JOUR-MAS: a multi-agent system approach to help journalism management

Alberto L. Barriuso[a], Fernando de La Prieta[a], Álvaro Lozano Murciego[a], Daniel Hernández[a] and Jorge Revuelta Herrero[a]

albarriuso@acm.org, fdelaprieta@acm.org, murciego@acm.org, ahernandezdelaiglesia@acm.org, jrevuelta@acm.org

[a] ACM student member

KEYWORD	ABSTRACT
Journalism; Multi-agent Systems; Cloud Computing	*Nowadays, the vertiginous progress of technology has led society to demand lower response times in general, but in press in particular. For this reason, media groups have to face a big challenge in order to facilitate their employees the technical means that enable professionals in the media sector to face the new communication models. In this article, we face this problem and present a novel system based on Virtual Organizations of agents and Cloud Computing in order to improve the organization, management and distribution of media contents.*

1. Introduction

The vertiginous progress of technology has led to a rapid change of the newspaper writing conception, as well as the role of communicators. By cause of this progress, this profession is being framed in a new context that integrates a novel set of elements that diametrically change the way of work. The tools which have been used by the media (online newspapers, television and radio websites, etc.) up to now, have a classical structure, which has been inherited from the last century. Usually, there are different levels of hierarchy in the treatment of information before it is published (collaborators, correspondents, writers, editors, etc.). However, these archaic procedures, which were effective time ago, do not allow to manage information in a flexible way, with the promptitude which is demanded by users.

In this situation, we must not forget that the traditional business model of communication is bound to its end due to the low demand of traditional media: online newspapers against the press, radios on the Internet at the expense of traditional radios, online television and digital terrestrial television, availability of information in real time through social networks, etc. Today, there is a change in the consumption patterns of printed newspapers, with a continuous reduction due to the introduction of new technologies.

This fact joins other fact: online newspapers must not only compete with other web media, but also with other online news providers (Rubio, J. L., 2009).

The main challenge for media groups is to put at the disposal of their employees the technical tools that enable newspaper industry professionals to address the new communication models, allowing them to response the low publication time requirements demanded by society, and providing the tools that enable new business models to make the online media profitable.

Since there are currently no integral and dynamic multichannel information management tools, the main objective of this system consists in responding to this demand in the telecommunications sector, by developing an effective system to organize and manage with-held generating media (written information, photographs, videos, audio, graphics, etc.) holistically.

Thus, this system is divided into three technological pillars: a new Communication model based on Social Computing, Cloud Computing and Artificial Intelligence. First, it is developed a model for effective communication, capable of receiving information from multiple heterogeneous sources, allow to the journalism to work in collaboration with the tools, and then publish on different platforms. Second, the deployment in a cloud computing environment that allows to address the needs of scalable computing (persistence, communication, etc.), offering applications through a web platform and multi-device environment, allowing to apply a pay per use business model. Finally, it makes use of Artificial Intelligence techniques, such as multi-agent systems to model and govern the system, developing algorithms that can be used by these agents to manage and capture information from heterogeneous data sources.

Given the characteristics of dynamism on the system environment, the use of organizational MAS (Multi-agent sytem) is proposed, following the approach proposed by the Virtual Organizations (VO) of agents. GORMAS (Guidelines for Organization-based MultiAgent Systems) (Argente *et al*, 2009) is chosen as the design methodology, allowing to model an open-MAS system. This agents include different techniques to improve the capabilities of the system (recommendation, web mining, etc.). Regarding the Cloud Computing (CC) environment, it was decided to use an existing system. In this sense, the Open-Stack platform (OpenStack, 2015) is selected as the deployment environment. This platform allows both to display the set of components, as well as the integration of all the MAS components. In short, the combination of both the MAS technology and the Cloud Computing paradigm, allows to model an intelligent system for journalism with advanced capabilities.

This paper is organized as follows. In the following section the main architecture concepts are described, and in Sect. 3 the proposed architecture model is described. In Sect. 4 the Case study is presented, and Sect. 6 presents the conclusions of this work.

2. Architecture Concepts

Throughout this section, we analyze the key technologies to be used in the platform. Starting from this analysis of the state of the art, it is necessary to take stock of the following steps on the development of the system.

2.1. Artificial societies and virtual organizations of agents

Artificial societies and VO are closely related terms, along this section, the relationship between these concepts will be clarified, according to the existing literature. First, the concept of artificial society is in-

troduced, which has been previously defined by a wide variety of authors (Davidsson *et al.*, 2006). For instance, (Artikis *et al.*, 2001) formally characterize an artificial society according a set of agents, restrictions in the society, a communication language, the roles that agents can develop, and the joint owners of the agents. Nevertheless, the most accurate definition, from our point of view, is the one proposed by Annunziato et Pierucci (Annunziato *et al*, 2003): *An artificial society is defined as a set of interrelated and interacting artificial entities, which are governed by certain rules and conditions.*

Within literature, Davidsson et Johansson (Davidsson *et al*, 2006) propose a classification of artificial societies based on (i) openness, which is the possibility that an agent can join the society; (ii) flexibility, indicating the restriction degree that the agent has, regarding its behavior according to the rules of the society; (iii) stability, which is a measure of predictability of its actions and, finally, (iv) reliance, which measures the degree to which the agent can rely on the society.

Once the concept of artificial society has been defined, the organization concept can be seen as a set of entities which are regulated by social mechanisms, so this entities try to achieve common goals. As with corporations, there is a wide variety of literature on agent organizations (Ferber *et al*, 2004) or (Wooldridge *et al*, 2000). However, from our point of view, the best way to define an organization of agents is attending to its characteristics (Ferber *et al*, 2004): (i) an organization is made up of agents (individual) who express their behavior; (ii) the organization can be divided into organizations that may be overlapped (sub-organizations or groups); (iii) the agents' behavior is functionally related to the organization as a whole, defining its functionality as a role to have within the organization; (iv) agents relate each other dynamically; (v) the types of behavior are related by links between roles, tasks and protocols.

Among these features, the key concept is role, which is an abstract description of the agents' behavior, including obligations, restrictions and skills. Likewise, the role must include a description of the interaction patterns between the different actors in the system.

2.2. Cloud computing

The CC concept has been strengthening both at the enterprise level, as in the research field. For this reason, many definitions have emerged (Mell *et al*, 2011), (Vaquero *et al*, 2008), (Foster *et al*, 2008). In each definition, authors try to highlight those features which, in his opinion, are most relevant. At the end of 2011, NIST (National Institute of Standards and Technology) (Ishida *et al*, 2008), proposes a very accurate definition from a technical and functional point of view: *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.*

In addition, this definition includes 5 mandatory features that any CC environment should have: (i) **automatic on-demand services**, which must be provided automatically, according to the demand, without human interaction; (ii) **availability of services through the network**, customers must access services through the network, and therefore, suppliers must use this mean to provide their services; (iii) **availability of resources**, the provider must have the ability to offer services regardless the demand, reassigning physical or virtual hardware resources dynamically to each service according to demand; (iv) **elasticity**, different resources must be provided elastically and even automatically depending on demand; (v) **tai-**

lored services, provided services must be fully monitored, being the control tasks performed automatically.

As stated above, a CC platform that provides computational support to deploy all the platform components is necessary. As a basis of this platform, we use OpenStack (Openstack, 2015), a Project founded by Rackspace Hosting and NASA in July 2010, which aims to create a free and open-source software platform to create private and public Cloud infrastructures.

3. Proposed system

To design a system with the proposed features using traditional software design techniques would be very complex, which would entail a lot of time, and would give uncertain results. Therefore, it was decided to use an innovative design model, based on VO of intelligent agents.

Overall, this system aims to solve a wide variety of needs of the media: the dynamic and standardized management of information distributed in digital format, automatic layout and organization of the information, dissemination of in-formation via web portals –accessible from any device, regardless of time and place–, agile content management, or 'pay-per-use' features.

In Sect. 3.1 we present the design and components of the MAS, and in Sect. 3.1, the design of the cloud computing system is described.

3.1. The Proposed Multi-agent Model

As stated before, to design the MAS model, GORMAS has been chosen as methodology. Following the phases of GORMAS design methodology, an analysis of the motivation pursued by the organization is done, in order to define the organization (e.g. the reason for the existence of the organization). As the results that are expected to achieve, the environment in which the MAS exists, is defined, detailing the products and/or services which the organization offers, what the groups of interest are, as well as their location. We can see the result of analysis of the organization environment in Fig. 1.

The **mission** of the organization is to minimize the tasks to be performed by the users (in terms of management and production of the information) and maximize the quality of services offered to readers. Overall, the system offers **two basic products**: *a system management* environment and *news*. The system management environment includes a set of services and applications (edition mobile application, management web, activity monitor, cloud management web, activity monitor, cloud management web, and online newspaper) which are aimed to be used by the workers of media companies, so they can manage the underlying system as the content to be subsequently disseminated. The news product represents both the content and services that end users (readers) consume.
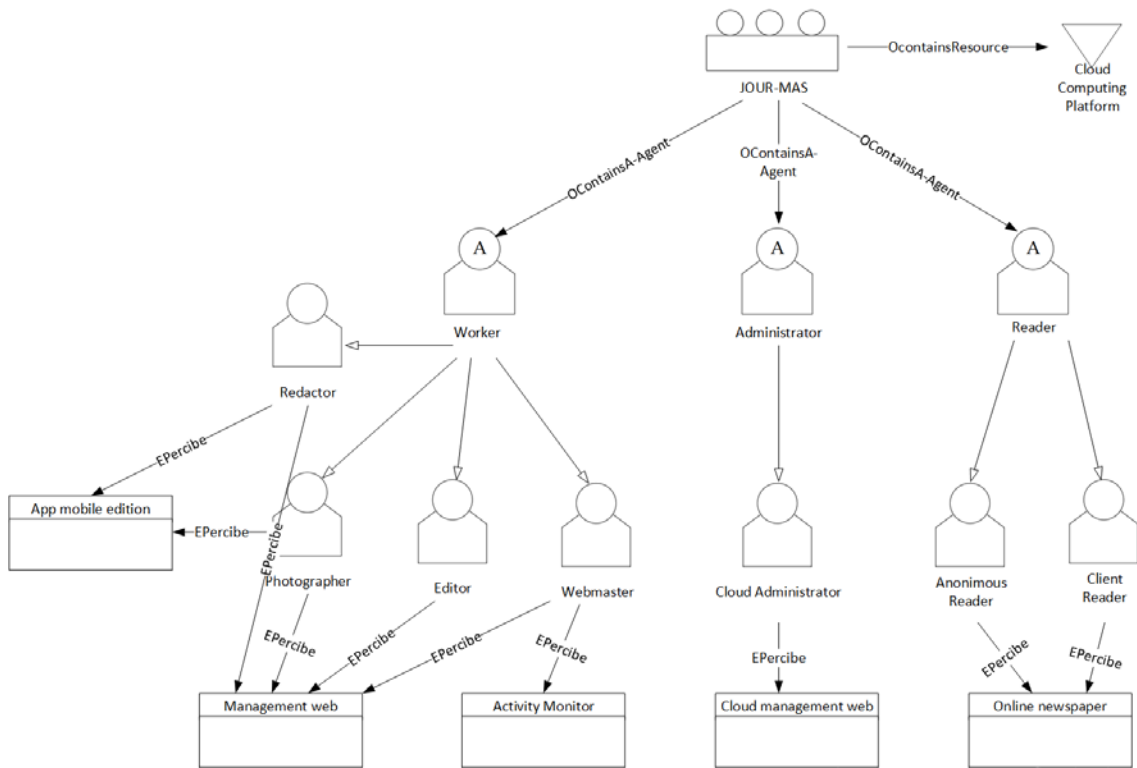
Fig. 1 – MAS-JOUR environment model

First, the interface agents that perform management tasks in the management environment platform, and tasks within the newspaper views and statistics tool are described:

- **Users Manager**: Manages the access to the platform, allowing to create and delete users, as well as defining different access levels to the services offered by the platform.
- **News Manager**: Is responsible to provide a range of services that allow you manage information relating to news, including searching, editing and deleting of news, and the inclusion of any additional content managed in other modules, such as videos, images, maps or comments.
- **Gallery Manager**: Manages galleries, which subsequently allows the inclusion of multimedia content in the news.
- **Modules Manager**: It allows you to manage additional content that may be as-sociated with the news. It can be of different types, such as static HTML content, surveys, videos or maps.

- **News Archive Manager**: It is the role responsible for controlling the retrieval of information from the historical record news. This archive service will enable to retrieve news from two separate databases, a document-oriented one which keeps a record of the previous ten years, and another one with the latest news.

- **Payment Manager**: It is the role entrusted to manage payments of the reader's subscriptions. Every report has a specific value associated to it, and every access to this content by the user is recorded, so the calculation of the price to pay for the use of the system is calculated based on this. Once calculated, the agent receives a payment order, so the payment can be performed through a gateway.

  In addition to the above agents, we have developed a set of independent agents who make use of techniques, tools and advanced features. The design and development of these agents is the basis of the functional algorithms of the system.

- **Comments Manager**: It provides the necessary services to manage the user's comments on the different news. It allows to certain users within the system to determine which comments are published, providing a supervised publishing process. Beside of this services, it also includes a service which will provide an automatic publishing process for the comments, as the result of the analysis of the comments polarity.

- **Content retrieval agent**: Agent in charge, on the one hand, to extract news from a predefined list of media, and on the other, to perform a recovery process of the extracted news. This is done with two objectives: the recommendation of news from other media, and the analysis of possible content that may be plagiarism of contents generated by the media.

  The news extraction process begins Reading the RSS feeds from the defined media, from which several features are extracted, as the deadline, date of publication or the URL of the full content. Since through the RSS, the body of the new is not available, it has been necessary to implement crawlers to extract this information from the URL.

  After getting the news, an XML document is generated for each news item and stored in a directory on which the index of the documents contained in the directory is created using Apache Lucene. Any software agent may use a search service by specifying query terms to be searched using this index, obtaining a JSON response with all the information about the obtained news items.

- **Recommendation provider**: This role implements the web mining process through which content recommendations are made, in response to user interactions with the system. The development of this recommender system is based on Apache Mahout. In this case, we do not have explicit ratings from the users. The used algorithm is within the group of user-based collaborative filtering algorithms, whose similarity model is based on a *Tanimoto coefficient* ratio,–Eq. (1)-., which, given two sets, A and B, with common elements, is expressed as ratio of items that match in both sets, and the total elements of both sets:

$$T_c = \frac{K_{11}}{(K_{01} + K_{10} + K_{11})} \quad (1)$$

Where K_11 is the total amount of elements which are shared by both sets, K_01 the number of elements which belong to B but not to A, and K_10 the number of elements which belong to A but not to B.

- **Marketing manager**: This agent is a component that offers its services to the Online Journal application, which can generate useful information for positioning the different media websites. To do this, this agent makes use of rich snippets, a schema.org convention of the three major search engines (Google, Bing and Yahoo) to label the contents of the websites and facilitate search results.
  The service offered by this agent allows digital publications pages to be indexed easily, thus presenting better quality search results, and making them better positioned and more attractive.

- **Translation provider**: It is the role responsible for providing translation services. Agents that use this services will indicate the text, the language of the text and the language that is to be translated, obtaining the translated text in response to the specified language. The process of translating the texts is done through to the Languaje Grid project API, which builds on the concept of Language Grid (Ishida *et al*, 2008). As a result, the content generated by journalists will be automatically translated into different languages.

- **Storage Manager**: This role is responsible for integrating the persistent information with the cloud computing system. The managed information may have a different nature, as it handles both content stored in the database as multimedia content.

Data storage, both database and multimedia data, is distributed on different machines of the cloud computing platform, so this role will manage the distribution of these data in the corresponding machines, always offering the same communication methods to the other agents, so they can access transparently, regardless of possible changes in the way in which the information remains in the underlying infrastructure.

- **Cloud Manager**: This role is responsible for controlling the cloud computing infrastructure on which operates the platform, allowing: managing computing resources, establish a scalable, and redundant storage system, define the topology of the network through which the different nodes that make up the base infrastructure communicate, or managing cloud computing users who have access to the functionality offered by the cloud manager.

- **Monitoring agent**: This role examines the different behavior of the agents involved users on the platform in order to carry out an analysis of the processes taking place in the organization. It also provides a set of visualization tools to provide all this information to the cloud computing managers.

3.2. Cloud computing platform

This section describes the design of the cloud computing platform that will give computational and deployment support to the components of the platform. As a basis for this platform, an OpenStack based environment has been selected. The objective of this platform is:

- Developing an infrastructure that allows the deployment of virtual servers on demand.

- Ensuring growth or scalability possibilities through the architectural design and the employed technologies. Thus, by increasing the hardware resources, more virtual machines can be supported, so that they make use of higher processing and storage possibilities.

The initial version of the platform will have four nodes: a controller node and three computing nodes. According to the design of the CC platform, Fig. 2 shows all the services or components that will are installed within each node of the cloud platform, as well as the network interfaces used by each node.
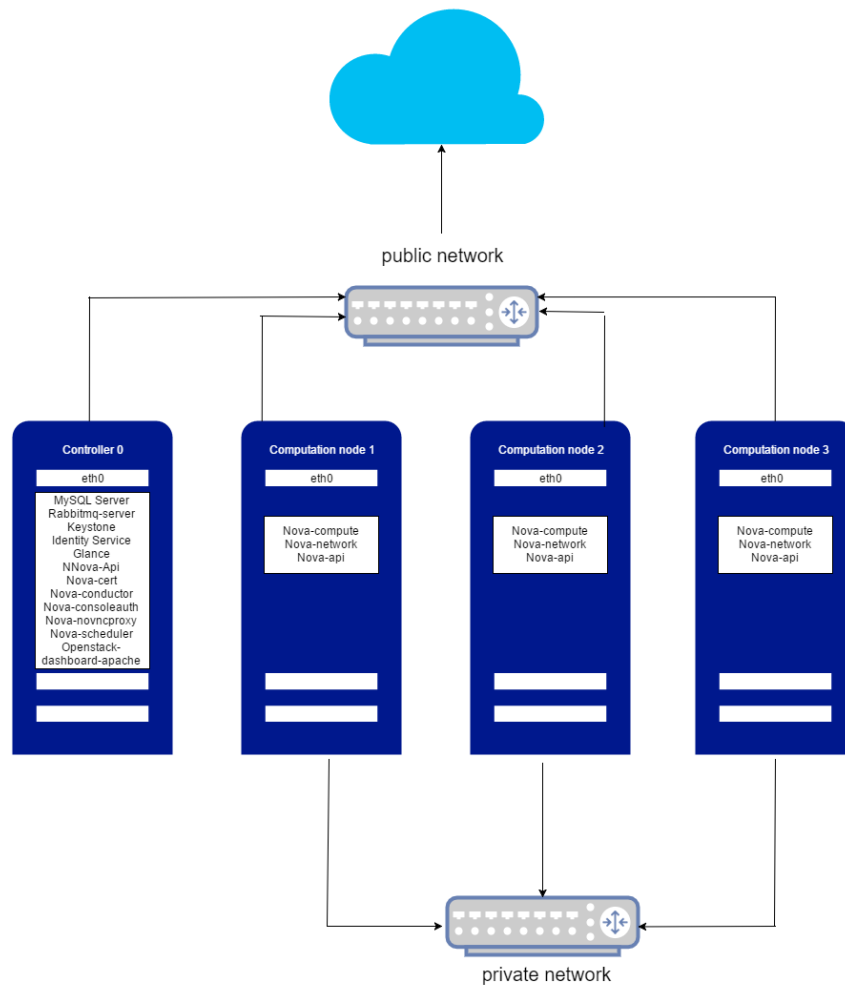

Fig. 2 Cloud computing platform components

In relation to the multi-agent system, the different agents that compose it are distributed in the three computation nodes:

- **Computation node 1** –database server-: storage manager and database manager.

- **Computation node 2** -backend server-: user manager, news archive manager, modules manager, content retrieval agent, news manager, gallery manager, monitoring agent and social media manager.
- **Computation node 3** –front-end server-: comments manager, payment manager, recommendation provider.

4. Case study, preliminary results

The presented system was tested on a Spanish media group, *Promecal*, which gathers various newspapers, both on printed and online editions, such as *Diario de Burgos*, *Diario Palentino.or Diario de Ávila.* This experimental phase aims to validate: (i) the proper operation of the system; (ii) the ability of the VO to model a human organization, by defining roles, tasks and restrictions; (iii) the correct operation of the CC platform that supports the system; (iv) the integration of the different components; (v) the scalability possibilities of the whole system.

In order to perform this validation, an experimental deployment phase has been performed, dividing it into two stages.

- The first stage is a set-up of the system for *Diario de Burgos* online edition, where a group of three journalists, two photographers, two editors, one web-master and one administrator have used the different components of the management system , which embraces the inclusion of new contents on the platform, the edition and layout of the news information, the publishing process, and the final generation of the online newspaper web content.
- Once the operation and integration of the different elements has been proved, a second stage of validation is performed. This stage is focused on validating the scalability of the system . For that purpose, the number of computational nodes has been duplicated, from the three initial nodes, to a total amount of six. At this point, the system gives support to the whole set of newspapers of the media group, including: *La Tribuna de Albacete, Diario de Ávila, Diario de Burgos, La Tribuna de Ciudad Real, Diario Palentino, El Adelantado de Segovia, La Tribuna de Cuenca, La Tribuna de Toledo and El Día de Valladolid.*

  At this point, besides a tenfold increase of the number of workers who have access to the management tool, a set of users have also had access to the online newspaper edition, through the client user role, so the comments manager and the recommendation provider performance could also have been tested.

  In the case of the recommendation system, we have used the activity registered on the production websites of the different newspapers, so we had a valid binary dataset to be used by the recommender system, where two user-based similitude metrics were compared: *Log Likelihood* and *Tanimoto*, being *Tanimoto* the one which presented a lower error when providing recommendations, as shown in Table 1, which presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Algorithm	MAE	RMSE
UserBasedTanimotoCoefficient	0.4836	0.6802

UserBasedLogLikehoodRatio	0.5792	0.8373

Table 1. Results of the errors: MAE and RMSE

5. Conclusions

The appearance of a new fast, cheap and reachable space for exchanging information that internet arising has meant, has brought profound structural changes in journalism, mainly on the way that information and materials are produced, on how their distribution is performed and on the management of companies. Besides of this changes, the production, edition and dissemination processes of their own information has radically been modified.

At this point, and more especially during the last ten years, both employers and workers of the different media groups have intensified efforts to adapt themselves to the new scenario, using Internet and the tools and applications that arise in their environment to improve his position and regain that space of privilege in the information context.

In conclusion at the development of the system, it has led to the development of an innovative functional prototype, with a high technological component that improves the operational capacities of the different companies of media, within the introduced context, that provides a technical solution which is very close to the needs of the media market.

6. References

Annunziato, M., & Pierucci, P. (2003). The emergence of social learning in artificial societies. In *Applications of evolutionary computing* (pp. 467-478). Springer Berlin Heidelberg.

Argente, E., Julian, V., and Botti, V. (2009). Mas modeling based on organizations. In Luck, M. and Gomez-Sanz, J., editors, Agent-Oriented Software Engineering IX, volume 5386 of Lecture Notes in Computer Science, pages 16–30. Springer Berlin Heidelberg.

Artikis, A., Kamara, L., & Pitt, J. (2001). Towards an open agent society model and animation. In Proceedings of the Agent-Based Simulation II workshop, Passau (pp. 48-55).

Davidsson, P., & Johansson, S. (2006). On the potential of norm-governed behavior in different categories of artificial societies. Computational & Mathematical Organization Theory, 12(2-3), 169-180.

Ferber, J., Gutknecht, O., & Michel, F. (2004). From agents to organizations: an organizational view of multi-agent systems. In Agent-Oriented Software Engineering IV (pp. 214-230). Springer Berlin Heidelberg.

Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008, November). Cloud computing and grid computing 360-degree compared. In Grid Computing Environments Workshop, 2008. GCE'08 (pp. 1-10). Ieee.

Ishida, T., Nadamoto, A., Murakami, Y., Inaba, R., Shigenobu, T., Matsubara, S., & Tsunokawa, E. (2008). A non-profit operation model for the language grid. In International Conference on Global Interoperability for Language Resources (pp. 114-121).

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing (draft). NIST special publication, 800(145), 7.

OpenStack documentation page. In: http://docs.openstack.org/ [Last access 20/01/2015].

Rubio, J. L. (2009). La demanda de los periódicos impresos y online: un análisis eco-nómico. Ambitos: Revista internacional de comunicación, (18), 29-39.

Vaquero, L. M., Rodero-Merino, L., Caceres, J., & Lindner, M. (2008). A break in the clouds: towards a cloud definition. ACM SIGCOMM Computer Communication Re-view, 39(1), 50-55.

Wooldridge, M., Jennings, N. R., & Kinny, D. (2000). The Gaia methodology for agent-oriented analysis and design. Autonomous Agents and Multi-Agent Systems, 3(3), 285-312.