

BioMen: an information system to herbarium

M. Delgado, W. Fajardo, E. Gibaja, R. Pérez-Pérez*

*Department of Computer Science and Artificial Intelligence, Universidad de Granada,
E.T.S.I. Informática. C\ Periodista Daniel Saucedo Aranda, 18071 Granada, Spain*

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. The services are managed by means of a Multi-Agent System, i.e. an Input/Output system, which interacts with the web server. 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 be accessed from anywhere in the world thanks to Internet technology. In this paper, we shall present the problems we encountered when building this tool and how we managed to overcome them.

© 2004 Published by Elsevier Ltd.

Keywords: Information system; Botanist; Biodiversity; Multi-agent system; Semantic web; Service ontology

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).

After a time when the declining interest in global questions entailed a certain neglect of floral studies,

and consequently of the work of herbariums, the current preoccupation for the deterioration of the environment has resulted in new interests being defined which has again brought to the forefront the concern for the particular composition of our surroundings. The study and conservation of plant biodiversity is undoubtedly one of the most important challenges facing botanists, naturalists, environmentalists, etc. in the next millennium. Let us take the example of GBIF (Global Biodiversity Information Facility) (GBIF, 2004), an international project which aims to make all important data about biodiversity freely and universally available on Internet. The approach of GBIF shall contribute to economic growth, ecological sustainability, social effects and scientific research by increasing the usefulness, availability and completeness of the main scientific source of available biodiversity information on Internet. Most of the information necessary for GBIF comes from information stored in the herbariums all over the world.

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,

* Corresponding author. Tel.: +34 95 824 9562; fax: +34 95 824 6251.

E-mail addresses: mdelgado@ugr.es (M. Delgado), aragom@ugr.es (W. Fajardo), gibaja@decsai.ugr.es (E. Gibaja), ramon@decsai.ugr.es (R. Pérez-Pérez).

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).

One first approach at tackling this problem of management was developed by Pando (1991) with the Herbar program. Due to the fact that the management problem is one shared by all herbariums, individual software packages have been developed for each institution in order to solve their problems, and there is no single, duly developed, standard solution. Generally speaking, the solutions advocated by each herbarium are partial solutions for their problems and have been developed by the herbarium's own staff with little experience of system development.

Below, we shall present some of the most important software packages and their features:

1. *HERBAR* (Real Jardín Botánico de Madrid, 1991): free software, entirely developed in Microsoft Access. This reflects some of the special characteristics common to herbarium management such as for example, the entering of information, genera, countries, provinces, information filtering. It creates labels.
2. *Virtual Herbarium Express* (the New York Botanical Garden, 1996): free software which enables data input by means of forms created in Microsoft Access XP. This software allows labels and certain reports to be created.
3. *BRAHUS* (University of Oxford, 2002): free software developed in Visual FoxPro using DBASE as the databases. This is one of the most complete software packages and allows a better control of information, printing of labels, user policy, etc.

The following characteristics are common to all of the above software packages:

1. Free software.
2. Use of not particularly powerful database managers (Access, Dbase).
3. Data entered using templates.
4. Information filtering.
5. Label creation.
6. Decentralized use of software.

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 advantage of modern communication technologies, the information is available online. The center's information can therefore be consulted without having to be requested from the center itself, in line with the philosophy of GBIF (2004) whereby all information should be public knowledge. This system administers all herbarium tasks.

In particular, this has been implemented in the prototype stage in the herbarium at the University of Granada. This center was chosen for its complexity and completeness. It meets all the characteristics to be modeled and presents all the problems to be resolved of an extremely important center historically among herbariums. The University of Granada's herbarium currently belongs to the executive of the Iberian and Macaronesian Herbarium Association.

After this brief introduction, we shall go on to explain the work carried out. In Section 2, we shall present an analysis of the problem. In Section 3, we shall show the characteristics of the developed system. In Section 4, we shall show how the information about the treatment of biodiversity has been obtained. Finally, Section 5 ends with a series of conclusions about the system.

2. System analysis

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 (Lane et al., 2000).

The need therefore arises for the information to be available in a suitable, standardized form so that it may be studied by researchers. These specimens are constantly consulted both by researchers from the same center and from other centers, so that relevant studies can be made, i.e. hot-biodiversity, specific richness, etc.

The specimens stored in the center come from the collection campaigns carried out by the center and from exchanges with other centers. Once the specimen has been collected, it undergoes a process of preparation in folds and of identification. The identification process reflects the classification of the specimen and how it will be entered into the system databases.

The specimens are lent to other researchers so that these may carry out their research work. These loans must be administered by the center's administrative staff. It is therefore necessary to be able to control the loans made and their period of validity. Lending material entails:

1. The possibility of making a discovery in the samples which the center makes available to researchers.
2. Time being wasted for the center staff in charge of administering and controlling these loans and returns.

Once the material has been studied, the researchers issue rectifications, if necessary, about its identification. Because of the complexity of nature, there are samples for which there are discrepancies in the identification or simply due to changes in the way the specimen has been identified. Consequently, the rectifications issued must be entered by

the center's operators, indicating that the specimen has been updated.

Due to the value of the information in these centers, the problem arises of restricted access for users to certain types of information. As a result, different levels of access are created and each user can be allocated a level thereby enabling or refusing access to certain types of information.

Having arrived at this point, we can see a series of characteristics which the system must offer in order for the objectives marked by a herbarium to be fulfilled from the administrative point of view:

1. Information administration (introduction of folds, revisions, genera, etc.)
2. Information consultation (folds, etc.)
3. Multimedia administration
4. User administration
5. Report creation
6. Issuing of labels
7. Loan control

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.

This would entail a complex treatment of the data since this information is not directly accessible and would require a data acquisition and processing process in order to reach the desired objectives (Research Directions in Biodiversity and Ecosystem Informatics, 2001).

3. System design

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 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

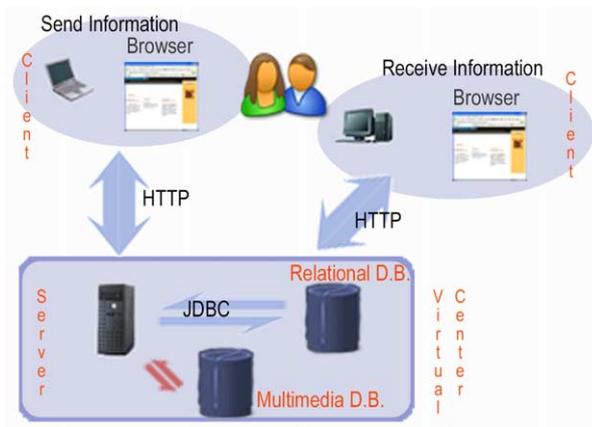


Fig. 1. Client/server architecture of the system.

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 (Fig. 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.

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

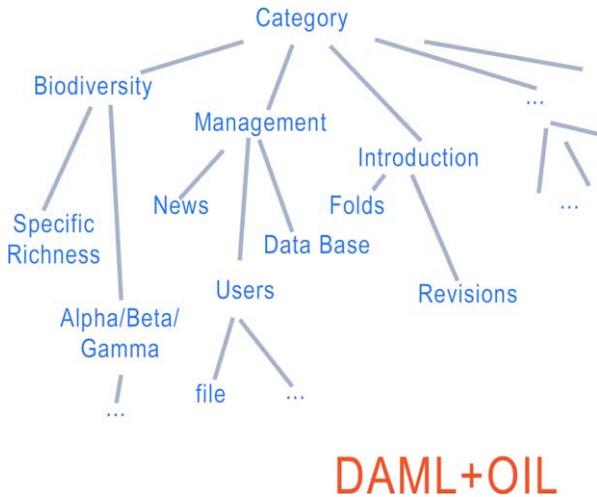


Fig. 2. Hierarchy of services.

organize the services on a graph and to provide a description of the services including the characteristics of each service (Fig. 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 (Fig. 3).

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,

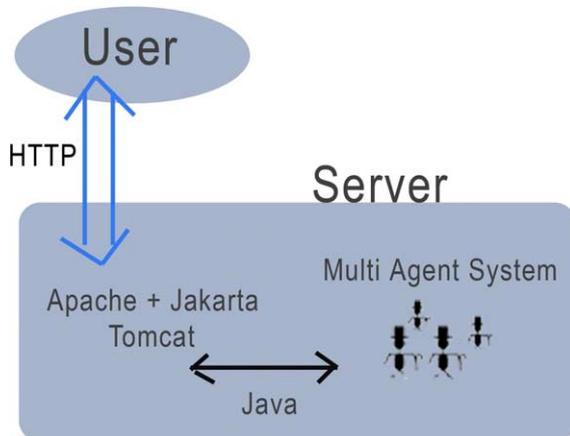


Fig. 3. Way system acts given user interaction.

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.

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.

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 (Fig. 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

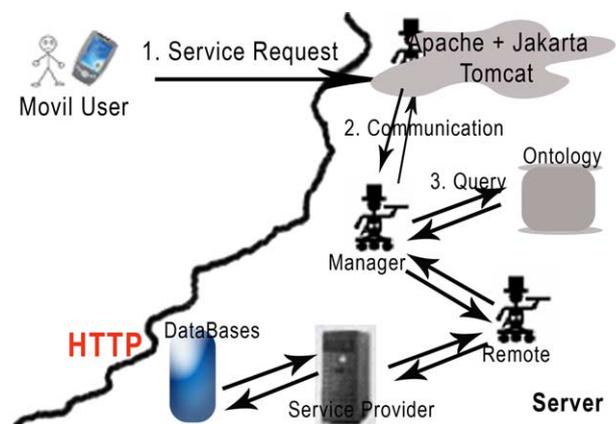


Fig. 4. Way of acting in a remote operation.

ii. Hybrid

The agents which form the multi-agent system use a blackboard architecture (Hayes-Roth, 1985; Kowalski and Kim, 1991; Nii, 1986a,b) 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.

Given the security restrictions called for by the complexity and the intrinsic value of the information managed by the center, it is necessary to create at least three different levels of access, which will match the following user types:

- *Internet user.* Anyone wishing to consult the center's data. The information available for this group of users is limited and controlled by the center's staff. This is due to the quantity of useful information which can be misused. For example, knowing the location of protected species by conducting environmental impact studies without the respective authorization. We can therefore speak of a data user who does not receive information processed by the system.
- *Researcher.* System user who can access more information than the Internet user and who in turn cannot perform any of the characteristic operations of the center management, such as the modification of information, insertion of loans in the system, etc. This user has access to the knowledge obtained by the system such as the representation of a specimen in the center, hot-biodiversity, etc. obtained automatically by the system, by means of the application of artificial intelligence techniques.
- *Center staff.* Within this role we can in turn distinguish two different users according to their functions within the system:
 - i. *Operators.* All the data contained in the system must be entered by someone, this would be the case of the operators. These will be responsible for inserting new folds into the system, new revisions, managing the images associated to the folds so that they can be consulted online, issuing labels for the preparation of folds, etc.
 - ii. *Managers.* User in charge of supervising and managing the center. This user will be authorized to access all the parts of the system, manage the accesses of the different users, obtain useful information for the center, such as for example, help in the planning of the collection campaigns by knowing which are the least represented specimens in the center.

3.1. 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,

Fig. 5. Consultation without identification.

the HTTP protocol will show the desired information and enter the parameters.

There are two ways to access the system:

- *Without identification* (see Fig. 5). The system is accessed as an Internet user and therefore access to the information is severely limited.
- *With identification* (see Fig. 6). 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.

By means of this access, whether identified or not, the client connects to the virtual center and the number of operations which can be performed will depend on the authorizations granted by the center managers. In this virtually centralized way, we allow the system user to overcome the problems of distance and time-wasting:

1. *Loss of time.* If we need any of the material contained in the center we will have two options:
 - a. Ask the center managers for a loan. These loans will only be granted to institutions or researchers with

Fig. 6. Identified access in the system.

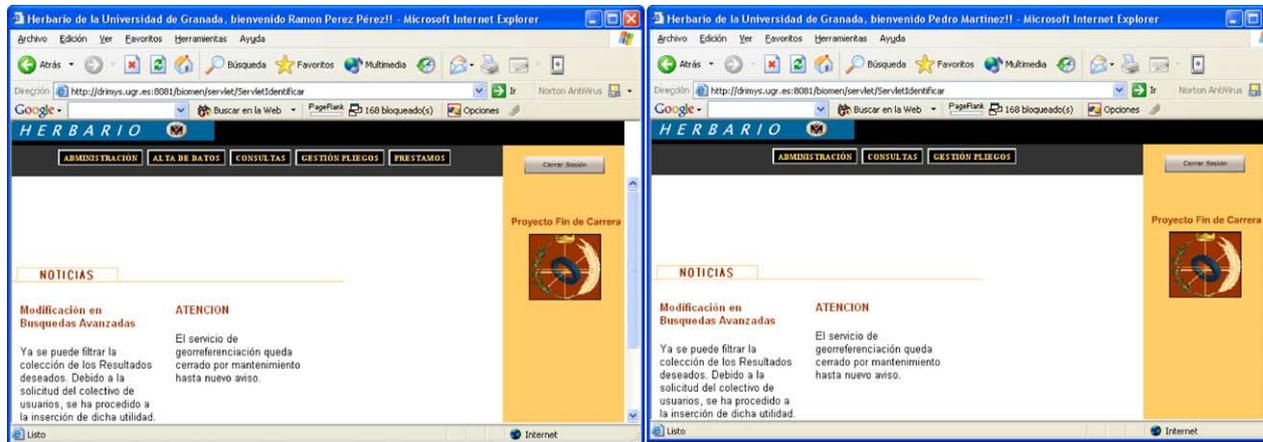


Fig. 7. Different options according to the user.

- a certain degree of confidence in the treatment of specimens.
- b. Visit the center so as to personally consult the desired specimens. As a result, the second of the centralization problems arises—that of distance.
2. *Distance problems.* The user/researcher will not always be in the same physical place as the center where the information needed is.

Another of the problems which exists in both options is that only those materials which are physically available and which have not already been lent can be borrowed or accessed.

In order to solve these problems, the need arises to centralize all the center's information so that any user/researcher can access the information (with more or fewer restrictions) simply, quickly, and comfortably. Consequently, a client/server has been used where all the information is virtually centralized, so that any user can consult the information needed without having to waste time by visiting the center or waiting for a loan.

3.2. 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.

Any system of these characteristics must present a series of security features which guarantee the durability of the system. Merely thinking of the operator hours needed to enter the 130,000 records in the database from the large

quantity of intrinsic knowledge to these records, which in their day have been provided by taxonomical experts who have identified and checked the specimens.

For this reason, and as a safety measure, both to guarantee access to the information and to guarantee the safety of this in view of severe, enormous problems, the system is replicated identically on two servers which work with a policy of a daily update. This policy has been chosen rather than a real-time update as in this way we can take action if errors have been entered by the users. Errors are not therefore propagated in real time, and so there is a better treatment of the information.

Let us look at the description of each of the services managed by the server. These will be accessible according to the user's profile. From these services, we shall jointly deal with all the information system.

3.2.1. Access without authentication

In this case, it will only be possible to access the system as a data user. This user will only be able to obtain from the system a set of lists obtained from the database. From the access without identification, any Internet user can consult the system and obtain a set of data entirely online. Not all the specimen information is shown as there is information about the specimens classified in the center which due to misuse of the information, and given its quality, could endanger the conservation of the species referenced.

3.2.2. Authenticated access

In order to enable more important services to be performed within the system, an identified access has been provided. From the server, the security certificate is issued which the user must accept before continuing with the access. Once this has been accepted, the information is exchanged between the client and the server using a 128-bit SSL secure protocol. In this way, we avoid the possibility of data being filtered by malicious people.

Once the user has been identified in the system, the system returns the series of services permitted.

Below we shall describe some of the more important tasks which can be carried out within the virtual center. The presentation of the different services is organized from the introduction of the data in the system until the information processed by the information system is obtained. The following modules are described below:

1. Introduction of folds
2. Management of revisions
3. Label creation.
4. Management of images
5. Advanced consultation and/or consultation of folds
6. Management of loans
7. Administration
8. Treatment for the biodiversity

Introduction of folds

The herbarium works with plants and consequently it is necessary for these to be treated and conserved. The container with one or more specimens is called a *fold*. The fold univocally identifies the sample in the center and can contain one or more specimens. The information about the fold (number of specimens, taxonomic names, collectors, exicata, location, UTM, height, etc.) is entered into the system by means of the corresponding service. Once the user has selected the desired service, the set of parameters to be provided is returned. When these parameters are entered and returned to the system, the agent in charge performs the remote service, returning the result of the operation to the user.

Management of revisions

Because of the interest required by other researchers for samples from other centers, folds are lent to other centers or researchers. The taxonomic name of a specimen is not easy to identify due to the complexity of nature. For this reason, there may be discrepancies between different researchers about the taxonomic denomination given to the specimen. Each of the researchers' proposals, in time, is recorded in the system as the current taxonomic denomination (the last in the instant t). This information will be very important

when it comes to studying the taxonomic changes in order to be able to make the biodiversity studies.

Creation of labels

Once the fold has been entered into the system, the center staff can obtain a label which is placed on the fold, so that part of the information and the fold identifier can be seen.

For the identification of the fold, an alphanumeric chain has been used. Each fold is assigned a barcode (Code 39) so that the subsequent treatment of the fold in the system will be a much simpler process (see Fig. 8).

These labels will mostly be used for loan management. By adding the barcodes, certain hybrid services (barcode reader and server) can be carried out.

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.

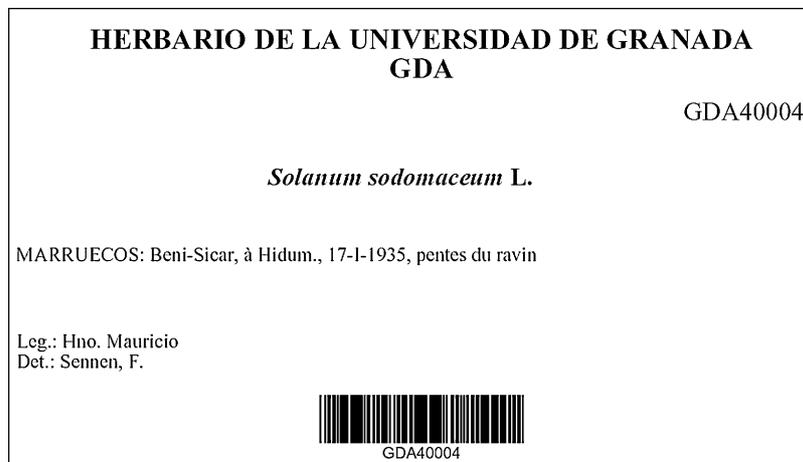


Fig. 8. Sample of a label.

3. In turn, there is a better conservation of the center’s material.

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 bacciano*, 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.

Loan management

As we have already mentioned, the center’s material is vitally important for everyone, center and researchers alike,

and it is therefore essential to control the loans made so that these are returned within the stipulated period.

Therefore, the loan service in turn carries out the following services:

1. *Loan registration.* The loan recipient and all the materials lent are recorded in the system. This new loan has an identifier within the system so as to be able to control the loan in consecutive processes.
2. *List of loans.* List of loaned material for a loan identifier.
3. *Loan return.* This enables the complete or partial return of the loan to be entered into the system. From this section, we control which of the loaned folds have not been returned.
4. *Loan control.* Control of the validity period of the loan. In this way, we will be able to demand that the loan be returned. It is possible that this operation is performed automatically by the system once the return period has ended and if there is no record of the return. The system sends an e-mail to the loan recipient informing them of the need to process the return of the loan.

Administration

Another just as important service is that of system administration. Given the architecture of the system and the way in which it is accessed, it is clearly necessary to have an exhaustive control of the operations allowed to each user within the system.

The system contains a list of operations and users. Using the checkbox, the system’s service administrator can control the services which these can make in the system (Fig. 9).

From this service, we can also request services to import data from other databases, user registrations/cancellations, etc.

Treatment for biodiversity

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

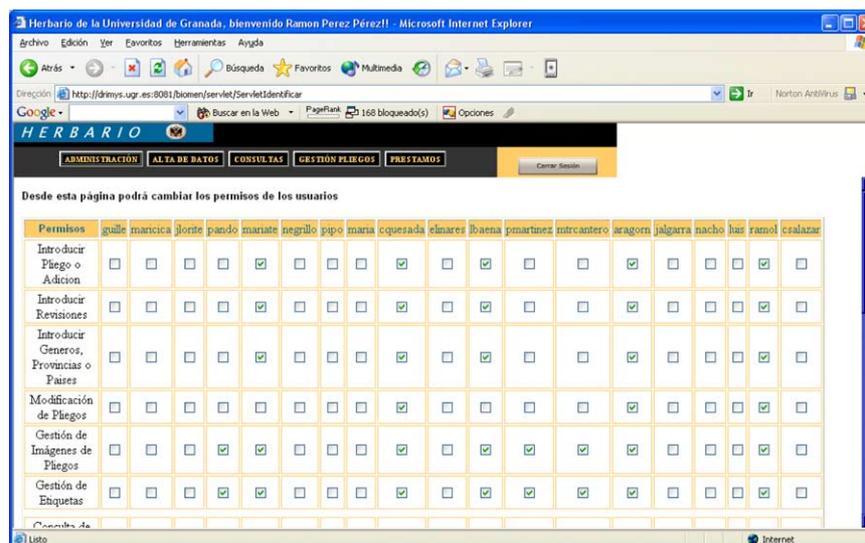


Fig. 9. User administration.

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.

4. 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 three specimens. However, according to Flora Ibérica (1996), the three names refer to the same specimen (Cruciferae *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) (Konolige, 1982), 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.

4.1. Description of the agents

The main innovation of the proposed solution is the treatment of the synonyms existing in the database. This is an important problem and one covered by an international project—*Species 2000*. In order to solve this problem locally, we shall study the concept of synonymy and how this can be solved. The method we propose does not need the expert to intervene directly as it is based on the study of the history of the specimens contained in the database.

In order to identify the synonyms, we use the concept of evolution. We shall define evolution as the change in the identification of a specimen. This evolution means that specimens are stored in the database with different names (i.e. synonyms). This means that the possibility of counting how many different species there are in the database is a complicated task without any real values.

In order to obtain the correct number of different specimens, we use the following agents.

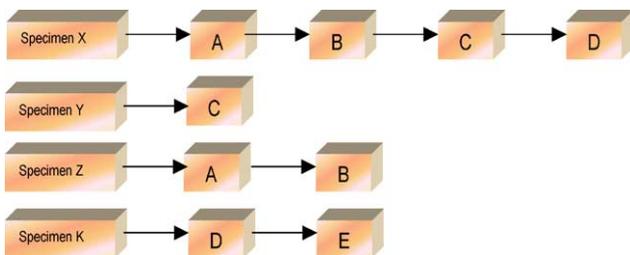
4.1.1. Revision agent

This agent is responsible for studying the number of revisions which each specimen has. In this way, the material most studied by botanists is obtained. Once we have the number of revisions for each specimen, we can conduct the relevant evolution studies. The agent places this information on the blackboard. The *AgenteRevisiones* table is used to enter the results.

4.1.2. Taxonomic complexity and specific richness agent

Information about taxonomic complexity

The solution to the synonymy problem was found by studying the evolutions which a certain specimen undergoes. Due to the formal complexity of the evolution and synonymy problems, a botany-based complexity which is therefore outside the field of this work, we shall present the concept using an example:



The different blocks marked as A, B, C, D and E are the different identifications which a specimen undergoes.

The arrows indicate the evolution to another identification. Let us examine the example more closely.

Specimen 1 was identified as A but later, after being researched by other researchers, it has been identified as B, C and D. These identifications are treated as Evolutions and therefore, the identifications A, B and C are synonyms of D. These values are entered in the *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:

1. *Taxon*. This indicates the number of the specimen. In this case 1.
2. *Antecedent*. Previous names of the specimen. For example: A, B or C.
3. *Consequent*. This indicates the name of the evolution. In this case D.
4. *Evolution*. This field can take two values:
 - a. When *evolution* is 0. This indicates that the specimen will not be identified in another way.
 - b. When *evolution* is 1. This indicates that if the specimen is studied, it may change its identification to that indicated by the Consequent field. This enables us to inform the botanist of the specimens to be revised so that the center’s material is completely up-to-date. This information is based on the revisions which have already been made to other specimens.

Below we shall place this information on the blackboard, in particular in the *taxones* table, so that it may be consulted by other agents. In this way, we shall enter the following taxon in this table:

Specimen 1: D.

Specimen 2 is a synonym since there is another specimen which has changed from state C to state D.

Therefore, we would enter in the *AlertasEvolución* table:

Specimen 2: C → D, evolution 1.

In this case, the evolution value of 1 indicates that if specimen 2 were studied, it would certainly be identified as D. Therefore, we enter it as a possible evolution.

The following specimen, specimen 3, has some revisions identified as A and B. As in the *AlertasEvolución* table, we have stored the fact that state B can become state D, due to the sequence of evolutions which specimen 1 has. We therefore add the following tuple to the *alertasEvolución* table:

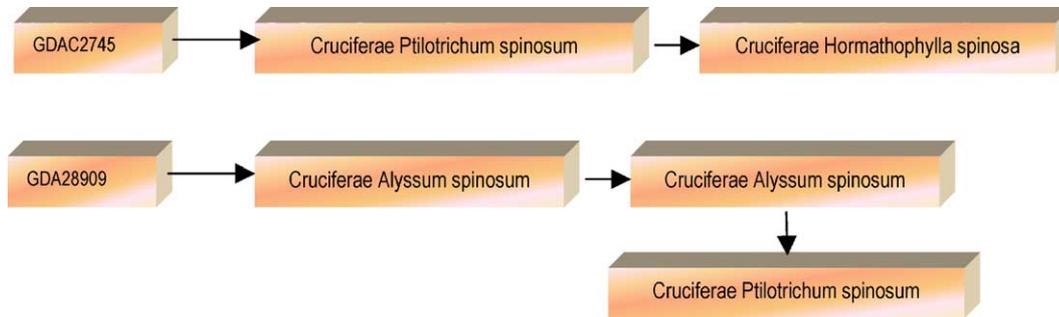
Specimen 3: B → D, evolution 1.

We should remember that in the *taxones* table there is a single tuple, indicating that of the specimens studied we only have one different taxon.

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

<i>AlertaEvolución</i> table	<i>taxones</i> table
Specimen 1: A → E, evolution 0	Specimen 4: E
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	

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.) K upfer

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.

AlertaEvoluci n table:

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

Information about specific richness

Once the synonym problem has been solved, we can study the specific richness of the areas covered by the center. In this way, botanists have a large amount of useful information for their hot-biodiversity studies.

In turn, the information which the specific richness agent provides can be crossed with the previously mentioned

project macro, *Species2000*. This project attempts to solve the problem of synonymy as a whole. If we cross the information developed by the agent with the information provided by *Species2000*, the category which a research center has can be established, indicating the number of series of species which this center covers. Suitably planned centers can therefore be combined and in accordance with the needs of the project to be embraced and covering the desired species.

4.1.3. Collection campaign orientation agent

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.

When the crossed information is obtained with a geographical information system, we obtain different ways of solving the problem of having a complete center:

1. Obtaining information from the areas which have not been visited for collection, thereby enabling each geographical point to be reflected in our center.
2. To obtain the necessary routes and itineraries in order to collect specimens which are little represented in the center.

5. 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, 2001)
- *Apache Server* (The Apache Software Foundation, 2001a)
- *Apache Jakarta Project Tomcat* (The Apache Software Foundation, 2001b)
- *Java Agents* (Borland Corporation, 2001; Eckel, 2000).
- *JDBC* for communication with the databases.
- Any *JDBC-compatible* database manager (MySQL, Oracle).
- *Semantic web and ontologies*. DAML + OIL and OWL-S

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

In this document, we propose a solution to the problem of studying hot-biodiversity. By means of artificial intelligence techniques, such as agents, complex and extremely interesting tasks have been performed such as the study of hot-biodiversity, cost minimization and orientation in the collection campaigns.

We are currently working on transferring the concept of collection from the virtual center where it is not necessary for there to be a physical place to store the specimens, thereby creating a totally virtual center. In order to do so, the latest technologies are being studied: PDA's, GPRS, GPS, multimedia transmission using mobile devices, etc.

References

- Borland Corporation. (2001). *JBuilder Enterprise 6.0.438.0 Documentation*. Available at: <http://www.borland.com>, last visited December 26th 2003.
- Castroviejo, et al. (1996). *Flora Ibérica. Real Jardín Botánico, CSIC*, Vol. II. Cruciferae-Monotropaceae pp. 193–195.
- Eckel, B. (2000). *Thinking in Java* 2nd ed. Prentice-Hall, Englewood Cliffs, NJ.. Available at: <http://planetpdf.com/>.
- Halfiter, Moreno y Pineda, *Manual para evaluación de la biodiversidad en Reservas de la Biosfera, MT and SEA*.
- Hall, M. (2001). *Core Servlets and JavaServlet pages*. Englewood Cliffs, NJ: Sun Microsystems Press/Prentice Hall.
- Hayes-Roth (1985). A blackboard architecture from control. *Artificial Intelligence*, 26(3), 251–321.
- Henry, C. & Lucas, J. R. (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.
- Konolige. (1982). *A first-order formalization of knowledge and action for a multi-agent planning system*.
- Kowalski, K. (1991). A metalogic programming approach to multi-agent belief. In V. de Lifschitz (Ed.), *Artificial intelligence and mathematical theory of computation: Papers in Honor of John McCarthy* (pp. 231–246). Boston, MA: Academic Press, 231–246.
- Lane, M. A., Edwards, J. L., & Nielsen, E. (2000). Biodiversity informatics: the challenge of rapid development, large databases, and complex data (keynote). In *VLDB 2000, Proceeding of the 26th international conference on very large data bases*.
- 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*.
- Pando, F. (1991). El Herbario de Criptógamas del Real Jardín Botánico y sus bases de datos. *IX Simposio Nacional de Botánica Criptogámica, Salamanca*.
- 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.
- Research Directions in Biodiversity and Ecosystem Informatics. (2001). *Report of an NSF, USGS, NASA workshop on biodiversity and ecosystem informatics*.
- Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge: Cambridge University Press.
- Species2000, Available at: www.species2000.org, last visited March 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. (2001a). *Apache HTTP Server Version 1.3.29 Documentation*. Available at: <http://www.apache.org>, last visited December 26th 2003.
- The Apache Software Foundation. (2001b). *Apache Jakarta Project Version 4.0 Documentation*. Available at: <http://jakarta.apache.org>, last visited December 26th 2003.
- Weiss, G. (1999). *MultiAgent systems. A modern approach to distributed artificial intelligence*. Cambridge, MA: MIT Press.
- Wooldridge, M. (2002). *An introduction to multiagent systems*. New York: Wiley.