

# Chapter 20

## An Enhanced Approach to Retrieve Learning Resources Over the Cloud

Fernando De la Prieta, Ana Belén Gil, Sara Rodríguez,  
Javier Bajo Pérez, José Alberto García Coria and Juan M. Corchado

**Abstract** This study proposes AIREH (Architecture for Intelligent Retrieval of Educational content in Heterogeneous Environments) that is a model for the development of digital content retrieval based on the paradigm of virtual organizations of intelligent agents Learning objects have made it possible to create digital resources that can be reused in various didactic units. These resources are stored in repositories, and thus require a search process that allows them to be located and retrieved.

---

F. De la Prieta (✉) · A. B. Gil · S. Rodríguez · J. A. G. Coria · J. M. Corchado  
Department of Computer Science and Automation Control, University of Salamanca,  
Plaza de la merced s/n 37007 Salamanca, Spain  
e-mail: fer@usal.es

A. B. Gil  
e-mail: abg@usal.es

S. Rodríguez  
e-mail: srg@usal.es

J. A. G. Coria  
e-mail: jalberto@usal.es

J. M. Corchado  
e-mail: corchado@usal.es

J. B. Pérez  
Department of Artificial Intelligence, Technical University of Madrid, Bloque 2,  
Despacho 2102, Campus Montegancedo, Boadilla del Monte 28660 Madrid, Spain  
e-mail: jbajo@fi.upm.es

## 20.1 Introduction

The paradigm of Learning Objects (LO), which is one of the most widely accepted approaches in the study of distance education today, is based on the fragmentation of self-contained learning units that can be reused in different educational contexts and different platforms. This systematic management of learning resources makes possible their dissemination.

To facilitate these dissemination tasks, the LO are stored in educational repositories. In the current educational context there is a significant growth of learning object repositories (LOR) as part of the hidden web in large databases.

The user interacts with the information contained in the repository via their web browser. But through the traditional web interfaces, it is not possible to manage all information directly labeled by these LO, nor to even know if that information is labeled or not according to any standard because the user does not know how that educational content is in the repository. Although there is a large volume of educational content on those repositories, they present problems at different levels:

- It is usually that they have low-level performance, unavailability, security, reliability, etc.; in other words, they have many technical problems that makes difficult their usage.
- The content is not directly accessible through conventional search engines.

At the same time, the paradigm as a whole also shows many deficiencies like the existence of too many schemas of metadata or interoperability specifications, or even internal architecture of the repositories.

It is necessary to work towards and develop solutions that enable the efficient search of heterogeneous content with regards to the distributed context where they lodge. Distributed information retrieval or federated searches, [1] attempt to respond to the problem of information retrieval in the hidden Web. The main aim of federated search is to develop models and strategies to get the most benefit from these distributed sources. The process is completely transparent to the user, who does not perceive the complexity involved, and gives a uniform treatment to the information retrieved. The main contribution of federated search is that the process is done through search mechanisms in individual information sources. In addition, the search refers to the location of each source and provides a distributed control of information related to the different sources of hidden information.

Thus, the objective of this study is to present AIREH tool (*Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments*) [2] which makes it possible to search and recover educational resources encapsulated in the form of a LO. Similarly, a system can use a Case-Based Reasoning (CBR) system to recommend which educational resources might be of particular interest to the user, based on information from previous searches. This system is based on Multi-Agent Systems (MAS) based on Virtual Organizations (VO). In this context, Cloud Computing paradigm [3] is the key to offering effective and efficient services such as storage, and the search and retrieval of educational resources.

This study is organized as follows, next section establishes the state of the and the related work, Sect. 20.3 shows the observed problems and the key opportunities of Cloud Computing, Sect. 20.4 shows the proposal system, and finally, Sect. 20.5 presents the experiments and the conclusions.

## 20.2 The Learning Object Paradigm

Many authors have recently been presenting their vision regarding this concept [4–6], which has led to the appearance of a number of definitions. The IEEE’s Learning Technology Standards Committee (LTSC) defines a LO in general terms as *any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning*. In short, practically any educational resource can be considered a LO, a fact heavily criticized by various authors [7–9] who have tried to delimit the concept as much as possible. As a result, there is a clear consensus that an LO must be the minimal reusable unit of learning content with a specific objective and also there is a consensus that each LO has to be associated with an external structure of metadata.

This metadata allows making a first approach to the educational resource. In other words, the metadata permits improving the utility of the resource, since it makes its retrieval, search, exchange, and hence, its reutilization, possible. The metadata schema is standardized. In fact, there are currently many standards. The most relevant standards are Dublin Core [10], which is more oriented towards digital resources in general and is quite widespread within a library context; IEEE Learning Object Metadata (LOM) [6], which is the most commonly used standard for describing LOs; and finally SCORM [11], which is oriented to packaging and distributing complete educational activities. Although at first sight these standards can be seen as an advantage, reality shows that in some cases they are the problem, as many existing standards are not compatible among themselves. It is important to note that not only is the existence of metadata standards necessary in order to reuse contents, but the data that the authors assign to each descriptor is very important as well. To this end, it is necessary to follow a traceable process from the creation of an educational resource to the creation of its metadata in order to establish a metadata structure that is consistent, relevant and interpretable. [12].

LOs are commonly stored in repositories, which are characterized by their heterogeneity [13]. The deployment infrastructure can basically be either distributed or centralized. Taking into account that an LO is formed by a digital resource and its metadata, there are four kinds of possible infrastructures [11]: (i) centralized resources and centralized metadata, (ii) centralized resources and distributed metadata, (iii) distributed resources and centralized metadata and (iv) distributed resources and distributed metadata. Furthermore, three kinds of storage strategies can be distinguished [11]: (i) *File-based*, which uses files with predefined formats and an index-based management; (ii) *Database-based*, which uses any kind of

database, and is the most extended method; and (iii) *Persistent objects-based*, where the LO are stored as serialized objects.

The heterogeneity itself is not a problem, since there are different tools that can isolate the internal logic of the LOR from the exterior, which in fact makes it possible to automatically search different repositories simultaneously using a single query application. Most notable among these tools, which serve as a middleware layer between the repository and the clients, are (i) Open Archives Initiative Protocol for Metadata harvesting (OAI-MPH) [14] which is a protocol that provides a technology-independent framework for retrieving documents or resources, thus enabling interoperability among systems; and (ii) Simple Query Interface (SQI) [8] that is formed by a set of abstract methods based on web services. SQI is also neutral in terms of the format of results as well as query language. This interfaces supports synchronous/asynchronous and stateful/stateless queries.

### 20.3 Observed Weakness and Possible Solutions

The birth of what can already be considered as the LO paradigm has produced many advantages when it comes to reusing learning content. The innovation process has also produced different challenges that have not yet been solved.

The main problem is that LORs still do not implement any abstraction layer that can encapsulate the internal logic of the repository. Consequently, consultation and LO extraction is a slow process which requires the manual intervention of users who must reuse the learning resources. But the problems are not limited to those associated with not using any middleware layer of abstraction. The systems in which this layer is included, also suffer from various problems such as:

- The problems associated with the monolithic structure of LOR, which does not allow external management with the flexibility and power necessary to ensure easy interoperability, and dispersed and heterogeneous sources.
- the absence of automatic mechanisms that control the technical quality, semantics and syntax of LO, ensuring the correct specification of such LOs in any of the metadata schemas that describe them.

These problems require solutions that are adapted to the heterogeneity. The solution should enable a centralized global search and the effective reuse of resources by the end user. This requires raising the level of abstraction and looking at the classification of systems storing and searching for LOs.

One of the possible solutions is to adopt the Cloud Computing paradigm that is emerging as the key paradigm of the present century. There are incipient developments that broach the topic of cloud technology and e-learning [15], however few studies incorporate both concepts. As demonstrated in [15], cloud computing and e-learning are fundamentally centered on the SaaS layer. Other research such as [16] has studied the applicability in different fields such as education, but also focuses on the SaaS layer, ubiquitously offering a set of applications to users. Due

to these circumstances, and given the possibilities offered by cloud computing in the field of e-learning for managing applications in the SaaS layer, it has become necessary to develop applications in cloud architectures that can form part of the PaaS layer, and that can manage the LOs that are stored within the architecture, thus facilitating the interaction and access of the applications developed in the system.

According to National Institute of Standards and Technology (NIST), *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models* [3]. This definition includes three levels of computational services (Software, Platform and Infrastructure).

The key characteristic of this new paradigm is the quality of services. Cloud services are able to offer the same level of quality independently of instant demand. In practice, end users make use of Cloud services that are always available and unlimited.

Taking into account the weakness that has been demonstrated in this study with regard to the performance, availability and interoperability of existing LO paradigm, this study proposes a new deployment architecture based on Cloud Computing.

This architecture will be detailed as follows and it will make use of the services that +Cloud platform [17] provides, such as storage and databases. This platform is based on the Cloud Computing paradigm. This platform allows offering services at the PaaS and SaaS levels. The IaaS layer is composed of a physical environment that allows the abstraction of resources into virtual machines. The SaaS layer is composed of the management applications for the environment (virtual desktop, control of users, installed applications, etc.), and other more general third party applications that use the services from the PaaS layer. The components of this layer are: (i) an *IdentityManager*, which is the module of +Cloud in charge of offering authentication services to clients and applications; (ii) the *File Storage Service (FSS)*, which provides an interface for a container of files, emulating a directory structure in which the files are stored with a set of metadata, thus facilitating retrieval, indexing, search, etc.; and finally, (iii) the *Object Storage Service (OSS)*, which provides a simple and flexible schemaless data base service oriented towards documents.

## 20.4 The Proposed System

This paper proposes a solution: the federated search architecture for educational content, particularly Los; deployed over +Cloud infrastructure.

Conflicting needs must be met in order to create the federation and thus integrate several components: the repositories must be distributed across physically dispersed locations; access to these repositories is read-only; information systems

are operating under different platforms; the types used are distributed repositories (metadata stored in one place and LO in other) and they are managed by autonomous systems; in some cases access to the repositories requires the establishment of a session, in others direct access, etc. In this sense, the use of a system architecture based on the use of intelligent agents is ideal, since they can apply their characteristics (autonomy, status, reactivity, rationality, intelligence, coordination, mobility and learning) to develop a stable system with the ability to react intelligently to the needs of the environment. The idea to model the architecture as a virtual organization, on the assumption that an organization can adapt its actions to achieve its goals and interact with heterogeneous components, is proposed as a theoretically efficient solution.

The proposed architecture is seen as a communication point midway between the LOR, the LO that they store and users who use them. The system provides a federated search system that makes it possible to search multiple geographically dispersed repositories simultaneously. In addition, once the results of the different repositories have been received, an identification phase is developed and filtered in order to adapt the results to the user preferences.

The organizational model used in the architecture of the platform uses THOMAS [18], to define the structure and rules. Adaptation in this type of model is based on coordination between the participants of the organization. To establish the interaction model it is necessary to analyze the needs and expectations of potential system users. From this analysis it is possible to deduce the roles of users who reuse the educational resources, and how they are going to exchange information. The interaction process is shown in Fig. 20.1, the initial identified roles are:

- **User.** It represents the system user or customer and it is the responsible for initiating the federated search process by sending a search pattern to the Query Manager role (Step 1). It gets the results from the query and is able to assess both the LO, and the order in which these results are presented.
- **Query Manager.** This agent is responsible for overseeing the entire federated search process. It gets the natural language query from the user agent (Step 1) and is responsible for finalizing the query by propositional logic. It queries federated search Repository Manager according to the pattern received (Step 2). Once the Repository Manager agent indicates the end of the federated search, the Query Manager asks the agent to apply the cataloguing techniques and collaborative filtering on the results (Step 8), which will handle the agent results. Once the agent has notified the Cataloguer that has completed the management process of search results, it orders their transfer to the user agent for consultation and valuation (Step 9).
- **Repository Manager.** It is the agent that has specific control over the queries that are made at different LOR (Step 4). This type of agent receives the query from the formalized Query Manager agent (Step 2) and checks the repositories that are active at the moment of the query (Step 3).
- **Translator.** It is the agent responsible for transforming the formalized language query language required by the repository to which the query is directed.

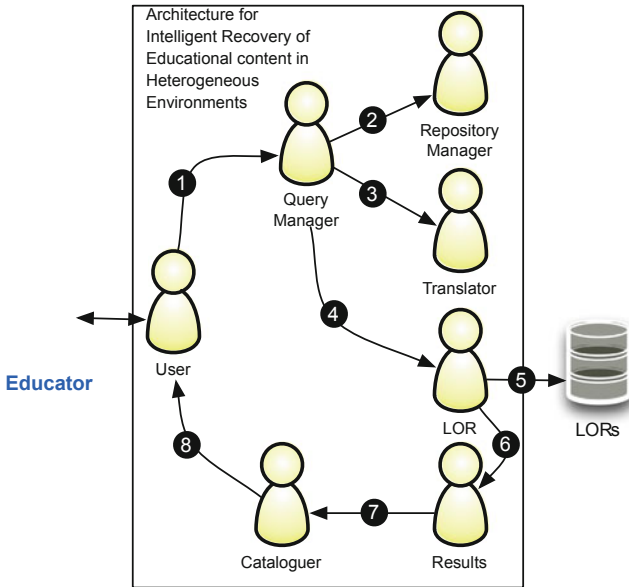


Fig. 20.1 Interaction process among agents/roles within AIREH

Receives the query expressed by propositional logic from the LOR agent (Step 5). Performs the conversion of this query from a formal language to a language that is capable of dealing with the LOR on which the query is directed. Once the transformation had been carried out a query will be sent to the LOR agent in the language in which it operates.

- **LOR.** It is the agent responsible for conducting the consultation to each repository (Step 6). Client code implements each of the possible middleware layers (SQL, OAI-MPH, among others). Although it only envisioned as a single role, in practice there are different types of agents who acquire the LOR role, so that each agent type implements a different middleware layer. In a federated search there will be many agents with LOR role as a repository that is directed towards the query. This agent has the responsibility of making the request to each individual LOR, so different instances of this agent will work simultaneously. The agent performs the LOR query conducting all the necessary process, as defined in the specification of the middleware layer that implements the specific repository. It is the agent responsible for sending the LO results that have been received in response to the query to Result Agent (Step 7).
- **Results.** The agent receives the LO results from each of the LOR agents in each federated search (Step 7). Automatically extracts metadata schema information and eliminates those that are not valid LO. Although in theory there is only one results role, in practice there are different types of result agents so that each type of agent implements a different scheme of metadata. It is responsible for receiving the federated search results from each one of the instantiated LOR

agents in each federated search. It extracts useful information from each diagram that describes the LO, therefore, will need to have access to metadata extraction and data structures needed to store the information provided by each of the schemes defines by the LO. An initial filter eliminates those defects that prevents the LO from receiving proper treatment and/or use by system users. This role stores a minimum data set suitable for each of the schemes, and provides description of levels of abstraction in the comprehensive management resource.

- **Cataloguer.** This agent is responsible for preparing the ranking of LO that have been returned as federated search results. The results will be stored by the agent who in turn automatically applies pre-filtering on LOs, obtaining useful information from them, and removes those that are incomplete. Cataloguer agent implements a Case-Based Reasoning (CBR) that uses previous search information to rank the items that best suit the needs of the application user based on previously obtained information. It uses user profile information as well as their educational information (content-based filtering). It also subsequently uses information from the user's vote for LO and the appropriateness of the previously created ranking of the results (collaborative filtering). The agent orders the LO results to be stored in the agent according to the preferences of both general (profile) and user education. To carry out this process, it requests the agent statistics for the voting of the LO and previous ordinations, and the feedback provided by the user. With all this information it produces the ranking of LO that best matches the user who made the query and send back to User Agent (Step 9). It provides the Statistics agent with the ranking information to be stored, which can then be used to produce future ranking. This agent stores ranking will be validated by the user during the results of the consultation process.
- **Statistics.** This agent is responsible for collecting statistical data from other agents and provides this information to increase system performance
- **Supervisor.** An agent carrying out this role will have overall control of the system. Analyzes the structure and syntax of all messages entering and leaving the system. Supervises the proper functioning of the other players in the system (sends a ping to periodically check the status of all agents of the architecture).

### *20.4.1 A Cloud Computing Deployment*

Finally, AIREH should be noted that this application was deployed in a Cloud Computing environment, which allows users to store information about the recovered resources in a cloud. This Cloud Computing platform provides a set of functionalities in terms of web services to the upper layers, that is, to the end user interfaces:

- FSS will be used to store the educational resources. FSS also encapsulates the traditional complexity of the file system storage; this component only has to call web services in order to retrieve or store resources. Furthermore, because of other FSS functionalities, such as file versions, metadata associated with each



resource, etc., it will be possible to increase the power of the service. Finally, it should be noted that the elasticity of the FSS implies no limitation regarding storage capacity.

- OSS will be used to store the metadata associated with each learning resource and work information for make the ranking of LOs. OSS makes use of a nonSQL database that permits storing the metadata en JSON format. The main advantage is that it permits storing any kind of metadata independent of its structure or schema, that is, its standard. Furthermore, queries about the LO will be performed very quickly thanks to the underlying database.

## 20.5 Experiments and Conclusions

AIREH was evaluated by performing a battery of tests to validate their efficiency in real environments. Evaluation metrics from information retrieval field have been adopted. The two most used evaluation measures are precision (the fraction of documents retrieved by the system that are also relevant to the query) and recall (the fraction of the relevant documents present in the database that are retrieved by the system). Precision,  $P$ , for any LOR  $j$  is the fraction of the retrieved Metadata Documents ( $O$ ) that are relevant to the query. Relative Recall,  $E$ , for each LOR is the fraction of retrieved relevant documents respect to the total number of relevant documents in the accessed repositories according to a priori knowledge. If we denote  $R$  the set of returned documents from each one of the LORs while  $T$  the consulting and retrieval time, we define a new evaluation metric called temporary gain. These parameters characterize the ability of the system to retrieve relevant documents and avoid irrelevant ones.

To estimate these issues, a set of queries were analyzed, in two repositories: LORNET<sup>1</sup> and Merlot.<sup>2</sup> For this reason the search patterns or topics were chosen at random among topics in English for Science and Technology of UNESCO codes developed by experts. These topics are mainly composed of sets of words (longer queries) due to the specific nature of this terminology, but single words (short queries) were also used to assess different behaviors on the search engines repositories.

The results reveal that the proposed architecture significantly increases the number of LO to recover in time, and increases the temporary gain value of the system by about 15 % on average as compared to isolated repositories such as Merlot or LORNET. Figure 20.2 shows the average temporary gains for the series of tested queries.

All the results are relevant consultations with the relevant criteria that have been established in this work. The retrieval of content by the proposed architecture system depends on the LO returned by the isolated repositories, so in the event that

---

<sup>1</sup> <http://www.lornet.ca>

<sup>2</sup> <http://www.merlot.org/>

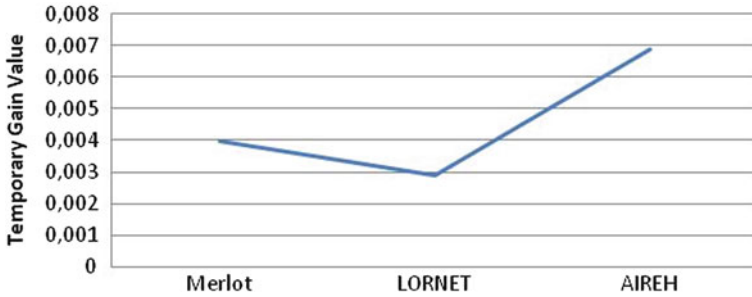


Fig. 20.2 Temporary gain values

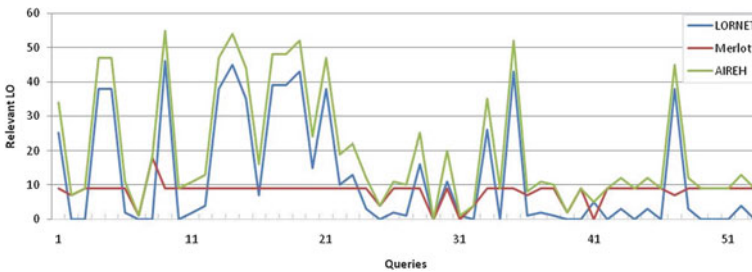


Fig. 20.3 Comparison of relevant metadata documents (LO)

there is no LO to respond to the user’s consultation, the AIREH system cannot resolve the lack of content. However, the system ensures that any content related to the user’s request will be recovered as shown in Fig. 20.3.

As a conclusion this study has presented an innovative architecture that constitutes an evolution over current storage system for educational resources. This new model, will enable the observed problems to be solved:

- *High heterogeneity in terms of number and characteristics of existing standards.* The proposed model allows dealing with the heterogeneity of current and future standards since it is based on virtual organization of a multiagent systems and it makes use of a non-relational database.
- *Low performance.* Cloud computing paradigm allows offering services with the same level of quality independently of its demand. The development of the LOR based on this paradigm will make it possible not only to offer an effective service effective, but to offer an unlimited storage capacity as well.
- *Interoperability among repositories.* The low linkage among components permits implementing many interoperability layers without needing to upgrade to other modules.
- *Complementary services.* This model will make it possible to include other functionalities in its own repository that until now were not possible, such as

recommendation model, space of storage for each user in the cloud, a collaborative model for creating learning resources and metadata, etc.

**Acknowledgments** This research has been supported by the project CLOUD-IO (IDI-20111471) funded by Centre of Technical Industrial Development (CDTI)

## References

1. Callan J (2000) Distributed information retrieval. In: Croft WB (ed) *Advances in information retrieval*, chapter 5. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 127–150
2. Gil AB, De La Prieta F, Rodríguez S (2011) Automatic learning object extraction and classification in heterogeneous environment. In: Perez JB et al. (eds) *Highlights in PAAMS, AISC 89*. Springer-verlag, Berlin, pp. 109–116
3. Mell P, Grance T (2011) The NIST definition of cloud computing. In: NIST Special Publication 800-145. Sep 2011
4. European Committee for standardization–Cen Workshop Agreement (2005) A simple query interface specification for learning repositories. Ref. No.: CWA 15454:2005 E
5. Frango I, Omar N (2007) Architecture of learning objects repositories. *Learning Objects: standards, metadata, repositories & LMS*, pp. 131–155
6. IEEE Learning Object Metadata (LOM) (2002) Institute of electrical and electronics engineers. <http://ltsc.ieee.org>
7. Chiappe A, Segovia Y, Rincon HY (2007) Toward an instructional design model based on learning objects. *Education Tech Research Dev* 55(6):671–681
8. Simon B, Massart D, Van Assche F, Ternier S, Duval E, Brantner S, Olmedilla D, Miklos Z (2005) A simple query interface for interoperable learning repositories. In: *Proceedings of the 1st workshop on interoperability of web-based educational systems*, Chiba, Japan
9. Lujara SK, Kissaka MM, Bhalaluseca EP, Trojer L (2007) Learning objects: a new paradigm for e-learning resource development for secondary schools in Tanzania. *World Acad Sci Eng Technol* 102–106
10. Dublin Core Metadata Initiative. DCMI Metadata Terms <http://dublincore.org/>
11. SCORM (2004) 4th Edition. March 2009
12. Berlanga AJ, López C, Morales E, García Peñalvo FJ (2005) Consideraciones para reforzar el valor de los metadatos en los objetos de aprendizaje (OA). Salamanca. Universidad de Salamanca. Depto. de Informática y Automática, Salamanca, Spain
13. De la Prieta F, Gil A-B (2010) A multi-agent system that searches for learning object in heterogeneous repositories. In: Demazeu Y et al. (eds) *Trends in PAAMS, AISC 71*. Springer Verlag, pp. 355–362
14. Lagoze C, Van De Sompel H, Nelson M, Warner S (2002) The open archives initiative protocol for metadata harvesting. Open Archive Initiative. Version 2.2
15. Ercan T (2010) Effective use of cloud computing in educational institutions. *Procedia Soc Behav Sci* 2:938–942
16. Sultan N (2010) Cloud computing for education: a new dawn? *Int J Inf Manage* 30(2):109–116
17. Heras S, De la Prieta F, Julian V, Rodríguez S, Botti V, Bajo J, Corchado JM (2012) Agreement technologies and their use in cloud computing environments. *Prog Artif Intell* 1(4):277–290
18. Carrascosa C, Giret A, Julian V, Rebollo M, Argente E, Botti V (2009) Service oriented MAS: an open architecture (Short Paper). In: *Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds), May 10–15, 2009, Budapest, Hungary, pp. 1291–1292