Configuration System for the Apache Nutch Spider: Practical Application in the Orion Search Engine

Yulio Aleman Jimenez[1], Yoniel Jorge Thomas Sosa[1], Aylin Estrada Velazco[1], Eyeris Rodríguez Rueda[1]
[1]University of Informatics Sciences, La Habana, Cuba. yulioaj@uci.cu, yjthomas@uci.cu, erueda@uci.cu, avelazco@uci.cu

Abstract– ***The steady increase in the amount of information in digital format public on computer networks around the world, has caused the difficulty of users to find what they really need at any given time. To locate the required information, the Information Retrieval Systems were designed; whose functionalities, have a large number of configuration options and difficult to administer. Apache Nutch is a free spiders with big advantages for collection and finding information on the web; however lacks a system that enables visually configuration without using console commands and conducive working with multiple instances simultaneously. At the University of Informatics Sciences of The Havana, Cuba, Orion search engine was developed, but it has many disadvantages that prevent optimal performance of the process of setting up its tracking mechanism based on Nutch. In this paper are shown the essential elements taken into account in the implementation of a system that improves the usability and makes easy the work of administrators in the configuration tasks. The system implemented, has a set of features and functionalities that contribute, through the availability of web interfaces, increased control of configuration changes and streamlining the process; also providing information on the settings, that previously impossible or difficult to obtain.***

Keywords-- ***Apache Nutch, configuration, Information Retrieval System, Orión, web interface.***

Digital Object Identifier (DOI): http://dx.doi.org/10.18687/LACCEI2015.1.1.027
ISBN: 13 978-0-9822896-8-6
ISSN: 2414-6668

Configuration system for the Apache Nutch spider: practical application in the Orion search engine

Yulio Aleman Jimenez[1], Yoniel Jorge Thomas Sosa[1], Aylin Estrada Velazco[1], Eyeris Rodríguez Rueda[1]

[1]University of Informatics Sciences, La Habana, Cuba. yulioaj@uci.cu, yjthomas@uci.cu, erueda@uci.cu, avelazco@uci.cu

Abstract– The steady increase in the amount of information in digital format public on computer networks around the world, has caused the difficulty of users to find what they really need at any given time. To locate the required information, the Information Retrieval Systems were designed; whose functionalities, have a large number of configuration options and difficult to administer. Apache Nutch is a free spiders with big advantages for collection and finding information on the web; however lacks a system that enables visually configuration without using console commands and conducive working with multiple instances simultaneously. At the University of Informatics Sciences of The Havana, Cuba, Orion search engine was developed, but it has many disadvantages that prevent optimal performance of the process of setting up its tracking mechanism based on Nutch. In this paper are shown the essential elements taken into account in the implementation of a system that improves the usability and makes easy the work of administrators in the configuration tasks. The system implemented, has a set of features and functionalities that contribute, through the availability of web interfaces, increased control of configuration changes and streamlining the process; also providing information on the settings, that previously impossible or difficult to obtain.

Keywords-- Apache Nutch, configuration, Information Retrieval System, Orión, web interface.

I. INTRODUCTION

The increase of the amount of information published on Internet; caused by the indiscriminate use of social networks, the constant generation of knowledge and global scientific publications and the need to save the personal data on the cloud; have induced the loss of accessibility or localization. In various occasions the users don't know how to find that they really need between among much existing documents on the Web.

In Cuba, the Network's Information Cuban Center (CUBANIC for acronym in Spanish) has registered a total of 4 857 unique domains under the superior level domain ".cu", distributed in different generics domains for different sectors of society and that have been delegated to other organizations, for example: com.cu, gob.cu, org.cu, and other [1]. Each one of those domains have a lot of web pages and digital resources that have relevant information.

One of the sectors that give more information, especially in the field of science, is the educational one. The Ministry of Superior Education (MES from its acronym in Spanish) in Cuba has registered 68 educational institutions of superior level that include 3 150 municipal university campuses [2]. Those institutions, in collaboration with other centers of advanced research, are in the front, together with other countries such as Brasil, Argentina, Mexico, Chile and Venezuela; accordingly to production and publishing of scientific results on Latin America and the Caribbean [3]. Besides, is necessary to stand out that in many cases, the institutions have their own web sites or intranets where they share some information of common interest.

The University of Informatics Sciences (UCI), created on 2002 as part of one of the programs of the Revolution for the superior education and the Cuban society computerization, has a privileged technological infrastructure, connecting on its internal computer network a great lot of computers. Thanks to the technological development the university has and the creation of research and develop software centers, are performed various research works and educational and scientific events that generate a great lot of documentation.

According to a study of the evolution of university's Web from the analysis of the webmetrics studies on the period of 2008 – 2011, the UCI had at the time a list around 300 addresses in the domain "uci.cu" and more than two millions of contents published, such as web pages with portables documents, plain text, images and other [4].

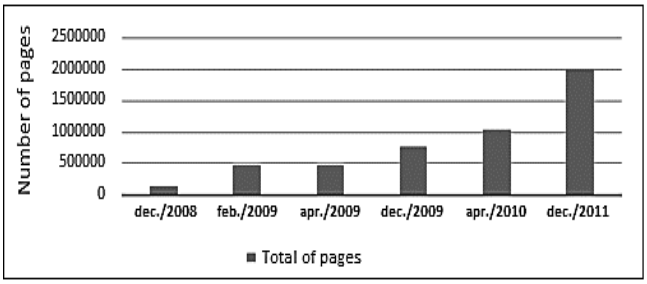

Figure 1: Results of the webmetrics studies on the period of 2008 – 2011 [4].

As shown in Figure 1, the Web of the UCI has been growing quickly during those 4 years at a level that about 90% of published pages were created or updated in this time period. As a consequence, there are difficulties that prevent people to find some information at the just time.

The Orion search engine was developed in the UCI, and at this moment it is deployed in servers of the university and MES. This system offers services of search and information retrieval, allowing the users a quick access to the resources published on

the internal computer network of the university, under the domain "uci.cu" and the network of the MES, respectively; further to have it as proposal as a Cuban search engine for deploying it in the near future in the superior level domain ".cu".

Nutch, as the web crawler of Orion, is an open source software, developed totally with Java by The Apache Software Foundation. It have been developed based on Apache Lucene, a library of high performance for searching based on text and using a modification of Vector Space Model algorithm, with a Boolean approach that restricts the estimates of results, but at the same time does not alter those results respect to the user's consults [5] [6]. This means it does not prevent the explanation of why a web page is occupying a relevant place on results, as do its private homologues with using complex formulas of ranking [7].

The more prominent characteristics of Nutch are the facility for recollection of documents in standard formats of Microsoft Office, Adobe Reader, HTML, TXT, RTF, images, and other. This multiplatform system, allows the execution in parallel of multiples instances at the same time based in its distributed architecture. Further, is easily extensible through development of plugins and highly configurable, reason why it is suitable for a great lot of development projects of information retrieval [6].

However, one of the weaknesses of Nutch that hasn't been solved yet for the international community, is the administration and configuration process through the console of the host operating system, directly on the server where is deployed the informatics solution. This carries out a considerable spending of time and effort in these processes. The connection with the host server is made using the SSH protocol and editing each file separately, this means risks of emerging threats that compromise the system stability and violations of informatics security policy of the university. There is no control of changes in the configuration, this causes high probabilities of occurrence of a critical error in the system.

This robot has more than of 150 configuration options and there are many instances of the same. Each instance doubles options to keep in mind by the administrators; further, they should manipulate many distributed files for all the operating system with a great lot of properties in each one; and it has repercussion in the quality of tracking and obviously, in the searching of users.

The configuration is always made through the administrator's experience and there is not a mechanism to allow to reutilize the acquired experience, transferring it to other persons that assume the task. Is contradictory that with the progress in the Information Technologies and Communication, this process does not perform of a better way.

By the previously expressed, this work is proposing to develop a system to allow management in a centralized manner, through web interfaces, the configurations of Apache Nutch, to reduce the time and insufficient control of changes.

II. MATERIALS AND METHODS

A. *Theoretical fundaments*

With the objective to check theoretically how have evolved on time the Information Retrieval Systems (IRS), analyze the essential elements referent to theories, documents and general literature about these, and specially of tracking mechanisms on Web; was employed the scientist methods historic – logic and analytic – synthetic.

In a performed study was taken as a reference for this work, the definition of IRS exposed by Chowdhury on 2010 [9], this it says textually: *"An Information Retrieval System is designed to analyze, process and store the fonts of information and recuperate those matching with the need of a particular user".* Considering to previous definition and the researches of Baeza-Yates and contributors [10] about the IRS, where his results are complemented with the comparison established by Parrilla [11] on his undergraduate thesis, this kind of system can be classified attending to many criteria according to the function it makes, how these operate, the scope these have, the kind of documents that it can recuperate, and other criteria. However, between the most popular classifications are: the searchers or search engines, the meta searchers and the directories. Considering Orion is a search engine, this work is related with the architecture and functioning of these kind of IRS.

Nutch, as the tracking mechanism of the Orion search engine, is responsible to tour the web through her hipertextual structure, looking for new published documents and progressively storing them in a database of inverted links. Next the contents are indexed by Solr, in an inverted index, whose information can be consulted by the users using a web interface developed with Symfony2.

B. *Precedent Works*

The administration and configuration manually and through the command line is one of more big disadvantages that Nutch has in front of his main competitors. This causes some discouragement and difficulties to users that start on world of search and Information Retrieval.

By these reasons, on 2006, Hanze, Bauhard and Grischupf start the development of an administration web interface for the web crawler. The main purpose was to enlarge the functional capacities of Nutch with a comfortable user interface for monitoring, configuring and administrating of one or more instances of the spider through of an API-REST [12]. Between the main functionalities proposed were: the monitoring of system status and its functions; the configuration of instances

and parameters involved in tours for the Web and the administration of tracking tasks.

Due to various problems of integration between Nutch and first versions of the administration interface, the development was discontinued on 2010, according to some conversations between members of the community and authors, reflected in the project web site and on GitHub[1], where is the last update of the source code [13] [14].

By the reasons previously mentioned and that have a much longer scope than this work, the project previously described is not a possible solution to the problem of this research. However, some functionalities and features of the interface design were taken into account. Among them are: the system organization using tabs, the manage of the instances and them configuration parameters, the history of configuration changes and the manage of the URLs seed.

C. Homologue studies

With the target to find an existing or adaptable solution to the problem previously described, attending to the way to make the configurations of Nutch, was made a study of different search engines or IRS that have mechanism for their configuration. Due that there are a great diversity of this systems in the world, it determined to study those that was developed under a free software model and with a major relation with Orión and Nutch as its spider.

To continue, is exposed on the table 1, a comparison of analyzed systems, with the purpose to illustrate much better the similarities and differences between these and their configuration mechanisms (see Table 1).

According to a performed study, it was determined that there are many spiders that does not have visual interfaces for their configuration; and the interfaces developed for specific systems of this kind, are not potential solutions to the problem described on this work, because each one respond to particular requirements of the web crawler for which they were designed.

However, were identified aspects of special interest that can be used on conception of the informatics solution that is proposed in this work. Among the positive elements to consider are: the distribution of content by tabs, the information architecture, the availability of specialized interfaces for some configuration options and the different ways of interaction between the user and system that contribute to the usability of the application.

Table 1: Comparison between the configuration mechanisms of some spiders.

Spider / Features	**Lucene [15]**	**Open Search Server [16]**	**mnoGoSearch [17]**	**Nutch [6]**
Way for settings	Setting interface	XML files, Setting interface	Command line, Text files Setting interface	XML files
Kind of interface	Desktop	Web	Desktop	Nothing
Usability level	Medium	High	Medium	Low
Content distribution	Menu bar and tabs	Tabs	Menu bar and tabs	Nothing
Setting level	Superficial	Deep	Intermediate	Deep
Platforms supported by the interface	Windows, Mac OS, GNU/Linux Unix	Windows, GNU/Linux, Mac OS, Solaris.	Windows.	Nothing
License	License Apache 2.0	GNU Public License (GPLv3)	GPLv2, Mnogosearch Search Engine For Windows License Agreement	Apache 2.0 License

D. Selecting development technologies

In order to make better use of existing technologies for the development of web application it was proposed the use of Client-Server architecture. Is achieved in this way separating the logic processing of Client of the Server, where the communication is established by a communication protocol based on request-answer (E.g. HTTP, HTTPS).

E. Client side

General technologies of the Web: Due that the system that is to be obtained is corresponding with a web application, is very common the presence of own technologies of this field that are essential and further, that are standards of the Web specified by the W3C [18]. Between them can be mentioned: **HTML** (acronym of Hypertext Markup Language); **CSS** (acronym of Cascade Style Sheet); **JavaScript**, interpreted language of client side [19] and **AJAX** (acronym of Asynchronous JavaScript and XM) [20].

Bootstrap: Developed on its start for Twitter, by Mark Otto and Jacob Thornton; is a CSS framework of a modular structure, compound by a set of style sheets LESS[2] that allow to create interfaces of web sites and web applications using CSS and JavaScript on a simple way, obtaining certain homogeneity in the graphic style [21]. Thanks to it, the system will have

[1] **GitHub:** Online repository of open source code.

[2] **LESS (*Stylesheet Language*):** It is a Dynamic Style Sheet Language that allow to use variables, functions, operators, nested selectors and *mixin* classes to specify the Cascade Style Sheet for web applications.

simple designs, clean and intuitive; and it's going to load easily by different devices and in multiples resolutions.

jqPlot: It is based on jQuery, and it is a strong JavaScript library of open source code and without cost of commercial use, employed for generating statistic graphics. This library will be used for generating graphics that allow to analyze the stability of configurations represented in the changes history of Nutch [22].

F. Server side

Apache 2: It is a web server of open source code that implements the HTTP 1.1 protocol, mainly characterized by its high level of configuration, modularity, stronghold and stability. As Apache2 is free software and multi-platform, of an extensible and modular architecture, it has maintained as the web server most used in the world since middle of 1996 [23] [24] [25]. It will offer the appropriate environment for the deployment of the web application.

PHP: Language of high level and interpreted, used mainly for dynamic processing of information on the Web. Its meaning is conferred to PHP Hypertext Preprocessor, it can be encrusted on HTML documents, but it only can be executed in the server side [26]. It is used thanks to the advantages that it offers on development of web application, its simplicity and speed at the time to code a project and establish its architecture.

Symfony2: PHP Framework based on components created basically to help programmers in the quick development and construction of web applications stronger and with high performance [27]. With this framework, it will obtain a configuration system with a better quality and in less time, thank to this the programmers can focus only on new functionalities.

Doctrine: It is a potent library for ORM (acronym of Object Relational Mapping), that has an Abstraction Layer of Database and allows the simple and dynamic interaction with different database managers. In the development of the application it will employ the ODM (acronym of Object Document Mapping) of Doctrine to communicate with the no relational database in MongoDB.

G. Other technologies

Doctrine MongoDB ODM: It is a library that integrates with Doctrine and provides options for the PHP object mapping for MongoDB. Its main objective is to manage the persistence of its domain model as a way to not interfere in other functions [28].

Doctrine MongoDB Bundle: This package integrates MongoDB ODM with Symfony2, so that it can persist and recuperate objects from and to MongoDB. This library is necessary because Doctrine MongoDB ODM is an external library, and is fundamental to have a mechanism to allow the joint use of it with the Symfony2 functions [28].

MongoDB: It is a system of open source code wrote and oriented to store data with the JSON style with dynamic schemas. The structure is compound mainly by Collections of Documents, compound this last by Fields that store the interested data [29]. It was decided to use MongoDB as a Database Management System (DBMS) because it is a scalable solution and of high performance for information management in front of great data volumes; so it will store the history of changes in the configuration files and other information about configuration profiles.

III. RESULTS AND DISCUSSION

The computer solution that was got provides a set of functionalities aimed to agile the work of the administrators as they gain more experience over time. Next are described the main functionalities to a macro level:

- **Administration of the Nutch instances:**
  It allows the management of the Nutch instances deployed in a same server, which can be directed to track different kind of contents.

- **Administration of plugins:**
  It allows to manage the plugins of each Nutch instance, which have specific functions in the tracking tasks, extracting information of interest about the resources found in the computer networks.

- **Administration of configuration files:**
  It guarantees the correct edition of the configuration properties in the files, which may impact in the tracking quality; allows too, that the properties can be treated in a particular way without the administrator confusing them or losing control about these properties.

- **Administration of changes in the configuration files:**
  It facilitates the suitable control in the configuration files of Nutch, with the target to provide facilities of consulting that will allow understand why a determinate change was done in a given moment; avoiding the excessive use of human memory.

- **Administration of configuration profiles:**
  It allows to apply certain configurations quickly in a given moment and to all the tracking mechanism or one part of this, taking into account previous requirements. With this can be reused the configurations and the work will be expedited significantly.

- **Generation of reports:**
  It offers some functions to obtain reports that will allow the analysis of the changes history, and decision making in a future.

- **Generation of statistics graphics:**
  It allows the obtaining of statistic graphics where is reflected the stability of the configurations about the changes registered by the system; further it provides other kinds of analysis to help the administrators to make decisions on different occasions.

The construction of this configuration system will contribute to expedite the configuration tasks of the tracking mechanism of the Orion search engine; to establish a better control about the historic changes in the configurations and decrease the complexity of this process.

IV. VALIDATION AND VERIFICATION OF THE SOFTWARE

The next stage is known, inside the software development process, as validation of the system and in it, can make different kinds of tests depending of the targets of them [30].

A. Functional testing

To verify the correct functioning of the implemented software, from functional testing through of the black box method, were designed a total of 42 testing cases based on use cases using the Equivalent Partition technique. It was identified a set of 44 nonconformities, distributed in the 3 testing iterations and all them were gradually corrected. In figure 3 can be appreciated in a most illustrative way the behavior of the nonconformities by each testing iteration executed.

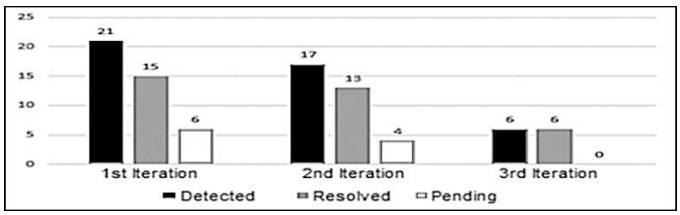

Figure 2: Behavior of the nonconformities by each iteration of functional testing.

B. Security testing

The specialists of the Security Group of the Software Products Evaluation Department (DEPSW for acronym in Spanish), belonging to the National Center of Software Quality (CALISOFT) located in the UCI, have established two main levels for security testing. With the purpose of evaluate this aspect in the first level, it was defined a checking list that establishes 14 indicators categorized in 4 kind of test.

After applyng the checking list to the configuration system, it was obtained the following results for the 4 kinds of tests (see Table 2):

Table 2: Results of applying the checking list for security testing of first level.

Kind of test	Indicators	Corrects
Authorization test	2	2
Session management test	2	1
Checking of authentication system	6	3
Validation of data	4	4

The indicators of the checking list applied that were evaluated as incorrect, were corrected opportunely according to its describing in each indicator; providing greater safety for the implemented system. Among the actions made to improve the security on level 1 are:

- Deactivation of the Symfony2 cache to avoid the access to system after user logout, pulsing the "Back" button of the web browser.
- Establishing an inactivity time limit for users sessions authenticated.

For the evaluation of security in a second level was used the Acunetix Web Vulnerability Scanner 8.0 tool, which is used on the university for detecting the vulnerabilities in web sites and web applications. The results obtained with this tool are the following (see Table 3):

Table 3: Results of the vulnerability scanning with Acunetix Web Vulnerability Scanner 8.0 for security testing of second level.

Categories of vulnerabilities	Quantity
Blind attack by SQL injection (Blind SQL)	0
HTML forms without protection against CSRF attacks	24
Sending passwords by a non-encrypted connection channel (HTTPS)	1
Password fields with auto completion activated	1
Broken links	0
Scripting Cross-site scripting (XSS)	0
Total Vulnerabilities	26

After analyzing the results of the tests, it was proceeded to correct deficiencies found. For this purpose, there were conducted a series of actions in the source code of the application, which helped to strengthen the security of the computer solution. To continue below are exposed a list to some made actions:

- Inclusion of fields with unique identifiers in HTML forms to prevent CSRF attacks.
- Exchange of sensitive data between client and server using an encrypted communication channel using SSL encryption protocol.
- Disabling the auto completion of form fields for username and password.

C. Evaluation of the time of the configuration process

In order to assess the indicator regarding the time of the configuration process, a pre-experiment was performed, demonstrating the use of the scientific method hypothetical - deductive, where it compares the results obtained considering

the use of operating system console and secondly the implemented system.

For this task was taken a sample of 5 people with essential skills in working with the operating system console and also are familiar with the use of web applications. The execution of five basic configuration tasks yielded the following results in minutes (see Table 4):

Table 4: Results of the pre-experiment.

	Console	System
Task 1	0:04:36	0:01:12
Task 2	0:04:36	0:03:12
Task 3	0:03:36	0:02:00
Task 4	0:01:12	0:01:48
Task 5	0:04:48	0:01:24
Average	0:18:48	0:09:39
Standard deviation	0,003622	0,001162

From the analysis of the comparison of averages and standard deviations obtained in the tests, a significant reduction in the time of the configuration process is evidenced by the use of the implemented system. Furthermore, by applying a T Student test, one can affirm with a 95% certainty the veracity of the above results; was obtained 0.004 as a result of this test, verifying the feasibility of the configuration system implemented.

V. CONCLUSIONS

According to the theoretical foundations discussed in this investigation, it can argue that there is many kinds of the Information Retrieval Systems. In most cases, the work with these computer systems is very complex, because it has a great lot of configuration options. This is the main reason why these systems require the integration of certain applications to help their administrators in the configuration tasks, for ensuring greater control of changes and focus on other tasks of higher priority.

Obtaining a configuration system, based on the use of a web development framework, will allow to manage the instances of the web crawler Apache Nutch as part of the Orion search engine. A key feature, is the obtaining of certain information about settings made, previously impossible or very difficult to obtain and that contributes to the statistical analysis to support decision making. Furthermore, applying predetermined parameters in a distributed configuration, may be performed by defining favorite files and configuration profiles.

With help of pre-experimental method, the measurements and statistical calculations, it was checked the feasibility of the implemented system. In addition, evaluation of the tests allowed to identify a set of common problems that threatened the proper functioning of software and were settled in full.

REFERENCES

[1] CUBANIC, "Estadísticas, ¿Cuántos dominios hay bajo .cu?," http://www.cubanic.cu/estadisticas.php.

[2] MES, "Cantidad de centros de Educación Superior en Cuba," http://www.mes.edu.cu/index.php?option=com_content&task=view&id=14&Itemid=30.

[3] C. A. MACÍAS-CHAPULA, "Hacia un modelo de comunicación en salud pública en América Latina y el Caribe," Revista Panamericana de Salud Pública, nº p. 427-438, 2005.

[4] Y. H. MONDELO, et al., "Estudio de evolución de la Web de la Universidad de las Ciencias Informáticas, a partir de los estudios Webmétricos del proyecto productivo Geweb en el periodo 2008 – 2011", in XII Information International Congress, INFO2012, 2012, 11 pp.

[5] I. A. N. MAHECHA, "Buscador web open-source: Nutch," http://dis.unal.edu.co/profesores/eleon/cursos/tamd/presentaciones/nutch.pdf.

[6] APACHE NUTCH, "Official web site of Apache Nutch," http://nutch.apache.org/index.html.

[7] R. KHARE, et al., "Nutch: A flexible and scalable open-source web search engine," Oregon State University, 2004.

[8] S. E. SEKER, "Performance Evaluation of a Regular Expression Crawler and Indexer," http://www.researchgate.net/publication/236622946_Performance_Evaluation_of_a_Regular_Expression_Crawler_and_Indexer/file/e0b4951877458dc6a2.pdf.

[9] G. CHOWDHURY, "Introduction to modern information retrieval," Facet publishing, 2010. 488 pp.

[10] R. BAEZA-YATES, et al., "Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering," 2005. 15 pp.

[11] M, M. PARRILLA, "La Internet que no aparece en los buscadores," Madrid, España: University Carlos III, 2012.

[12] NUTCH WIKI, "Nutch Administration User Interface - Nutch Wiki," http://wiki.apache.org/nutch/NutchAdministrationUserInterface.

[13] GITHUB, "Nutch GUI," https://github.com/101tec/nutch.

[14] THE APACHE SOFTWARE FOUNDATION. [NUTCH-251], "Administration GUI - ASF JIRA," https://issues.apache.org/jira/browse/NUTCH-251.

[15] APACHE LUCENE, "Apache Lucene - Apache Lucene Core," https://lucene.apache.org/core/.

[16] OPENSEARCHSERVER, "Open Source Search Engine | OpenSearchServer," http://www.open-search-server.com/open-source-search-engine/.

[17] LAVTECH.COM CORP, "mnoGoSearch 3.3.15 reference manual," http://www.mnogosearch.org/doc33/.

[18] WORLD WIDE WEB CONSORTIUM (W3C), "HTML/Specifications," http://www.w3.org/community/webed/wiki/HTML/Specifications#HTML.

[19] MOZILLA PROJECT, "JavaScript Overview," https://developer.mozilla.org/en-US/docs/Web/JavaScript/Overview.

[20] A. K. SINGH, "Ajax complexity," International Journal of Engineering Science Paradigms and Researches, vol. 1, no 1, pp. 2319-6564, 2012.

[21] TWITTER BOOTSTRAP, "Official web site of Bootstrap," http://getbootstrap.com/.

[22] JQPLOT, "jqPlot Charts and Graphs for jQuery," http://www.jqplot.com/.

[23] NETCRAFT, "Web servers most used since 1995," http://www.netcraft.com.

[24] INTECO – CERT, "Guía básica para la securización del servidor web Apache," Spain, INTECO – CERT, 2012. 35 pp.

[25] APACHE HTTP SERVER, "Official web site of the HTTP web server Apache 2," http://httpd.apache.org/.

[26] S. S. BAKKEN, et al., "Manual of PHP," The PHP Documentation Group, 2013. 1063 pp.

[27] SENSIOLABS, "The Book for Symfony 2.4," SensioLabs, 2013. 254 pp.

[28] SYMFONY.ES, "DoctrineMongoDBBundle (bundle de Symfony2)," http://symfony.es/bundles/doctrine/doctrinemongodbbundle/.

[29] MONGODB, "Características de MongoDB," http://www.mongodb.com.

[30] I. SOMMERVILLE, "Software Engineering," 7 ed. Pearson Education, 2005. 712 pp.