu.edu.hk/WI05/
http://www.hds.utc.fr/WI05
Intelligent Agent Technology (IAT'05):
http://www.comp.hkbu.edu.hk/IAT05/
http://www.hds.utc.fr/IAT05
WI'05 and IAT'05 will include tutorials providing in-depth background on subjects that are of broad interest to the Web intelligence and intelligent agent community. Both short (2 hours) and long (half day) tutorials will be considered.

The following is a *non-exclusive* list of preferred topic areas for tutorial proposals:

Web Intelligence:
Agents for Web Services; Intelligent/Semantic Web Services; Intelligent Wireless Web and Ubiquitous Computing; Knowledge Grids; Natural Language Processing for the Semantic Web; Ontology Engineering; Rules and Inference Engines for the Web; Semantic Web Concepts and Techniques for Security and Trust; Semantic Web Languages; Service-Oriented Computing; Social Networks and Social Intelligence; Web Mining, Farming and Reasoning

Intelligent Agent Technology:
Communication in Multiagent Systems; Learning and Evolution in Multiagent Systems; Cooperative Problem Solving; Auctions and Negotiation; Agent-Based Modeling and Simulation; Agent-Based Software Engineering Methodology; Agent Technologies in e-Business Systems; Agent-Mediated Knowledge Management; Peer-to-Peer Models for Multi-Agent Systems; Agent-Based Grid Computing

SUBMISSION DETAILS

Proposals for tutorials should consist of an outline and background information on the presenter(s). The tutorial outline should be limited to 2 pages and contain the following information:

- Title and abstract of the tutorial
- Proposed duration: 2 hours or half-day
- Intended audience: to whom is the tutorial of interest
- Prerequisite knowledge: what the attendees should already know
- Detailed outline

The background information on the presenter(s) should be limited to 1-2 pages and contain

- Names, affiliations, home-pages and contact details
- Short biographies
- Information about previous tutorials given by the same presenters (title, location, number of attendees, etc.)

Tutorial materials such as handouts and slides should be included if already available, but are not required for submission.

Please send your proposal to rineke@ai.rug.nl

IMPORTANT DATES

June 10, 2005: Tutorial submissions
June 25, 2005: Acceptance notices
August 10, 2005: Camera-ready copy of tutorial handouts

TUTORIAL CHAIR

Rineke Verbrugge, University of Groningen, The Netherlands. Email: rineke@ai.rug.nl

Call for Papers
5th International Conference on Computer and Information Technology (CIT2005)

September 21-23, 2005
Shanghai, China

AIM AND SCOPE

Information technology has been widely considered as one of the most important emerging technologies for the 21st century. It is already making great impact on our society, changing the way people work and live.

In 2005, as an important academic activity associated with the centenary celebration of Fudan University, this conference will serve as one of the major forums for researchers and practitioners from academia and industry to discuss and exchange ideas and opinions, and envision the future of computer and information technology. The technical program will consist of 9 technical tracks, preceded by tutorials.

The 2005 International Conference on Computer and Information Technology (CIT2005) will be held in Shanghai, China, September 21-23, 2005.

IMPORTANT DATES

Paper Submission Deadline: April 20, 2005
Acceptance Notification: May 30, 2005
Camera Ready Due: June 25, 2005
Author Registration: June 25, 2005

REFEREED PAPERS TRACK

CIT2005 seeks original papers describing research in computer and information technology within the following 9 technical tracks. databases track, web data processing track, cooperative information and grid computing track, networking and mobile computing track, multimedia processing track, information security track, embedded systems track, software engineering track, e-technologies track.

Submissions should present original reports of substantive new work. Authors must indicate which track (only one track) the submission is made to. Papers should properly place the work within the field, cite related work, and clearly indicate the innovative aspects of the work and its contribution to the field. We will not accept any paper that, at the time of submission, is under review for or has already been published or accepted for publication in a journal or another conference.

Papers will be peer-reviewed by at least 3 reviewers from an International Program Committee. The accepted papers have to be presented orally within 25 minutes including discussion at the conference and will be published in proceedings of CIT2005 by IEEE Computer Society Press (cited by EI, INSPEC, cross-searchable by SCI).

SUBMISSION GUIDELINES

Submissions should include an abstract, 5-7 keywords, the e-mail address of the corresponding author, and must not exceed 8 pages (IEEE format), including tables and figures, with PDF, PostScript, or MS Word format.

Submission of a paper should be regarded as an undertaking that, should the paper be accepted, at least one of the authors will attend the conference to present the work. Submissions to CIT2005 will be conducted electronically on the CIT2005 website. Authors unable to access e-mail may send 4 hard copies of their papers to:

Dr. Zhipeng Xie
Department of Computing and Information Technology
Fudan University, Shanghai, 200433, China
Email: cit2005@fudan.edu.cn
Tel: +86-21-55664497
Fax: +86-21-65643502
URL: http://www.cit.fudan.edu.cn/cit2005/

Call for Papers
17th Belgian-Dutch Conference on Artificial Intelligence

October 17-18, 2005
Brussels, Belgium

The BNAIC'05 venue will be held at the Royal Flemish Academy of Belgium for Science and the Arts in Brussels and is organized by the Computational Modeling Lab of the Vrije Universiteit Brussel and the Theoretical Computer Science Group of the Limburgs Universitair Centrum, under the auspices of the Belgian-Dutch Association for Artificial Intelligence.

The conference aims at presenting an overview of state-of-the art research in artificial intelligence in Belgium and The Netherlands. Submissions of the following three types are invited:

TYPE A: REGULAR PAPERS

Papers presenting new original work. Submitted papers should not exceed a length of 8 pages. These papers will be reviewed on overall quality and relevance. A-Papers will be accepted for either oral or poster presentation. All accepted papers will be fully published in the proceedings.

TYPE B: COMPRESSED CONTRIBUTIONS

AI papers that have been accepted after June 1st, 2004 at other refereed conferences or journals can be resubmitted and will be accepted as compressed contributions. Authors are invited to submit the officially published version (without page restriction) together with a one or two-page abstract. B-Papers will be accepted for either oral or poster presentation. The abstract of the paper will be published in the proceedings. Every author may submit at most 1 B-paper of which they are the corresponding author, and only if they do not submit any A-paper as corresponding author. Note that, as for BNAIC'04, a separate author registration is required for each B-type contribution.

TYPE C: DEMONSTTRATIONS AND APPLICATIONS

Proposals for demonstrations will be evaluated based on submitted demonstration summaries (in English) stating the following: the purpose of the system to be demonstrated, its user groups, the organisation or project for which it is developed, the developers, and the technology used. In addition, the system requirements and the duration (not exceeding 30 minutes) should be mentioned. Especially researchers from industry are encouraged to submit papers presenting their applications and experiences. The maximum size of demonstration summaries is 2 pages.

For all submission types, possible topics of submissions include, but are not limited to:
multi-agent systems; intelligent agents; robotics; logic in AI; games; search; verification and validation; logic programming; knowledge-based systems; knowledge representation; knowledge management; ontologies; machine learning; optimisation; evolutionary algorithms; neural networks; knowledge discovery and data mining; natural language processing; cognitive modelling; speech recognition; handwriting recognition; applications; AI in law, music & art

Papers and demonstration summaries should be submitted electronically. More details can be found at: http://como.vub.ac.be/bnaic2005

Submissions should be accompanied by a message stating the submission type (A, B, or C) and an abstract of the paper in plain text. Proper receipt of submissions will be acknowledged by e-mail. The deadline for submissions is June 15th, 2005. Submission implies willingness of at least one author to register for BNAIC and present the paper. For each paper, a separate author registration is required. Authors keep the copyright of their submissions.

Since last year, the BNAIC Proceedings carry an ISSN series number, just like journals, magazines and series of technical reports.

IMPORTANT DATES

Deadline for submissions: June 15th, 2005
Notification of acceptance: August 20th, 2005
Deadline for camera-ready papers: September 12th, 2005

CONFERENCES, SYMPOSIA, WORKSHOPS

Below, the reader finds a list of conferences and websites or addresses for further information.

APRIL 1, 2005
CABS colloquium Modeling Complex Multi-issue Negotiations Using Utility. Delft University of Technology, Delft.
http://cabs.ewi.tudelft.nl/Members/location

APRIL 2-10, 2005
8th Intl. Conf. on Fundamental Approaches to Software Engineering (FASE 2005). Edinburgh, Scotland.
http://fase05.disi.unige.it/

APRIL 4-6, 2005
IEEE Symposium on Computational Intelligence and Games. Essex University, Colchester, UK.
http://www.cigames.org

APRIL 11-13, 2005
10th Annual Euromedia Conference (Euromedia 2005). Toulouse, France.
http://biomath.ugent.be/~eurosis/conf/euromedia/euromedia2005/

MAY 4-6, 2005
ACM International Conference on Computing Frontiers (CF'05). Ischia, Italy.
http://cf05.ac.upc.es

MAY 10-14, 2005
The Fourteenth International World Wide Web Conference (WWW2005). Chiba, Japan.
http://www2005.org/

MAY 22-25, 2005
International Conference on Computational Science (ICCS 2005). Atlanta, USA.
http://www.iccs-meeting.org/

MAY 24-28, 2005
International Conference on Enterprise Information Systems (ICEIS-2005). Miami, USA.
http://www.iceis.org

JUNE 6, 2005
Legal Ontologies and Artificial Intelligence Techniques (LOAIT) Workshop. Bologna, Italy.
http://www.ittig.cnr.it/loait/loait.html#top
Held in conjunction with ICAIL-05
http://www.wogli.unibo.it/icail05/

JUNE 6-8, 2005
Worskhop Business Processes: Models, Examples and Purposes, Eindhoven University of Technology Eindhoven.

JUNE 6-10, 2005
Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005). Bologna, Italy.
http://www.iaail.org

JUNE 10, 2005
Fourth Worskhop on Law and Electronic Agents (LEA 2005) as part of ICAIL 2005. Bologna, Italy.
http://www.lea-online.net/

JUNE 13-17, 2005
The 17th Conference on Advanced Information Systems Engineering (CaiSE'05). Porto, Portugal.
http://www.fe.up.pt/caise2005/

JUNE 29, 2005
Workshop at the Annual Meeting of the Association of Computational Linguistics (ACL 2005). Ann Arbor, Michigan.
http://research.microsoft.com/~ringger/FeatureEngineeringWorkshop/

SEPTEMBER 19-22, 2005
The 2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence (WI'05) and Intelligent Agent Technology (IAT'05), Compiègne University of Technology, Compiègne, France.
http://www.comp.hkbu.edu.hk/WI05/
http://www.hds.utc.fr/WI05
http://www.comp.hkbu.edu.hk/IAT05/
http://www.hds.utc.fr/IAT05

SEPTEMBER 19-23, 2005
Ninth International IEEE EDOC Conference (EDOC 2005). Enschede, The Netherlands.
http://www.edocconference.org

OCTOBER 17-18, 2005
Seventeenth Belgian-Dutch Conference on Artificial Intelligence. Brussels, Belgium.
http://como.vub.ac.be/bnaic2005

NOVEMBER 10-12, 2005
International Symposium on Health Informatics and Bioinformatics (HIBIT'05). Belek, Antalya, Turkey.
http://hibit05.ii.metu.edu.tr

ADDRESSES
BOARD MEMBERS BNVKI

Prof.dr.ir. J.A. La Poutré (chair)
Centrum voor Wiskunde en Informatica
P.O. Box 94079
1090 GB Amsterdam
Tel.: + 31 20 592 9333. E-mail: Han.La.Poutre@cwi.nl

Dr. A. van den Bosch (secretary)
Universiteit van Tilburg, Faculteit der Letteren
Taal en Informatica, Postbus 90153, 5000 LE Tilburg
Tel.: + 31 13 4663117. E-mail: Antal.vdnBosch@uvt.nl

Dr. C. Witteveen (treasurer)
TU Delft, ITS
P.O. Box 5031, 2600 GA Delft
Tel.: + 31 15 2782521. Email: c.witteveen@its.tudelft.nl

Prof.dr. M. Denecker
Katholieke Universiteit Leuven
Dept. of Computer Science, Celestijnenlaan 200A
3001 Heverlee, België
Tel.: + 32 16327544. E-mail: marcd@cs.kuleuven.ac.be

Dr. C. Jonker
University of Nijmegen, Division Cognitive Engineering
Nijmegen Institute for Cognition and Information
Spinoza Building, Montessorilaan 3, 6525 HR Nijmegen
The NetherlandsTel.: +31 24 3612646.
E-mail: C.Jonker@nici.kun.nl

Dr. J.W.H.M. Uiterwijk
Universiteit Maastricht, IKAT
Postbus 616, 6200 MD Maastricht
Tel: + 31 43 3883490. E-mail: uiterwijk@cs.unimaas.nl

Dr. E.D. de Jong
Universiteit Utrecht, Inst. for Information & Computing Science
P.O. Box 80089, 3508 TB Utrecht
Tel.: + 31 30 253.9049. E-mail: dejong@cs.uu.nl

EDITORS BNVKI NEWSLETTER

Dr. J.W.H.M. Uiterwijk (editor-in-chief)
Universiteit Maastricht, IKAT
Postbus 616, 6200 MD Maastricht
Tel: + 31 43 3883490. E-mail: uiterwijk@cs.unimaas.nl

Prof.dr. E.O. Postma
Universiteit Maastricht, IKAT
Postbus 616, 6200 MD Maastricht
Tel: + 31 43 3883493. E-mail: postma@cs.unimaas.nl

Prof. dr. H.J. van den Herik
Universiteit Maastricht, IKAT
Postbus 616, 6200 MD Maastricht
Tel.: + 31 43 3883485. E-mail: herik@cs.unimaas.nl

M. van Otterlo, M.Sc.
University of Twente, Dept. of Computer Science
P.O. Box 217, 7500 AE Enschede
Tel.: + 31 53 4894111. E-mail: otterlo@cs.utwente.nl

Dr. M.F. Moens (section editor)
KU Leuven, Interdisciplinair Centrum voor Recht & Informatica
Tiensestraat 41, 3000 Leuven, België
Tel.: + 32 16 325383
E-mail: marie-france.moens@law.kuleuven.ac.be

Dr. J. van Looveren (editor Belgium)
Vrije Universiteit Brussel, AI Lab
Pleinlaan 2, 1050 Brussel, Belgium
Tel.: + 32 6293702. E-mail: joris@arti.vub.ac.be

Dr. R.J.C.M. Starmans (section editor)
Manager Research school SIKS, P.O. Box 80089
3508 TB Utrecht
Tel.: + 31 30 2534083/1454. E-mail: office@siks.nl

Ir. E.M. van de Vrie (section editor)
Open Universiteit Nederland, Opleiding Informatica
Postbus 2960
6401 DL Heerlen
Tel: + 31 45 5762366. Email: Evert.vandeVrie@ou.nl

HOW TO SUBSCRIBE

The BNVKI/AIABN Newsletter is a direct benefit of membership of the BNVKI/AIABN. Membership dues are € 40,-- for regular members; € 25,-- for doctoral students (AIO's); and € 20,-- for students. In addition members will receive access to the electronic version of the European journal *AI Communications*. The Newsletter appears bimonthly and contains information about conferences, research projects, job opportunities, funding opportunities, etc., provided enough information is supplied. Therefore, all members are encouraged to send news and items they consider worthwhile to the editorial office of the BNVKI/AIABN Newsletter. Subscription is done by payment of the membership due to RABO-Bank no. 11.66.34.200 or Postbank no. 3102697 for the Netherlands, or KBC Bank Veldwezelt No. 457-6423559-31, 2e Carabinierslaan 104, Veldwezelt, Belgium. In both cases, specify *BNVKI/AIABN in Maastricht* as the recipient, and please do not forget to mention your name and address. Sending of the BNVKI/AIABN Newsletter will only commence after your payment has been received. If you wish to conclude your membership, please send a written notification to the editorial office before December 1, 2005.

COPY

The editorial board welcomes product announcements, book reviews, product reviews, overviews of AI education, AI research in business, and interviews. Contributions stating controversial opinions or otherwise stimulating discussions are highly encouraged. Please send your submission by E-mail (MS Word or text) to newsletter@cs.unimaas.nl.

ADVERTISING

It is possible to have your advertisement included in the BNVKI/AIABN Newsletter. For further information about pricing etc., see elsewhere in the Newsletter or contact the editorial office.

CHANGE OF ADDRESS

The BNVKI/AIABN Newsletter is sent from Maastricht. The BNVKI/AIABN board has decided that the BNVKI/AIABN membership administration takes place at the editorial office of the Newsletter. Therefore, please send address changes to:

Editorial Office BNVKI/AIABN Newsletter
Universiteit Maastricht, Marlies van der Mee,
Dept. Computer Science, P.O. Box 616, 6200 MD
Maastricht, The Netherlands
E-mail: newsletter@cs.unimaas.nl
http://www.cs.unimaas.nl/~bnvk