AN INFORMATION SYSTEM TO PERFORM SERVICES REMOTELY FROM A WEB BROWSER

M.P. Cuellar, M. Delgado, W. Fajardo, R. Pérez-Pérez*
Escuela Superior de Ingeniería Informática, University of Granada,C. Periodista Daniel Saucedo, s/n, Granada 18072, Spain

Keywords: Information System, Botanist, Biodiversity, Multi-Agent System, Semantic Web, Service Ontology

Abstract: This paper presents the development of BioMen (Biological Management Executed over Network), an Internet-managed system. By using service ontologies, the user is able to perform services remotely from a web browser. In addition, artificial intelligence techniques have been incorporated so that the necessary information may be obtained for the study of biodiversity. We have built a tool which will be of particular use to botanists and which can by accessed from anywhere in the world thanks to Internet technology. In this paper, we shall present the results and how we developed the tool.

1 INTRODUCTION

A **herbarium** is defined as a place where collections of dried, classified plants are stored before being used as material for the study of botany. The specimens contained in herbariums are and always have been the essential base for performing systematic, floral and biogeographical studies; in addition, as a collection of perfectly identified and ordered dried plants these represent a permanent record of biodiversity.

It is currently calculated that more than 2.5 billion specimens are to be found in natural history museum collections and herbariums throughout the world (DUCKWORTH, GENOWAYS & ROSE, 1993). Biological diversity research and study requires satisfactory access to this biological information. As this complex information is currently distributed among herbariums all over the world, this makes it practically inaccessible (BERENDSOHN et al., 1999).

Therefore, by its own nature, the herbarium once again becomes an essential piece for the development of these objectives and those in charge of it are responsible for providing the response called for by the research community.
Consequently, one of the prime current needs is to acquire updated, relevant, scientifically contrasted and easily accessible information as the basis for conservation, the handling and the sustainable use of biodiversity. However, the complexity and variability of studies carried out in this field has forced these institutions to adopt new techniques and protocols which are capable of responding to the ever growing demands (BERENDSOHN, 2001).

Actualy, there are some system that solves the herbarium management problem: Hebar (Pando, 1991), Virtual Herbarium Express (The New York Botanical Garden, 1996), Brahus (University of Oxford, 2002). The following characteristics are common to all of the above software packages:

- Free software.
- Use of not particularly powerful database managers (Access, Dbase).
- Data entered using templates.
- Information filtering.
- Label creation.
- Decentralized use of software.

These systems show a set of deficiencies:

- The have no user supervision neither the operations they perform.
- They don't incorporate powerful DBMS that allow a great amount of registries neither concurrency access of the users.
- They don't have a friendly interface.

After analyzing a herbarium's needs, it can be seen that the systems developed so far have not been able to incorporate a large number of requirements. For this reason, BioMen was developed, and by taking

P. Cuellar M., Delgado M., Fajardo W. and Pérez-Pérez R. (2005).
AN INFORMATION SYSTEM TO PERFORM SERVICES REMOTELY FROM A WEB BROWSER.
In *Proceedings of the Seventh International Conference on Enterprise Information Systems*, pages 91-99
DOI: 10.5220/0002512400910099

advantage of modern communication technologies, the information is available online.

2 SYSTEM ANALYSIS AND DESIGN

A herbarium, as we have already mentioned, is a place where collections of dried, classified plants are stored, so that these can later be used as material for botanical study. From this definition, we can highlight the concepts of storage and study. The stored material is studied in order to obtain information which will be used for the conservation, handling and sustainable use of biodiversity.

The need therefore arises for the information to be available in a suitable, standardized form so that it may be studied by researchers.

We shall show some features of BioMen:

- Management of loans (information about specimens that are bellow minimum levels).
- Management of Taxonomic (new information, revisions of folds, …).
- Supervisión of the Users.
- Information consultation (Folds, etc.)
- Multimedia Administration.
- Report Creation.
- Issuing of labels.

Another of the most important points for a herbarium, besides administration, is to be able to satisfy the demands of biodiversity studies. These studies use databases, and this can at times pose quite a complex task because of the way the data is displayed. We should therefore point out that the system must provide a series of services so that studies may be obtained about:

1. Specific richness (this is the number of species in a certain region or location)
2. Taxonomic complexity (complexity when it comes to identifying a specimen)
3. Study of the alpha/beta/gamma diversity (diversity within the habitats: alpha diversity; between the habitats: beta diversity; and for all the habitats being studied: gamma diversity) (Rosenzweig, 1995).
4. Orientation in the collection campaigns.

Among all the possibilities which currently exist to tackle the problem, the convenience of information systems was thought of because of the intrinsic nature of the problem. According to Henry C. Lucas (1987), we can define *information system* as a set of organized procedures, which when performed, provide information for decision-making and/or control of the organization. The general theory of systems on which the information system analysis and design is based, indicates that it is necessary to consider the system to comprise smaller subsystems. The connection of the smaller systems with the larger systems forms a hierarchy which is characteristic of the theory of systems. It also shows us that we must have an overall view of the system, knowing that all the system components are interrelated and interdependent, with this being one of the most important tasks.

We can therefore say that BioMen is an information system with a client-server architecture (Figure. 1) designed for herbarium management. Researchers and those interested in this subject matter can gain online access to a virtual center which models the real behavior of the units which comprise the research center, and they are able to obtain all the information offered totally dynamically. The virtual users request (remote and/or hybrid) services which will enable them to perform all the intended operations within the system.

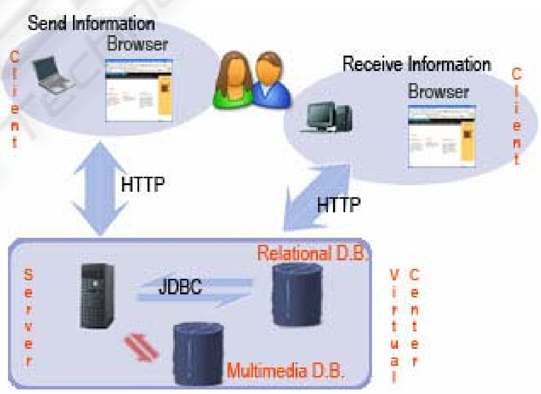

Figure 1: Client/Server Architecture of the System

BioMen offers a series of services which enable the users to have:

- All the centralized information
- Security protocols
- Greater computational power

The services offered might be:

1. Remote services: remote execution of processes and return of the results to the user.

2. Hybrid services: interaction between local and remote processes. e.g. integration of barcode readers.

The majority of the services are remote, although there are some multimedia services which will need hybrid services (remote image processing and inclusion).

As BioMen needs a representation of the domain knowledge, our system uses a service ontology described by means of the DAML+OIL terminological system, and the services are described using OWL-S (OWL-S is an OWL-based web service ontology). This enables us to organize the services on a graph and to provide a description of the services including the characteristics of each service (Figure 2).

The ontology is used by the system to enable the user to select the desired service, provide the necessary parameters, and the system is therefore in charge of executing the selected service. Due to the characteristics of the system, implementation has been carried out using agent technology (Figure 3).

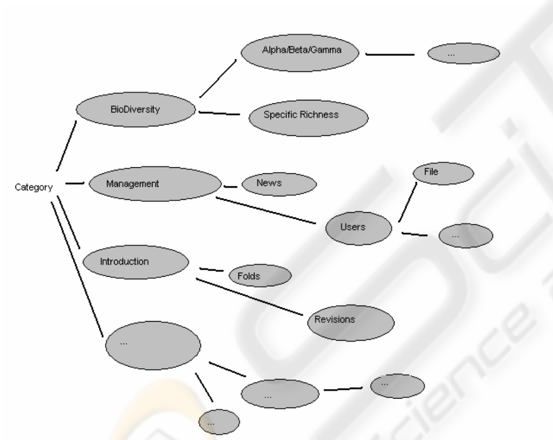

Figure 2: Hierarchy of services

In the last 25 years, we have seen the appearance of several paradigms to design software systems such as procedural programming, structured programming, object orientation and component-based software. Agents (Weiss, 1999; Wooldridge, 2002) are now being championed as the next generation paradigm to design and build complex and distributed software systems. An agent-based architecture provides additional robustness, scalability, flexibility, and is particularly appropriate for problems with a dynamic, uncertain, and distributed nature. In particular, they seem to be the ideal computational model for developing software for Internet, and open networked system with no single controlling organization (Jennings, 2000). Lastly, architectures allow the incremental development of modular systems not only because of the modular nature of the agents, but also because of the possibility to incorporate legacy code by wrapping it within an agent interface.

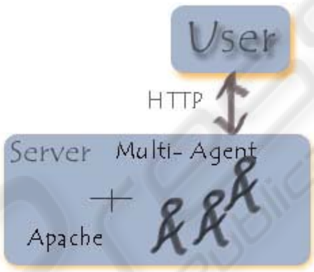

Figure 3: Way system acts given user interaction

In a multi-agent system (MAS), agents interact with one another to achieve their individual objectives by exchanging information, cooperating to achieve common objectives, or negotiating to resolve conflicts. Alternative flexible patterns of interaction have been used such as the Contract Net Protocol (Reid & Smith 1980), where a task is advertised by a coordinating agent and is assigned to the agent that makes the best bid. However, details of all possible interactions between agents cannot be foreseen *a priori* and consequently:

1. Agents need to be able to make decisions about their interactions at run-time, and
2. the organizational relationships between agents need to be represented explicitly (e.g. peer member in a team, manager, coordinator) by means of constructs such as roles, norms, and social laws.

An agent is anything which can be observed sensing its environment using sensors and acting on this environment by means of effectors/actuators. The programming language which has been used is Java, enabling us to include a greater number of mobile devices and operating systems.

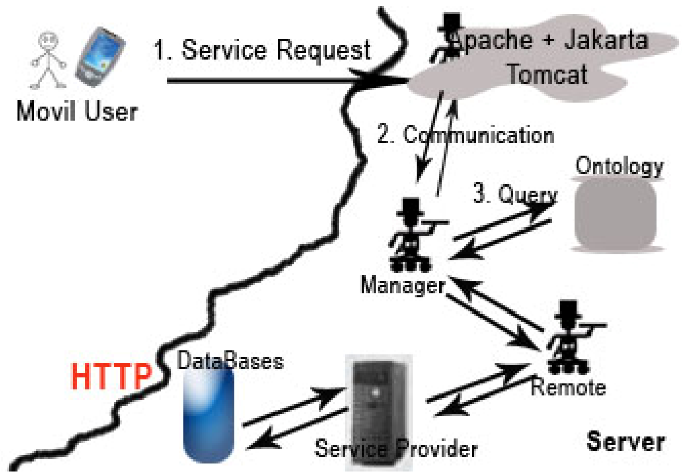

Figure 4: Way of action in a remote operation

The way the user interacts with the user is even simpler. The user interacts with the server using the HTTP protocol, performing the operations desired by means of a totally pleasing interface and without needing to have any additional tool installed. Once the web server has gathered the user's request, it interacts with the multi-agent system in order to carry out the service requested by the user (Figure 4) and returns the results of the service. The multi-agent system is made up of the following agents:

- User
- Request manager or coordinator
- Service execution agents:
  - i. Remote
  - ii. Hybrid

The agents which form the multi-agent system use a blackboard architecture (Nii 1986[a], 1986b)(Hayes-Roth,1985)(Kowalski & Kim, 1991) for communication. The blackboard is implemented by a series of tables. The agents use the blackboard to exchange the necessary information.

Having looked at the operational logic of the system, we shall describe the characteristics of the system and how the requirements of a system for a herbarium have been solved.

2.1 From the client's point of view

As we have already mentioned, the client communicates with the server using the HTTP protocol in order to execute the desired (remote and/or hybrid) services. In addition, the HTTP protocol will show the desired information and enter the parameters.

There are two ways to access the system:

- without identification: the system is accessed as an Internet user and therefore access to the information is severely limited.
- with identification: the system controls the different authenticated user types using login/password and allows more or less sophisticated services to be carried out according to the level of security allocated by the managers.

The identified users access the system from the authorization window (see Fig. 6), beginning a new session. Having been identified in the system, menu systems are created which show the services allowed according to the level of security (see Fig. 7). Through the menu systems and the I/O interfaces, the system will receive requests and will provide the user with the requested information.

Figure 5: Different options according to the user

2.2 From the server's point of view

From the server, pages are dynamically generated for each of the users, enabling all of the services required of the center to be performed. By maintaining a client/server structure, we provide solutions to the location problems which have previously been mentioned. Therefore, the server will act as a virtual center enabling as many services as those allowed to each user by the center managers.

All these services are carried out and are managed by the multi-agent system totally transparently to the user so that a dynamic system is obtained with excellent features from the client's and the server's point of view.

Bellow we shall describe a few of the more important tasks which can be carried out within the virtual center.

2.2.1 Multimedia management

As we have already mentioned, one of the features of the service which distinguishes it from others is the incorporation of a multimedia service which enables the user to obtain much more detailed information. There is a multimedia element associated to each fold, for example, fold images, video of the habitat, etc. which any authorized user can consult. This service would require the execution of local services (reproducer, image viewer, etc.).

From the point of view of information management and incorporation, users can pre-process the information which they wish to incorporate for a given fold. For example, scanning the center's folds and associating an identifier with the scanned image, recording a video about the collection of the specimen in the field, etc.

This is one of the solutions offered by the system to improve the work of the center's staff. This multimedia service is available to all system users. In this way, the multimedia consultation of a specimen is made possible without the need for a corresponding loan request. Therefore,

1. The researcher can consult the specimen's multimedia information the moment the center's staff carry out the operation in the system.
2. There is a reduction in the number of loans which the center must make to the researchers.
3. In turn, there is a better conservation of the center's material.

2.2.2 Advanced consultation and/or consultation of folds

If we make a consultation using the fold's identifier, we will only obtain information about that fold, but what happens with the information contained in these? From the researcher's point of view it is much more necessary to be able to make a consultation using the information contained within a fold than for the existence of a fold. For example, existing specimens for a UTM and/or above a certain height, etc.

This is, as we have already mentioned, the main distinction between a library and a herbarium, in that the information which is useful to the researcher is the information which there is within the fold and not the fold in itself. It is therefore as if we were asking about the information contained in each book in a library. From this service, any type of information existing in the system can be searched for, and the result can be obtained both in HTML and PDF so that it can be easily exported. The online access to the information when consultations are made bestows the power that the herbarium staff and researchers need.

From the advanced consultation, the center's staff can begin to make a loan to the center requesting it. If the center needs to lend all the folds containing the specimen *Pinaceae Pinus baciano*, in a normal, non-computerized process, the staff would need to go to the storeroom, look through the folds one by one in order to select those requested by the researcher. If the system is used, the fold identifiers containing this specimen can be obtained and in turn, the loan service of the system can be activated merely by entering the loan recipient's data and recording this in the system.

2.2.3 Treatment for biodiversity

By means of a series of remote services, the user can request information about:

1. Taxonomic complexity (Magurran, Moreno)(Halffter et al)
2. Specific richness (Magurran, Moreno).
3. Orientation in collection campaigns
4. Study of the alpha/beta/gamma diversity

These remote services show the user the desired information, using the information provided by other agents who are constantly processing the data contained in the databases.
We shall now see how this information can be obtained and we solve the existing problems.

3 BIODIVERSITY

As we have mentioned already, there is a large amount of interesting information in the center's databases. This information enables important improvements to be made in the quality of botanists' work.

Nevertheless, the information is not usually directly accessible since it needs to be processed from the database. As a first approach to the solution of this problem, we can recover and process information in order to obtain new knowledge and determine:

1. Taxonomic complexity (Magurran, Moreno)(Halffter et al)
2. Specific richness (Magurran, Moreno).
3. Orientation in collection campaigns
4. Study of the alpha/beta/gamma diversity

The main disadvantages of obtaining the taxonomic complexity are:

1. Existence of a large volume of data
2. Redundancies present in the information
3. The existence of synonyms in the database

Because of these problems, it is not possible to perform the taxonomic complexity studies directly. In order to look at this problem in more detail, we shall consider the following example:

Below we shall show the identification of a small sample of the specimens contained in the database. The specimen's name, in this case, comprises the family, genus and species:

- Cruciferae Alyssum spinosum
- Cruciferae Hormathophylla spinosa
- Cruciferae Ptilotrichum spinosum

If we want to know the number of different specimens, when the count is made in the database, we would obtain 3 specimens. However, according to Flora Ibérica (1996), the three names refer to the same specimen (Cruficerae Hormathophylla spinosa). In addition, the order by which the name (identification) has evolved (Alyssum → Ptilotrichum → Hormathophylla) is established. For this reason, as we mentioned before, there are synonyms in the database. This makes it impossible for us to obtain the information necessary for biodiversity studies (different number of specimens in one area e.g. for specific richness studies).

Below, and in view of the importance which the problem of synonymy has within the research center, we shall attempt to resolve the problem. In order to do so, there are two possible courses of action:

1. **By creating a synonym database**. This alternative accelerates the processing work. However, it offers a series of drawbacks:
   a. The size of the synonym table is very large, since there is a great variety of species.
   b. The table would have to be compiled by an expert. The expert would have to carry out a repetitive and tedious task.

2. **By studying the evolutions**. The name we give to the change in the denomination of a specimen is evolution. We shall explore this in greater depth later. This task can be carried out without an expert having to intervene and enables us to obtain the sequence of the change in the identification.

Another piece of extremely interesting information relates to orientation in the collection campaigns. This provides the center with advantages both in terms of finances and documentation. The idea is to provide information about the types of specimens needed for the center to be complete and well represented. For example, if the number of specimens in the center is low, it might be that:

1. there really are few specimens.
2. the specimen has been lent to other research centers.

It is therefore necessary to inform the center of the specimens which need to be collected so that the center is complete and well-represented. The information might be:

1. specimens which are not particularly represented and/or below minimum levels.
2. the best path to follow in order to collect the specimens.

In order to obtain this knowledge, three intelligent agents have been used (according to Wiener's definition of intelligence) which will act in turn within the multi-agent system described above. These agents would constantly be observing the media (databases), acquiring and processing the information in order to achieve the necessary information. The agents deposit the information in the system, using the blackboard, so that the users who so desire can access it by means of the previously described corresponding services.

The first of the agents, called the **revision agent**, is responsible for studying all the revisions for a specimen. This result is taken advantage of by the agent called the **specific richness agent.** This agent obtains the set of synonyms contained in the database. This information is necessary in order to count the different specimens which there are in the database. In turn, the information obtained by the two previous agents is used by the agent called the **collection campaign orientation agent**. This agent issues a report of those little represented specimens in the center.

We found the solution to the synonymy problem by studying the evolutions which a certain specimen goes through. We shall present the concept using an example (see Figure 6):

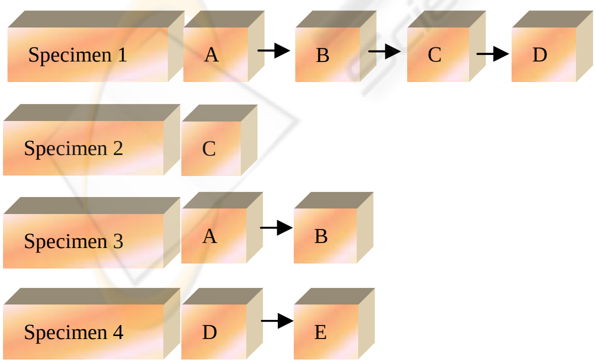

Figure 6: Example of Synonyms

The different blocks determined as A, B, C, D and E are the different determinations through which a specimen passes. The arrows indicate the evolution to another determination.

Specimen 1 was first determined as A but is later determined as B, C and D. These determinations are processed as evolutions and therefore, the determinations, A, B and C are synonyms of D. These values are inserted into ***EvolutionAlert*** table**.** This table is managed by the agent. The following information is entered:

Specimen 1: A→D, evolution 0.
Specimen 1: B→D, evolution 0.
Specimen 1: C→D, evolution 0.

The agent enters the following fields:
- *Taxon*: indicates the number of the specimen. In this case 1.
- *Antecedent*: Previous denominations of the specimen, for example A, B or C.
- *Consequent*: indicates the denomination of the evolution. In this case D.
- *Evolution.* This field can take 2 values:
  · *evolution* is 0: this indicates that the specimen will not be determined in another way.
  · *evolution* is 1: this indicates that if the specimen is studied, it may change its determination to the one indicated by the Consequent field. This enables us to inform the botanist of the specimens to be revised so that the center material is totally updated. This information centers on the revisions which have already been made to other specimens.

We shall now put this knowledge on the blackboard, and in particular, the *taxones* table, so that it may be consulted by other agents. The following taxon is therefore entered in the table:

Specimen 1: D.

Specimen 2 is a synonym as there is another specimen in which it has gone from a state C to a state D.

We would therefore enter the following in the *EvolutionAlert* table:

Specimen 2: C→D, evolution 1.

In this case, the evolution value to 1 indicates that if Specimen 2 were studied, it would probably be determined as D. We therefore enter it as a possible evolution.

The following Specimen, Specimen 3, has some revisions determined as A and B. As in the *EvolutionAlert* table, the fact that state B can pass to state D is stored, due to the sequence of evolutions which Specimen 1 possesses. We therefore add the following tuple to the *EvolutionAlert* table:

Specimen 3: B→D, evolution 1.

We should remember that in the *taxones* table, there is only one tuple, which indicates that of the specimens studied, there is only one different taxon.

In the next specimen, number 4, we can see that state D is revised to state E, and therefore state D becomes a synonym of E and consequently, it is necessary to revise the data stored in the *EvolutionAlert* table, and in turn, to update the taxons identified. The tables would therefore remain as follows:

- *EvolutionAlert* table:
  - Specimen 1: A→E, evolution 0.
  - Specimen 1: B→E, evolution 0.
  - Specimen 1: C→E, evolution 0.
  - Specimen 1: D→E, evolution 1.
  - Specimen 2: C→E, evolution 1.
  - Specimen 3: B→E, evolution 1.
  - Specimen 4: D→E, evolution 0.
- *taxones* table:
  - Specimen 4: E.

Having taken a general look at the example, we shall use the data which we have shown previously in order to see how the agents would act:

- Fold GDAC2745:
  - Revision 1: Cruciferae Ptilotrichum spinosum (L.) Boiss.
  - Revision 2: Cruciferae Hormathophylla spinosa (L.) Küpfer
- Fold GDA28909:
  - Revision 1: Cruciferae Alyssum spinosum L.
  - Revision 2: Cruciferae Alyssum spinosum L.
  - Revision 3: Cruciferae Ptilotrichum spinosum Boiss.

When the multi-agent system acts, we would obtain the following final situation:

AlertaEvolución Table:

- Fold GDA28909: Cruciferae Alyssum spinosum → Cruciferae Hormathophylla spinosa, evolution 0.
- Fold GDA28909: Cruciferae Ptilotrichum spinosum → Cruciferae Hormathophylla spinosa, evolution 1.
- Fold GDAC2745: Cruficerae Ptilotrichum spinosum → Cruciferae Hormathophylla spinosa, evolution 0.

Taxa table:

- Fold GDAC2745: **Cruciferae Hormathophylla spinosa**

As we can see, we obtain the desired result without needing to produce any table which contains these synonyms and which would involve a great deal of work for the specialist.. So far, all the solutions provided for the synonymy problem have involved the construction of the synonym table Species2000. We therefore believe that we have provided an easy and innovative way for the researcher to obtain very important information which does not entail any expense for the center using it. This information determines the taxonomic complexity and the richness of species for any area and consequently enables biodiversity studies to be made.

If we combine the information obtained by the existing agents, we can obtain yet more advantages. We obtain the necessary specimens to be collected so that we may have a complete center. The systems issues reports filtering the synonymy problems.
In order to achieve a complete center, both on a geographical level and in terms of plant groups, we cross the information obtained by the agents with a *Geographical Information System*. This fact offers a number of advantages which will enable the center to obtain better results from the information provided by the multi-agent system.

4 CONCLUSIONS

In this paper, we have described our experiences of constructing BioMen, an information system executed on Internet and developed for herbariums. The constructed system incorporates all the center's needs, uses a multi-agent system which makes the system much more dynamic and easy to maintain. In addition, this is done entirely independently of the user who does not need to know how an ontology operates or how the agents must communicate with one another.

BioMen is a totally operational system which uses the newest technologies:

- Access to the system by means of a web browser.
- *Java™ Servlet technology* (Sun Microsystems Corporation, 2001; Hall, 2001a,b)
- *Apache Server* (The Apache Software Foundation, 2001a)
- *Apache Jakarta Proyect Tomcat* (The Apache Software Foundation, 2001b)
- *Java Agents.*
- JDBC for communication with the databases.

The majority of software tools are free and this makes the system much more attractive and enables it to be more standardized.

ACKNOWLEDGEMENTS

This work was founded through Project TIC2003-08807-C02-01 from the Spanish Ministry of Science and Technology (MCYT).

REFERENCES

Berendsohn W., Anagnostopolous A., Hagerdorn G., Jakupovic J., Nimis P., Valdés B., Güntsch A., Pankhurst R., White R. (1999). A comprehensive reference model for biological collections and surveys. *Taxon 48: 511-562.*

Berendsohn W. (2001) Biodiversity Informatics Available online at http://www.bgbm.org/BioDivInf/def-e.htm

Castroviejo et al. (1996). Flora Ibérica. Real Jardín Botánico, CSIC . *Vol II Cruciferae-Monotropaceae. Pp. 193-195.*

Duckworth, W.D., Genoways, H.H. & Rose, C.L. (1993). *Preserving natural science collections:chronicle of our environmental heritage*. Washington, D.C.:iii+140 pp.

Halffter, Moreno y Pineda, Manual para evaluación de la biodiversidad en Reservas de la Biosfera, *MT&SEA.*

Hayes-Roth. A Blackboard Arquitecture from Control", *Artificial Intelligence, 26(3):251-321, 1985*

Henry C. Lucas, JR. (1987). Sistemas de Información, Análisis, Diseño y Puesta a punto. *Paraninfo.*

Jennings, N.R. (2000). On agent-based software engineering. *Artificial Intelligence, 177, 277-296.*

Magurran. Diversidad. Ecología y su Medición.

Moreno. Métodos para medir la biodiversidad, *MT&SEA*.

Nii, (1986a). Blackboard Systems: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures.

Nii, (1986b). Blackboard Systems (Part Two): Blackboard Application Systems, Blackboard Systems from a Knowledge Engineering Perspective.

Louis, R., (1999). Software agents activities. In *ICEIS'99, 1st International Conference on Enterprise Information Systems*. ICEIS Press.

Pando, F. (1991). El Herbario de Criptógamas del Real Jardín Botánico y sus bases de datos. *IX Simposio Nacional de Botánica Criptogámica, Salamanca, 1991.*

Reid, G., & Smith, T. (1980). The contract net protocol: high load communication and control in a distributed problem solver. *IEEE Transactions on Computers, 29(12), 1104-1113.*

Species2000, Available at: www.species2000.org, *last visited September 26th , 2004.*

Sun Microsystems Corporation (2001). Servlet Specification v 2.3. Available at: http://www.sun.com/servlet/, *last visited December 26th, 2003.*

The Apache Software Foundation (2001). Apache HTTP Server Version 1.3.29 Documentation. Available at: http://www.apache.org, *last visited December 26th, 2003.*

The Apache Software Foundation (2001). Apache Jakarta Project Version 4.0 Documentation. Available at: http://jakarta.apache.org, *last visited December 26th, 2003.*

Wooldridge, M. (2002). An introduction to multiagent systems. *New York: Wiley*.