An Electronically Based Horticultural Information Retrieval System

G.L. Davis¹, Edward F. Gilman¹, and Howard W. Beck²

SUMMARY. A large horticultural database and an electronic retrieval system for extension education programs were developed using compact disk-read only memory (CD-ROM) and World Wide Web (WWW) as the medium for information delivery. Object-oriented database techniques were used to organize the information. Conventional retrieval techniques including hypertext, full text searching, and expert systems were integrated into a complete package for accessing information stored in the database. A multimedia user interface was developed to provide a variety of capabilities including computer graphics and high resolution digitized images. Information for the CD-ROM was gathered from extension publications that were tagged using the standard generalized markup language (SGML)-based document markup language (International Standards Organization, 1986). Combining funds from the state legislator with grants from the USDA and other institutions, the CD-ROM system has been implemented in all 67 county extension offices in Florida and is available to the public as a for-sale CD-ROM. Public access is also available to most of the database through the WWW.

ADDITIONAL INDEX WORDS. database, World Wide Web, CD-ROM, information retrieval

One of the challenges of the cooperative extension service is to deliver information in a timely manner to its customers. With reduced operating budgets and increasing population, distribution of printed information is becoming increasingly difficult. Institutions are implementing electronic delivery systems to serve their customers better. USDA handbooks (White et al., 1990), expert systems (Beck et al., 1993; Gilman et al., 1996, Plant et al., 1989), diagnostic software (Ferguson et al., 1995), and other horticultural information (Anonymous, 1995; Beck and Jones, 1995) are available on CD-ROM, video disc, or through the WWW. This provides increased access for professionals, hobbyists, and the public to horticultural information generated by the Univ. of Florida and others. In the past year there has been a tremendous growth in the Internet, the WWW particularly, for delivery of hypermedia applications consisting of text and images.

¹Department of Environmental Horticulture, P.O. Box 110670, University of Florida, Gainesville, FL 32611.

²Agricultural and Biological Engineering, P.O. Box 110570, University of Florida, Gainesville, FL 32611.

Florida Agricultural Experiment Station journal series no. R-05175.

The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked advertisement solely to indicate this fact.

It would appear that the WWW offers capabilities identical to those of CD-ROM, but with the advantages of providing updates instantaneously and having virtually limitless storage capacity. The WWW is currently available to millions of users and continues to grow in capacity. Unfortunately, currently Internet access in rural areas is severely limited by lack of adequate telecommunications. Most users must access the Internet through dial-up lines and modems at speeds no greater than 14.4 kB/s. This is too slow to access multimedia resources such as photographs or video, but is adequate for text.

An information retrieval system has been developed in Florida for delivery of horticultural and other agricultural information. We recognized the importance of developing information in a way that is platform independent, which means that widely distributed applications must run under DOS, MS Windows, and Apple Macintosh. Now they must also run under the WWW.

The Florida Cooperative Extension Service at the Univ. of Florida has been developing electronic information retrieval systems since 1982, beginning with the Florida Agricultural Information Retrieval System (FAIRS) (Johnson and Beck, 1986). In the earliest efforts (1989–94), all information generated through this system was delivered on CD-ROM only. More than 2 gigabytes of agricultural information is currently available on this electronic media representing >4000 documents and 23,000 pages including text, tables, line art, and images. Almost 20 discs have been developed to date. The discs are updated and distributed to all 67 county extension offices in Florida regularly. Now, much of the information on the CD-ROMs is accessible on the WWW and it is being accessed at least 100,000 times each week.

Electronic delivery advantages

One of the advantages of delivering information electronically is that it is virtually impossible for users to lose the information. Unlike a pamphlet or fact sheet, which can be lost or accidentally thrown away, an electronic document can always be found in the same place you left it—in the computer. Well-designed retrieval software also organizes information as you would in a filing cabinet. This allows quick access without having to leave your seat and eliminates looking in an actual filing cabinet or in a number of reference books for the information.

Our experience has been that the long-term costs of publication in an electronic format are significantly lower than in a printed format. Few horticultural reference books, for example, contain even 1000 color photo-

graphs because of the high cost of the color separation process. In contrast, several plant reference CD-ROMs are available with >2500 color photographs each (e.g., Gilman et al., 1996; Green Thumb Software, 1995; Hortocopia, Inc., 1995, 1996). Electronic media can also provide audio and video for the user.

Perhaps the greatest advantage of electronic media is the ability to package information in a modular format. This allows the user to extract pieces or modules from various documents to form a new, custom document including only the desired information. For example, let's say you were making a presentation to a homeowners association on care of red maple (*Acer rubrum*) trees along streets in a subdivision. You could search the database for all documents having the words red maple or tree and create a new document including all information in the data base on fertilization, irrigation, planting techniques, pruning, important pests, control measures, chemical safety, and pertinent regulations. Although several authors, each with a different specialty, originally prepared each of these individual documents, you can easily assemble parts of them into a new document and could use this as a handout for the meeting. Extensive graphics, compactness, and portability add to the advantages of electronic delivery mechanisms. As with all creative works, users of copyrighted information are expected to cite their sources and obey all applicable copyright laws.

Content of databases

The materials in the Univ. of Florida database are comprised primarily of complete extension handbooks or compendia. The current discs contain >36 handbooks, including the Master Gardener handbook, Citrus handbooks, Disaster handbook, Rice handbook, Florida pest control guides (including aquatic weed, insect, nematode, plant disease, and weed control guides), Energy handbook, Florida lawn handbook, Vegetable handbook, and many more (Table 1).

A single handbook may contain >200 extension documents (Table 2). A document may be >20 pages long. The handbook concept was developed to assemble a comprehensive set of documents. We found it important to obtain a critical mass of documents covering enough information on a particular commodity or subject area to be useful to county agents and others. The database must cover enough topics so that the user is confident about finding answers to a particular question. Most of this information is compiled into our WWW site. Individual topics such as urban trees, soils, citrus, agricultural safety, and landscape plants each have an

individual CD-ROM. These individual CDs are not necessarily accessible via the WWW yet.

Each document within a database consists of text, tables and often figures; figures may be black-and-white drawings or color photographs. Viewing the photographs requires a monitor with at least VGA resolution, though SVGA is recommended. For each document, there is also a WordPerfect file and a print file formatted for most printers. Thus, each document can be printed on demand or captured to your hard drive.

In addition to documents, record-oriented databases can be imported to your computer. Examples of record-oriented databases include the Univ. of Florida Institute of Food and Agricultural Sciences Faculty Directory containing a summary (record) of each faculty member, which can be searched by name or specialty. Another example is plant databases that contain original text, line drawings, range maps, detailed attributes, and >5000 color photographs of >2000 trees, shrubs, vines, ground covers, annuals, perennials, garden crops, and more. Plants can be searched over various characteristics. These plant databases are currently available only on CD-ROM.

Application programs include the entire software library of >60 programs from the Univ. of Florida software support office. These cover topics such as irrigation system design and models of soil and groundwater pollution. In addition, expert systems provide decision making for diagnosis and recommendations (Crassweller et al., 1993). For example, southern trees, an expert system for urban forestry, provides advice on selecting

species and planting trees in an urban environment (Beck et al., 1993; Gilman et al., 1996). The southern trees expert system helps users select trees for planting at a particular site by asking the user a series of multiple-choice questions. The program selects questions based on how previous questions were answered, making this a truly interactive expert system. Facilities for asking questions about the site were developed using a graphical user interface, including displaying photos to illustrate site characteristics. The object-oriented database management system (OODBMS; Beck et al., 1994) was used to store a taxonomy of site characteristics (including questions to ask and rules associated with these characteristics) and a taxonomy of trees organized under various characteristics. The expert system uses about 300 rules to determine which tree characteristics are desirable for a particular site, based on the user's answers to questions. Then a search of the database of 887 trees is made to find trees having these characteristics.

Creating, storing, retrieving, and managing this diversity of materials requires a comprehensive, flexible information retrieval system. Not only must this software handle text, but also images, record-oriented databases, expert systems, and application programs, each with specialized requirements. The architecture of the information retrieval software for supporting this broad range of functions is described elsewhere (Beck et al., 1994).

Full-text searching

Full-text searching is the most widely used technique for accessing text databases on CD-ROM or the WWW. In full text searching, documents can be retrieved that contain a particular word or combination of words or phrases specified by the user. Before we place materials on the WWW or on CD-ROM, every word appearing in every document is scanned and placed into an index during a preprocessing phase. Once the index, which is keyed by word, has been created, it can be used to identify quickly every document containing a particular word.

Table 1. Major content areas in Univ. of Florida electronic databases.

4-H plant identification guide	Pest control guides
Agricultural engineering	Aquatic weed control guide
Agricultural extension	Insect control guide
Agricultural safety	Nematode control guide
Alternative crops	Plant disease control guide
Animal science	Weed control guide
Beneficial insects	Pesticide poisoning
Citrus diagnosis	Program development and evolution
Citrus handbooks	Rice handbook
Commercial nursery	Services and resources directory
Community development	Soil Science
Dairy Science	Specialist directory
Disaster handbook	Sugarcane handbook
Energy topics	Turfgrass information
Florida citrus pest management guide	Urban trees
Florida sea grant college program	Vegetable gardening
Florida weeds	Veterinary medicine
Home economics	Water and Irrigation
Landscape plants	Water quality
Master Gardener notebook	Wildlife resources
Agent handbook	Woody ornamentalist

The full-text searching function can be used to search all of the handbooks simultaneously. For example, let's say you want to know about scale insects and their effects on landscape plants. Simply select the search button and type in the key words scale and landscape plants. Upon completion of the search, which takes several seconds, a list of documents containing both key words is displayed. Next, to each document title is a number, which indicates the number of times the words in the search query appeared in the document (number of hits).

Documents are sorted in descending order of number of hits. Presumably the document at the top of the list is most relevant to the user's request because it contains the most hits. The user can select any document for inspection. When selected, the document is displayed in a window with the key words highlighted. You also have the option of listing documents containing either the word landscape plants or scale. This would display a longer list of documents.

Full-text searching does present a number of known problems, however. The retrieved documents may be inapplicable to the user's needs. The search may be imprecise in that documents contained in the database that are relevant to the user's needs are not located. Various attempts have been made to improve full-text searching, such as use of a thesaurus. The main problem with full-text searching is that it is based on string matching (words) rather than concept matching (concepts). Research is underway in using the OODBMS to give a richer representation of word meaning and ultimately improve searching through concept matching (Beck et al., 1994).

Hypertext

One advantageous feature of the full text searching function is that it includes complete integration with hypertext. The document retrieved during full-text searching is a hypertext document, the format of which is described below. Users can browse forward or backward through the horticulture electronic database with the hypertext network by jumping from one concept to the next along connecting links. Each topic or concept is highlighted in a contrasting color. The link is activated with a click of the mouse. Topics or concepts can contain not only text, but also line drawings, color photographs, software programs, or sound.

Distribution and evaluation

The Florida Institute of Food and Agricultural Sciences CD-ROMs are distributed regularly to each of 67 county extension offices in Florida. Most of these counties have at least one 80386 MS-DOS computer with

VGA+ graphics capability, a 100-megabyte hard disk system, and a CD-ROM player. The CD-ROMs, including installation software, are mailed to each office whenever a new release is available, typically twice each year.

The plant selector programs have received more favorable comments than other applications in the system. Users like the searching capability and the thousands of photographs of plants (Gilman et al, 1996; Gilman et al, 1991). They have requested more images and more databases in general. The programs are also used by many customers in other states because much of the information has application elsewhere. A broader array of images and plants are added periodically and are integrated into the master database as the electronic media programs expand.

The FAIRS electronic media have been in use for more than 7 years. The latest version was distributed to >500 sites. Feedback obtained by talking with county agents indicates that the electronic media are used regularly. Master gardeners working with cooperative extension service offices in the counties have been one of the most active user groups of the system. The CD-ROMs are used to provide training as well as an information resource for answering questions.

Because of the interest shown by county agents, more extension specialists have expressed interest in placing their materials into the electronic system. This, coupled with improvements in the publication handling procedures at the university, have helped to gain acceptance for electronic technology. Because the WWW interface has been in place for only about 2 years, evaluation of this medium is more difficult. Users are anonymous and their reactions to the usefulness of our electronic databases are difficult to measure.

Table 2. Major subject areas in the Florida Master Gardener handbook section of electronic media.

Subject area	Documents (no.)
Communications	2
Entomology	96
Fruit Crops	45
Houseplants and Flowers	13
Landscape	17
Nematology	8
Plant Pathology	5
Shrubs and Trees	22
Soils and Fertilizers	4
Turfgrass Information	68
Vegetables	13
Wildlife Resources	13

Conclusions

Future research will continue to enhance the capabilities of object-oriented databases, which will improve the usefulness of this system. In particular, object matching and machine learning techniques will eventually provide the capability for more precise queries, automatic evolution and organization of the databases for users, as well as those accessing the wealth of information via the WWW. We believe that it will take several years before the services provided by the Internet can replace CD-ROM. For now, we have demonstrated that it is possible to prepare materials once so that they can be distributed through both channels. Thus, we have the best of both worlds and have secured our sizeable investment in building databases for whatever future technical developments may occur.

Literature Cited

- Anonymous.* 1995. Putting knowledge to work CD 95B. Colorado State Coop. Ext. Serv., Fort Collins.
- Beck, H.W., P. Jones, and D.G. Watson.* 1994. A CD-ROM based information retrieval system. *Applied Eng. Agr.* 10(2):127-132.
- Beck, H.W. and P. Jones.* 1995. FAIRS disc 9. Florida Coop. Ext. Serv., Gainesville.
- Beck, H.W., E.F. Gilman, and P.A. Fowler.* 1993. An expert system for tree selection in urban forestry. *Applied Eng. Agr.* 10(5):743-747.
- Crassweller, R.M., J.W. Travis, P.H. Heinemann, and E.G. Rajotte.* 1993. Future use and development of expert system technology in horticulture. *HortTechnology* 3:203-205.
- Ferguson, J.J., F.S. Zaueta, and J.I. Valiente.* 1995. CITPATH: Diagnostic and hypertext software for fungal diseases of citrus foliage and fruit. *HortTechnology* 5:277-278.
- Gilman, E.F., H.W. Beck, R.J. Black, and K.C. Ruppert.* 1991. Florida plant selector CD-ROM. Florida Coop. Ext. Serv., Gainesville.
- Gilman, E.F., H.W. Beck, D.G. Watson, P.A. Fowler, D. Weigle, and N.R. Morgan.* 1996. Southern trees (2nd ed.): An expert system for selecting trees. Monograph 3, Florida Coop. Ext. Serv., Gainesville.
- Green Thumb Software.* 1995. The pro series plant database for CD-ROM. Boulder, Colo.
- Johnson, F. and H. Beck.* 1986. The design and use of the Florida agricultural information retrieval system (FAIRS). *Proc. Intl. Conf. Computers Agr. Ext. Prog.* 633-638.
- Horticopia, Inc.* 1995. Horticopia: Perennials and annuals. Purcellville, Va.
- Horticopia, Inc.* 1996. Horticopia: Trees, shrubs and ground covers. Purcellville, Va.
- International Standards Organization.* 1986. Information processing—Text and office systems—Standard generalized markup language (SGML). Rpt. ISO8879-1986, Geneva, Switzerland.
- Plant, R.E., F.G. Salom, J.A. Young, and R.E. Rice.* 1989. CALEX/Peaches, an expert system for the diagnosis of peach and nectarine disorders. *HortScience* 24:700.
- White, J.W., D.J. Beatie, and P. Kubek.* 1990. Inquiry learning with video discs and computers: an innovative teaching method for horticulture courses. *HortScience* 25:385-388.