Visual Plants - An image based tool for plant diversity research.

Visual Plants - Un instrumento basando en imagines para la investigación en ecologia de plantas.

Helmut Dalitz1* & Jürgen Homeier2

1University of Hohenheim, Institute of Botany (210), 70593
Stuttgart, Germany, Email: Helmut.Dalitz@uni-bielefeld.de;
2University of Bielefeld, Faculty of Biology, Department
of Ecology, Universitätsstr. 25, 33604 Bielefeld, Germany,
Email: Juergen.Homeier@uni-bielefeld.de,
Tel.: +49 521 106 5522, Fax: +49 521 106 2963
* corresponding author

Visual Plants - An image based tool for plant diversity research.

Resumen

Hay un proceso comenzando de hacer accessible información de biodiversidad tanto con programas para la computadora como en la red. El espectro se extiende de sistemas de información para herbarios y sistemas taxonomicas hasta sistemas de información visual mas faciles para el usario. Visual Plants es un exemplo para la ultima categoría presentando imagenes digitales junto con la información correspondiente y con caracteristicas para la identificación. Visual Plants es una base de datos basando en especímenes con función automatica para la compilación de una lista de especies. Se puede utilizar lo sin conección a la red para la educación de estudiantes, para la propagación de información botanica y para la investigación taxonomica y ecologica. Palabras clave: Imagen, digital, plantas, ecología

Abstract

There is an ongoing process to make biodiversity information available using computer programs as local as well as worldwide accessible Internet-based systems. The range of solutions reaches from information systems for herbaria, taxonomic diagnostic systems to more easy to use visual information systems. Visual Plants is one of the last category providing digitalized images from plants with all attached label information together with some characters for the identification of individual plants. Visual Plants is at first a specimen-based database with automatic function to compile the information as species-based. It can be used on-site without having an Internet connection for the education of students, for dissemination of information and for taxonomic and ecological research. Key words: Imagen, digital, plants, ecology

Introduction

Although the diversity of plants is quite smaller than of animals, especially insects, information about plant diversity is still not complete and disseminated completely. Especially the treatment of tropical regions still remains incomplete (Prance et al. 2000). Knowledge about the determination of certain plant groups (e.g. large tropical plant families like the orchids) is often concentrated as the knowledge of a small number of researchers. Research in tropical regions is therefore due to the high number of plant species in a given area catchy for young researchers, when information is not jet published and available in the country. But it is the most crucial question not only for taxonomical, but more for ecological research, that information about the plant taxonomy is available for local researchers. Not only the availability of information is a critical issue, the ongoing and rapid destruction of tropical rain forests needs a comprehensive and integrated effort of scientists for the inventory of species in the tropics (Hamilton et al. 1994). Otherwise species will disappear without being known by science.

Not only the lack of availability of information causes problems for researchers but also the rapid evolving plant taxonomy and systematic. Systematic of plant species now uses modern taxonomic tools adopted from genetics, and relations between large plant groups are seen in different manner: families and genera are rearranged (see Judd et al. 1999) and genera are grouped into several smaller or larger genera. The effect of renaming or differentiation of taxa into smaller groups is not a new problem. It follows new insights into the relations between certain taxa. As a consequence the knowledge on plant taxa follows a certain notion which sometimes causes serious confusion. Berendsohn (1995) mentioned three different categories under which biological information can be seen: first the purely nomenclature point of view, where a name can be understood as a hypothesis, "as a set of biological objects within a classification unit supposedly linked by phylogenetic descent, or as a set of criteria applying to such objects". In the second category practising taxonomists follow a more vague concept of a taxon, which was described as a collection of objects and the sum of herbarium specimen and living plants with the same set of criteria. The third group are all users of botanical names with the demand of an operational definition of a taxon. The resulting confusion, which may occur, can be described as the lack of description or references to which a specimen (!)

not a taxon refers to. The misunderstanding is, that not all specimen which may have the same name, really belong to the same taxa due to differences in nomenclature or the underlying concept. As a consequence Berendsohn (1995) stated the necessity of specimen-based information with all attached information about the context, the references etc.

Asking a simple question: How many plant species exist on our planet? At the moment this question cannot be answered satisfactorily, because we do not know exactly how many species exist on our planet or, caused by different opinions of researchers and taxonomists, the numbers differs. But the question, how many species do we already know from our planet cannot be answered very easily because the information is stored regionally. This means that we know more or less exactly the number of plant species in Germany or Middle Europe, but it is quite problematic to have access to this information from another regions.

[[subheadinh text="What must be concluded from these problems?"]]

1. Information should be stored in a central institution, where the state of the art is transparent for all interested researchers (see Godfray 2002). This new idea of a more or less virtual taxonomical institute is a little bit revolutionary, but has a lot of advantages, if the existing specialists agree and contribute to this idea. Several main aspects must be fulfilled: New information about the taxonomical grouping of taxa must be updated regularly including the information about their former systematic state (see Güntsch et al. 2002). Information should be automatically mirrored to other institutions for a more rapid access, and updating of information must be automatically proceeded.

2. Information must be easily accessible. This demand has at least two main implications:

a) Information must be accessible for all researchers using the Internet; this technique is now accessible from most places in the world. But, to be honest, in developing countries the access to the Internet is sometimes rather slow and not very stable which means that the dissemination of information is somewhat restricted.

b) Information must be easily accessible due to an easy to use interface without having special knowledge about data basing of information. This issue is likewise important to the others, because the user-interface can be understood as the window to have an insight to the information itself.

3. Information must be extractable for the use in the field when no Internet is available. In case of developing countries information must be extractable from this central system to carry the information to remote places and use it in the field, where no Internet is available.

What kind of information should be stored?

This question can be discussed controversially, because it depends on the needs and the main intentions, for what purposes it is meant and for which persons the information is thought. Managers of herbaria may have other needs; they may prefer to store specimen-based information about their herbarium specimen, geo-referenced collecting sites etc. Research scientists working in ecological fields may prefer to store information about the ecological needs of the plant species and may need therefore a species-based system. Additionally a key system for the identification may be helpful for ecologists and related research areas or neighbouring disciplines. This, of course should be easy to use and, in case of tropical trees, based more on vegetative information, if possible. But, "information on plants is used and processed by many people outside the taxonomic community, who often will not know what a specimen looks like" (Berendsohn, 1995). Therefore, images of plants may be helpful to control the identification and give an "idea" on the habit of a plant. These images must show most of the relevant information, which is necessary to identify a plant. Because this is mostly impossible with only one image, a series of images (an assemblage) may show different aspects or parts of the plant.

Data-basing provides an excellent chance for managing and disseminating the vast biological data in natural history collections such as museums and at the same time making it reachable widely to specialists and the public. Capturing much of the data as images and text is also a good back-up of specimen data, which would otherwise completely vanish if the voucher specimens were to be destroyed by normal wear and tear or some other catastrophic factors such as fire, flooding, pest attacks, or war and lawlessness. Imaging and data-basing specimens is quite useful from the long term preservation point of view, because it allows users to access data without necessarily handling the original specimens and thus greatly reduces the hazard of destruction for the specimens and also drastically reduces the time required to access information.

The complex relationships in ecosystems between and among the various biotic and abiotic factors

necessitate the sharing of information generated by persons working on aspects of biodiversity. Capturing much of the already collected data in museums in electronic form is an essential step towards making taxonomic information easily accessible and also acts as a back-up of taxon data, that would otherwise completely vanish if the vouchers were to be destroyed by normal wear and tear or some other catastrophic factors.

Most of the initiatives in the field of biodiversity informatics concentrate on applications, which provide information on the Internet (see [[Table 1]]). Visual Plants (Dalitz 2002) follows another approach: information is provided locally as an on-site application, which can be used in the field for three main tasks. First it allows the data capturing including the visual information using plant images. Secondly it provides information about plant specimen, species and families. Thirdly a set-based key system is included which allows a low level determination using mainly vegetative characters. The user-interface is very easy to use and special knowledge about databases is not necessary. This means that Visual Plants may be useful as a tool for the work in the field, but it cannot fulfil all needs which are mentioned above.

Problems of ecologists and vegetation researchers in the tropics

Determination of plants is sometimes a task only for specialists, especially when difficult groups are involved. But also for easy to determine plant species there are some specific problems when working in the tropics. The diversity of plants is rather high and there is often a lack of fertile characters because plants, especially trees, do not flower every time or every year (Newstrom et al. 1994). Getting the material as well can be a difficult task when tree flowers hang in 30 or 40 m height. Within difficult taxonomic groups specimens from herbaria and specialists are needed for a correct determination. Literature is often very comprehensive, incomplete or in revision. Herbarium specimens are often not stored in the tropical field stations because of the high humidity in the field stations. Field guides are often not available or incomplete. All these difficulties can be solved with technical solutions at a field station, but research in really remote areas will still be under the proviso that taxonomical literature or herbaria cannot be taken into these places. This means that long-term experience is necessary for a correct determination in the field.

Visual Plants was developed as an on-site database program to give researchers in the field a tool for their work. On-site means, that the researcher can carry all the information stored in Visual Plants to his remote research places on a mobile computer. Different types of images can be stored in the database with all attached label information: flat-bed scans of living plant material (see [[Figure 6]]), digitized slides (Figure 4), digitized herbarium specimens (Figure 5) or illustrations can provide information about the habit, the leaf and floral characters of a plant species in a series of images of these different types. This information can be used for comparisons between the collected specimen and the information in Visual Plants either in the field or in a herbarium. With the help of specialists the information may be updated regularly, to ensure that this information can be used as a reference.

Biodiversity information management

In several attempts different projects try to make biodiversity information accessible for a broader audience, e.g. the DELTA-project (CSIRO division of Entomology, Canberra, Australia; see Dallwitz 1974, 1980), GBIF (Global Biodiversity Information Facility), SysTax (University of Ulm, Germany; see Hoppe et al. 1996) or the program "Biodiversity informatics" funded by the German Ministry of Science and Technology (Berendsohn et al. 1999) with two main goals: first to make accessible the information which is stored in herbaria and secondly to provide taxonomical software, which allows the determination of selected groups, especially for insects. Most of these attempts are Internet-based, and require therefore a fast net access.

Advantages of these attempts which have developed since longer time are on the one hand the compatibility between the systems (see Standard of the International Taxonomic Databases Working Group TDWG, Conn 1996), and on the other hand the possibilities for the identification of species through different key systems. Additionally these systems are used for the systematic acquisition of information for a further processing. A comparative study about seven systems with their different approaches can be found in Dallwitz (2000). One problem, which still occurs, is the discrepancy between the general characters of one species and the existence or non-existence of a character, which can be found of a single individual of this species.

A particular attribute of computer aided information systems is the user-interface, which allows the user to access the information. Especially for taxonomic solutions there are often many steps needed to

come to an accepted identification of a specimen or species (see Dallwitz 2000). This is an important criticism, because a difficult to use interface, which is not consistent, may frustrate the user or the information cannot fully be reached.

In [[Table 1]] some important Internet sites were listed, which provide a substantial overview on current systems.

Conception and description of Visual Plants

Information about plant diversity is often given as a check-list of species in a given area or in floras. Species keys, usually dichotomous keys, are given to assist in the identification of taxa. Such information is of limited practical use to non-specialists, because most people have no idea about what the many species look like. The process of finding out what the species look like can be quite taxing. Therefore identification keys and descriptions alone are not sufficient for most people to identify species. The use of images of voucher specimens in addition to descriptions and keys greatly simplify the identification of plants by local researchers, non-specialists, students, and the general public. Images speak for themselves and users may not need to understand the complex terminology that may be used in keys and descriptions in order to identify a species. Imaging of plant species is thus important in making information widely available and easily understood by participants, local researchers and the public.

For this reasons Visual Plants is an image-based system, which can contain an unlimited number of images of one species. These images may show different aspects of information. A flatbed-scan of fresh collected material shows the specific colour of the plant parts (see (Figure 6). After the process of herbarisation a second image may show the plant after drying the voucher specimen, which demonstrates in some cases different colours (see. (Figure 5). Digitized slides provide an "idea" of the three-dimensional arrangement of plant parts, especially flowers, and may help to identify a plant (in (Figure 4) the characteristic bark of *Bursera simarouba* is shown). Illustrations show extracted and interpreted information of the whole plant or of parts of it with all relevant taxonomic features of the plants. All images were coded with the name of the photographer (see (Figure 1), the country (see (Figure 2) and type of the image, but not with taxonomical information. All images were prepared in the same manner and a copyright sign is included in one corner of the image due to international copyright behaviour.

This means that Visual Plants is therefore not specimen-based in the common sense, because there may be several images of the same specimen. But, browsing through these images gives a clear impression about the appearance of the plant. Additionally the context of information, finding places, references, habitat, altitude are stored together with each image.

All images were stored with the complete set of label information including the habitat as geo-referenced information (either with UTM or latitude and longitude coordinates), which allows a further processing in a GIS-system. Names of genera and families are provided from internal lists (using trigger functions), to avoid writing errors, which would make it impossible to find these images.

The main list therefore is called "List pictures" and shows a list window in two views with different information. The header of the list window is multi-functional with buttons for searching genera, families and selected records. An input window allows the description of the image and provides the information separated in topics in so-called register cards (see [[Figure 1 to 4]]). One of these register cards provides information about the family, if the image is already determined. This gives a possibility to test the correctness of the determination.

Each record gives the possibility to state characters for the same plant with different character statements ((Figure 2). As characters easy detectable features of plants are used such as "Life form", "Leaf characters", "Stipules" (absent or present) or "Colour of flower". Some of the character statements can be chosen from an internal list, other must be typed with additional information. In case of "Stipules" the user can state the position or the shape of stipules, when they are present. This gives the opportunity to describe the habit of a species for each specimen (or image) separately.

The information is used in a special search dialogue, which allows a set-based search for these character statements (see (Figure 7). This set-based search dialogue provides a rapid and easy to use interface as a helping tool for the identification of plants. The results are shown in the lower part of the "Search"-dialogue and images can be easily browsed for a visual identification.

Depending on the completeness of the character information, specimens can be found using their character information and compared with already determined specimens. At the moment 7 characters

with 123 different combinations are programmed, each containing up to 10 character statements. In the next version ten characters will be used with 1023 combinations of characters. This will raise the chances for better identifications.

The command "List families" shows a list of families with some of their most common attributes and the possibility of inclusion of up to 12 characteristic images. The information of one family (Figure 8) will be displayed in all records, which belong to this family.

A second feature of Visual Plants tries to come over the separation between the concepts "Specimen-based" or "Species-based". Automatic functions create new records in the table "Species" (see (Figure 9), whenever a new species is detected in the image table. The resulting record entry will hold general information about the attributes of the species, which may differ from the found specimen. For instance it may be possible that a specimen is only a small treelet or shrub in one region, but a big tree in another region or altitude. This information can be observed and a list of all connected specimen can be examined.

There are several possibilities for printing the information out of Visual Plants. As an example in [[Figure 10 a report is shown which is generated by the program. This report can be chosen by the user and is not fixed. This solution follows the idea that the program should be as flexible as possible for the needs different users may have.

The already included menu items are shown in (Figure 11).

Future plans will implement the concept of "potential taxa" (see Berendsohn, 1995) and the compatibility with the TDWG standard will be improved. As a second main aim Visual Plants will be searchable on the internet and data-sharing with other institutions will be provided using the technique of Webservices.

Advantages and disadvantages of Visual Plants

Main objective for the development of Visual Plants was the creation of an on-site system, which can be used not only by specialist, but only by students and untrained, but interested persons. Therefore the effort laid on the design of the user-interface. The use of Visual Plants should be possible without understanding the technical aspects of database systems. The program should run on the most common platforms (Windows and Macintosh), which makes it necessary to program it with a Rapid Application Development tool, which supports both platforms. Although 4th-Dimension (see www.4D.com) is not a common RAD-tool, it serves as a rather easy to use platform for developing the application.

The main functional aspects, which should be realised, are the concept of an image-based database with attached information about family and species characteristics.

As one of the main advantages of Visual Plants the easy to use user-interface may be seen which has been tested in student courses and on field stations (ECSF, Loja, see: www.bergregenwald.de, or RBAB, Costa Rica, see: http://www.ucr.ac.cr/otros/r-alberto.htm)

The rapid database engine enables the user to search in large series of records and the images were shown at each level of information. No additional window will open to see the information stored in the images.

As described in the introduction, the necessity is quite obvious to concentrate the efforts of all taxonomic and systematic work in the world to support an online system, which stores all the information about plant diversity at one place. This will enable all interested researchers (and the public) to use this information for their research and for making decision, which may have influence the diversity (see Güntsch et al. 2002 for the description of an example for taxonomical revisions which may be concentrated on the Internet). For this purpose, Visual Plants is at the moment not a proper tool, because the data-exchange is still restricted although an Internet-interface is now available (www.visualplants.de). Plans for the near future will focus on this and the technique of so-called "web services" will be used for international data-exchange.

References

Berendsohn, W.G. 1995. The concept of "potential taxa" in databases. *Taxon*, 44: 207-212.

Berendsohn, W.G.; C. Häuser & K.H. Lampe. 1999. Biodiversitätsinformatik in Deutschland: Bestandsaufnahme und Perspektiven. *Bonner Zoologische Monographien*, 45: 61 pp.

Conn, B.J. 1996. HISPID3. Herbarium Information Standards and Protocols for Interchange of Data. -

version 3 - Royal Botanical Gardens, Sydney. Electronic version under: http://www.rbgsyd.gov.au/hiscom/

Dallwitz, M.J. 1974. A flexible computer program for generating identification keys. *Syst. Zool.*, 23: 50-57.

Dalitz, H. 2002. Visual Plants - Bildbasierte Datenbank für die vegetationskundliche oder ökologische Forschung in den Tropen. *Berichte der Reinhold-Tüxen-Gesellschaft*, 14: 119-129.

Dallwitz, M.J. 1980. A general system for coding taxonomic descriptions. *Taxon*, 29: 41-46.

Dallwitz, M.J. 2000. A comparison of interactive identification programs. http://biodiversity.uno.edu/delta/

Godfrey, H.C.J. 2002. Challenges for taxonomy. *Nature* 417: 17-19.

Güntsch, A.; J. Li & W. Berendsohn. 2002. Euro+Med PlantBase. Design of the Internet taxonomic sector editor. Electronic version under: http://www.bgbm.org/BioDivInf/Projects/ Euro+Med/default.htm

Hamilton, L.S.; J.O. Juvik & F.N. Scatena. 1994. *The Puerto Rico Tropical Cloud Forest Symposium: Introduction and Workshop Synthesis*, 1-23. In: Hamilton, L.S.; J.O. Juvik & F.N. Scatena. (eds.). Tropical montane cloud forests. *Ecological Studies* 110. Springer, New York, Berlin, Heidelberg.

Hoppe, J.R.; E. Boos & G. Gottsberger. 1996. The database system SysTax - an aid for systematics and taxonomy and the management of botanical gardens and herbaria. *Albertoa* 4, 107-108.

Judd, W.S.; C.S. Campbell; E.A. Kellogg & P.F Stevens. 1999. *Plant Systematics: A phylogenetic approach*. Sinauer Associates, Sunderland, Massachusetts, USA. 464 pp.

Newstrom, L.E.; G.W. Frankie; H.G. Baker & R.K. Colwell. 1994. Diversity of long-term flowering patterns. In: McDade, L.A.; K.S. Bawa; H.A. Hespenheide & G.S. Hartshorn. (eds.). *La Selva - Ecology and natural history of a neotropical rain forest*. University of Chicago Press, Chicago, 142-160.

Prance, G.T.; H. Beentje; j. Dransfield & R. Johns. 2000. The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden*, 87: 67-71.

Figure 1. Main data entry in Visual Plants. In a window with six register cards all relevant information about a specimen is stored. The first register card shows the general information.
Figura 1. Registro de datos en Visual Plants. En una ventana de seis tarjetas de registro se graba toda la información de una especies. El registro primero representa la información general.

Figure 2: Main data entry in Visual Plants. The second register card allows the input of specimen characters.
Figura 2. Registro de datos en Visual Plants. Este segundo registro permite la entrada de caracteristicas del espécimen.
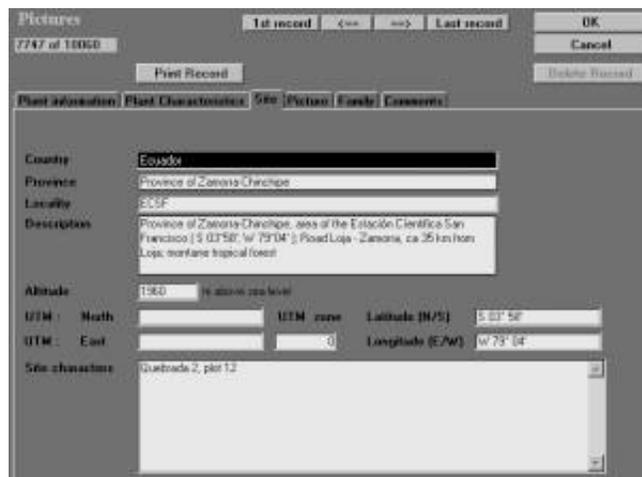
Figure 3. Main data entry of Visual Plants. The third register card shows the information about the collecting site.
**Figura 3. Registro de datos en Visual Plants. La tercera ventana indica la información sobre el lugar de colleción.**
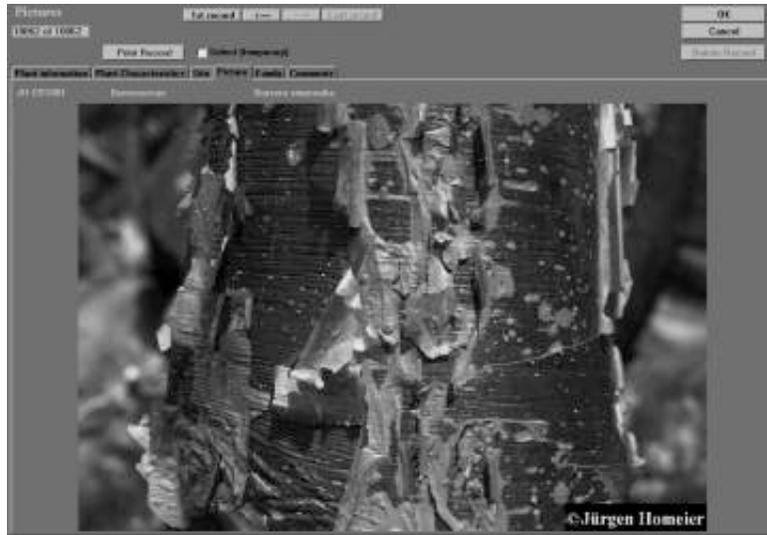
Figure 4. Main data entry of Visual Plants. The forth register cards shows the image or picture, in this example the bark of *Bursera simarouba* (Burseraceae) from a digitized slide.
Figura 4. Registro de datos en Visual Plants. El cuarto registro muestra la imagen of dibujo de la planta, en este ejemplo la corteza de *Bursera simarouba* (Burseraceae) de un slide digitalizado.

Figure 5. Main data entry of Visual Plants. In this example an image of a herbarium specimen of *Parinari* spec. (Chrysobalanaceae) is shown.
Figura 5. Registro de datos en Visual Plants. En este ejemplo una imagen de un espécimen de herbario de *Parinari* spec. (Chrysobalanaceae).

Figure 6. Main data entry of Visual Plants. The inclusion of flatbed scans can help to store and present a lot of relevant information from fresh material. As an example a still undetermined specimen from the Simaroubaceae is shown.
Figura 6. Registro de datos en Visual Plants. Los scans pueden ayudar para almacenar y presentar mucha información importante de material vegetal fresco. Como un ejemplo la especie aún no-determinada de Simaroubaceae.

Figure 7. The "Search"-dialogue. In the upper part character statements can be chosen from internal lists. The program searches through all record and stores the information for each character statement in a so-called "set". The button "Make Intersection" combines the information stored in these sets and presents a list of records for which all chosen character statements are true. This list is clickable and shows the data entry window with all specimen-based information. In this example with only four characters a list is presented with specimens which all belong to the family Rubiaceae. The main advantage of this dialogue is that the user can "play" with the information, because every query result for a single character statement is stored separately, the user can switch between different character statements rapidly, whereas the program executes only the query for the new character statement decision.
Figura 7. El dialogo de buscar. En la parte alta se puede elegir de diferentes caracteres de una lista interna. El programa busca todos los registros y graba la información para cada caracter en un set. El boton "maker intersection" combina la información de estos sets y presenta un listado de especímenes que tienen estos caracteres. Se puede elegir

de este listado y se muestra la ventana con todos los datos. En este ejemplo se presentan las especies de Rubiaceae. La mayor ventaja de este dialogo es que el usario puede jugar con la información y cambiar los caracteres facilmente.
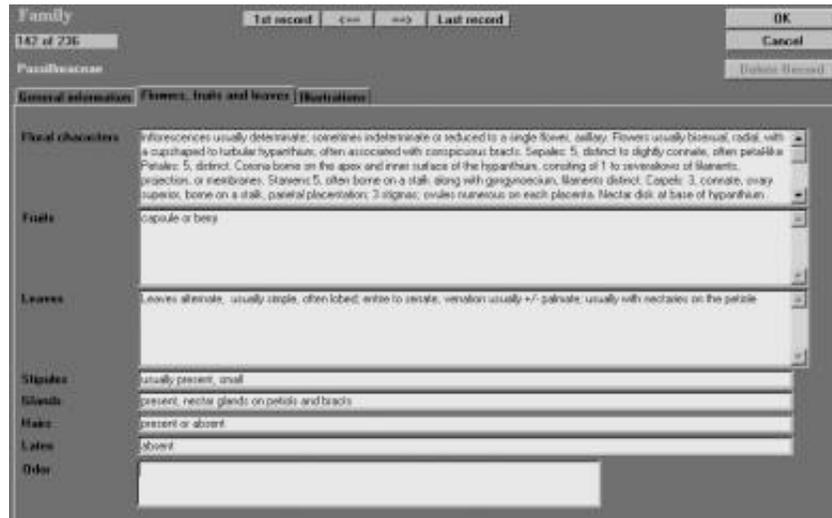
Figure 8. Information about family characteristics of the Passifloraceae. As an example is shown one of the three register cards, holding the information about this family .
Figura 8. Información sobre caracteres de la familia Passifloraceae. Un ejemplo muestra una de las tres tarjetas de registro que contienen la información de la familia.

Figure 9. List window of the menu item "List Species". The records are clickable and the data entry window shows most of the relevant information about the species (not shown). The list is generated automatically with each record in the picture list, which is determined to the species. For each record in the species list there is provided a clickable list for all available specimen in the database.
Figura 9. Ventana de "Lista de Epecies". Se puede elegir de los refistros y la ventana de datos musetra la información mas importante de especies (no se puede ver). El listado se genera automaticamente con enlace a las figures. Para cada registro en el listado de especies se presenta un listado con posibilidad de elegir especímenes en la base de datos.
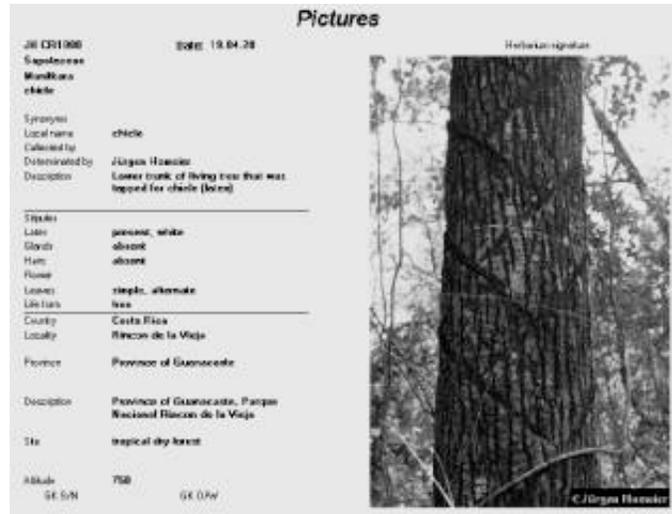
Figure 10. One example for a possible output of data in a report.
Figura 10. Un ejemplo para un posible informe de datos.
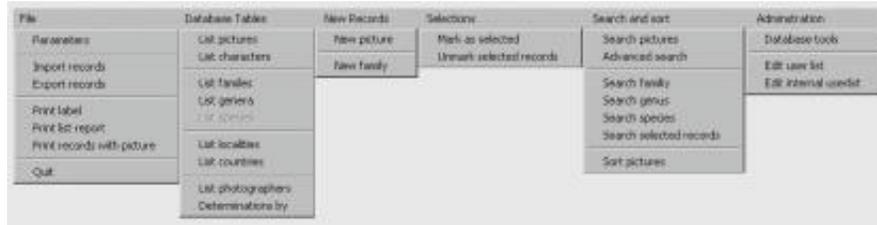
Figure 11 The menu items of Visual Plants.
Figura 11. El menu de Visual Plants.

Organisation	Adress	Topic
Missouri Botanical Garden, St. Luis, Missouri, USA	http://www.mobot.org/manual.plantas/lista.html	Plant list of Mesoamerica in Spanish, no taxonomical approach
Missouri Botanical Garden, St. Luis, Missouri, USA	http://mobot.mobot.org/W3T/search/vast.html	Plant list of Mesoamerica, no taxonomical approach
CSIRO Division of Entomology, Canberra, Australia	http://biodiversity.uno.edu/delta	DELTA-format and DELTA programs for taxonomic and systematic work on taxa
Digital Taxonomy, an Internet site of Cavalcanti, Rio DeJaneiro, Brasil	http://www.geocities.com/Rainforest/Vines/8695	List of programs for taxonomical informationsystems
Global Biodiversity Information Facility GBIF, Stuttgart, Germany	http://www.gbif.org	Worldwide network of databased information on biodiversity
Botanical Garden and Botanical Museum Berlin Dahlem, Berlin, Germany	http://www.bgdm.org	Biodiversity informatics, information about different systems with various approaches
Visual Plants, Germany	www.visualplants.de	Example of an on-site solution with some determination capabilities

Table 1 Important Internet addresses for biodiversity informatics and realisations of information systems.
Tabla 1. Direciones de internet importantes para información sobre informatica de biodiversidad y sistemas de información.