
From Paper and Ink to CD-ROM: Digitizing the *World Book* Image

ROBERT J. JANUS

ABSTRACT

THIS ARTICLE DESCRIBES THE PROCESS by which the print version of *The World Book Encyclopedia* was initially converted to electronic form and then enhanced over several subsequent editions. The process involves much more than just the digitization of text and graphic elements and the ultimate addition of nonprint resources. Creating an electronic version of the encyclopedia meant capturing the image *World Book* had established as a unique reference work. The article discusses some of the major issues designers faced during the conversion process. It describes how the designers applied a combination of *World Book* tradition and technology to resolve such issues.

INTRODUCTION

The World Book Encyclopedia is a general reference set whose objective is to present selected information from the vast reservoir of knowledge in the most accessible and usable form. The editors design *World Book* articles especially to meet the reference and study needs of students in elementary school, junior high school, and high school. *World Book* also serves as a general family reference tool. Librarians, teachers, and the general public likewise turn to *World Book* to satisfy their everyday reference needs.

By the time digital technology began to change the very nature of the publishing industry, *The World Book Encyclopedia* had appeared in more than seventy editions. First published in eight volumes in 1917, *World*

Robert J. Janus, Word Book Publishing, Inc., 525 West Monroe Street, 20th Floor, Chicago, IL 60661

LIBRARY TRENDS, Vol. 45, No. 4, Spring 1997, pp. 602-622

© 1997 The Board of Trustees, University of Illinois

Book bore the subtitle “Organized Knowledge in Story and Picture.” At a time when supplementary learning materials were either hard to come by or largely arcane, *World Book* introduced a new concept: reference learning that emphasized interesting content, simple language, and many informative pictures.

The technology benchmark for the era in which *World Book* was born is evident in the list of the encyclopedia’s entries. For example, there appeared in Volume 1 the entry, “Aeroplane. See Flying Machine.” The “Flying Machine” article, in turn, conveyed another benchmark—a benchmark on the attitude toward technology. In an introductory section on “Faith and Unbelief” it reported:

It is a matter of record that when Wilbur and Orville Wright, the brothers who in 1903 made the first successful flier, announced to their family their intention to build a heavy machine which would fly in the air, their father took immediate steps to discourage so foolish an idea. He wrote to a professor in a great university, outlining the “crazy proposition” and asking him whether he believed it could be done. This professor in his reply declared a man-made flier to be a physical impossibility, and Mr. Wright handed the letter to his sons with the remark, “I told you so.”

The World Book Encyclopedia gradually grew from its original eight volumes in 1917 to twenty volumes in 1960. It reached its present count of twenty-two volumes in 1971 when the publisher added an index compiled with the aid of a computer-based information retrieval system.

During that long period of development, the editors had begun a system of continuous revision in which they updated articles throughout the set on an annual basis. As early as 1936, the editorial and research staffs had begun to compile and analyze contents of courses of study and to base content and readability decisions on such analyses. In 1955, *World Book* initiated its “classroom research” program. In this program, students in several hundred classrooms in the United States and Canada report on how they use the encyclopedia: what they look up, where they look for it, and whether they find what they look for. Every year, *World Book* researchers receive over 100,000 such bits of information and compile reports that show how many times each article in the encyclopedia is looked up at every grade level from elementary school through senior high. Editors use the classroom research data along with the courses of study data to keep tabs on what young people need to know and want to know. Grade-level data gathered in these research projects help the editors tailor vocabulary in a given article to the grades at which that article is most needed and most used.

As *World Book* grew and developed, it continued to employ emerging print and illustration technologies to improve the effectiveness of its resources. In 1960, *World Book* incorporated its first Trans-Visions in the

articles on the "Frog" and the "Human Body." By turning those acetate pages, printed in full color, readers peeled away layer after anatomical layer to reveal ever increasing detail. In 1961, *World Book* became the first encyclopedia published in braille. In 1980, it became the first encyclopedia reproduced as a voice recording. This recorded edition included audio cassette tapes of the encyclopedia, a special audio cassette player, and indexes reproduced in braille and large type.

When digital technology came of age, *World Book* faced the prospect of interpreting and adopting the new technology in a manner consistent with the image the encyclopedia had developed over the preceding seventy years. Alternately, the prospect could be considered one of reinterpreting the *World Book* image in the light of capabilities provided by the new technology. For example, digitizing the text would capture the carefully structured and tightly edited articles already designed to make information easy to find in print. But the book metaphor does not always succeed on a computer screen. What could *World Book* do to endow its digital incarnation with preeminent search capabilities made available by the new technology?

Even more fundamentally, for what medium should the digital version of the encyclopedia be designed to provide the greatest access for its users? At the time, CD-ROM had not yet achieved its position of superiority over other delivery systems. What minimum computer requirements would make the digital version most accessible? Should the software "push the envelope" and require the latest in computer memory, disk space, and screen resolution? Or should the program be designed for the more modestly configured computer systems that were then available in many schools and libraries? How important would it be to have a network version?

This article describes some of the steps taken to resolve the issues listed above for the first several releases of the digital *World Book Encyclopedia*. These editions consist of:

- 1990 *Information Finder*—text only; DOS stand-alone and network versions; still available.
- 1994 *Illustrated Information Finder*—pictures and maps added; Windows and Macintosh, stand-alone and network versions.
- 1995/96 *The World Book Multimedia Encyclopedia*—sounds, animations, and videos added; Windows and Macintosh, stand-alone and network versions.
- 1997 *The World Book Multimedia Encyclopedia*—developed jointly with IBM; virtual reality segments, interactive simulations, new videos and animations, and online components added; operates with two disks; a work in progress at the time of this writing.

OVERALL DESIGN AND INTERFACE

World Book chose to make its first release—*Information Finder*—a text-only, DOS-based operating system on CD-ROM, available in both stand-alone and network versions. System specifications were pegged at a modest level consistent with the capabilities of computer systems in many schools and libraries.

Always noted for its ease of use, *World Book* took great care in designing the user interface. The basic uncluttered screen displays the outline of an article and the text of the article side-by-side. This parallel display evolved naturally from a long-standing tradition in the print set: the presentation of an outline had been a standard part of the study aids attached to major articles since the first edition in 1917. Rendered digitally, the presentation assumed and conveyed new power. Looking at the outline, the user sees at a glance the contents of the article. Moreover, because the outline is linked with the text, scrolling the outline or clicking on a heading takes the user to the corresponding place in the text.

Always present in *Information Finder's* basic display is a box for entering a search term. All the user has to do is to start typing, and the keystrokes appear in this box. Special keystrokes or mouse clicks allow the user to access the various functions, including the dictionary, related topics, bookmark, and notepad. The original DOS release was controlled by keystrokes. Mouse capability was added in a subsequent release. Many schools and libraries, particularly those operating a local area network, still use *Information Finder* with an annually updated database.

DATA PREPARATION

It was decided without debate that *Information Finder* would include all the text of *The World Book Encyclopedia*. The text had been digitized several years earlier as a development in the print production process. The annual updating operation produced changes in the digital database, which was then used to photocompose revised pages for the print edition. However, although the print database was in digital form, it was not thereby automatically suited for incorporation in a completely electronic version of the encyclopedia.

When the database was first digitized, it was put into a special format, with codes that identified article titles, various heading levels, cross-references, extract copy, and the like. In all, the database used over 400 codes, but more than 300 of them had meaning only for the page composition process. These codes had to be stripped from the database, leaving only the ninety or so required for electronic display and retrieval.

Database specialists also had to clean up numerous other peculiarities that had not bothered the photocomposition process but could stall or even crash an electronic retrieval system. For example, consider such a basic language unit as the sentence. As will be shown later, the sentence

is an important unit in the operation of an electronic retrieval system. One might expect that at least the end of a sentence would be easily identifiable by virtue of the period. But a database has many periods—e.g., in abbreviations and as decimal points—that do not designate the end of a sentence. The end of a sentence in fact is defined by a period *and two spaces*. Although most of the sentences in the print database ended in this fashion, many did not because the photocomposition system operated just as well without the extra space. Finding and correcting those faulty endings became an important task in converting the print database to what was called Retrieval Data Base (RDB) format.

In the RDB, a unique six-digit number identifies each article title. These numbers, called MTL numbers (Master Title List numbers) have been in use for many years for managing production in the editorial and prepress departments. Their precise function is to determine, by means of their numerical sequence, the alphabetical order in which titles should appear. *World Book* needs such an alphabetizing device because its preferred word-by-word order does not match the letter-by-letter system used by computers. MTL numbers took on new importance in the RDB—i.e., they established links between articles and their tables, cross-references, and other resources.

Tables presented a formidable formatting task. *Information Finder* would include over 1,700 tables, almost every table in the print set. Only a few specialized tables could not be formatted for the computer screen. *World Book* tables fall into several general types: some have many columns with numbers or text, others have only one column consisting of sentences or paragraphs of information. Database specialists created programs that processed each group of tables. The programs reformatted each table so that it could be displayed to best advantage on the screen, and so that its content would be accessible via the search system. The programs were largely automatic, but they required critical input from a human operator.

DOS technology imposed a limitation on displaying text: the type fonts available in DOS did not include the specialized characters that appeared in some *World Book* articles. In such cases, an editor deleted the special character and “wrote around” it. This occurred, for example, in the article on “Shorthand” and in some mathematics articles, and represents the only instances in which *Information Finder* does not contain the full text of *The World Book Encyclopedia*.

Having been put into RDB format, the text components could be displayed properly and also searched. While the RDB formatting operation was in process, the search engine was being designed and built.

IN SEARCH OF A SEARCH ENGINE

The discussion of the search capabilities of *Information Finder* in the

next sections also applies to all subsequent versions of the digital *World Book*, which utilize the same search systems.

The basic *Information Finder* interface made it easy for the user to enter a search term by always displaying the search-term entry box. The key issue for *World Book's* designers was what happens after the user types a search term. To some developers, the solution would be a "no brainer." Digital technology, after all, brings with it its own set of tools, among them standard tools for searching a database. To be sure, *Information Finder* would employ Boolean searching. But not one off-the-shelf search system worked fast enough to meet the stringent response-time requirement set by the designers. The designers therefore developed *World Book's* own search engine. This engine incorporates specially developed data handling routines that make it the fastest search system available in a CD-ROM encyclopedia.

But Boolean—or what *World Book* calls "keyword"—searching is only half a solution. An encyclopedia also needs topic searching. The typical result of a topic search is a list of encyclopedia article titles that are associated with the search term. If the results list is lengthy, the user—particularly a young student—may have difficulty determining which of the titles has the information being sought. On the other hand, the list might be too short—or it might even be blank—because the user did not key an appropriate search term. *World Book* designers adjudged such a result not acceptable, and they set about developing a unique topic search system. Once more, they solved the problem with a blend of tradition and new technology. The *World Book* index, volume 22 of the print set, provided the traditional component.

The *World Book* index, consisting of more than 150,000 entries prepared by professional indexers, provides a guide to topics discussed in various places under different titles throughout the encyclopedia. The designers of *Information Finder* hit upon the idea of making the index the basis of a topic search engine. For example, when the user searches for "Environmental pollution," the engine consults the index and then takes the user directly to the "Environmental pollution" article. A keypress or mouse click then displays more than thirty related article titles, including such topics as "Pesticide," "Recycling," "Sanitation," and "Waste disposal." The user can select any or all of these additional hits, assured of their pertinence because a human intelligence has already associated each of these topics with the search term during the indexing process.

INFORMATION FIRST THEN LISTS

In *Information Finder*, both keyword and topic searching take the user directly to the article with the highest probability of containing the sought-for information. The user can always elect to see the results list and to choose other items from it. In this regard, *World Book* differs from other

reference works that present not an article but a list from which the user must make a selection. *World Book* thus operates on the bold assumption that it knows what the reader is looking for.

For example, suppose a student types the search term "elephant" and chooses keyword searching. The keyword search engine takes the user to the "Elephant" article, in which the word "elephant" is used eighty-seven times. The "Elephant" article happens to be the first article in a list of sixty-two. The designers chose to take the user directly to the first article in the list because they expect that the user would choose that article first if presented with the list rather than the article. The user may indeed consult the list and choose from among the sixty-one other articles in which "elephant" appears. Table 1 shows the first few and other sample items from the results list.

Table 1. Results of keyword search for "elephant" (abbreviated)

Titles	# Hits
Elephant	87
Seal [sea animal]	7
Animal	6
Elephant bird	5
Circus	4
...	...
Butler, Samuel [1612-1680...]	1
...	...
Cuvier, Baron	1

Two articles in Table 1 help illustrate the difference between keyword searching and topic searching: the article on the French naturalist Baron Cuvier states that he studied the rhinoceros and the elephant; the article on the English poet and satirist Samuel Butler (1613-1680) alludes to his work "The Elephant in the Moon."

Suppose the user now repeats the search but chooses topic searching. Table 2 shows the topic-search results list. The topic-search engine takes the user to the "Elephant" article because (1) Elephant [animal] is the first topic in the alphabetical index (left-hand column of the list, under Topics), and (2) because the "Elephant" article (right-hand column under Article Titles) occupies the first position in the title list by virtue of being placed there by an indexer.

Table 2. Results of topic search for "elephant" (abbreviated)

Titles associated with the topic "Elephant [animal]"

Topics	Article Titles
Elephant [animal]	Elephant
Elephant, Order of the [award]	Animal (table: Names of Animals and Their Young)
Elephant seal [sea animal]	Circus (At the Circus)
Elephant Butte Dam [Rio Grande, New Mexico]	Ear (The ears of Animals)
Elephant graveyards	Heart (Birds and Mammals)
Elephant man's disease	Ivory
Elephant trunk snake	Mastodon
...	
"Elephant in the Moon, The" [story by Butler]	

Table 3. Results of topic search for "elephant" (abbreviated)

Titles associated with the topic "Elephant Butte Dam"

Topics	Article Titles
Elephant [animal]	Elephant Butte Dam
Elephant, Order of the [award]	Rio Grande (Upper Course)
Elephant seal [sea animal]	Rio Grande Project
Elephant Butte Dam [Rio Grande, New Mexico]	
Elephant graveyards	
Elephant man's disease	
Elephant trunk snake	
...	
"Elephant in the Moon, The" [story by Butler]	

The list of topics in the left-hand column of Table 2 shows how effective topic searching can be in locating significant information within a topic. Because an indexer had deemed Butler's book indexable, that book appears in the topic-search results. Cuvier's study of elephants, on the other hand, was not indexed and so it does not appear in the topic-search results.

Note that the article titles shown in the right-hand column are those associated with the highlighted topic "Elephant" [animal] in the left-hand column. If the user were to highlight the fourth topic in the left-hand column—"Elephant Butte Dam"—the table would appear with appropriate titles in the right-hand column as shown in Table 3.

World Book's testing showed that topic searching often provided the youngest users of the software with the most direct answers to their questions, and so the designers made it the default search mode. Designers even programmed the search engine to find "elephant" when the user typed "elephants."

SERIOUS RESEARCH

That is not to say that serious Boolean searching was neglected. Far from it. The specially designed keyword engine was equipped to handle *and*, *or*, and *not* Boolean operators; allowed the user to set the proximity for Boolean searches to *word*, *sentence*, *paragraph*, *heading*, and *article* levels; and included wild-card capabilities.

One *World Book* representative liked to illustrate wild-card searching by performing a search on the term "murder*." The program responded to the asterisk in the search term by asking which terms beginning with "murder" were to be included. The demonstrator selected "murder," "murdered," "murderers," "murdering," "murderous," and "murders," but before launching the search, challenged those viewing the demonstration to guess what article would include the most mentions of those terms. No one ever got the correct answer—Shakespeare, William—but everyone understood that result after realizing that the "Shakespeare" article presents synopses of his plays. The "murder*" terms are mentioned nineteen times in the "Shakespeare" article, six more times than in the "Crime" article.

Other *World Book* representatives delighted in demonstrating how quickly *Information Finder* responded with an answer to the question "Who was the first woman senator?" The demonstrator typed "first & woman & senator" and the software quickly responded by displaying a portion of the history section of the "Georgia" article:

In 1922, Rebecca L. Felton became the first woman U.S. senator. She was appointed by Governor Thomas W. Hardwick after Senator Thomas E. Watson died. Felton served only one day while Congress was in session. Walter F. George, who had been elected to complete Watson's term, took her place.

The demonstration didn't stop there. The demonstrator pointed out that the computer interprets "&" as "and" with the default proximity "sentence" (which can also be written "&s"). The results of the search therefore included the single instance in the encyclopedia in which the terms "first" and "woman" and "senator" occurred in the same sentence. The demonstrator then modified the search term, setting the proximity to "paragraph" by typing "first &p woman &p senator," and launched the search again.

Once more, the program responded quickly, but this time the results were different: the article displayed on the screen was "Smith, Margaret Chase," who was the first woman to be elected to both houses of the U.S. Congress. The article on Rebecca Felton (the single result of the "sentence" search) also appeared in the results, along with two other articles: "Arkansas" and "Kassebaum, Nancy Landon." The "Arkansas" article said:

In 1932, an Arkansas woman, Hattie Caraway, became the first woman elected to the United States Senate. She won a special election held shortly after the death of her husband, U.S. Senator Thaddeus Caraway.

The "Kassebaum" article said:

Kassebaum...was the first woman elected to a full term in the United States Senate who did not succeed her husband in either the Senate or the House of Representatives.

Boolean searching had now led to four women senators who had some claim to being "first." Were there more? One more step could be taken to find out. A close look at the "Smith, Margaret Chase" article shows that her inclusion in the search result was serendipitous: the article states that "she became one of the first senators to oppose tactics used by Wisconsin Senator Joseph R. McCarthy," but otherwise it does not refer to her being first and being a senator in the sense intended by the search term. The article does, however, refer to her being first and being elected to the U.S. Senate. That's the clue: *senate*, not necessarily *senator*. *World Book's* wild card capability provides the final tweak to the search term.

The demonstrator types "first &p woman &p senat*" and launches the search. The software responds by asking the demonstrator to choose from among all the words in the database beginning with the characters "senat": "senate," "senates," "senator," "senatorial," "senators." The demonstrator selects all of them. The program fetches this article:

Caraway...Hattie Ophelia Wyatt...was the first woman elected to the United States Senate. Caraway, a Democrat, was also the first woman to head a Senate committee. In 1943, she became the first woman to preside over a Senate session.

One of the previous searches had already found information on Hattie Caraway in the article on Arkansas. The latest search finds it again, this time in the article on Caraway and with added information as a bonus.

The new results list also includes all the previous hits and adds three more. One of those three additional results is a repeat of the Margaret Chase Smith information, this time found in the article on Maine. The other two results consist of paragraphs in which a “senat*” term appears with the terms “first” and “woman” but is not related to them. That last search, then, did not add any new members to the group of first woman senators, but it did provide additional information, and it remains the best of the three ways to query the database. A researcher might tabulate the results as shown in Table 4.

Table 4. Tabulated results of search for “first & p woman & p senat*”

Individual	Notability
Felton, Rebecca L.	First woman to serve in the U.S. Senate (appointed 1922; served 1 day)
Caraway, Hattie Ophelia Wyatt	First woman elected to U.S. Senate (1932, succeeding her husband); first woman to head a Senate committee; first woman to preside over a Senate session.
Smith, Margaret Chase	First woman to be elected to both houses of the U.S. Congress (replaced her husband in U.S. House, 1940; subsequently elected to U.S. House; elected to U.S. Senate, 1948)
Kassebaum, Nancy Landon	First woman to serve a full term in the U.S. Senate who did not succeed her husband... (elected 1978)

An important point to be made is that Boolean searching is only as good as the database it operates on. In this regard, the digital *Information Finder* had the comprehensive text of *The World Book Encyclopedia* at its disposal and could provide the rich results tabulated above.

The Boolean-search story has an interesting natural-language sidebar. When *Information Finder*'s designers created the search system, they developed the following protocol—default search mode would be topic search. If the user's search term was not found in topic search, the program would ask the user if the search should be “broadened,” meaning should it be conducted in keyword mode. If the user said “Yes” but the term could not be found in keyword mode, the program would look for the term in the dictionary and display the result.

Let's say a user who knows nothing about Boolean searching begins in topic search and types "who was the first woman senator." The program responds "No topics found— Broaden search?" The user chooses "Broaden search." The program then eliminates the nonessential words ("who," "was," "the") from the user's search term, links the essential words ("first," "woman," "senator") with "&s" to form the search term "first &s woman &s senator." The program launches and completes the search and displays the "Rebecca Felton" article. What has happened is that the program has provided an answer to a question *typed in natural language*.

The digital version of *The World Book Encyclopedia* from its inception as *Information Finder* thus has had the capability of processing questions in natural language. On the one hand, as can be seen from the examples that appear in the research tabulation above, the results achieved from the natural language inquiry (one hit) may not be as complete as the results from a carefully crafted Boolean search term (four hits). On the other hand, the results may be far too extensive.

For example, suppose the user types "why is the sky blue" in topic search and elects to broaden search when no results are found. The program will list hits in twenty-three articles. However, only five of those hits actually deal with why the sky is blue. The others all involve some combination of the words "sky" and "blue" unrelated to the question being asked. The user, and especially a young user, would have difficulty discerning which of the twenty-three hits would be most likely to answer the question. The latent natural-language capability simply did not produce results that were good enough by *World Book* standards. As a result, *World Book* has never touted it.

SEARCHING AND VIEWING TABLES

Information Finder, like its print parent, is rich in tabular material, containing over 1,700 tables. The tables in the Illinois article illustrate the kinds of tabular material presented in all the states of the United States and the provinces of Canada:

- Illinois in Brief
- Interesting facts about Illinois
- Universities and Colleges
- Places to Visit
- Annual Events
- Average Monthly Weather
- Economic Production

Many other types of tables present valuable information throughout the database. *World Book's* designers knew that this information had to be accessible to topic and keyword searching.

Topic searching was not a problem: once the topic search engine had been developed, terms could be found by topic searching as long as they had been indexed. Thus, a topic search for "Cahokia Mounds" takes the user to the "Places to Visit" table in the "Illinois" article, and searching for "Thomas Carlin" locates his entry in the "Governors" table.

Keyword searching also was easy for non-Boolean searches. For example, the keyword search engine had no trouble finding "Cahokia Mounds" and "Thomas Carlin" in their respective tables in the "Illinois" article. But suppose the user wanted to know what the weather was like in Cairo, Illinois, in April. The search term would be "weather & cairo & illinois & april." But "&" stands for "and" with the proximity "sentence," and the "Average Monthly Weather" table does not have any sentences, just names and numbers in columns.

This time *World Book's* unique solution lay in instructing the search engine to consider the elements of a table as parts of a sentence. The elements include all the titles, subtitles, column heads, and column entries (except for perishable numerical data). Launching the search for "weather & cairo & illinois & april" then takes the user to the portion of the "Weather" table in the "Illinois" article, shown in Table 5.

Table 5. Results of search for "weather & cairo & illinois & april"

AVERAGE MONTHLY WEATHER (in Illinois)					
<i>Month</i>	<i>High Temp. (F.)</i>	<i>Low Temp. (F.)</i>	<i>High Temp. (C.)</i>	<i>Low Temp. (C.)</i>	<i>Days of Rain or Snow</i>
			Cairo		
March	58	40	14	4	12
April	69	51	21	11	12
May	78	59	26	15	12
June	87	68	31	20	10
July	91	72	33	22	9
August	89	70	32	21	7
September	82	64	28	18	8
October	72	52	22	11	7
November	57	40	14	4	9
December	47	32	8	0	9

Another problem had to do with some tables' sheer horizontal expanse. The table of Presidents of the United States, for example, has data on forty-two presidents in thirteen columns that spread out over two pages in the print version of the encyclopedia. By *World Book* standards, displaying such a table on a computer screen presents special problems. In scrolling the table vertically, the user must not lose sight of the column heads or else the data in the columns will lose their meaning. Similarly, in scrolling horizontally, the user should not lose sight of the President's name in the left-hand "master" column. *World Book* designers therefore developed a unique display routine that keeps column heads in place during vertical scrolling and keeps "master" column entries in place during horizontal scrolling.

TESTING, TESTING,...

After making many choices and compromises, *Information Finder's* designers approached the moment of truth—actually not a moment but a full year—when their creation would actually be put to use by real people, both students and teachers. In fall 1988, the designers set up fifteen centers in schools across the United States to test the product. Special software tracked every search performed in the testing centers. Designers followed up with personal interviews. What do you think of the interface? Did you like the way text and outline worked together side by side? How well did topic searching work? Did searching work fast enough? Was the dictionary easy to use? And on and on.

Data collected during testing guided the designers in putting the finishing touches on *Information Finder*. The speed and accuracy of searching remained a key concern. One of the test results involved the "Michigan factor." This had to do with a search protocol designers had built into topic search. The protocol established a priority for the names of people so that if a user typed "Lincoln" the topic search result would be a display of the "Lincoln, Abraham" article and not the "Lincoln [Nebraska]" article. In general, if the user typed a single term and that term appeared in the index followed by a comma and a first name, the protocol would give all such hits precedence. The protocol was meant to help young users who might not know the full name of the person they wanted to know about.

During testing, students at a center in Michigan frequently typed the topic-search term "Michigan" because they wanted information about their state. The first result displayed on the screen was "Michigan, University of." In the index, the term "Michigan" was followed by a comma followed by what the computer thought was a first name. "Michigan, University of" looked like a person and, according to the protocol, that title jumped to first position in the results list. And so was born an adjustment to the protocol called "exact match override." In this adjustment, a select

group of nonpersonal search terms—such as Michigan—were electronically tagged so that they would override the people-first hierarchy. With the tag set, a user who types “Michigan” goes to the “Michigan” article and not to “Michigan, University of.”

Other test data confirmed that some peculiarities of usage could only be partially assisted. In topic search, people often type a plural search term when the target is singular (for example, “elephants” vs. “elephant”) or a singular search term when the target is plural (for example, “Slav” vs. “Slavs”). The best the software can do is first to look for the term exactly as typed, and if it can’t be found in the index to change it from singular to plural (or plural to singular as the case may be) and look again. “Elephants” will take the user to “elephant” because “elephants” is not found in the index and “Slav” will take the user to “Slavs” because “Slav” is not found in the index. But the term “seasons” is found in the index and so the program will take the user to the section in the “Haydn, Joseph” article that talks about his composition *The Seasons*. The article on “Season” does appear later in the results list.

ILLUSTRATED INFORMATION FINDER

As soon as *Information Finder* was launched, many schools began requesting a Macintosh version. In addition, competitive pressure mounted for a version with pictures, if not full multimedia. The time came to deliver a Macintosh version and also to take advantage of Macintosh’s indigenous graphics-handling power by including pictures and maps. That meant that a corresponding Windows version had to be created to enable graphics to be displayed on IBM-compatible computers. This amounted to a major development task to be completed in a short time. The interface had to be redesigned to accommodate access to the new graphic resources. Many decisions had to be made. How many pictures? Which ones? How big would they be? How sharp should they be? How many maps? Which ones? How detailed should they be? Each answer had a dollar sign in front of it. And each answer claimed its own bit of “real estate” on the CD-ROM.

Some things would not change: the product would continue to use the text of *The World Book Encyclopedia*, and it would continue to use *Information Finder*’s topic search and keyword search. The interface would still display the outline and text side-by-side, but icons for the graphic resources would be placed in the outline for easy access to those resources. The text components—captions and labels—of the new graphics would have to be put into proper format. Of course, every caption would be accessible to keyword searching. The *World Book* extra was that every label on every diagram and every label on every map would also be searchable. This was not something available in other CD-ROM encyclopedias.

In order for a label to be searchable, the computer has to know where it is within the boundaries of the image. If the computer doesn't know where the label is, it can't highlight it, and the user can't see the hit. *World Book* designers had to create an electronic tool that enabled a person working at a computer to capture the x,y coordinates of a label and then associate that position with the appropriate label selected from a list. And then someone had to use that tool on every label on every diagram and on every map.

Choosing the pictures meant first selecting almost all of the teaching illustrations that had been prepared over the years for the print set. *World Book* owns these pictures and could incorporate them freely and permit them to be printed from the CD-ROM. Pictures not owned by *World Book* had to be selected more carefully. New license fees would have to be paid. For the most part, the right to print such pictures was not available. Designers chose to display pictures at an unusually large size and with a degree of resolution superior to that of other CD-ROM encyclopedias. Each picture had a small-sized "preview" version that could be displayed quickly on the screen and expanded to full size by clicking a button.

Maps present a special problem in digital software. The *World Book* print set contains many highly detailed maps of continents and countries, many of which are saturated with the names of places and physical features. Unfortunately, such maps cannot simply be digitized and displayed on a computer screen. The computer screen cannot match the high resolution available on a printed page. As a result, maps have to be created especially for computer display. Such maps typically have fewer place names than print maps. Despite the fact that *World Book* owned its own collection of excellent print maps, it had to license over 360 low-resolution maps created to its specifications by a cartographic producer.

In the *Illustrated Information Finder*, the maps are linked to form a tiered atlas. The map of the world forms the top tier. Here the user can click on a continent to display its map. At the continent level, the user clicks on a country to see its map. At the continent and country levels, the user may click on labels in surrounding areas to navigate to those places. If the place is a capital or other major city, the program takes the user to the article on that city. All maps in the atlas are linked to their articles: when viewing the Africa map, the user clicks on a "See article" button to go to the Africa article. Conversely, all maps are displayed within their "home" articles, appearing in preview size, that can be expanded just like other illustrations.

World Book designers chose to integrate map labels into the main database so that searching always includes the maps. In other CD-ROM encyclopedias, the user searches the atlas separately. An interesting demonstration of the value of integrating the map search is a keyword search for "Springfield." The results consist of sixty hits of which seven are to

maps: Illinois, Massachusetts, Missouri, Ohio, Colorado, Oregon, and Vermont.

Designers created three new features to provide ways to browse the enriched resources of the *Illustrated Information Finder*—Gallery, Timeline, and InfoTree.

World Book approached the browsing function gingerly. Other CD-ROM encyclopedias attracted much attention because of the various browse modes they offered. *World Book* knew the value of enticing the user into its rich resources by offering browsing tools but wanted to make sure that the user did not confuse browsing with directed research. Designers gave research preeminence over browsing by locating the “Search” and “Results” icons at the beginning of the set of main icons in the basic interface.

The Gallery browser consists of a set of postage-stamp sized icons representing categories of pictures. For example, a click on the Human Body icon displays a list of all the pictures dealing with the human body. Clicking on an item in that list displays that picture. Once the picture is displayed, the user can go to the associated article by clicking a “See Article” button.

Timeline offers access to selected resources associated with earliest times up to the present. Designers created Timeline especially for the CD-ROM version of the encyclopedia by extracting and adapting information from the many individual timelines in the print set. Also displayed in chronological context are selected images from the graphic database. The display helps the student visualize the time-relatedness of historic events and the lives of notable people. Clicking on elements in the timeline results in the presentation of a nugget of information or takes the user to an article or pictures.

InfoTree allows the user to browse through articles by content, starting with eight basic subject areas: Geography, History, Humanities, Industry and Technology, Life Science, Physical Science and Math, Recreation, and Social Science. Clicking on a subject opens up successive lists of subcategories and, eventually, displays article titles within a category. InfoTree is the CD-ROM embodiment of a tool that *World Book* editors have been using for over thirty years. Called the “Classified Title List,” this breakdown of articles into subject areas serves as the basis for apportioning updating responsibilities among subject editors. To create InfoTree, the editorial tool was simply transferred to the CD-ROM.

Designers wanted the *Illustrated Information Finder* to serve as many users as possible, and so minimum platform specifications again were modest. This meant that some sizing compromises had to be made to allow acceptable screen display on the Macintosh. In addition, the network versions became a concern because of the inclusion of graphics with their notorious ability to degrade response time.

The Macintosh network posed special response-time problems. To document and help solve these problems, *World Book* designers enlisted the aid of Tom Bookler, assistant to the superintendent for Technology at Addison [Illinois] School District 4. The district's computer lab, with its more than thirty Macintosh computers, provided an excellent proving ground for the alpha and beta versions of *Illustrated Information Finder*. Bookler also made the lab available to testers who launched searches with stopwatch in hand to document the quality of performance. As a result of such testing, designers developed procedures for transferring certain databases from the CD-ROM to the hard drive of the network server in order to bring performance across a network to an acceptable level.

THE WORLD BOOK MULTIMEDIA ENCYCLOPEDIA

Information Finder had started with text and then had added "sights"—pictures and maps. It was now time for "sound" and "action." Once again, designers approached the addition of new resources gingerly. Some other products with much sound and action were viewed somewhat as toys. *World Book* had to add meaningful and effective resources that expanded the research and learning opportunities afforded by the existing text, pictures, and maps.

Once again questions were asked. Which sounds are most important? How long will a sound clip be? What quality will it have? What subjects can best be explained by animation? What events will be dramatized in video? Again, the answers came with dollar signs and with claims on CD-ROM real estate.

But adding multimedia was fun. Of course, it was work to find out from whom a collection of animal sounds or symphonic music could be licensed, and work to struggle through the licensing process. It was work for editors to write scripts and then to cooperate with outside producers in developing animations and videos. The scripts arose out of existing materials already present in *World Book* text and illustrations, and—to use a well worn but still accurate cliché—they brought information to life in a way that text and still pictures never could. That's what made working on multimedia fun.

In its animation of the human circulatory system, *World Book* added something new to the technology by developing linked animations. While viewing the animation on the circulatory system, the user could click on a button to see five "detail" animations that dealt with blood, how the heart pumps blood, respiration, nutrition, and waste removal. These animations were all based on existing illustrations, which continue to exist as still illustrations that can be printed.

Designers inserted icons for the sound, animation, and video resources into the outlines and redesigned the Gallery screens to

accommodate the new multimedia features. They added selected videos to the Timeline presentation. All scripts became accessible to keyword searching. When all was done, one designer reveled in typing the query "why do dogs chase their tails" *in keyword mode*. Almost immediately, there appeared on the screen the first frame of the video on "Dog Behavior." Superimposed on it were the first words of the script: "There are reasons why dogs chase their tails..." If you ask the right question, natural language works perfectly!

JOINING FORCES WITH IBM

The ability to produce innovative and effective multimedia depends on having the right combination of the right resources. One of the basic resources is *content*, such as text, tables, and illustrations. Another is technological development capabilities, such as production facilities for creating animations, videos, virtual reality segments, online components, and so on. Yet another basic resource is a mechanism for distribution of product into the consumer market. The word *alliance* became a byword to many multimedia publishers who lacked one or more of these basic resources. Partner sought potential partner and, in such a milieu, *World Book* and IBM joined forces to produce the next release of *The World Book Multimedia Encyclopedia*.

The prospect facing the joint effort: so many opportunities to enhance learning by technology, so little disk space! In the beginning, the CD-ROM revolution had engendered amazement that a single disk could hold an entire encyclopedia. Almost immediately, the disk had shrunk when measured against the resources publishers wanted to add to it. There was irony here as well: publishers' designs had outstripped technology! The technological development that would solve the disk-space problem was still at least a year away. Called DVD (initials that officially don't stand for anything but are commonly interpreted as meaning Digital Video Disc), this method of formatting a CD-ROM is expected to make one disk the equivalent of about fifteen disks. In addition, DVD-formatted disks can only be played on a special drive. It will take time for the "installed base" of computers with such drives to become large enough to justify DVD development.

In the meantime, the solution to the space problem became: make the encyclopedia a two-disk product. But how could this be done and still preserve the ease-of-use standard? The basic elements and functions of the encyclopedia had to be integral on the main disk. The second disk would hold many of the complementary and supplementary videos and the virtual-reality segments. It was a risky decision not easily reached.

With the added "real estate," *World Book* and IBM designers added animations and videos and made them longer and better than if they had to fit them on a single disk. The newest virtual-reality technology was

incorporated to give students such experiences as walking around Chicago's Picasso sculpture and viewing it from different angles. Designers and editors created interactive simulations, among them one in which students drop water balloons from various heights to experiment with the laws of falling bodies.

The entire interface underwent metamorphosis, adopting a sleek modern look. Still easy to use, still functional, but more in tune with the added resources. The Gallery feature became "Just Looking," a control center from which the user browses through the encyclopedia's resources, perhaps selecting them by type or category. A new Time Frame feature presents access to *every* resource dealing with a user-selected year, century, millenium, or era. This feature replaces the previous Timeline, which provided links only to preselected items. The atlas was enlarged with numerous additional maps showing topography, climate, and land use—maps that many users of the CD-ROM encyclopedia had asked for.

Online linkage became an essential resource, providing for informative monthly updates. The encyclopedia also contains thousands of additional online links. *World Book's* Online Library, one of the linked resources, offers a collection of timely feature articles dealing with subjects ranging from genetic science to the greenhouse effect. An online feature called "Our Century" presents articles dating back to 1922 on continents, countries, industries, and inventions. These "time specific" expositions enable the student to view the world through the eyes of the authors writing at various times during the 1900s. Special links in many encyclopedia articles provide access to selected Web sites that offer particularly useful information relative to the topic at hand.

The school and library version of *The World Book Multimedia Encyclopedia* was endowed with *World Book's* unique "Guide to Better Writing, Speaking, and Research Skills." This feature consisted of an HTML (Hyper Text Markup Language) adaptation of helpful materials at the front of the print set's volume 22.

Consistently, the *World Book* tradition guided decisions. Consistently, compromises had to be made. In earlier versions, designers feared "falling off" one disk by having too many resources and too little space. Now they faced the prospect of falling off the second disk as well. Everything is reexamined. A designer laments: "It looks like the video on Dog Behavior with that neat natural language hit won't make the cut."

CONCLUSION

When *World Book's* founders coined the subtitle "Organized Knowledge in Story and Picture," they had no idea of how many ways knowledge would one day be organized. They had no idea of what technologies "story and picture" would eventually embrace. But they probably did know that their revolutionary concept, "reference learning that

emphasize(s) interesting content, simple language, and many informative pictures," would stand forever regardless of reinterpretation within new technological milieus.

Digitizing the *World Book* image remains a challenge as technological capabilities continue to burgeon. Cherishers of the *World Book* tradition have in fact been effecting and will continue to effect reinterpretations of that tradition in the light of new technologies. Those who know the tradition should have no trouble recognizing the *World Book* touch in present and future products.