

Flora Croatica Database Application

Krešimir Fertalj¹, Toni Nikolač², Tomo Helman¹, Vedran Mornar¹, Damir Kalpić¹

¹Chair of Computer Science, Department of Applied Mathematics
Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, HR-10000 Zagreb, CROATIA

²Department of Botany
Faculty of Science, University of Zagreb
Marulićev trg 20/2, HR-10000 Zagreb, CROATIA

Abstract: A client-server database application, named CROFlora has been developed to foster the endeavour for the systematic and taxonomic classification of the Croatian flora. The software consists of several modules to encompass the taxonomy, herbarium, literature, horology and ecology. The paper presents the underlying database structure, querying and reporting. Standard reports cover the taxon sheets, taxa checklists, herbarium labels and bibliography labels. Rather complex reports are also available. A connection to the geographical information system (GIS) has made easier the production of distribution maps and has enabled comprehensive spatial analysis.

Key-Words: taxonomy, vascular flora, distribution, database application, geographical data, Croatia

1. Introduction

Development of new technologies like databases, multimedia and Internet has had its impact on the treatment of botanical data. It gave rise to devising of various approaches regarding the construction of research-oriented databases [1],[2],[5],[9],[18],[19].

This paper describes Flora Croatica database application (CROFlora), which enables the storage of taxa and specimen oriented data, updating, and data analysing of Croatian vascular flora. CROFlora was developed in collaboration between the Department of Applied Mathematics, Faculty of Electrical Engineering and Computing and the Department of Botany, Faculty of Science, both from University of Zagreb, Croatia.

The first version of CROFlora [25] was developed as a multi-user database handled by a client application and supplemented with web interface. The taxonomic backbone is newly developed Checklist of Croatian Flora (CCF), produced by the Department of Botany [26],[27],[28]. The data is deriving from taxonomy (nomenclature, synonyms, authorisation), horology (distribution based on literature data, herbaria collections, field investigations, oral reports), bibliography, etymology, ecology (ecological indexes) etc. The database supports the processing of numerical,

text and multimedia data and can be connected to GIS applications.

CROFlora is a part of broader Croatian Information Service for Biodiversity - Database (CIS-B Database) [30], which also contains raw species list for all other groups of living organisms in Croatia (*Animalia*, *Lichens*, *Monera*, *Mycota*, *Plantae non vascular* and *Protista*). In addition to data management, the main purpose of CROFlora is to be used as a tool for flora analysis and preparation of the Atlas of the vascular flora, as well as Flora Croatica, which need to be produced in the near future.

2. CROFlora Database

2.1. Taxonomy

Taxonomically, the database deals with vascular flora, i.e. *Pteridophyta* and *Magnoliophyta* divisions. The lowest level that can be defined in the database is subspecies. The lower levels, as varieties and forms, are included only as part of synonyms. CROFlora has the systematic arrangement according to [8], with minor modification.

The taxa are classified in nine hierarchical levels (*Kingdom* to *Genus*) as presented by the conceptual data sub-model in Fig. 1.

Besides the officially recognised levels, the database contains aggregates (also known as *complex*) – a group of mostly small species, to facilitate work with difficult groups, as commonly practised [13].

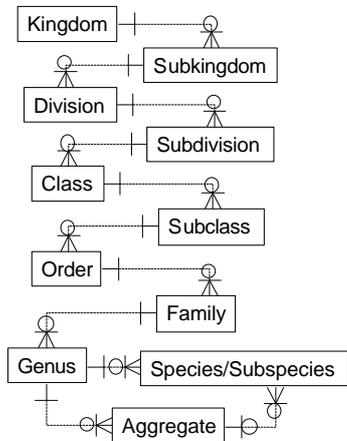


Fig. 1 – The model of taxa hierarchy

The essential data about species or subspecies (*Species/Subspecies* in Fig. 1 and Fig. 2) comprise the values such as species name and subspecies name, place of publishing, taxa name abbreviation and free-form taxa description. The data stored in the database were collected from several sources. Nomenclature and related data that form a basic list of about 12000 taxa, originate from the database for middle Europe FLOREIN [4]. This original data contained the taxa list of Central European Flora according to [13].

As the work on CCF was going on, this original list has been successively modified according to new proprietary results [26],[27],[28]. The taxa names were checked out by comparing with Names in Current Use (NCU) [17] and with taxa list from digital version of *Flora Europaea Database*, which is part of PANDORA taxonomic database system at the Royal Botanic Garden Edinburgh [32],[33].

Information on doubtful data (taxonomically or horologically), as well as on endemic, cultivated and naturalised taxa, are derived from the CCF. Data on endangered taxa (*ConservationStatus*) are following marks according to IUCN Conservation Monitoring Centre [3]. Besides the standard IUCN marks, there is also a special mark for taxa protected by the Law for protection of Nature in the Republic of Croatia.

The data on endangered taxa (273) originated from Red data book [37], and were modified (mostly supplemented, now more than 400) with data from the new CCF. The complete data set contains information about the endangered taxa in other European countries.

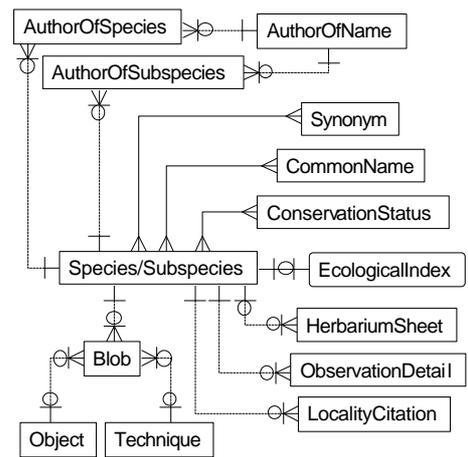


Fig. 2 – Species/Subspecies and related data

The data on the authors of scientific names (*AuthorOfName*) is stored according to [7]. Initially, the data were obtained in digital form and were imported into CROFlora.

The abbreviated author's name is used as basis to form valid names of taxa. Additional information about the authors of scientific names (*AuthorOfSpecies*, *AuthorOfSubspecies*) contains a prefix and a suffix (parenthesis, *et*, *non*, etc.) that are used to create the full scientific name of taxa. For example, full scientific name *Vitis vinifera L. ssp. sylvestris (C. C. Gmelin) Hegi* is automatically created based on the values for genus: *Vitis*, species: *vinifera*, author of species: *L.*, subspecies: *sylvestris* and authors of subspecies: *C. C. Gmelin* (prefixed and suffixed by a rounded parenthesis) and *Hegi*.

Relationship between taxa (*Species/Subspecies*) and an universal set of synonyms (*Synonym*) carries information about the type of the synonym (*basionym*, *exclusive*, *inclusive*, *nomen ambygum*, *nomen nudum*, *nomen illegitimum*, *pro parte* and *doubtful*), information about the author of non-valid name and place of publishing.

The initial data on the vernacular names (*CommonName*) were included in database according to [11]. The data entry for common names from several other sources is currently in progress.

The multimedia data (images and video clips) are stored as binary large objects (*Blob*). The additional data describe multimedia contents (*Object*), such as whole plant, fruit, leaf, blossom, pollen, etc. The applied method (*Technique*) can be for instance digital camera acquiring photography/slide scanning or

microphotography. Photo documentation is related to the CCF. It was mostly produced by the associates of the Department of Botany. The production of multimedia is still being in progress.

Binding to the DDescription Language for TAXonomy (DELTA) standards [10] is foreseen for future development.

2.2. Ecology

Ecological data include index parameters (see Fig. 3) according to [12], [14], [15], [16], [20], [21], [22], [23], [31], [34], [35] and [36].

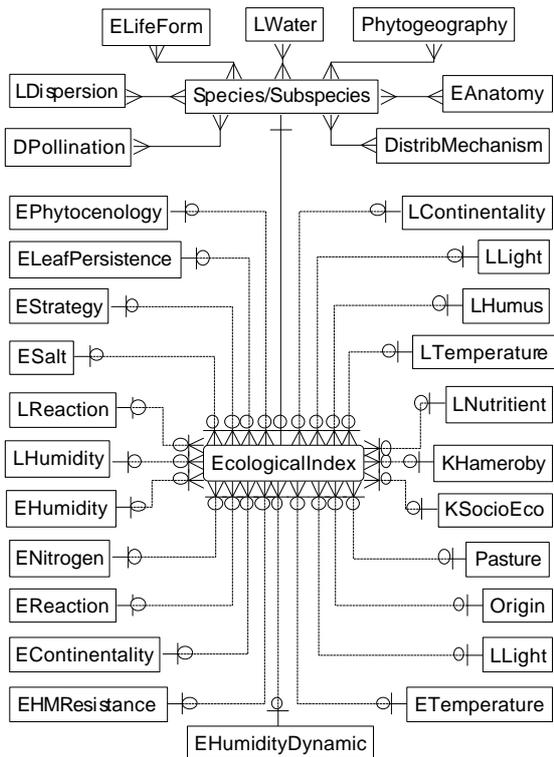


Fig. 3 - Model of ecological index parameters according to several authors

Altogether, thirty ecological parameters for 7300 taxa were obtained in digital form from two sources. The first part of data was obtained on the commercial basis from Verlag Erich Goltz Goltze / Co. KG, as a part of database SCRIPTA GEOBOTANICA XVIII. The second part of data was imported from database PHANART and PHANSYS [24].

2.3. Herbarium

Herbarium module (Fig. 4) stores the data about the herbarium collections (*Collection*) and sheets (*HerbariumSheet*), specimens (*HerbariumSpecimen*)

and persons (*Author*) who collected (*Collector*), determined (*Determinator*) or re-determined (*Redeterminator*) the species. The re-determination data (*Redetermination*) include the re-determination date and the author's comments.

The *HerbariumSheet* stores all the data necessary for the management of herbarium. For the flexibility reasons, the user is allowed to store only the information about the genus and/or the aggregate of the specimen (i.e. temporarily not determined below genus).

The information about the type of collected specimen (*SpecimenType*) is codified as *holotype*, *isotype*, *lectotype*, *neotype*, *paratype* and *sintype*. The origin of the specimens (*SpecimenOrigin*) can be designated as a deposit, a substitution, a purchase or a gift.

Collected specimen (*HerbariumSpecimen*) can be described (*SpecimenPart*) as fruit (carpological collection), wood, root, seed, leaf, whole plant, etc. Predefined modes of preservation (*PreservationMode*) include dried material, liquid media, silica gel media etc.

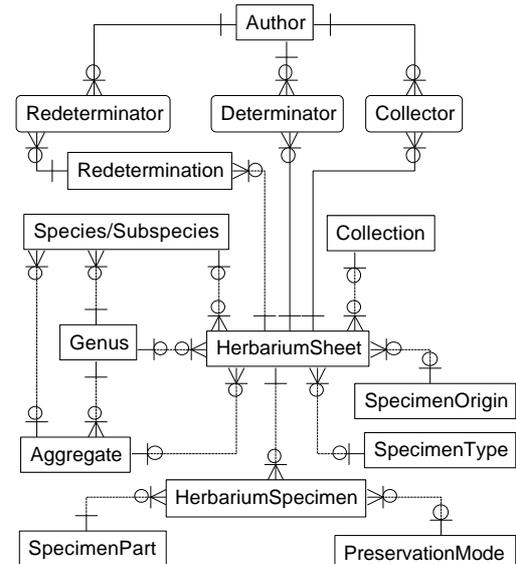


Fig. 4 - Herbarium sub-model aimed to handle information about collections

2.4. Literature

Bibliographical data are stored in a database module presented by the data sub-model shown in Fig. 5.

Similarly to the aforementioned generation of the full scientific name of the taxa, the full name of the reference (*Literature*) is generated automatically, based on the data on authors (*Author*, *Authorship*), year, title, volume, page range, publisher etc.

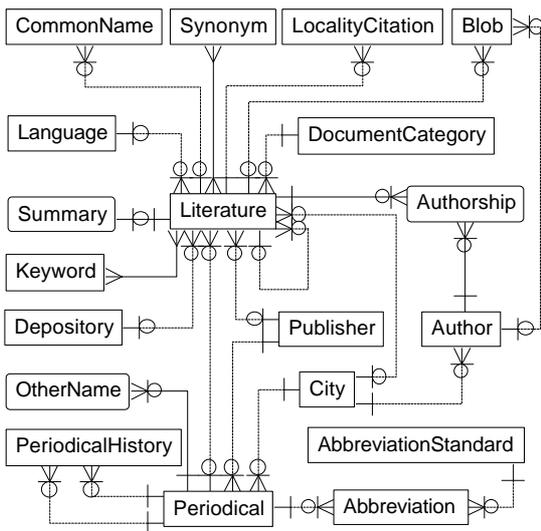


Fig. 5 - Sub-model aimed to handle information about literature

As a reference can be part of another reference, a recursive relationship has been implemented. The application preserves the data consistency by use of predefined document types (*DocumentType*). For example, an article can be published in conference proceedings or in a book.

Common names, synonyms, locality citations, as well as the blob data can be referenced by literature (see Fig. 5). The relationship between literature and synonyms holds the information about the citation type (citation, original description and original rejection).

The data about the periodicals are stored separately, extended by the history of periodical names (*PeriodicalHistory*), a set of abbreviated names and a standard (*AbbreviationStandard*) which was applied on abbreviations (BPH, ANSI, DIN, TS/10 etc.).

The public part of bibliography contains 5330 references. Most of the journal title abbreviations are synchronised with the well-known *Botanico-Periodicum-Huntianum/Supplementum* [6].

The bibliographical contents have been collected and processed by various collaborators during the last five years. Many active and retired botanists have given their contribution to the bibliography by sending their papers or their lists of publications. The ulterior necessary corrections and supplementation are in progress.

2.5. Horology

The part of the data model presented in Fig. 6 defines data about the on-site observations of the species (*FieldObservation*), issued by persons (*Author*). Every observation is further described by a set of gathering events (*ObservationDetail*) that take place on a wider geographical area (*District*). More precise information about the geographical positions is geocoded in separate database tables (*MTBSquare*, *UTMSquare*) (predefined according [29]). The same codes are applied to herbarium sheets and locality citations found in literature. In this way an extensive cross checking can be performed. Both the *ObservationDetail* and *HerbariumSheet* hold the information about the altitudes at which taxa resides (*AltitudeClass*), such as *planar/mediterranean*, *colin/submediterranean*, etc.

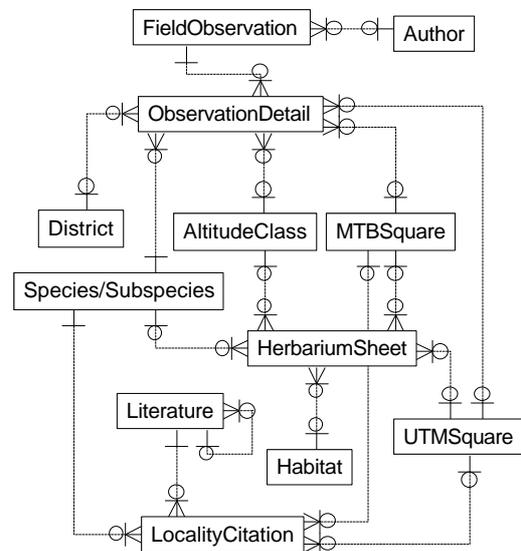


Fig. 6 - Horology sub-model to manage geographical data from herbarium, literature and filed observations

Habitat preferences of the taxa can be taken from a predefined set of values (*Habitat* in Fig. 6), such as different forest types, different types of meadows, marshes, etc. In addition, altitude is available in explicit values (from-to in meters) and/or in form of predefined altitude classes.

CROFlora can be connected to a GIS application (*ArcView*, *ArcInfo*) via ODBC link, which makes possible to search for data in the database from the GIS side by using SQL statements. Easy production of distribution maps is available. Usage of ecological indexes and other data with different GIS layers have made complex spatial analysis possible. Geographically, the data correspond to the area within the Croatian borders.

3. The Client Application

The client application is divided into several modules, following the database organisation: Taxa hierarchy, Species/Subspecies, Herbarium, Field observations, Bibliography, Basic tables, Reports and System administration. The management of taxa can be performed by using the screen forms to handle separate database tables and an additional module has been developed to help users in navigation through the taxa hierarchy. Every application module supports a set of standard functions such as query by example (QBE), a data entry and data update, a deletion function, a browse function and a standard report.

QBE function allows a user to search for data by entering patterns in almost every form field. The patterns can include wildcards. The patterns are used to dynamically create SQL statements, which are further used as record sources for the data. For example, when searching for the species/subspecies, the user can type **fera* into species name field and **Gmelin* into the field that represents the short name of *Author* that is related to *AuthorOfSubspecies*. The following SQL statement is produced, where the word *Taxon* stands for Species/Subspecies:

```
SELECT DISTINCTROW Taxon.*
FROM (Taxon INNER JOIN AuthorOfSubspecies
ON AuthorOfSubspecies.TaxonId = Taxon.TaxonId)
INNER JOIN
AuthorOfName AS AuthorOfSubspeciesName
ON AuthorOfSubspecies.AuthorId =
AuthorOfSubspeciesName.AuthorId
WHERE Taxon.SpeciesName LIKE "*fera"
AND AuthorOfSubspeciesName.ShortName
LIKE "*Gmelin"
```

The resulted data set contains taxa having full scientific name *Vitis vinifera L. ssp. sylvestris (C. C. Gmelin) Hegi*. Of course, for the given example it would be more appropriate to search for all species/subspecies having the full scientific name that matches the pattern **fera*Gmelin**. In that case, the SQL statement would be as follows.

```
SELECT DISTINCTROW Taxon.*
FROM Taxon
WHERE Taxon.FullName
LIKE "*fera*Gmelin*"
```

Query results data can be browsed record by record on predefined screen forms or the data can be browsed in a

customisable grid. Queried data can be dynamically sorted by almost every attribute.

As described in previous chapters, the application automatically generates full scientific taxa names and literature references. The re-determination data (*Redetermination* and *Redeterminator*) can be handled manually or the data can be automatically generated when the name of the taxa changes in the main species/subspecies form. In the latter case, the program generates a synonym that is equal to the previously stored valid name of the taxa. In addition, the user decides whether she/he wants to generate the re-determination data. Finally, the user decides about an automated modification of the related data (herbarium, literature citations and field observations). In addition, the user of the application is asked to select the type of the synonym.

Sophisticated reporting is provided for major groups of data (Herbarium, Species/Subspecies, and Literature). Sophisticated reports can be dynamically customised by selecting the groups of attributes for which the values are going to be printed. In addition, the user can decide whether to report the current record or all records selected by the last query.

The administration module provides information about users of the application. The proprietary data transfer protocol has been defined and related data transfer routine has been implemented.

4. The Web Site

The web site (<http://hirc.botanic.hr/croflora>) was built to enable quick access to the CROFlora database via Intranet/Internet. The site is divided into three logical sections: CIS-B Database, CROFlora Taxonomy and CROFlora Bibliography. A navigation system enables users to find and browse for the species and bibliographies quickly and easily.

CIS-B database can be searched and browsed. Search fields can be combined and wildcards are allowed (e.g. division: *Ascomycota* AND genus:*Arthop** AND *Endemic in Croatia*). The resulted species are listed in the table, and one can choose to see the details about any of them. Browsing the database provides the users with quick access to any part of the species hierarchy. One frame shows the currently selected species hierarchy node and links to its child and parent nodes. The other frame shows the search form that is

automatically filled with the currently selected hierarchy nodes.

CROFlora Taxonomy deals with data on vascular flora and can be searched and browsed by using the same web forms as for the CIS-B database. However, some additional fields are shown such as tags for spurious, naturalised and cultivated species, species with pictures, IUCN category, etc. There can be several pictures shown for species/subspecies (habitus, inflorescence, and other photo details...).

CROFlora Bibliography can be searched by author, title, keywords, reference identifier and date of publication. Search results are displayed in a list, which can be viewed one page at the time. The details about a single reference can be shown on a separate page.

5. Technical Considerations

The server runs under Microsoft Windows NT. On the server side, Microsoft SQL Server manages the data.

The client application is written in Microsoft Access, enhanced by some ActiveX controls [38], such as Microsoft Common Controls and Microsoft Multimedia Control, and by some dynamic link libraries, for instance Microsoft Office Library and Microsoft Word Library.

The client connects to the server by using the ODBC data source. Some database tables are replicated locally to ensure better response times of the user interface. The synchronisation of the local data is done automatically or the synchronisation can be forced by user's request.

Internet Information Server 4.0 has been chosen as the platform for the web site. Web site supports multilingual features. Each web page contains phrases in all supported languages, and displays only phrases in the user-preferred language when the page is requested.

Some parts of the software were written in Visual Basic, for instance a program to acquire the pictures and video clips and a program to handle batch conversion of the pictures in various formats.

6. Conclusion

To enhance the research capabilities of botanical scientists, such as taxonomists, ecologists, collection managers and bio-geographers a database application

CROFlora has been developed. Its multi-user capabilities and Internet access make it appropriate also for co-operation by scientific teams on the global level.

The information to be handled consists of nomenclature, distribution, classification, ecology and literature. The representation makes use of multimedia. Information about specimen covers the sites, collectors, time of collection, storage locations, loans and accession and catalogue numbers. A number of prepared reports deals with the material for a taxon, list of sites for species, reference cards and list of synonyms. The end users can customise more complex reports.

Application of a database as the storage of information has enabled some floristic work, which had formerly been lacking for this region. Connection to a GIS application resulted in the Atlas of Croatian Vascular Flora with different types of spatial analyses connected with ecological indexes. Complex and customised reports make possible different sorts of publication such as checklists, catalogues, revisions and monographs, as well as Flora Croatica, which need to be produced in the near future. It also provides for data exchange with other similar systems on the global level.

Acknowledgements

Authors wish to thank for kind support and provision of the data in digital format to Mr. W. Greuter for NCU data, to Mr. R. J. Pankhurst for data from Flora Europaea Database, to Mr. R. K. Brummitta for data about authors names and to Mr. R. Lindacher for ecological data from PHANART and PHANSYS databases.

References

- [1] L. A. Abbot, F.A. Bisby, D.J. Rogers, *Taxonomic analysis in biology*, Columbia University Press, New York, 1985
- [2] R. Allkin, F.A. Bisby, *Databases in systematics*, Academic Press, London, 1984
- [3] Anonymus, *IUCN Red List Categories*, The World Conservation Union, Gland, 1993
- [4] Anonymus, *FLOREIN, Interaktive Programm zur Bearbeitung floristischer Daten, Version 4.1. Benutzerhandbuch*, Zentralstelle für die Floristische Kartierung Deutschlands, Regensburg, 1995
- [5] W.G. Berendsohn, A taxonomic information model for botanical databases: the IOPI model, *Taxon*, No.46, 1997, pp. 283-308.

- [6] G. D. R. Bridson, *B-P-H/S. Botanico-Periodicum-Huntianum/Supplementum*, Hunt Botanical Institute, Pittsburg, 1991
- [7] R. K. Brummitt, C. E. Powell, *Authors of plant names*, Royal Botanic Garden, Kew, 1992
- [8] A. Cronquist, *An integrated system of classification of flowering plants*, Columbia Univ. Press, New York, 1981
- [9] T. J. Crevello, R. D. MacDonald, Index of EDP-IR projects in systematics, *Taxon*, No.19, 1974, pp. 63-76.
- [10] M. J. Dallwitz, T. A. Paine, E. J. Zurcher, *User's Guide to the DELTA System: a General System for Processing Taxonomic Descriptions. 4th edition*, CSIRO Division of Entomology, Canberra, 1993
- [11] R. Domac, *Croatian Flora*, Školska knjiga, Zagreb, 1994 (in Croatian)
- [12] R. Düll, H. Kutzelnigg, *Neues botanisch-ökologisches Exkursionstaschenbuch (2., erw. Aufl.)*, IDH, Rheurdt., 1986
- [13] F. Ehrendorfer, *Liste der Gefäßpflanzen Mitteleuropas*, Gustav Fischer Verl., Stuttgart, 1973
- [14] H. E. Ellenberg, R. Düll, V. Wirth, W. Werner, D. Paulißen, *Scripta Geobotanica, XVIII Datenbank V 1.4. Zeigwerte von Pflanzen in Mitteleuropa*, Erich Goltze KG, Göttingen, 1991-1996
- [15] H. Ellenberg, Zeigewerte von Pflanzen in Mitteleuropa, *Scripta Geobot.*, No.9, 1979, pp. 1-122.
- [16] H. Ellenberg, H. E. Weber, R. Düll, V. Wirth, W. Werner, D. Paulißen, Zeigewerte von Pflanzen in Mitteleuropa, *Scripta Geobot.*, No.18, 1991, pp. 1-248.
- [17] W. Greuter, R. K. Brummitt, E. Farr, N. Kilian, P. M. Kirk, P. C. Silva eds, NCU-3. Names in Current Use for Extant Plant Genera, *Regnum Vegetabile*, Vol.26, No.129, 1993, pp. 1-1464.
- [18] V. H. Heywood, Electronic data processing in taxonomy and systematics. In *Databases in systematics* (R. Allkin, F. A. Bisby Eds.), pp. 1-15. Academic Press, London, 1984
- [19] S. L. Jury, Some recent computer-based developments in plant taxonomy, *Bot. J. Linn. Soc.*, No.106, 1991, pp. 121-128.
- [20] E. Klapp, P. Boeker, F. König, A. Stählin, Wertzahlen der Grünlandpflanzen, *Grünland*, No.2, 1953, pp. 38-40.
- [21] I. Kowarik, Zum menschlichen Einfluß auf Flora und Vegetation. Theoretische Konzepte und ein Quantifizierungsansatz am Beispiel von Berlin (West), *Landsch. Umwelt. Sch. FB Landsch. TU Berlin*, No.56, 1988, pp. 1-280.
- [22] W. Kunick, *Veränderungen von Flora und Vegetation einer Großstadt dargestellt am Beispiel von Berlin (West)*, Diss, TU Berlin, Berlin, 1974.
- [23] E. Landolt, *Ökologische Zeigerwerte zur Schweizer Flora*. Veröff. Geobot. Inst. ETH, Stiftung Rübel, Zürich, No.64, 1977, pp. 1-208.
- [24] R. Lindacher, *PHANART, Datenbank der Gefäßpflanzen Mitteleuropas. Erklärung der Kennzahlen, Aufbau und Inhalt*. Veröff. Geobot. Inst. ETH, Stiftung Rübel, Zürich, No.125, 1995, pp. 3-436.
- [25] T. Nikolija, CROFlora: Specimen and species orientated database. Presented at the *Book of Abstract, Workshop Disseminating Biodiversity Information*, Amsterdam, 1996
- [26] T. Nikolija ed., *Flora Croatica, Index florae Croaticae Pars 1*, *Nat. Croat.*, No.3, 1994, pp. 1-116.
- [27] T. Nikolija ed., *Flora Croatica, Index florae Croaticae Pars 2.*, *Nat. Croat.*, Vol. 6, No.1, 1997, pp. 1-232.
- [28] T. Nikolija ed., *Flora Croatica, Index florae Croaticae Pars 3*. *Nat. Croat. Suppl.*, 2000 (in press)
- [29] T. Nikolija, D. Bukovec, J. Šopf, S. D. Jelaska, Mapping of Croatian Flora: possibilities and standards. *Nat. Croat.*, Vol. 7, No.1, 1998, pp. 1-62. (in Croatian with extended abstract in English)
- [30] T. Nikolija, S. D. Jelaska, D. Holcer, Croatian information service for biodiversity on WWW. In *Book of Abstract, Workshop Disseminating Biodiversity Information*, Amsterdam, 1996
- [31] E. Oberdorfer, *Pflanzensoziologische Exkursionsflora (5. Aufl.)*, Ulmer E., Stuttgart, 1983
- [32] R. J. Pankhurst, Database design for monographs and floras, *Taxon*, No.37, 1988, pp. 733-746.
- [33] R. J. Pankhurst, An interactive program for construction of identification keys, *Taxon*, No.37, 1988, pp. 747-755.
- [34] W. Rothmaler, *Exkursionsflora für die Gebiete der DDR und der BRD. Band 3: Atlas der Gefäßpflanzen*, Volk und Wissen, Berlin, 1987
- [35] W. Rothmaler, *Exkursionsflora für die Gebiete der DDR und der BRD. Band 2: Gefäßpflanzen*, Volk und Wissen, Berlin, 1987
- [36] H. Sukopp, A. Auhagen, W. Bennert, R. Böcker, U. Hennig, W. Kunick, H. Kutschkau, C. Schneider, H. Scholz, F. Zimmermann, Liste der wildwachsenden Farn- und Blütenpflanzen von Berlin (West) mit Angaben zur Gefährdung der Sippen, In *Rote Listen der gefährdeten Pflanzen und Tiere in Berlin, Schwerpunkt Berlin (West)* (Sukopp, H.; Elvers, H. Eds.). No. 11, pp. 19-58. Landsch. Umwelt.
- [37] I. Šugar, *Red databook of plant species in Croatia*, State Agency for Environment Protection, Zagreb, 1994 (in Croatian)
- [38] A. Williams, Visual Basic 5 and ActiveX Controls, *Dr. Dobb's Journal of Software Tools*, No.3, 1997, pp. 74-74.