HUNABC'98
First Hungarian National Conference on Agent Based Computing
Budapest, Hungary, May 29-31, 1998

The Role of Emotions in Multi-Agent Systems

Tamás Mészáros, Tadeusz P. Dobrowiecki
Dept of Measurement and Instrument Eng.
Technical University of Budapest
Mûegyetem rkp. 9., H-1521 Budapest, Hungary
meszaros@mmt.bme.hu, dobrowiecki@mmt.bme.hu
tel: (+36)-1-463-2057

Abstract

It is widely accepted that the introduction of emotions means necessarily an upgrading of the cognitive abilities of an agent system [2, 10, 11, 16]. Although everybody is aware that there is an essential gap between the artificial nature of an agent system and the human character of implemented emotions, this problem lacked a closer look, and some authors warned even against a too strict resemblance [4]. In this paper the authors, concluding that emotions mean a very useful information for deliberative systems equipped with perception, propose a new approach to broad emotional agents [16]. It is based on a strictly 'engineering stance' and aimed at a purposeful emotional system tailored to the agent's needs. Authors argue that system designers can profit from such stance but then they should accept and undertake certain nontrivial design decisions, leading to features not considered until now.

The paper deliberately does not address the issue of formal description. It attempts to clarify the problems and the principal design decisions in a possibly general way [9]. Analysis of representation is left after at least some of the open questions are worked out, considering that the choice of the representation and the concrete implementation will influence many particular decisions.

I. Prologue

'... Jo-2 arrived at the end of the page and Joker stopped him. Jo-2 would have done it by himself, but Joker lacked confidence. Too much of unexpected happened in this passage. Jo-1 and four Leo-s had been lost here, no wonder that Einstein is nervous and this influenced Joker. Six Leo-s were securing Jo-2's safety and Jo-3 was waiting in the room behind. In case of trouble Einstein's anxiety will turn into resentment and Joker wanted to avoid it. He will be responsible if he makes an error advancing swiftly, on the other hand Einstein will be impatient if progress is slow. And the command to open up this passage had been of course issued by Einstein...'

In the cited S-F story a few people, dedicated to save humanity on the verge of a nuclear disaster, established on planet Pluto a colony of 'polimorphic brains' [21]. Events interrupted and 'brains' were left untrained, without any knowledge and memory of the humans. As their programs evolved, ‘brains’ (more exactly their clones) got off toward the Sun and inner planets encountering finally people, who in the meantime, lacking energy to support technology, declined into an almost medieval conditions.

From the point of view of the discussion it is essential to note that the author equipped his 'brains' with a developed system of emotions, influencing their own and mutual behavior. However, it was a system significantly different from that of the humans. ‘Brains’ emotions were deeply rooted in their energy demands, and the coordinating Einstein, controlling the energy supplies, exercised an emotional leverage upon the others.

II. Emotions as the Source of Useful Knowledge

Relation between emotions and intelligence is obscure. What can be learned from psychology textbooks is that emotions influence thought processes, direct or influence behavior and urge people into actions [1, 15]. In the long-term emotions lead to new emotional manifestations, in the short term they control the interaction with the environment. To an engineer familiar with the concept of a dynamic system even such shallow notion suggests the existence of state, which then can be subjected to typical engineer tasks of 'state observation', 'state estimation' and the like, considering that the knowledge of the state yields information of how a particular system is going to behave or how it will respond to a contact.

Output = f (Input, State)

(Dynamic System Theory)

Emotion = f (External Events, Emotional State)

(Psychology)

Fig. 1. Analogy between dynamic and emotional systems.

An interesting question is that of the intensity of emotions. To a system observer the difference between an 'emotionally overheated' and an 'emotionless' individual appears primarily in the extent of how the reaction to the environment affects every aspect of behavior. If the emotions are intense, everything will be impressed, influenced by the state of mind. In general, the observer will qualify the observed system as being in a particular emotional state if the intensity of perceived emotion exceeds certain level. In consequence emotions serve as compact modeling language for systems of otherwise intractable complexity of structure and behavior [7, 9, 13].

Perceived emotions introduce knowledge about the state of the other system, which in turn makes the prediction of its actions more reliable. On the other hand showing emotions influences the 'knowledge base' of the observer, affecting thus indirectly also its own activities. Within such context emotions are used by humans in a honest or dishonest way, fooling, e.g., somebody with feigned emotions into reactions never otherwise anticipated. Human society values emotionally stable, i.e. predictably behaving individuals, emotionally 'difficult' ones are tolerated only if they do not endanger the group's interests.

III. Emotions in Agents

To perceive and demonstrate emotions is hence a valuable property, but what about software agents? An intelligent agent is an *autonomous* software system. Weak notion of agency [6, 19] equips it with concurrency, social contacts, reactivity, proactivity and the like. Secondary characteristics involve being e.g. rational and veritable. It is inevitable, that there is a strong emphasis on the communication and cooperation abilities.

Emotional agents until now were considered almost solely in the context of human-agent interactions. Introduced emotions meant at most an addition to user interface [2], influencing primarily not the agent's but the users' perception and behavior. Development of interconnected global information systems and the growing complexity of information processing tasks [18] shifts the attention towards the intense agent-agent interactions.

The research on multi-agent systems focused until now primarily upon the questions of logical models [20], protocols for communication and cooperation [5, 8] and deliberative and mixed architectures [3]. The question of emotions hasn't been considered, especially from the point of view of inter-agent usage and designing emotion-based protocols.

The cognition-based emotion model of Ortony, Collins, and Clore (OCC), widely used in broad agent development [16], is based on a human system of emotions. Although OCC system is heavily abstracted and some universal conceptions like goals, standards, social relationship can be identified, it is still a purely human system of emotions. Thus the principal problem considered in this paper is the confrontation of two seemingly incompatible concepts:

Emotional System (OCC and the like):

abstracted from systems sharing common physiology and developed in a lengthy and yet not fully understood evolutionary process;

Intelligent Software Agent:

computer system decomposable in terms of hardware and software, with the sole activity that of information processing, purposefully designed starting from a formal specification and designed using formally specified methods

The knowledge aspect of emotions is attractive and potentially useful also in the agent world. There are however concrete reservations voiced in the literature against the acceptance of the human emotion model in the whole [4]. There is, e.g., no need for agents (robots in the original) to bring up emotions (and thus reactions) in an uncontrolled way (although we observe panic, anger, etc. in humans and animals, there is no reason why say a stampede should be advantageous in purely software systems).

The concept of a believable agent found in the literature is strongly related to software agent - human agent interactions [2]. People however need not to be a necessary ingredient of the agent's environment, especially in the future (Fig.2).

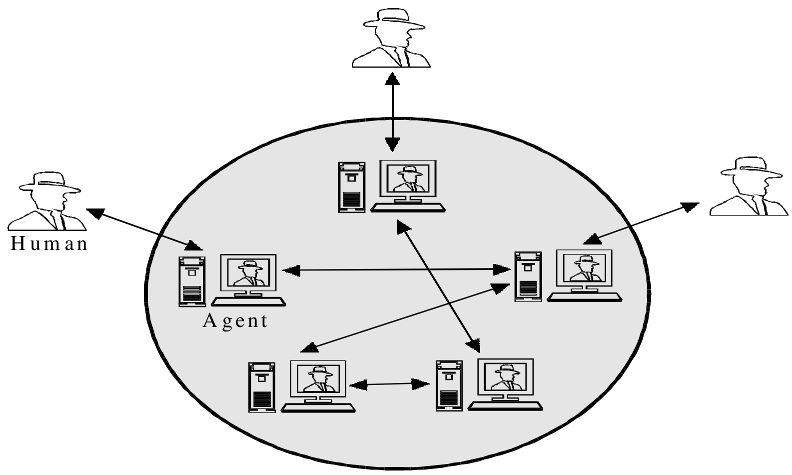

Fig. 2. A pure agent-agent and a mixed agent-human society.

When an agent is a part of a society composed solely of the like individuals, there is no need (one could argue) to involve a human-like emotional system at all. Situation will differ when an intelligent agent, by the nature of its services or its implementation, will communicate and cooperate also with humans. In this case some (but perhaps not necessarily all) of its emotions should be shared and thus have to be compatible in meaning for all of the involved parties (Fig.3.).

One of the aims of using emotions in agents is to make them believable, i.e. easily interpretable in human terms [2]. An easy interpretation of the state of the society, an intelligent agent is part of, makes perfect sense as a goal also when no human users are involved. Our question in the following is thus how to design emotions for software systems to make sense.

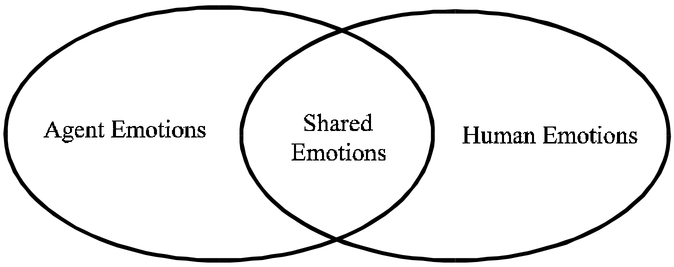

Fig. 3. Emotions in a mixed agent-human system.

Summarizing the above discussion we can say that (technically) emotion is a 'global state variable' and is used by humans as such. We assume the need for a similar descriptor in any sophisticated intelligent system involved in cooperation with similar entities. Systems can use 'emotion state variables' to:

- Get a feeling (i.e. a highly compressed information) of their own state for their own (internal cognitive) purposes, using to their advantage the compactness of such model,
- Pass such information to others, drawing from the advantage that this model is common to all, and that using it does not require the detailed knowledge of architecture, knowledge representation, etc. (emotional description can be used even as a meta-knowledge level protocol in information exchange among intelligent systems, particularly if they had been designed to different ends using different means),
- Influence directly other systems,
- Make communication with such system more palatable to the humans.

IV. Designing Emotions

A number of questions must be solved to design an emotional agent. We must deal in particular with the following issues:

- How to built emotional model, i.e. which emotions to define and how to relate them?
- What should be the source of emotion in case of a machine (program)?
- What are the possible emotional states of a finite resource system which supports certain services or accomplishes certain tasks? How the state of system's resources should influence its emotions? Is such emotional system enough for a machine?

McCarthy pointed that people do not need mechanical companions but rather ideal servicemen (not necessarily slavish servants but "servants notwithstanding") [4]. In our view this should be the basic design motive to create artificial emotional systems. **The aim is thus not to mimic human emotions, but to create emotions which will be useful also to the agents, which will simplify or make more effective the agent-agent or the agent-human communication.**

To this end we propose the following grossly oversimplified framework (similar elements can be found widely in the literature, e.g. in [7,8,14,18]):

Intelligent agent

intentions: **primarily to maintain service, i.e. to keep a continuous correct and reliable task processing capability as it had been specified in the agent design;**
attitudes: **agent likes every resource needed or facilitating the accomplishment of its intentions;**
emotions: **a kind of 'satisfaction' when its intentions are not endangered and 'uneasiness' otherwise, with subsequent typization**

Actions induced by emotions should be strongly oriented toward the cooperative behavior. Knowledge encapsulated in the observed emotions can be used consequently to initiate and support the dialogue and to direct activities in such way that the overall cooperativness of the team increases. This aim can be consistently achieved by a thorough evaluation of what resources are available to an agent.

Resources - things a normal agent likes:

direct (private) physical resources (memory, coprocessor, etc.)
direct (private) abstract resources (time, data, library of better methods, etc.)
remote physical resources (other systems available and ready for help)
remote abstract resources (knowledge distributed in the society, which can be fetched or learned)

What could be hence the 'chemistry' of the agent’s emotions? Everything, which disturbs the balanced spending of its resources, e.g. difference between the specified (normal) state and the actual situation (errors, loads, etc.), relation between the requested service and the available resources, concern of how to support a service in the actual state, relation between the internal and the external world (e.g. is the support of the other systems needed, is the cooperation at the actual moment to the system's advantage, is some support requested by the others, etc.), etc.. All these can give rise to 'excitement' and subsequently to the visible 'emotions'. Based on the above one can formulate some general design considerations:

1. **Language of emotions should be independent from the concrete realization of the agents. State information conveyed by the emotions should be universally valid.**
2. **System description produced by the emotional model should be easily processed and understood also by the humans. It should be compatible with the usual system state description.**
3. **The model should be suitable for agent's control.**
4. **Emotion model should be hierarchical, the basic emotions at a higher level should be easily interpretable by the other agents with minimum of information.**
5. **The model should be suitable to maintain the dynamic view of emotions.**

Resource-based emotional system makes it a suitable tool for both analysis and design of agent systems. Such emotions can be even identified in already existing, non-emotional systems to test their performance in an extremely compact way. Observed positive emotions will verify the system, lasting negative emotions signal e.g. design problems, actual formation of emotions reflects the efficiency of the service or the influence of the environment. The idea can be extended to self-verification and self-diagnosis: the running system checks its own emotions in a particular state and on this basis it forms conclusions and plans for further actions. Emotions can serve a similar role in system design. A proper shaping of emotions and activities induced by emotional states leads to the design of a 'good', 'balanced' system, reliable in a sense that it will be able to maintain high efficiency of its specified services.

The pursuit to unify emotions and make their language independent from the realization aspect serves the purpose of judging quickly the state (e.g. willingness to cooperate), even in case of highly sophisticated and diverse systems. If however a more detailed information is needed, i.e. when the background and the fine structure of emotions is investigated, a system dependent interpretation of emotions must be supplied or inferred from the profile of the particular system. Emotional framework in agents means thus (1) a common language, and (2) a system dependent interpretation, which will require realization of a specialized ‘emotion interpreter module’, similar to the subsystems foreseen to interface stand-alone agent to the communication environment [7,8].

How well a particular system can ‘read’ emotions of the others will depend upon multiple factors. Interpretation skill can be pre-designed and placed in the system’s knowledge base, alternatively the system can be equipped with a mechanism to learn the interpretation of have-not-seen emotions, e.g. based on the statistics and the effectiveness of communication.

V. Dimensions of Emotions

Human emotions and the emotions tailored to the needs of the artificially designed systems contrast fundamentally with respect to their polarity and their intensity. Humans demonstrate and understand both positive and negative emotions (see e.g. OCC system [16]), which both can vary in intensity.

In a purposefully designed artificial systems there is no place for the performance better than the nominal. System which can perform in a given moment services much over its nominal efficiency is simply badly designed and thus not a subject of the present discussion. Good system design (design trade-offs) should use up the available resources to the sensible (sub-optimal) limit. In consequence a well designed agent should be able to demonstrate even very intense negative emotions (e.g. during a serious break-down), its positive emotions, particularly those which in humans are related to the over-fulfillment of expectations, should be mild and blurred into a common ‘satisfied/ everything as usual’ state of affairs.

Viewing the OCC model from this perspective some emotions like joy, distress, hope, fear, disappointment, resentment, etc. can be sensibly interpreted in terms of increasing or mitigated load, actual and predicted resource shortages or team support. What would mean however pride, shame, admiration, pity, satisfaction and the like?

The primary structure of the agent’s emotions can be illustrated in a suitable emotion space (Fig.4.). Concrete dimensions of such space are of course disputable and possibly even application dependent. We feel however that the choice in Fig.4 is generic enough to give a rough and yet a correct view and to expose the principal differences of such emotional system, when compared to the humans.

Space of emotions can be spanned by three basic sources of agent's events:

(1) observed deficiencies in the agent’s own physical setting,
(2) problems coming from the environment, and
(3) a general feeling that even when there is nothing to blame, things are not going as expected, expressed e.g. by the perceived difference between expected and observed rate of progress.

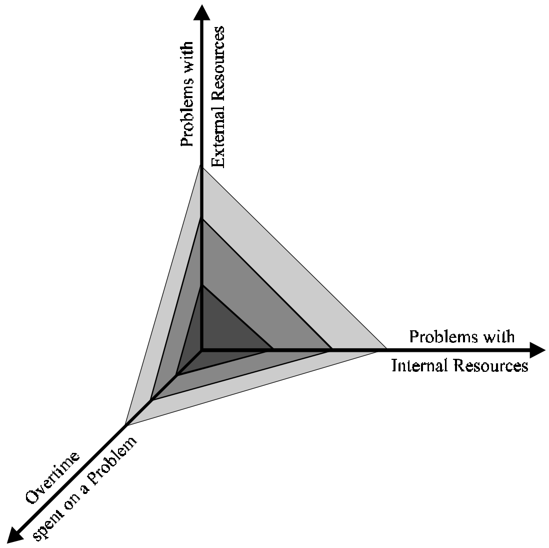

Fig. 4. Resource-based emotion space of an agent.

It is worth to observe that the distance from the origin is always related to the increasing level of stress and that positive emotional state is placed close to the origin. The mechanism to 'sense' emotions can be thus implemented as the measurement of a suitable metric and an agent can judge to what extend its state is 'pleasurable' or 'desirable' by measuring its distance from the origin. The principal emotions of the OCC model, mentioned earlier, can be strictly related to the actual and predicted movement in this space. For example:

	is felt when	
joy	the agent	can move towards the origin,
distress		is pushed away from it,
hope		does expect to be able to move toward the origin,
fear		does expect to be pushed away from it,
reproach	others	had promised but did not helped to ameliorate the agent's position,
disappointment		....
resentment		.....
gratitude		such promises had been fulfilled,
relief	the agent	hadn't been forced against expectation to move away from the origin,
fears-confirmed		it happened so,
etc.		

The primary emotions identifiable in this emotion space are sets rather than particular 'point-like' emotions and in this respect they resemble the OCC model. In agent systems we assume that the more fine emotions should assemble into hierarchy. Strict hierarchy is not visible in the human emotional model, possibly due to the complexity and the shallowness of its notions. In machine (software) world strict hierarchy works always to everybody's advantage.

Basic emotional categories at the highest level are decomposed into more application related details (including the reasons of a particular emotional state) at the lower levels of description. It is enough to perceive that an agent is 'nervous' or 'morose' to give up the attempt to communication with it, but when offering or asking for help the nature of the deficiency, an agent is struggling with, may be important and influence the protocol. Working at a particular level and wandering in the emotion hierarchy is related to the long term behavior of the agent. Top-to-bottom approach, i.e. considering more and more detailed emotions is like getting familiar, discovering the personality of the others. Agent wanting to forget, to detach itself will 'move' in abstraction bottom-up toward the more general impressions.

VI. Open Questions

The number of open questions is abundant and only some of them are addressed below.

Perception of emotions

Agents and humans do not share sensing organs. Even the more serious is that even if robots can feel and see, software agents have only communication channels to their disposal. Humans perceive emotions in passive (I see that he is all right) or active (I asked him) way. In passive perception the observed system is usually not concerned with the fact that his/her emotions are actually observed.

In agent world even ‘passively observed’ emotions must be actively broadcast by their holders. This broadcast must be concise and directed to the whole society. Active perception resembles more closely what the humans do. Agent evaluates the request and conveys emotional information in the style of a protocol and how the answer is wrapped. Primary problem is the abstraction level and the frequency of emotional broadcast.

Introducing emotions into KQML

The question of perception of emotions is interesting in relation to KQML, the first general, expandable agent communication protocol [5, 12, 13]. An emotionally charged performative called ‘sorry’ is already included. Next step would be the specification of a new common ontology to represent the emotional information and the introduction of new performatives for seeking and offering assistance.

Simultaneity of emotions

A particular agent can demonstrate many different emotions at a time. The global view of its state will be thus some function (weighted combination?) of all more or less intensive emotions. The strongest primary emotion (at the highest level) should be propagated first. In addition an agent should be able to process emotions separately or as a mixture to decide upon its subsequent activities.

Gruffness

Although particular in itself the implementation of gruffness presents an excellent opportunity for an agent to get rid of insistent but not contributing communication, without giving up the communication at all. It can be particularly useful for agents charged with critical tasks, or task with random loads, fearing more for the resources and keeping more resources in reserve.

Boredom

An interesting question is that of 'boredom'. Boredom in a machine, although surely not positive, should not be so negatively charged as in humans. There should be of course some mild compulsion against boredom, helping 'to contract employment' and to take over tasks from the others, but the lack of work should not be paired with 'going mad from being bored'. Learning (improving knowledge base) should be a good primary remedy to the problem (when services are not requested). Machine version of a hobby can also be considered as an activity basically useful but different from the services supported by the system.

Dynamics of emotions

Emotions evolve in time. Why not also those of the agents? Although difficult to realize emotions should be equipped with their own dynamics, relating the intensity of emotion to the time scale. There are stable emotions, strongly related to the state of affairs in the agent's world. Such emotions will change following the changes in the eliciting reasons. Other emotions should intensify or deteriorate as the time flows by. The intensity of 'boredom' should grow pressing agent into activity, 'cheerfulness' should disappeared if there is no fresh evidence that the state of affairs makes it reasonable, and so on. There is however problems with the time scale. We accept the dynamics of the human emotions as perceived. For finite-resource agents the dynamics of emotions must be consciously designed into the system and for a moment there is no clue regarding the scale, neither the dependencies. Problem is serious because agents represent a set of systems of much more diverse architectures than the humans.

Individual personal trait

Variances in emotion hierarchy and dynamics, due to the differences in design and representation, in close analogy to the human emotions, can be abstracted into different kinds of personality traits, diversifying also further protocols of inter-agent communication. Should it be a necessary but actually unwanted feature? If not, how a balanced agent society should be composed?

Safety

Negative emotions could be designed to enforce actions toward personal and group safety [14]. An intriguing question is whether the cultural universals of human emotional experience, abstracted after all only from humans, should apply to agents? What should be the ethical code of the purely agent society? Is the development of diversified 'agent cultures' a sound aim or not?

VII. Conclusions

Authors propose a different perspective on the emotional stance in agency. Emotional information represents a concise view of an intelligent system and its environment, and as such it should be implemented in agents even if they do not collaborate directly with the humans. The rationale of the proposed approach is the artifact nature of the agents, where the governing concern is to keep the specified and designed services going. Such systems can be described in terms of finite-resource emotions, which take the role of global state variables, related to the state of the system resources.

A number of emotions can be shared between agents and the human users, the interpretation of them in terms of events and reactions will however differ. Certain emotions can be easily identified in the agent's emotional space, spanned by the stresses, due to the problems with resources. Agent-tailored finite-resource emotions differ in general in structure from those perceived in humans. Primary difference is the downplayed role of positive emotions. On the other hand performance oriented definition of such emotions makes them into a tool, useful for the design and even analysis of distributed intelligent systems.

VIII. References

[1] R.L. Atkinson, R.C. Atkinson, E.E. Smith, D.J. Bem: *Introduction to Psychology*, 11th Edition, Harcourt Brace Jovanovich College Publ., Forth Worth, 1993.

[2] J. Bates: *The Role of Emotion in Believable Agents*, Communications of the ACM, Special Issue on Agents, July 1994

[3] K.S. Decker, V.R. Lesser: *Coordination Assistance for Mixed Human and Computational Agent Systems*, Umass Computer Science Technical Report 95-31, University of Massachusetts, Amherst, USA, May 18, 1995.

[4] J. McCarthy: *Making Robots Conscious of their Mental States*, in John McCarthy's Home Page at http://www-formal.stanford.edu/jmc/consciouness-2e.ps

[5] T. Finin, R. Fritzon, D. McKay, R. McEntire: *KQML - A Lanquage and Protocol for Knowledge and Information Exchange*, Technical Report CS-94-02, Computer Science Dept, University of Maryland

[6] S. Franklin, A. Graesser: *Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents*, technical report, Insitute for Intelligent Systems, University of Memphis, March 5, 1996.

[7] N.R. Jennings: *Specification and Implementation of a Belief-Desire-Joint-Intention Architecture For Collaborative Problem Solving*, Int. Journal of Intelligent and Cooperative Information Systems, 2 (3), 1993, 289-318.

[8] N.R. Jennings, J.M. Corera, I. Laresgoiti: *Developing Industrial Multi-Agent Systems.* , (Invited Paper), First International Conference on Multi-Agent Systems (ICMAS'95), San Francisco, CA., June 12-14, 1995, 423-430.

[9] N.R. Jennings: *A Knowledge Level Approach to Collaborative Problem Solving*, Proc. of the AAAI Workshop on Cooperation among Heterogeneous Intelligent, San Jose, USA, 1992, 55-64.

[10] N.R. Jennings: *On Being Responsible*, Decentralized AI. 3, (eds. E. Werner & Y. Demazeau), North Holland Publishers, 1992, 93-102.

[11] S. Kalenka, N.R. Jennings: *On Social Attitudes: A Preliminary Report*, International Workshop on Decentralised Intelligent and Multi-Agent Systems, Krakow, Poland, Nov. 22-24, 1995

[12] Y. Labrou, T. Finin: *A semantics approach for KQML - a general purpose communication language for software agents*, Proc. of the Third International Conference on Information and Knowledge Management (CIKM'94), November 1994

[13] Y. Labrou, T. Finin: *A Proposal for a new KQML Specification*, Technical Report TR-CS-97-03, University of Maryland Baltimore Country, Baltimore, Maryland, 1997

[14] S. E. Lander, V.R. Lesser: *Understanding the Role of Negotiation in Distributed Search Among Heterogeneous Agents*, Proc. of the International Joint Conference on Artificial Intelligence, August/September 1993.

[15] Oxford Keys. *English Dictionary*, Oxford Univ. Press, 1980

[16] W. Scott Reilly, J. Bates: *Building Emotional Agents*, Technical Report CMU-CS-92-143, School of Computer Science, Carnegie Mellon University

[17] D. Weld, O. Etzioni: *The First Law of Robotics (a call to arms)*, Proc. of the AAAI-94, 1994

[18] D. Weld (ed.): *The Role of Intelligent Systems in the National Information Infrastructure*, The American Association for Artificial Intelligence, http://www.aaai.org/Policy/Papers/nii.html, 1997

[19] M. Wooldridge and N. R. Jennings: *Intelligent Agents: Theory and Practice*, Knowledge Engineering Review, Vol 10, No 2, jun, 1995

[20] M. Wooldridge: *The Logical Modelling of Computatuinal Multi-Agent Systems*, Ph.D. thesis, Department of Computation, University of Manchester, United Kingdom, 1992

[21] Péter Zsoldos: *Ellenpont (Counterpoint)*, Kozmosz Könyvek, Budapest, 1973, (Fragment translated by the authors)