Compact Optical Discs and the World Wide Web: Two Mediums in Digitized Information Delivery Services

Ziyu Lin
Assistant Professor
Systems Librarian
Seton Hall University
South Orange, New Jersey U.S.A.
E-mail: linziyu@shu.edu

【Abstract】

Compact optical discs (CDs) and the World Wide Web (the Web) are two mechanisms that contemporary libraries extensively use for digitized information storage, dissemination, and retrieval. The Web features an unparalleled global accessibility free from many previously known temporal and spatial restrictions. Its real-time update capability is impossible for CDs. Web-based information delivery can reduce the cost in hardware and software ownership and management of a local library, and provide one-to-one customization to better serve library’s clients. The current limitations of the Web include inadequate speed in data transmission, particularly for multimedia applications, and its insufficient reliability, search capabilities, and security. In comparison, speed, quality, portability, and reliability are the current advantages of CDs over the Web. These features, together with the trend in the PC industry and market, suggest that CDs will exist and continue to develop. CD/Web hybrids can combine the best of both developing mechanisms and offer optimal results. Through a comparison of CDs and the Web, it is argued that the functionality and unique features of a technology determine its future.

Keywords(關鍵詞)：
CD-ROM; DVD; World Wide Web; Digitized Information Delivery; CD/Web Hybrid; Internet Technology
1. Introduction

Compact optical discs and the World Wide Web are two mechanisms which contemporary libraries extensively use for digitized information storage, dissemination, and retrieval. Compact optical discs (CDs) have two formats: CD-ROM and DVD. Due to the increasing popularity and convenience of the Web and its fast-developing technologies, many database vendors, such as EBSCO, Information Access, SliverPlatter, Ovid, and UMI/Bell & Howell, have been moving steadily toward providing Web access to their products in lieu of the earlier CD-ROM formats. For information service managers and users, it is highly appropriate to wonder whether CDs will soon be, or are becoming, obsolete so that investment in CDs and the equipment to use them will no longer be worthwhile. Answers to such concerns are not simple, because many features and functions of these two types of mechanisms remain distinct, and such differences will probably be valid for some time to come. In many ways, the advantages of one mechanism at this stage complement the disadvantages of the other. For these reasons, it is significant to have an in-depth review and comparison of the major features and functions of CDs and the Web.

2. Major Characteristics and Concerns of Web-based Information Delivery

The most striking feature of the Web is its global accessibility that defies previously known temporal and spatial restrictions. Similar to television and radio networks, the Web offers real-time information delivery in multimedia forms to connected users anywhere and anytime. Therefore, gaining access to the Web has been a fad in both developed and developing countries. It is estimated that 66.6 million households worldwide will be online by 2000[1]. A late 1997 ARL (Association of Research Libraries) libraries survey shows that 97% of the surveyed libraries support patron access to the Web[2]. To a great extent, the Web equalizes the opportunity in information dissemination and retrieval, giving information providers the opportunity to extend their products throughout the world. This kind of global information distribution and the accompanying user conveniences are difficult, if not impossible, to achieve through CD products that are hosted on either a stand-alone workstation or a local area network.

Another outstanding feature of the Web is its real-time update capability, an attribute which is fundamentally different from CDs that “freeze” their data content in a polycarbonate snapshot at replication time[3]. At any time, and at the discretion of the information providers, Web-based databases can be updated and become immediately available to users. In this sense, Web-based databases are immune from the temporal gap that is inevitable in CD-based database distribution between the new information addition to the database and its lagged availability in the users’ hands. In addition, owing to the flexibility in real-time updating, the regulated intervals in distributing new versions of a database, which must be strictly observed in CD-based delivery, are now less important, if not irrelevant.

From the end-user’s point of view, another positive feature of Web-based information delivery is the simplicity of its hardware requirements.
All one requires is a decent desktop PC, or simply a net PC, or a Windows terminal, which is connected to the Internet. In terms of total cost of ownership (TCO), there are reduced hardware and software maintenance expenses at the local level. In the context of this paper, this applies to a library. On the one hand, vendors maintain the software and hardware of a remote system, and, on the other hand, the requirement for local hardware and software is less, thus reducing relevant costs. In this context, a system librarian no longer has to manage servers or server-clusters that are necessary to host CD-ROMs. In other words, the TCO in hardware and software, and in maintenance and management, decreases. This is illustrated in the following continuum:

<table>
<thead>
<tr>
<th>Server Cluster</th>
<th>Server</th>
<th>Workstation</th>
<th>Desktop PC</th>
<th>Portables</th>
<th>Net PC</th>
<th>Windows Terminal</th>
<th>Handheld PC</th>
</tr>
</thead>
</table>

+ TCO

Figure 1. Reducing TCO with Software and Hardware and Their Management[4]

The minimal TCO shown in Figure 1 is a handheld PC, which is a very thin client, but, in many cases, it is sufficient to take full advantage of the enormous wealth of Web-based information resources[5]. In information technology management, one of the tendencies is to make clients thinner. Web-based information delivery helps to realize this goal.

However, while Web-based information delivery can reduce a library’s TCO, it may increase the expenses of other departments outside the library, such as the computing services, since a greater investment is required to enhance the qualities of the overall information technology infrastructure, such as bandwidth, reliability, and user help.

Web-based information delivery also facilitates many customized one-to-one services that otherwise would not be possible. The means through which these personalized services can be implemented include the application of cookies, interactive profiles, and push technology. Web-based one-to-one services are extensively applicable to fields ranging from commerce to education, profoundly changing the manner of information acquisition and provision[6]. Web-based information providers can offer tools for building digital carrels that hold and organize customized information retrieved through the Web for a targeted audience. UMI’s ProQuest Direct SiteBuilder can help pack articles retrieved from their databases into a URL so that a specially selected audience, such as a group of students, can be directed later to it[7]. This type of online customized collection and reservation is highly desirable in teaching and learning. Another aspect of one-to-one customized services is that Web-based information providers can also provide their client with comprehensive database usage reports. For example, when using CD-based UMI
GPO and BPO products, a client library may receive reports that provide data about the number of documents printed out within a certain period of time. The Library’s management was not able to learn many details of how the databases were searched and displayed. With Web-based UMI ProQuest Direct, however, the monthly usage reports include both macro overviews of the total number of documents retrieved, as in Figure 2 and many micro-level counts of how each publication is accessed, as shown in Figure 3.

The type of data in the Figures 2 and 3 are valuable for decision-making.

While Web-based information delivery has many advantages, it is nevertheless subject to many restrictions that are caused by the current infrastructure and technology limitations in Internet development. The most noticeable complaints about the Web concern unsatisfactory data transmission speeds. The speed of an Internet connection is determined by multiple factors, including the connection of an end-user computer to the local network, and then to the Internet, and then the particular data transmission paths to the remote host, the processing power of that remote server, and number of simultaneous users. Most importantly, the speed for retrieval ultimately depends on a particular speed, or bandwidth, of the network. Figure 2 presents a comparison of times that are needed in downloading a 10 Mb video and audio file at VCR-quality compression rates using various bandwidths[9]:

<table>
<thead>
<tr>
<th>Delivery Type</th>
<th>Format</th>
<th>Within Subscription Quantity</th>
<th>Outside Subscription Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Mail</td>
<td>Abstract</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Electronic Mail</td>
<td>Citation</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Electronic Mail</td>
<td>Full Text</td>
<td>178</td>
<td>0</td>
</tr>
<tr>
<td>On-Line Display</td>
<td>Abstract</td>
<td>2535</td>
<td>0</td>
</tr>
<tr>
<td>On-Line Display</td>
<td>Citation</td>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td>On-Line Display</td>
<td>Full Text</td>
<td>4303</td>
<td>0</td>
</tr>
<tr>
<td>On-Line Display</td>
<td>Page Image</td>
<td>173</td>
<td>0</td>
</tr>
<tr>
<td>On-Line Display</td>
<td>Text+Graphics</td>
<td>1965</td>
<td>0</td>
</tr>
<tr>
<td>Total Documents</td>
<td></td>
<td>9291</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2. Summary of ProQuest Direct Usage

Database Name: The New York Times

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>Citations</th>
<th>Abstracts</th>
<th>Text Only</th>
<th>Text+Graphics/Page Image</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York T Silver</td>
<td>14</td>
<td>479</td>
<td>489</td>
<td>0/0</td>
<td>982</td>
</tr>
<tr>
<td>New York Times Book</td>
<td>0</td>
<td>11</td>
<td>5</td>
<td>0/0</td>
<td>16</td>
</tr>
<tr>
<td>New York Times Mag</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>0/0</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>499</td>
<td>495</td>
<td>0/0</td>
<td>1008</td>
</tr>
</tbody>
</table>

Figure 3. Number of Documents Accessed from the New York Times[8]
According to analysts at Gartner Group and analysts at the Yankee Group, about 65% of all Internet users are expected to be surfing with 28.8 Kbps modems through the year 2000[10]. This means that the majority of Internet Web users are not able to access large files, such as video and audio recordings, at a satisfactory speed. Even when accessing files through using a T1 line, which is unaffordable in the average household, the data transmission speed, with the aforementioned multiple factors being involved, can only roughly be an equivalent to a 1X CD-ROM drive, which is 150 Kpbs[11]. Due to the low Web speed, it is not unusual that only video with 5fps (frames per second) to 10fps shown in a tiny screen can be seen on the Web. As to static graphics, a picture as small as 35K takes 31 seconds to download at 14.4 Kbps or 15 second at 28.8 Kbps[12].

However, the good news is that new technologies are being developed to enhance Web speeds. Primary rate ISDN, for example, allows a data transfer rate of up to 2 Mbps, making it possible to transmit large data volumes within a reasonable time. With this technology, the transmission of 650 Mb for a CD-ROM would require about 47 minutes. In addition, a primary rate ISDN connection can be used for video confer-
Even for modern users, technologies such as streaming and compression are making Web video practical. Cable TV, telephone, and satellite companies are competing to bring megabit network speed to home computers. It is expected that one or more of them will succeed in the next several years, thus breaking the primary bottleneck for accessing the Web from home.

Reliability poses another difficulty in Web-based delivery. Data transmissions on the Web can be interrupted by a variety of problems that occur on a connection between a remote server and an end-user machine. Even worse, the end-user and the local institute can do little to correct the situation. Unlike operating CD-ROM towers on a local area network, the local institution, which can be a library, has no control over the remote server and little control over the paths leading to the institution. Hence, one of the consequences of TCO reduction is decreased local control over the reliability of a larger part of the whole information delivery system. Seen from a different perspective, Web-delivery increases the end user’s dependency on a larger system that is beyond the traditional local system, which typically has been a local area network managed by a library. This “larger system dependency” tends to become conspicuous with the advances in information delivery technology from paper to CD to the Web. In the paper format, a book delivered to a user’s hand contains a limited amount of information that requires no computer-assisted systems to deliver. One can open the book and read it, turning its pages or using its index to search for information. With CDs (650 Mb to 19 Gb of electronic storage), information must be retrieved through using a PC equipped with a CD-ROM drive or a DVD drive. Consequently, one has to depend on a relatively small system known as a personal computer. When many users intend to retrieve information from multiple CDs, they most likely have to depend on computers linked through a LAN. When users wish to obtain information from the Web, their retrieval success is contingent upon the Web, which is probably man’s largest network system made in history.

In terms of searching capabilities, it can be more challenging to locally customize the interface of a remote Web-based database with preferred features, such as loading local holdings, and adding the local search options. For instance, a system librarian can easily change the search and display options with the CD version of UMI’s ProQuest databases. However, the same job will be more complicated on the Web version. Furthermore, the current Web searches can only manage the most basic math. With complex math and tables, Web technology cannot handle the information. Special software is needed. For this purpose, software is being developed using Standard Generalized Markup Language (SGML) to allow complete searches of tables, figures, equations and sections. Java Client, launched by scientific information provider Ovid, for instance, is acclaimed for allowing users to access its Web databases with full search functionality.

Access security is an important issue in using subscription Web-based databases. Currently, accesses to a commercial database, such as UMI ProQuest Direct, are IP-controlled or/and password-controlled. An IP address is a sequence of four numerical byte addresses, each ranging from 1-253, such as 149.150.2.6 for the PIRATE.SHU.EDU server that leads to the home
interface customizability, search capabilities, and security. Yet, there is little doubt that Web-based information delivery is an upcoming and increasingly dominant vehicle of online services, offering an immensely attractive option to an earlier technology--CDs. The aura of CDs has been diminished. However, they and their related technologies and applications continue to be utilized and developed.

3. The Ubiquity, Uniqueness, and Survivability CDs

The ubiquitous, but now occasionally frowned-upon CD-ROM, provides up to 650 Mb of storage, enough capacity to accommodate an entire 28-volume encyclopedia, including sound, graphics, and full interaction that is impossible in a paper version[21]. 650 Mb of storage also represents about 60 minutes of MPEG-1 or AVI video (i.e., VCD). It's more recently arrived and growing cousin, DVD, yields up to 19 Gb of storage on a double-sided, double-layered disk, holding about 8 hours of MPEG-2 broad video[22].

Speed, quality, portability, and reliability in information delivery are the current advantages that CDs offer over the Web-based counterpart. The minimal speed of a CD-ROM drive recommended for purchase is X24[23], which gives a theoretical data transfer rate of about 3,600 Kps, depending on a variety of factors such as the computers CPU speed and bus width. With such a speed, almost any multimedia task can be performed with satisfactory quality. CDs are highly portable and can be used anywhere with a computer that is equipped with a CD-ROM or DVD drive. With a storage capability up to several gigabytes, a self-contained DVD can even support
the complete data of a Web site. When searching a CD on a stand-alone computer, a user is free from many anxieties that often result from Web searching, ranging from connection speed to downtime. The system-dependence is reduced from numerous junctions in hardware and software to keep connected with a remote host to a machine with a decent CD drive.

CD-based information delivery is also economical and durable. Costs to deliver 1 Mb of information are $17 online, $3.50 print, and $0.0024 CD-ROM[24]. The cost to duplicate a CD is about $0.7 to $1[25]. With CD-R (recordable) drives becoming popular and affordable, copying a CD costs about as much as a blank CD-R disk, and duplicating a CD is as easy as copying a floppy disk. CDs are durable, resistant to dropping and minor scratches. The digitized materials stored on a CD are durable for up to 15 years, a period long enough for many archival purposes[26]. This period can be doubled by making a new copy from the original so that the life of digitized materials can be prolonged indefinitely without deterioration in quality.

From the perspectives of the PC and CD industries, there is no convincing evidence that CDs are close to their demise. On the contrary, data show that new CD titles and CD sales are increasing. In 1996, the number of disc titles available commercially increased by 45%. It is believed that the growth is due to the falling prices of CD-ROMs and growth of general interest and recreation titles[27]. By the end of 1997, 46% of all commercial CD-ROM titles were under $50; 69% were under $100; 17% were under $25. Only 13% were priced over $500[28]. In 1998, the sales of CD-ROM software went up 150% over 1997. Figure 3 shows the buyers’ profiles and their purchasing behavior[29].

Interestingly, many CD-ROMs are sold on the Web through e-commerce, which makes inquires and purchases more convenient. According to statistics reported by the online CD-ROM vendor The CD-ROM Shop, the company is currently receiving in excess of 3,785,000 hits per months. The number represents the people who are interested in buying CD-ROM software[30]. This volume of traffic shows that CD-ROMs con-

<table>
<thead>
<tr>
<th>Age</th>
<th>30-55.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>66% Male, 34% Female.</td>
</tr>
<tr>
<td>Education</td>
<td>Some College.</td>
</tr>
<tr>
<td>Annual Incomes</td>
<td>&gt;$45,000 per person, &gt;$50,000 per household.</td>
</tr>
<tr>
<td>Buyers' Title Preferences</td>
<td>Shareware Collection (67%), Games (50%), Reference Guide (40%), Business Application (30%), Education titles (25%).</td>
</tr>
<tr>
<td># of Titles Purchased last 12 months</td>
<td>&gt;10&lt;200 (33%), &gt;5&lt;10 (33%), &lt;5 (34%).</td>
</tr>
<tr>
<td>Average Sale and Methods of Payment</td>
<td>$50, with credit cards.</td>
</tr>
</tbody>
</table>

Figure 5. CD-ROMs Buyers Profiles and Their Purchasing Behavior
to enjoy a vast market.

Another important factor that helps sustain and expand the demand for CDs must be attributed to the increased sale of PCs, now almost all of which are equipped with CD drives, be it for CD-ROM or DVD. In Mid-1997, it was estimated that there were 117 million computers equipped with CD-ROM drives in the United States[31]. In addition, the worldwide sale of PCs in the first quarter of 1999 increased 19%[32]. By the year 2000, it has been estimated that five hundred million PCs will be sold, 99% of these PCs will have CD drives[33]. It is difficult to image that these half billion CD drives can stay idle. Therefore, increasing demands for CDs must logically follow the vast availability of CD drives.

The unique functions of CDs and the growing number of CD-ready PCs have hence determined CDs’ continuing existence and currently irreplaceable role in information services. In fact, a trend is in progress to develop a CD/Web hybrid to optimally utilize the best of each world. This trend illustrates functional complements, not compromises.

4. CD/Web Hybrids: Simple Concepts, Effective Tools

The concept of a CD/Web hybrid is simple and straightforward: seamlessly integrate the speed, portability, and the extensive static storage capabilities of CDs with the real-time delivery, dynamic updating capability, and global accessibility of the Web. Typically, high-bandwidth video or multimedia files, and files that do not have to be frequently changed, can be stored on and served with CDs, and time-bound files can be periodically presented on the Web, allowing changes of information on CDs instantly via the Web. When a user works with CDs from his/her Web-connected workstation, the software automatically sends the requests to the CDs and the Web pages to retrieve all of the available information from the two sources, so that the findings will always be current[34].

Technically, data stored on the CDs and the Web of a CD/Web hybrid usually are published in Web standard formats, such as HTML, SGML, PDF, and XML, so that a common browser can be used to view data easily from both sources. Many CD/Web hybrid titles access Common Gateway Interface (CGI) programs on a Web server with multi-functions that generally allow the Web to operate dynamically in real time. Sometimes, a proxy Web server is embedded on a CD that can provide a variety of CGI interactions without an online connection. In a CD/Web hybrid situation, a local proxy server functions as a server to the client until an actual Web server is required for a connection outside the CD. In this case, the CD functions as a very thin server. It is believed that such CD-based portable client/server environments might have been essential in the mass popularization of the Internet public actions [35]. Microsoft FrontPage provides these client/proxy server environments for Web page developers.

In terms of the content of a CD/Web hybrid, it is a combination of features jointly offered by the CD and its corresponding Web part. Because the CD can be periodically updated by the new Web content, the total information on the nth edition of the CD, plus that on the live Web, forms an updated CD of (n+1)th edition. This relationship can be expressed as follows:
Total Information = CDn + Web
\[\rightarrow\] CD(n+1)

Where \(n\geq 1\).

The frequency of the \(n\) increment describes how often the CD part is updated.

Some special applications for developing CD/Web hybrids are now available. They integrate Web and CD-ROM publishing into a single production process, making it relatively straightforward to create and maintain contents on both CD-ROM and the Web. Contents can be published only once on either CD or the Web and have them exported to the other part[36].

In October 1998, Interleaf, Inc. (www.interleaf.com) introduced two CD/Web publishing applications: BladeRunner XML Publishers for publishing XML documents to CD-ROM with automated incremental Web updates, and Panorama CDWeb Publisher, a functionally equivalent product for SGML documents. Both are cost-effective and efficient for delivering large amounts of data to users who are challenged to access this information over the Web due to the network speed and availability. The information on the CD can be kept current by updating the CD’s content over the Web[37].

Similarly, CD-Web Publisher, an efficient, cross platform tool developed by Verity, Inc. (www.verity.com) allows users to create a Web-connected CD-ROM to distribute information such as Web site and technical documentation to an audience that may not always have Web access. Free from Web-traffic stresses, users can use a standard Web interface to search a local CD-ROM. If connected to the Web, the user can retrieve the latest updates for the Web seamlessly [38].

Data show that the CD/Web hybrid titles will grow. For example, in 1996, there were 756 hybrid titles, and an estimated 3,500 titles in 1997, a growth of 463%. Continued unsolved net congestion problems for multimedia file delivery, and the popularity of PCs equipped with CD drives support the further development in CD/Web hybrid publishing. In addition, the gradual popularization of CD-Rs is expected to stimulate CD/Web hybrid publishing.

5. Conclusion

When CD-ROM technology appeared the late 1980’s, its wide applications were immediately adopted in many areas, from audio entertainment to library and information services. Numerous database providers found them to be an excellent medium for their data. Within six or seven years, another emerging technology, the Internet’s Web, appeared and began to displace CD-ROM dominance in many application domains. The effectiveness of the Web in text-based database delivery continued to grow and dominate the market. This growth continues today and we see the increasing number of commercial databases that have migrated from CD-ROMs to embrace the Web and its technologies.

The growth and increasing popularity of the Web poses a question of whether or not Web technology and Web-based delivery doom CD as a medium. Sales data and the continued use of CD in many applications suggest that CD-ROM technology is alive and doing well. CD-ROM
applications that have been migrated to the Web represent a re-division of application responsibilities. Until the Web can assume all of the functions and unique features that CDs possess, the latter will not only flourish but will grow. The emergence and wide acceptance of DVD and the further development of higher speed CD-ROM and DVD drives for PCs evidence that CDs are on a fast-moving track to exploring more application values suitable for its growing functions. As CD and the Web technologies continue to develop, these two mediums represent the multifaceted information services vehicle. Each complements the shortcomings of the other. Owing to its many unique functions and features, CDs are expected to exist and be the choice in many digitized applications. In short, therefore, functionality of a technology determines its future.

Acknowledgement

The author wishes to thank Dr. Arthur Hafner for his valuable comments on the draft of this paper.

Notes


[18] A quick exposure to notions of IP and domain names can be found at Chris De Young’s TCP/IP, domain name, IP address, nameserver ... what?! some info and tips to help your computer communicate,” available at http://w3.arizona.edu/~cccinfo/newsletters/mayjun97/TCPI P.HTM, access 6/5/1999.


