

Chapter 2

Towards the Integration of Multiagent Applications and Data Mining

Célia Ghedini Ralha

Abstract This chapter has the objective to present research on combining two originally separated areas, agents including distributed multiagent systems and data mining, which are increasingly interrelated. Recent research has present that such interaction features are bilateral and complementary, since new approaches and techniques are developed to benefit from the synergetic enhancement of intelligence and infrastructure for information processing and systems. This chapter draws attention to illustrate agent-mining interaction with two different domain multiagent applications: BioAgents at the bioinformatics area and MADIK at the computer forensics area. The presented case studies are driving forces towards the integration of the agent-mining challenging area. As ongoing research works we discuss the prospects of both agent-mining projects.

2.1 Overview of Agents/Multiagent Systems and Data Mining Integration

In the past decade, agents/multiagent systems (MAS) and data mining (DM)/knowledge discovery have emerged as two increasingly interrelated research areas, opening space to the agent-mining interaction and integration (AMII) research field. This new field has driven efforts from both sides to find benefits and complementarity to both communities. The AMMI field has prove to be so promising in recent years that there is a special interest group, entitled the Agents and Data Mining Interaction and Integration (AMII-SIG) , which aims to foster a forum for boosting the research and development on AMII studies [4].

In general terms, an intelligent software agent (ISA) uses Artificial Intelligence (AI) in the pursuit of goals [41, 56]. Thus, an ISA is a computer system capable

Computer Science Department, University of Brasilia (UnB)
P.O.Box 4466, Brasilia/DF, Zip Code 70904-970, Brazil
e-mail: ghedini@cic.unb.br

of autonomous action in some environment in order to meet its design objectives [62]. According to the given agent definition, we may cite that an agent is a computing entity with four features: autonomy, reactivity, interaction and initiative. ISA properties frequently cited in the literature include: mobility, veracity, benevolence, rationality and learning/adaption. In the literature, we find many definitions for a MAS, but mostly they agree as referring to a computational system composed by more than one agent [61, 62]. Thus, a MAS is a system where many agents interact with the environment in a cooperative or competitive way, to achieve individual or group objectives.

As we already defined, agents will be acting on behalf of users with different goals, motivations and to successfully interact, they will require the ability to cooperate, coordinate, and negotiate with each other. In summary, the agents intelligence and infrastructure for information processing and systems can enhance data and knowledge treatment. MAS technology provides a powerful method for building and managing DM in a distributed system. Agents autonomously exchange data with other agents and discovering their interesting data. All these characteristics are very important to the AMII research area.

This chapter presents research work related to the Applications and Case Study AMII topic, more specifically to the Emergent Agent-Mining Organizations and Applications topic, since it presents an overview of two SMA applications: (i) BioAgents, defined as a bioinformatics application and services topic and (ii) MADIK, an artificial system and service at Computer Forensics. Thus, this chapter draws a high-level overview of two different domain multiagent applications, with the intention to promote AMII from the practical perspective of an application-oriented approach. We also present some related publications to our research and do an evaluation of our ongoing research work, with open issues and prospects of both agent-mining projects.

2.2 Related Work

Since, the research topics in the area of AMII are quite diverse and ubiquitous, there are some works that might be related to all streams of the emerging area, which were presented at [27]: [15, 57, 19]. Giving attention to integration issues of MAS and DM we can cite [49, 43]. In [49] we find a model integration proposal *MultiAgent System for Distributed Data Mining - SMAMDD*, where agents perform mining tasks locally and merge their results into a consistent global model. In order to achieve that, agents cooperate by exchanging messages, aiming to improve the process of knowledge discover generating accurate results. A Web document database integration technique is presented in [43], where mining agents were defined to information extraction from HTML files.

Considering the MAS application domains, we now present related work to Bioinformatics. The work of Santos and Bazzan [21, 26], deals with knowledge discovery and data mining. In this work, authors propose a MAS to the Bioinformatics area, where agents are responsible for applying different machine learning

algorithms and using subsets of the data to be mined, and are able to cooperate to discover knowledge from these subsets. Authors propose a case study to use cooperative negotiation to construct an integrated domain model from several sources. At the bioinformatics scenario the application of the approach was related to automated annotation of proteins' keywords. In this work, agents do not use any domain dependent information, as they just encapsulate data and machine learning algorithms used to induce models to predict the annotation, using data from biological databases. In a previous work [10, 2] we can find efforts to generate a system for automated annotation proteins related to the *Mycoplasmataceae* family data. A similar works on automatic annotation uses symbolic machine learning techniques with a ISA [36].

Referring to the intersection of bioinformatics and DM, we may say that this is a cutting edge research topic, with important forums to discuss this interaction, e.g. [1]: [35, 39, 14, 51]. Also another new research topic is multiagent projects developed at the bioinformatics domain [33, 8, 7, 48].

Now turning our attention to Computer Forensics research domain area, there are very interesting work using DM techniques. In Gary Warner spam research, they are applying the principles of DM and Grid Computing to establish the Spam DM for Law Enforcement project, but no AMII perspective is used [60, 59]. In the spam project, they have build a large corpus of spam emails, which were analyzed and clustered to provide significant forensic and investigative data to law enforcement [44]. This project has the objective to provide tools, techniques, and training to fight CyberCrime. The projects' approach is a three-pronged one: academics, awareness and research. There are some special focus area the group is doing research: spam, phishing and malware.

A more methodological work includes the Cross Industry Standard Process for Data Mining (CRISP-DM), which is specialized for evidence mining in [58, 22]. Authors use the term *evidence mining* to refer to the application of these techniques in the analysis phase of digital forensic investigations. Thus, this paper presents an approach to the specialization of CRISP-DM to CRISP-EM, an evidence mining methodology designed specifically for digital forensics.

Another work presents a novel DM method, called *AuthorMiner*, for determining the authorship of a malicious e-mail [32]. Since there is an alarming increase in the number of cybercrime incidents through anonymous e-mails, the problem of e-mail authorship attribution is to identify the most plausible author of an anonymous e-mail from a group of potential suspects. Most previous contributions employed a traditional classification approach, such as decision tree and Support Vector Machine (SVM). In this paper, authors introduce an innovative DM method to capture the write-print of every suspect and model it as combinations of features that occurred frequently in the suspects e-mails. This notion is called *frequent pattern*. In summary, there are many recent research work in DM integration to Bioinformatics and to Computer Forensics domains, but to the best of our knowledge, there is a lack of AMII research in both areas. Thus, we present our application-oriented approach to AMII field with a high-level overview of two different domain multiagent applications: BioAgents and MADIK.

2.3 BioAgents

Enormous volume of deoxyribonucleic acid (DNA) sequences of organisms are continuously being discovered by genome sequencing projects around the world. The task of identifying biological function prediction for the DNA sequences is a key activity in genome projects. This task is done in the annotation phase, which is divided into automatic and manual. The automatic annotation has the objective of finding, for each DNA sequence identified in the project, similar sequences among millions, stored in public databases, e.g. *GenBank* [13], by using approximated pattern matching algorithms (*BLAST* [5] and *FASTA* [50]). The manual annotation is done by the biologists, that use the results produced by the automatic annotation, and their knowledge and experience, to decide the function prediction to each DNA sequence. In this way, the biologists guarantee accuracy and correctness to each sequence function prediction.

BioAgents is a MAS for supporting manual annotation [54, 9, 38]. The system simulates the biologists' knowledge and experience for annotating DNA sequences in genome sequencing projects. The MAS cooperative approach, allows to create different specialized ISA that, working together, suggest proper manual annotation.

The architecture of *BioAgents* is divided into three layers: interface, collaborative and physical [54]. The interface layer receives the requests and returns the results to users. The collaborative layer is the architecture core, it is composed by the *conflict resolutions agent* (CR), the *manager agents* (MR) and the *analyst agents* (ANL). The specialized MR are responsible for executing particular algorithms, like *BLAST* and *FASTA*, that interact with ANL for treating specific databases, like *nr* or *kog*. Note that we defined specialized agents to deal with different algorithms and specific knowledge sources (KS). At last, the collaborative layer suggests annotations to be sent to the interface layer. The physical layer consists of different local databases containing the results of the automatic annotation.

BioAgents was implemented in *Java* with development environment *Eclipse SDK*, version 3.1.2. For the agent development framework we used *Java Agent Development Framework - JADE*, version 3.4.1 [12]. The parsers used by the ANL agents were implemented by adapting some libraries of the *framework BioJava*, version 1.4. For the rule-based motor, we have used *Java Expert System Shell - JESS*, version 6.1 [23] to allow agents reasoning in *BioAgents* system. With JESS we defined the biologists knowledge through the use of production rules (declarative rules) according to the parameters defined on a specific genome project. In all experiments we used the same production rules based on *BLAST* and *FASTA* results, following the biologists recommendations.

BioAgents rules use two parameters: the *expectation-value* (*e-value*) and *score*. These parameters express the similarity between each sequence generated on the project with each sequence stored on the database. As lower is the *e-value* as lower is the error probability between the correspondences of both sequence of nucleotides, and as higher is the *score* more close are the sequences. The annotation of each sequence is based on the similarity between both sequences, and the hypothesis is

that as much closer are the sequences as higher is the chance that both have the same biological function prediction.

In order to validate *BioAgents*, we used data from three genome sequencing projects developed at the MidWest Region of Brazil: Functional and Differential Genome from the *Paracoccidioides brasiliensis* (Pb) fungus [46], Genome Project of *Paullinia cupana* (guaraná) plant [47] and Genome Sequencing Project of the *Anaplasma marginale* rickettsia [45]. We used *BioAgents* to suggest annotations for both Genome Project Pb and Genome Project Guaraná using the results of *BLAST* and *FASTA*, and comparing the suggested annotations with the manual annotations previously done by the biologists. The results of *BioAgents* applied to the Genome Project Pb presented 44.1% of correct suggestions, while the results applied to the Genome Project Guaraná presented 45.35% of correct suggestions. Results of *BioAgents* applied on the Genome Anaplasma Project presents data with 2,759 suggested annotations for a total of 3,214 ORFs, which corresponds to 85.84% of suggestions.

2.4 MADIK

Computer Forensics consists of examination and analysis of computational systems, which demands a lot of resources due to the large amount of data involved. Thus, the success of computer forensics examinations depend on the ability to examine large amounts of digital forensic data, in search of important evidences. Forensic examination consists of several steps to preserve, collect and analyze evidences found in digital storage media, so they can be presented and used as evidence of unlawful actions. In this scenario, either distributed agent/multiagent architectures and DM techniques can be of great help. At real computer forensics cases, experts can't define at first what evidence is more relevant to the incident or crime under investigation. Thus, a pre-analysis of the suspect machines would limit the number of evidences collected for examination, reducing the time of investigation and analysis by the forensic experts. But the lack of intelligent and flexible tools to help forensic experts with the pre-analysis phase, and with a concrete cross-analysis of large number of potential correlated evidences, is a reality.

Thus, we propose the use of a MAS, to help digital forensics during forensic examination process. With our approach, it is possible to have different specialized ISA to suggest proper investigative actions, based in experts knowledge of technical domain. The process cited by [11] served as basis of analysis for the conception of new specialized agents. In real forensic examinations, a constant necessity in all investigations is the distribution and coordination of tasks amongst the team of specialists. Thus, in the proposed approach, different roles are played by different levels of agents, similarly to organizational hierarchy levels [52].

The **Multi-Agent Digital Investigation ToolKit - MADIK** is a SMA to assist the experts during the forensic examination process [31, 30, 29]. The architecture of MADIK is a four-layer one, defined as a metaphor to the organizational hierarchy levels: strategic, tactical, operational and specialist. With this architecture, we define

autonomous agents, each specialized in a small and distinct subset of the overall objectives and constraints: *HashSetAgent*, *FilePathAgent*, *FileSignatureAgent*, *TimelineAgent*, *WindowsRegistryAgent* and *KeywordAgent*. In this case, no centralized and rigid control is necessary, but a conflict resolution mechanism, to solve differences and keep the main objective trail. With this mechanism, agents can collaborate by observing and modifying one another's work, through the use of a common base named blackboard [42, 20, 17].

MADIK was implemented as an Open Source Software Project with a GNU General Public License (GPL) [53]. MADIK is all developed in Java [37]. MADIK uses as database environment PostgreSQL (pgsql). For the agent development framework we used *Java Agent DEvelopment Framework - JADE*, version 3.4.1. For the agents inference, we have used *Java Expert System Shell - JESS*, version 6.1. In order to validate MADIK we did two experiments using data from real investigation cases. The first case included data of 110 thousand files from two different hard drives, both belonging to the same case [30]. The second experiment is based on a retirement pension fraud investigation, involving workers in the public administration [29]. The objective was to observe the blackboard and evaluate the levels of conflict between the agents. Fourteen pieces of evidence, seized from the same location, were examined (10 hard drives and 4 removable media). The total number of files is 353,466 for a total of 75.502 GB, including recovered files but excluding free space fragments.

2.5 Evaluation and Future Work

There are many different ways to tackle the emergent AMII topic. Related to bioinformatics domain area, the MAS architecture defined to *BioAgents*, described in Section 2.3, is totally adequate to the integration of DM techniques. The existence of the physical layer, which is composed by many public biological databases, e.g. *nr*, *COG*, *GO KOG* and *Swiss-Prot*, can be automatically mined with many different algorithms defined into sets of ISA in a distributed and parallel way. Although automated annotation has not been the focus of *BioAgents* project so far, certainly this process can increase its performance through DM techniques integration. For this focus, a very good computational infra-structure is necessary to deal with real-world biological data scale volume, even counting on the cooperative skills of the agent society.

Another focus of our research would be related to the classification model functions using DM techniques, which can be used to classify ncRNAs (non-coding RNAs). In NONCODE [28], for example, they have concentrated classification effort on the process in which a given ncRNA takes part, along with its function in this process. In this matter, *BioAgents* can use Support Vector Machines (SVM) to learn classification and improve manual annotation recommendation, as a prediction of ncRNAs with different criteria. Some recent researches have been working in this direction [34, 40]. In the study of pathogenic organisms of [6], authors used as a

target for such approach the fungus *Paracoccidioides brasiliensis* (*Pb*), the ethyological agent of paracoccidioidomycosis, whose transcriptome has recently been elucidated.

Another very interesting focus is being directed in our project at the moment to improve *BioAgents*' manual suggestion, through the use of structured machine learning, as considered a emerging trend for the next ten years at the Bioinformatics area [18]. The idea is to define training annotation data for a very well annotated organisms, such as *Caenorhabditis elegans* genome for example [24]. This training data would be of the form (x, y) , where x is a feature vector describing a particular site and $y = 1$ is true if the annotated organism is present at a very good public biological databases, or site, e.g. *Swiss-Prot*), and $y = 0$ otherwise. This is a standard supervised learning problem. This is an interesting manual annotation prediction problem and we'd like to know how well *BioAgents* can act in such scenario.

Related to Computer Forensics, we have proposed a distributed digital forensics toolkit - MADIK. The work of [55] has defined a distributed tool too. Our proposal benefit from the distributed nature of a MAS and we already observed the performance gains, which helped us obtain better computational resource usage and reduce the time required to perform the examination. This approach can be normally extended to improve performance using DM techniques. In [25] we find two approaches for analyzing large data sets of forensic data called Forensic Feature Extraction (FFE) and Cross-Drive Analysis (CDA). We consider CDA to be the most interesting for MADIK, since it uses statistical techniques for correlating information within a single disk image and across multiple disk images. A recent work by [16] describes a tool called Forensics Automated Correlation Engine (FACE), whose objectives are similar to ours. They also present some scenarios, where an increased level of correlation of disparate evidences was achieved.

In [3], authors develop profiles to describe user or system behavior as a useful technique employed in Computer Forensics investigations. Information found in data obtained by investigators can often be used to establish a view of regular usage patterns, which can then be examined for unusual occurrences. Events compiled from potentially numerous sources are grouped according to some criteria and frequently occurring event sequences are established. The methodology and techniques to extract and contrast these sequences are then described and discussed. In this direction, we are planing to implement in MADIK an example-based method of DM technique, as cited in Section 2.4, CBR to improve the agents' reasoning process and the quality of recommendation.

This chapter presented research work related to the Emergent Agent-Mining Organizations and Applications topic. To better illustrate MAS applications, we presented two different study cases already defined and implemented: *BioAgents* and MADIK. We discussed some ideas to the *BioAgents* and MADIK AMII integration case, showing that both projects can be naturally extended to include DM techniques as an ongoing research works, as demanded by the emerging AMII perspective.

References

1. International journal of data mining and bioinformatics (ijdmb). <http://www.inderscience.com/browse/index.php?journalCODE=ijdmb>.
2. *Proceedings of the European Conference on Computational Biology (ECCB 2002)*, October 6-9, 2002, Saarbrücken, Germany, 2002.
3. Tamas Abraham. Event sequence mining to develop profiles for computer forensic investigation purposes. In *ACSW Frontiers '06: Proceedings of the 2006 Australasian workshops on Grid computing and e-research*, pages 145–153, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc.
4. The Agents, Data Mining Interaction, and Integration interest group (AMII-SIG). <http://www.agentmining.org/>.
5. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, October 1990.
6. Roberto T. Arrial, Roberto C. Togawa, and Marcelo M. Brigido. Outlining a strategy for screening non-coding rnas on a transcriptome through support vector machines. In Marie-France Sagot and Maria Emilia Telles Walter, editors, *BSB*, volume 4643 of *LNCS*, pages 149–152. Springer, 2007.
7. Ezio Bartocci, Flavio Corradini, and Emanuela Merelli. Building a multiagent system from a user workflow specification. In Paoli et al. [48].
8. Ezio Bartocci, Flavio Corradini, Emanuela Merelli, and Lorenzo Scortichini. BioWMS: a web-based Workflow Management System for bioinformatics. *BMC Bioinformatics*, 8(Suppl 1):S2, 2007.
9. Ana L. C. Bazzan, Mark Craven, and Natália F. Martins, editors. *Advances in Bioinformatics and Computational Biology, Third Brazilian Symposium on Bioinformatics, BSB 2008, Santo André, Brazil, August 28-30, 2008. Proceedings*, volume 5167 of *LNCS*. Springer, 2008.
10. Ana L. C. Bazzan, Paulo Martins Engel, Luciana F. Schroeder, and Sérgio C. da Silva. Automated annotation of keywords for proteins related to mycoplasmataceae using machine learning techniques. In *ECCB [2]*, pages 35–43.
11. Nicole Beebe and Jan Guynes Clark. A hierarchical, objectives-based framework for the digital investigations process. *Digital Investigation*, 2(2):147–167, 2005.
12. Fabio L. Bellifemine, Giovanni Caire, and Dominic Greenwood. *Developing Multi-Agent Systems with JADE (Wiley Series in Agent Technology)*. Wiley, April 2007.
13. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. Genbank. *Nucleic Acids Res*, 36(Database issue), January 2008.
14. Massimo Buscema and Enzo Grossi. The semantic connectivity map: an adapting self-organising knowledge discovery method in data bases. experience in gastro-oesophageal reflux disease. *Int. J. Data Min. Bioinformatics*, 2(4):362–404, 2008.
15. Longbing Cao, Chao Luo, and Chengqi Zhang. Agent-mining interaction: An emerging area. In Gorodetsky et al. [27], pages 60–73.
16. Andrew Case, Andrew Cristina, Lodovico Marziale, Golden G. Richard, and Vassil Roussev. Face: Automated digital evidence discovery and correlation. *Digital Investigation*, 5(Supplement-1):S1–S140, september 2008. The Proceedings of the Eighth Annual DFRWS Conference.
17. Daniel D Corkill. Collaborating Software: Blackboard and Multi-Agent Systems & the Future. In *Proceedings of the International Lisp Conference*, New York, October 2003.
18. Thomas Dietterich, Pedro Domingos, Lise Getoor, Stephen Muggleton, and Prasad Tadepalli. Structured machine learning: the next ten years. *Machine Learning*, 73(1):3–23, October 2008.
19. Christos Dimou, Andreas L. Symeonidis, and Pericles A. Mitkas. Evaluating knowledge intensive multi-agent systems. In Gorodetsky et al. [27], pages 74–87.
20. Rajendra Dodhiawala. *Blackboard Architectures and Applications*. Academic Press, Inc., Orlando, FL, USA, 1989.

21. Cássia Trojahn dos Santos and Ana L. C. Bazzan. Integrating knowledge through cooperative negotiation - a case study in bioinformatics. In Gorodetsky et al. [26], pages 277–288.
22. CRoss Industry Standard Process for Data Mining. Crisp-dm 2.0 special interest group. <http://www.crisp-dm.org/new.htm>.
23. Ernest Friedman-Hill. *Jess in Action : Java Rule-Based Systems (In Action series)*. Manning Publications, December 2002.
24. Ratsch G, Sonnenburg S, Srinivasan J, Witte H, Muller K-R, and et al. Improving the caenorhabditis elegans genome annotation using machine learning. *PLoS Comput Biol*, 3(2), 2008.
25. Simson L. Garfinkel. Forensic feature extraction and cross-drive analysis. *Digital Investigation*, 3(Supplement-1):71–81, 2006.
26. Vladimir Gorodetsky, Jiming Liu, and Victor A. Skormin, editors. *Autonomous Intelligent Systems: Agents and Data Mining, International Workshop, AIS-ADM 2005, St. Petersburg, Russia, June 6-8, 2005, Proceedings*, volume 3505 of *LNCS*. Springer, 2005.
27. Vladimir Gorodetsky, Chengqi Zhang, Victor A. Skormin, and Longbing Cao, editors. *Autonomous Intelligent Systems: Multi-Agents and Data Mining, Second International Workshop, AIS-ADM 2007, St. Petersburg, Russia, June 3-5, 2007, Proceedings*, volume 4476 of *LNCS*. Springer, 2007.
28. Shunmin He, Changning Liu, Geir Skogerbo, Haitao Zhao, Jie Wang, Tao Liu, Baoyan Bai, Yi Zhao, and Runsheng Chen. Noncode v2.0: decoding the non-coding. *Nucl. Acids Res.*, pages gkm1011, November 2007.
29. Bruno W. P. Hoelz, Célia G. Ralha, and Rajiv Geeverghese. Artificial intelligence applied to computer forensics. In *The 24th Annual ACM Symposium on Applied Computing, SAC'09, Computer Forensics Track*, pages 211–217, Honolulu, Hawaii, USA, 2009. ACM Press.
30. Bruno W. P. Hoelz, Célia G. Ralha, Rajiv Geeverghese, and Hugo C. Junior. A cooperative multi-agent approach to computer forensics. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 2, pages 477–483. IEEE Computer Society, 2008.
31. Bruno W. P. Hoelz, Célia G. Ralha, Rajiv Geeverghese, and Hugo C. Junior. Madik: A collaborative multi-agent toolkit to computer forensics. In Robert Meersman, Zahir Tari, and Pilar Herrero, editors, *OTM Workshops*, volume 5333 of *LNCS*, pages 20–21. Springer, 2008.
32. F. Iqbal, R. Hadjidj, B. C. M. Fung, and M. Debbabi. A novel approach of mining write-prints for authorship attribution in e-mail forensics. *Digital Investigation*, 5(1):42–51, 2008.
33. Konstantinos A. Karasavvas, Richard Baldock, and Albert Burger. A criticality-based framework for task composition in multi-agent bioinformatics integration systems. *Bioinformatics*, 21(14):3155–3163, 2005.
34. Y. A. N. Karklin, Richard F. Meraz, and Stephen R. Holbrook. Classification of non-coding rna using graph representations of secondary structure. In *Pacific Symposium on Biocomputing 2005*. Pacific Symposium on Biocomputing 2005, World Scientific Publishing Co. Pte. Ltd., 2005.
35. I. R. Konig, J. D. Malley, S. Pajevic, C. Weimar, H-C. Diener, and A. Ziegler. Patient-centered yes&no prognosis using learning machines. *Int. J. Data Min. Bioinformatics*, 2(4):289–341, 2008.
36. Ernst Kretschmann, Wolfgang Fleischmann, and Rolf Apweiler. Automatic rule generation for protein annotation with the c4.5 data mining algorithm applied on swiss-prot. *Bioinformatics*, 17(10):920–926, 2001.
37. Java language. <http://java.sun.com>.
38. Richardson Silva Lima, Célia Ghedini Ralha, Maria Emlia Machado T.Walter, Hugo Wruck Schneider, Anderson Gray F. Pereira, and Marcelo Macedo Brgido. Bioagents: Um sistema multiagente para anotação manual em projetos de seqüenciamento de genomas. In *XXVII Congresso da SBC, VI Encontro Nacional de Inteligência Artificial (ENIA'07)*, Rio de Janeiro, RJ, 30 Junho a 6 Julho. *Anais*, pages 1302–1310, 2007.
39. Cui Lin, Shiyong Lu, Xuwei Liang, Jing Hua, and Otto Muzik. Cocluster analysis of thalamo-cortical fibre tracts extracted from diffusion tensor mri. *Int. J. Data Min. Bioinformatics*, 2(4):342–361, 2008.

40. Hong-Wei Liu. Predicting non-protein-coding rna genes in escherichia coli using svm with signature descriptor. In *The Second International Symposium on Optimization and Systems Biology (OSB'08)*, pages 287–293, Lijiang, China, October 31–November 3 2008.
41. George F. Luger. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
42. H. Penny Nii. Blackboard systems, part one: The blackboard model of problem solving and the evolution of blackboard architectures. *AI Magazine*, 7(2):38–53, 1986.
43. Ayahiko Niimi, Hitomi Noji, and Osamu Konishi. Distributed web integration with multiagent data mining. In Rajiv Khosla, Robert J. Howlett, and Lakhmi C. Jain, editors, *KES (3)*, volume 3683 of *LNCS*, pages 513–519. Springer, 2005.
44. Department of Computer and USA Information Sciences, University of Alabama at Birmingham (UAB). Cis@uab computer forensics research project. <http://www.cis.uab.edu/forensics/>.
45. Genome Sequencing Project of the *Anaplasma marginale rickettsia*. <https://www.biomol.unb.br/anaplasma/servlet/IndexServlet>.
46. Genome Project of the *Paracoccidioides brasiliensis* (Pb) fungus. <https://dna.biomol.unb.br/Pb-eng/>.
47. Genome Project of the *Paullinia cupana* (guaraná) plant. <https://dna.biomol.unb.br/GR/>.
48. Flavio De Paoli, Antonella Di Stefano, Andrea Omicini, and Corrado Santoro, editors. *Proceedings of the 7th WOA 2006 Workshop, From Objects to Agents (Dagli Oggetti Agli Agenti), Catania, Italy, September 26-27, 2006*, volume 204 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.
49. Ana Carolina M. Pilatti de Paula, Braulio C. Avila, Edson Scalabrin, and Fabricio Enembreck. Multiagent-based model integration. In *WI-IATW '06: Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology*, pages 11–14, Washington, DC, USA, 2006. IEEE Computer Society.
50. W. R. Pearson and D. J. Lipman. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, 85(8):2444–2448, April 1988.
51. Tuan D. Pham. Computational prediction models for cancer classification using mass spectrometry data. *Int. J. Data Min. Bioinformatics*, 2(4):405–422, 2008.
52. S. Pinson and P. Moraitis. An intelligent distributed system for strategic decision making. *Group Decision and Negotiation*, 6:77–108, 1996.
53. MADIK's project Web hosting page. <http://madik.sourceforge.net/>.
54. Célia Ghedini Ralha, Hugo W. Schneider, Lucas O. da Fonseca, Maria Emilia M. T. Walter, and Marcelo M. Brígido. Using bioagents for supporting manual annotation on genome sequencing projects. In Bazzan et al. [9], pages 127–139.
55. Vassil Roussev and Golden G. Richard III. Breaking the performance wall: The case for distributed digital forensics. In *The fourth annual Digital Forensics Research Workshop*, Baltimore, Maryland, August 2004.
56. Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2003.
57. Ichiro Satoh. Self-organizing multi-agent systems for data mining. In Gorodetsky et al. [27], pages 165–177.
58. Jacobus Venter, Alta de Waal, and Cornelius Willers. *Specializing CRISP-DM for Evidence Mining*, volume 242/2007. Springer, Boston, 1st edition edition, 2007. Book Series-IFIP International Federation for Information Processing.
59. Roger L. Wainwright and Hisham Haddad, editors. *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC)*. ACM, 2008.
60. Chun Wei, Alan Sprague, Gary Warner, and Anthony Skjellum. Mining spam email to identify common origins for forensic application. In Wainwright and Haddad [59], pages 1433–1437.
61. Gerhard Weiss, editor. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT Press, Cambridge, MA, USA, 1999.
62. Michael Woolridge and Michael J. Wooldridge. *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., New York, USA, 2001.



<http://www.springer.com/978-1-4419-0521-5>

Data Mining and Multi-agent Integration

Cao, L. (Ed.)

2009, XIV, 334 p., Hardcover

ISBN: 978-1-4419-0521-5