

4

Digital Resources: Content and Language

Computers and the Internet are not much use without content and applications that serve people's needs. With the surge of material published in recent years on the World Wide Web—and millions of more Web pages added every month—it might seem that any shortage of online information and content has been long overcome. And from the point of view of a middle-class English-speaking American, that may well be the case. However, for those who live in different sociocultural environments or speak different languages, the situation is often very different. As this chapter argues, the massive amount of digital content being created on the Internet does not necessarily meet the needs of diverse communities around the world, and this has important consequences for issues of social inclusion. Fortunately, some excellent models exist of ways to develop relevant local content through active participation of diverse groups, and these models will be highlighted.

Global Web Content Production

It is impossible to determine the exact number of Web pages in the world, though current estimates top one billion.¹ It is somewhat easier, though still not exact, to count the number of Internet host domain names, such as www.harvard.edu, www.nytimes.com, or www.greenpeace.org, each of which, of course, can contain any number of pages within them. The number of Internet host domains rose nearly a hundredfold from 1993 to 2001 and now tops one hundred million.²

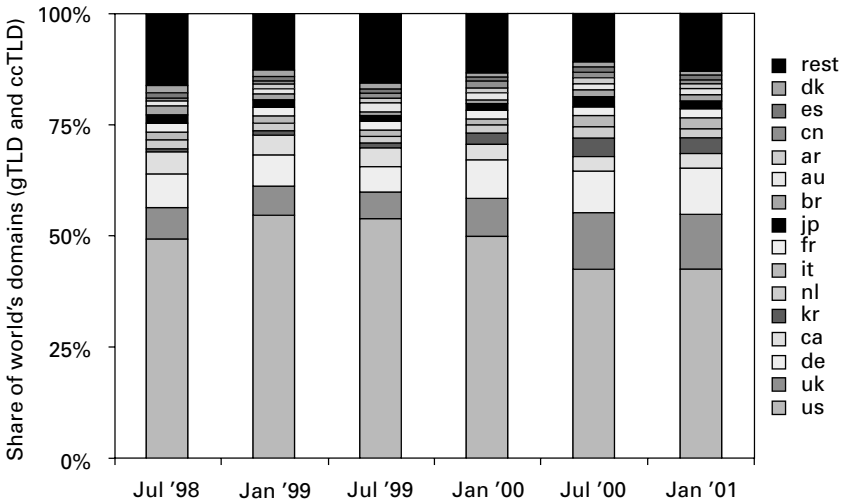


Figure 4.1

Share of world's Internet domains, by country, 1998–2001.

Source: Zook (2001a). Copyright 2001 by Matthew Zook. Used with permission.

Matthew Zook of the University of California, Berkeley, has conducted extensive analyses of the concentration of domain names by city and country around the world (e.g., Zook 2001c). According to his latest figures, some 65% of the domains in the world are located in the United States, the United Kingdom, and Germany—a figure that remained fairly steady from 1998 to 2001 (figure 4.1). This signals a great disparity in terms of number of domain names per person, even among the wealthy countries themselves (figure 4.2), let alone between the developed and developing countries. Within individual countries, the most domain names are located on servers in major cities. Thus Internet content is overwhelmingly concentrated in the major cities of the United States and Europe, with only a few other key Internet server sites located in East Asia, the Middle East, and Latin America, as illustrated by Zook's map of global domain names (figure 4.3). There is also great disparity in regard to representation of languages online; this is a pressing issue and is the focus of the second half of this chapter.

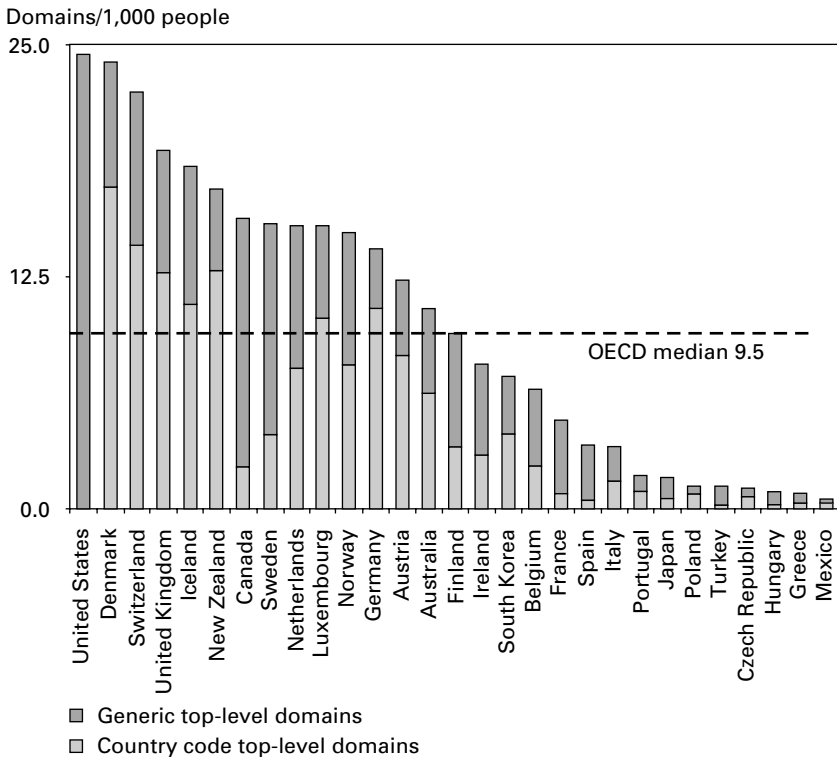


Figure 4.2

Internet domains per 1,000 people in OECD countries, January 2000.

Source: Zook (2001b). Used with permission from Matthew Zook.

Content and International Development

The geographic imbalance of Internet content production suggests that the content needs of diverse communities are not being met. For example, small-scale farmers and agricultural laborers in rural areas of Africa, Latin America, and Asia have little use for the types of material currently available on the Internet in their languages, and these rural areas of developing countries are almost completely uninvolved in production of Internet content. As a representative of M. S. Swaminathan Research Foundation told me, “The villagers in Kannivadi [in southern India] are not interested in what’s going on in the White House or even in Chennai

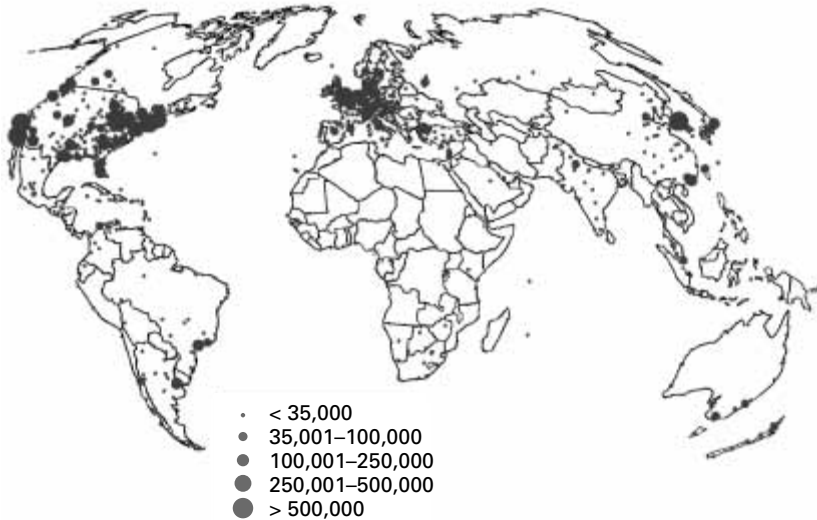


Figure 4.3

Total number of top-level domains (.com, .net, .org, country codes), by city, July 2000.

Source: Map based on methodology described in Zook (2001c). Used with permission from Matthew Zook. See also Zook’s Internet Geography Project, <http://www.zooknic.com/>.

[the state capital]; they are interested in the price of rice in the local market.” Governments, nongovernmental organizations, and community groups seeking to use the Internet for social development thus have to pay serious attention to the question of creating new digital content.

Some of the content and application areas that have been targeted as important for international development include the following.

Economic Development Information

In countries such as India, the population is made up mostly of small farmers raising a few crops each year on tiny plots of land. These farmers can benefit greatly from greater access to key information. For example, small-scale farmers often suffer financially from not knowing current crop prices in various markets in nearby cities. In response to this dilemma, several Internet projects for rural development in India collect crop prices and post this information as part of their intranets. For

example, in the Gyandoot Internet kiosks in the Dhar district of India, a farmer can come to a local kiosk and pay the equivalent of \$0.10 USD to receive the market prices on that day for a particular crop at several local, regional, and national sales points. The farmer can then make a better-informed decision about whether to harvest the crop soon or let it continue growing and, when harvested, where to sell it for the best price.

Other types of information of value to small-scale farmers include data on soil testing, crop management, crop rotation, local crop varieties, and composting. Local projects in India, such as the M. S. Swaminathan Foundation's village knowledge centers, have taken the initiative to gather this information, rewrite it in local languages, and make it available to small farmers through local and regional intranets. Farmers can drop into the village knowledge centers that are located in fourteen villages of Pondicherry and Tamil Nadu and request free information from the kiosk operators.

Health Care

Some of the most promising information and communication technology (ICT) applications for telecommunications development are in the area of health care. The village knowledge centers are also used in India to deliver health-related information to rural areas. This information includes topics such as prenatal care, postnatal care, child immunization, tropical diseases, and local and regional health care resources. Rural areas in India suffer from a lack of trained medical personnel. For this reason, community development groups are working to develop software applications that could assist health workers with early disease diagnosis and prevention efforts. For example, the George Foundation in Bangalore is developing a software application known as the Early Detections and Prevention System 2000 (EDPS2000). This application is designed to enable the early detection of diseases and nutritional deficiencies among the rural population. It is intended for use in primary health centers, which often are unable to deal with early detection of diseases because of a lack of multidisciplinary medical expertise and laboratory facilities. EDPS2000 consists of a database of disease characteristics and conditions and a symptom diagnosis program. The

program prompts the user for a step-by-step description of symptoms, followed by subsequent diagnostic questions. The software is designed to identify whether blood, urine, stool, or other laboratory tests are required in response to the answers keyed in by a health care worker. The system also indicates whether further investigation by a physician is warranted. In addition, the software maintains an exhaustive database of patient medical history and treatment visits to the health clinic and allows gathering of statistics about diseases, deaths, vaccinations and inoculations, pregnancies, and contagious diseases, thus enabling improved rural health care management.

Internet-based applications services also serve rural health care needs. One of the largest and most important is HealthNet,³ which is used by approximately 19,500 health care workers in more than 150 countries worldwide (Accenture et al. 2001). HealthNet offers e-mail connections through low-orbit satellites to medical personnel in various locations throughout Africa. Beyond that, it has created online content and applications that are of use to medical practitioners throughout the world. These include two weekly online newsletters that focus on health issues in developing countries (one newsletter deals with general health issues and the other reports on AIDS issues), links to disease-specific online information, discussion forums on topics related to medical and pharmaceutical issues in the developing world, and a GetWeb application that allows people to download this Internet-based information using basic e-mail functions. HealthNet has had its greatest impact in sub-Saharan Africa, where it has local affiliates in Eritria, Ethiopia, Ghana, Kenya, Sudan, Uganda, and Zimbabwe. HealthNet is used for long-distance consultation among doctors in different African countries, for the scheduling of medical appointments in rural areas, and for gathering medical data for clinical trials.

Education

Most people in rural areas need to travel outside their communities to further their education. Information about places to study and entrance requirements is often difficult to access in remote areas. The Internet offers a vehicle for gathering and making available information about schools, courses, fees, schedules, and sample examination questions.

One of the most popular uses of rural Internet telecenters in India has been to get information on examination results. This information is often lost or delayed in the mail, making it difficult for rural youth to make plans for following up on their education. The information is available on the Internet and can be downloaded free or for a small fee at village Internet centers.

In the long run, rural development organizations need to develop computer-based and online content and applications not just *about* educational opportunities but to be directly used in teaching and learning. The possibilities and limitations of online education are discussed in more detail in chapter 5.

Community Affairs and Culture

Low-income urban and rural groups in developing countries often lack resources to express and share their own community's culture. Since it is less expensive to produce on the Internet than via print, television, or radio, online publications can provide an excellent medium for sharing locally developed community content and can often contribute to minority language and culture preservation.

For example, São Paulo, Brazil, is one of the most socially, economically, and geographically divided cities in the world. The wealthy residents of the city, including many of Latin America's leading bankers, financiers, and media moguls, live a world apart from the working class residents of the city's infamous *favelas* (shantytowns). Newspaper reporters and television journalists also keep a distance from the *favelas* except to report on drug wars or murders. Slum dwellers thus lack cultural and news outlets that report on events directly concerning their own lives. A community coalition called Sampa.org has stepped into this gap and established a community news service that gathers and publishes online information about local affairs, community services, neighborhood news, and cultural events. Sampa.org has also established an MP3 (music file) server, so that local hip-hop bands can share their music with each other and with listeners. All of this, of course, involves the active participation of the community itself.

Online Content in the United States

Even the United States, which leads the world in Web site production, suffers from significant content gaps that affect underserved communities. An in-depth study of Internet content and diversity in the United States was carried out by the Children's Partnership (Lazarus and Mora 2000). The study combined discussions with user groups, interviews with community center directors, and the analysis of 1,000 Web sites linked to commercial and noncommercial portals to evaluate the extent to which currently available content meets the needs of diverse U.S. communities. They identified four main content-related barriers that affected large numbers of Americans.

Perhaps the greatest barrier was a lack of locally relevant information. According to the study, low-income users seek practical, relevant information that affects their daily lives, such as the following:

- *Education.* Adult high school degree programs, adult literacy programs, financial aid, homework assistance, telementoring
- *Family.* Low-cost child care, low-cost enrichment activities for children, public programs for families
- *Finances.* Public benefits news, consumer information, credit information
- *Government and advocacy.* Immigration assistance, legal services, tax filing support
- *Health.* Easy-to-understand health encyclopedias, local clinics, low-cost insurance resources
- *Housing.* Low-cost housing, low-cost utilities, neighborhood crime rates
- *Personal enrichment.* Foreign language newspapers and search engines, communities of interest for youth and adults
- *Vocational.* Low-cost career counseling programs, job training programs, job readiness programs; job listings

In some of these cases, the information may be located in print documents, but these documents are difficult to locate and obtain. In many other cases, general or partial information may exist online, but not

information that is particularly suited to low-income communities. For example, online job services generally target the higher-end market rather than entry-level jobs. Similarly, most online housing services focus on the higher end of the rental market rather than on low-rent apartments.

A second need was for information at a basic literacy level. For example, there are a large number of tutorials online that cover different computer and Internet skills, such as the use of spreadsheets, Web page design, or photo editing, but these generally demand a high level of literacy. Materials tailored to limited-literacy populations are badly needed by community technology centers, which often present computer instruction for those with limited English-language or literacy skills.

A third need is for content for non-English speakers. While there is a large amount of information on the Web in languages such as Spanish, there is little public information in Spanish directed towards U.S. audiences. Users seek information related to governmental programs that affect them, for example, Medicare, taxes, voting.

Finally, more diverse cultural resources are desired. Although some local U.S. communities are starting to build a cultural presence online (e.g., HarlemLive), users still feel that far more needs to be done to develop Web sites that reflect diverse cultural heritages and practices (Lazarus and Mora 2000).

The Children's Partnership has recently developed a major new portal in an attempt to address the lack of online information for low-income Americans. The portal includes sections for planning personnel, such as administrators of community technology centers, and for the lay public on topics like health, housing, employment, education, and culture.⁴

Content for the Disabled

Both developed and developing countries require content that addresses the special needs of the disabled in format and subject matter. As to formatting, the best and most up-to-date source of information is the Web Accessibility Initiative, supported by governmental and non-governmental bodies in the United States, Canada, and Europe. These Web sites describe how online content should be developed so that it can be accessed by the disabled.⁵ A principal requirement is to provide a

redundancy of output mechanisms, that is to ensure that all graphical content has a text equivalent (for the blind, who can then convert the text to speech); that all audio content has a text equivalent (for the deaf); and that animated graphics can be frozen (for those with attention deficit disorder or learning disabilities). It is also recommended that sites allow users to input via both keyboarding or pointing (e.g., a “submit” button can be designed to also accept the input of the letter *s*) and that sites use a clear, consistent, well-labeled format (to benefit all users, and especially those with disabilities). Accessibility criteria exist on a continuum, so it is difficult to determine the exact percentage of existing Web sites that are or are not accessible, but by any measure there is still a long way to go. For example, one report found that the Web sites of all nine U.S. presidential candidates in 2000 failed to fully meet even the first level of accessibility requirements for the disabled (Báthory-Kitz 1999).

Beyond formatting, it is also important to develop online content for the disabled. Some of the best work in this regard has been carried out in Europe, where several countries have Internet portals for the disabled, with information on rehabilitation programs, assistive technology, special education, workplace adaptations, legislation, and training (European Commission 2001a; 2001b).

Community Mobilization and Content Development

The successful development of online content demands the active participation of the communities that will make use of the materials. There are three principal ways that community participation is achieved: through needs assessment, database development, and content production.

Needs Assessment

The approach of Participatory Rural Appraisal (PRA) provides a model of how a community can be engaged in helping define and determine its own needs (Mukherjee 1993). PRA has been used in development projects throughout the world over the last two decades, evolving from a prior, similar approach known as Rapid Rural Appraisal (RRA) (Chambers 1992). PRA uses focus groups, interviews, door-to-door surveys, community meetings, and special participatory exercises to max-

imize a community's involvement in defining its own needs. Rural Internet projects in India, such as Gyandoot and the M. S. Swaminathan village knowledge centers, were launched through intensive PRA in local villages. This appraisal of needs helped determine what resources villagers already had access to and what resources they needed.

Database Development

One important area of online content is listings, maps, and databases of local community resources. The community itself ought to be centrally involved in gathering and mapping those data. For example, the Camfield Estates project in Massachusetts, which placed computers and Internet access in many apartments in a low-income housing area, brought together a team of community residents to survey residents' existing skills, capacities, and interests and to identify other local assets, such as businesses, churches, and child care facilities. This information became a key part of the Web portal that served the Camfield Estates community.⁶

Community members can also contribute to databases through online communication. The city of Muenster, Germany, for example, has published a database and interactive street map for mobility-impaired people, with detailed and easily accessible information on public institutions, recreation facilities, social services, doctors, and information bureaus with barrier-free or disability-compatible buildings and services (Neumann and Uhlenkücken 2001). One interactive feature being developed will allow community members to contribute directly to the database so that they themselves can point out urban locations that need to be altered or physical barriers that need to be removed. This system of community contribution is designed to make the database more comprehensive and informative while actively involving the disabled as consultants and partners in the project.

Content Production

The third major area of community involvement is through specific content development. Teams of community residents can be trained to develop Web-based information about their community that focuses on news, current events, culture, or any other items of interest or concern. One example is São Paulo's Sampa.org project, discussed earlier, which

involves teams of community residents in producing a local online news service. Another excellent example is HarlemLive in New York.⁷ HarlemLive is an Internet-based youth publication launched in 1996. It has a close relationship with the Playing2Win community technology center in Harlem, which hosts the publication on its Web site and provides office and production space for the publication's editorial team. Columbia University and a number of other local organizations provide additional support.

HarlemLive is a high-quality online publication, with general news reports, articles on community issues, arts and culture articles, photo galleries, a creative writing section, and a special women's section. The publication thus provides current, topical information by and for the Harlem community. Equally important, HarlemLive has trained several hundred Harlem young people as journalists, photographers, media administrators, Webmasters, and public speakers. The publication thus serves as a focal point for young people to develop and showcase their technical and communication skills while they address issues of concern to the community and create original content that helps give the community voice.

In summary, there are many types of online content of use to marginalized communities. Some of this content can be provided by outside agencies. But, for the most part, active involvement of the targeted populations—in defining their needs, collecting data, and authoring and publishing content—is usually required for success. This kind of approach, based on active community involvement, also helps guarantee the kinds of community training and mobilization necessary for long-term success.

Language

Language is one of the most complex and significant issues related to content and to broader issues of ICT and social inclusion. Language intersects with many other forms of social division related to nationality, economics, culture, education, and literacy. Language questions dramatically affect how diverse groups can access and publish information on the Web as well as the extent to which the Internet serves as a medium for expression of their cultural identities.

Language and Identity in the Age of Information

The critical role of language is situated in the broader social and economic transformation of recent decades. The information revolution, accompanied by the processes of international economic and media integration, has acted like a battering ram against traditional cornerstones of social authority and meaning. Throughout the world, shifts in economic and political power have weakened the role of the state, new forms of industrial organization have decreased the possibilities for long-term stable employment, and women's entry into the work force has shaken up the traditional patriarchal family (Castells 1997).

But every action brings a reaction. The last quarter-century has also witnessed a worldwide surge of movements of "collective identity" that "challenge globalization and cosmopolitanism" on behalf of people's control over their culture and their lives (Castells 1997, 2). These differ from earlier social movements, which in many parts of the world were based on struggles of organized workers. As Alain Touraine explains, "In a postindustrial society, in which cultural services have replaced material goods at the core of production, *it is the defense of the subject, in its personality and its culture, against the logic of apparatuses and markets, that replaces the idea of class struggle*" (quoted in Castells 2000b, 23, emphasis in original). Castells (1997) further explains the central role of identity in today's world:

In a world of global flows of wealth, power, and images, the search for identity, collective or individual, ascribed or constructed, becomes the fundamental source of social meaning. This is not a new trend, since identity, and particularly religious and ethnic identity, have been at the roots of meaning since the dawn of human society. Yet identity is becoming the main, and sometimes the only, source of meaning in a historical period characterized by widespread destructuring of organizations, delegitimation of institutions, fading away of major social movements, and ephemeral cultural expressions. People increasingly organize their meaning not around what they do but on the basis of what they are. (3)

Within this contradictory mix of global networks and local identities, language plays a critical role. With other cornerstones of social authority, such as nation, family, and career, battered by the processes of globalization, language can become "the trench of cultural resistance, the last bastion of self-control, the refuge of identifiable meaning" (Castells 1997, 52). The struggle over bilingual education in the United States; the

Québécois, Basque, and Kosovar separatist movements; the battles over language and citizenship in post-Soviet countries; and language revitalization movements in Ireland (Gaelic), New Zealand (Maori), Morocco (Tamazight), and many other countries indicate the powerful role of language-based identity in today's world.

It is not surprising that language and dialect have assumed such a critical role in identity formation. The process of becoming a member of a community has always been realized in large measure by acquiring knowledge of the functions, social distribution, and interpretation of language (Ochs and Shieffelin 1984). In most of the world, the ability to speak two or more languages or dialects is a given, and language choice by minority groups becomes a symbol of ethnic relations as well as a means of communication (Heller 1982). In the current era, language signifies historical and social boundaries that are less arbitrary than territory and more discriminating (but less exclusive) than race or ethnicity. Language-as-identity also intersects well with the nature of subjectivity in today's world. Identity in the postmodern era has been found to be multiple, dynamic, and conflictual, based not on a permanent sense of self but rather on the choices that individuals make in different circumstances over time (Henriquez et al. 1984; Schechter, Sharken-Taboada, and Bayley 1996; Weedon 1987). Language, though deeply rooted in personal and social history, allows a greater flexibility than race and ethnicity, with people able to consciously or unconsciously express dual identities by the linguistic choices they make even in a single sentence (e.g., through switching or combining languages; see Blom and Gumperz 1972). By means of choices concerning language and dialect, people constantly make and remake who they are. For example, a Yugoslav becomes a Croatian, a Soviet becomes a Lithuanian, and a Canadian becomes a Québécois.

Yet, at the very time that linguistic diversity is becoming more critical than ever in people's lives and identities, a new communication medium has emerged that has been dominated by a single language: English. The dominance of English, not just on the Internet but also in many other international media and communications forums, has led to the rise of new concepts such as "global English." In order to fully appreciate the issues associated with ICT and social inclusion, it is necessary also to

understand global English and how it has come to dominate digital telecommunications.

Global English

Although there have been many important international languages over time, including Latin, French, Russian, Chinese, Arabic, and Spanish, English is generally considered to be the first global language because of its current dominant role as a *lingua franca* in international communications. The rise of global English is the flip side of movements for local identity; it represents the need for an international medium of communication for global economic, political, and social exchange. According to information gathered by Crystal (1997), 85% of international organizations make use of English as at least one of their official languages, 85% of the world's film market is in English, and 90% of the published articles in leading journals of linguistics are in English.

Nevertheless, these statistics belie the fact that English is only spoken as a native language by a relatively small minority of people in the world. According to calculations, about 350 million people around the world speak English as a native language (Crystal 1997; Graddol 1997; 1999), representing some 6% of the world's population. This places English well behind Chinese in its number of native speakers, and not that far ahead of Spanish, Hindi, and Arabic, all of which may catch up or pass English in number of native speakers in the next sixty years (Graddol 1997). Another 350 million people are estimated to speak English as a second language, in countries such as India, Nigeria, the Philippines, and Singapore. There are also an estimated 700 million people who speak English as a foreign language, albeit with varying degrees of proficiency. Putting these numbers together, we see that three-quarters of the world's population knows almost no English, and even among the one-quarter who are said to speak it, the degree of competence varies markedly.

In many countries, unequal access to learning English overlaps with other social inequalities. Even though English is almost universally taught in secondary schools and universities, the majority of people in many developing countries never attend secondary school. Even those who do often face poorly trained teachers who do not speak English well themselves. Indeed, in many countries, the only reliable

route to learning English is through expensive private education. With knowledge of English a requirement for access to many professions and university programs, English becomes one more barrier to equal opportunity for the poor. And even many people who speak English well may not be happy with the thought of its supplanting their own local language in as important a medium as the Internet.

English on the Internet

One of the first published studies of language on the Internet, and conducted in 1997, indicated that some 81% of international Web sites were in English (“Cyberspeech” 1997). At the time these results were made public, the dominance of English on the Internet caused great consternation around the world. Anatoly Voronov, director of a Russian internet service provider, voiced the sentiments of many when he said:

It is just incredible when I hear people talking about how open the Web is. It is the ultimate act of intellectual colonialism. The product comes from America so we either must adapt to English or stop using it. That is the right of any business. But if you are talking about a technology that is supposed to open the world to hundreds of millions of people you are joking. This just makes the world into new sorts of haves and have nots. (Cited in Crystal 1997, 68)

Within three years of this study, the percentage of English Web sites had fallen to 68% (Pastore 2000), still a sizable majority and well out of proportion to the number of English speakers in the world. A calculation of the ratio of Web pages to speakers of leading languages indicates that English speakers are still better represented and served on the Internet than speakers of other languages (table 4.1).

The question remains as to whether this drop from 81% to 68% represents the beginning of the end of English dominance online, or whether it marks a continuation of that dominance at an unacceptably high level. To better understand and interpret these figures, and to predict the likely trend in international communication online, it is necessary to distinguish between the short-term and long-term advantages that English has in the computing and Internet realms.

The short-term advantages were principally two: the Internet first arose in the United States and speakers of English were its designers—they thus wrote programs that relied on an English-language interface;

Table 4.1
Ratio of Speakers of a Language to Web Pages in That Language, 2001

Rank	Language	No. of Web Pages	No. of Speakers (thousands)	Speakers/ Web Page
1	English	214,250,996	322,000	1.5
2	Icelandic	136,788	250	1.8
3	Sweden	2,929,241	9,000	3.1
4	Danish	1,374,886	5,292	3.9
5	Norwegian	1,259,189	5,000	3.9
6	Finnish	1,198,956	6,000	5.0
7	German	18,069,744	98,000	5.4
8	Dutch	3,161,844	20,000	6.3
9	Estonian	173,265	1,100	6.4
10	Japanese	18,335,739	125,000	6.8
11	Italian	4,883,497	37,000	7.6
12	French	9,262,663	72,000	7.8
13	Catalan	443,301	4,353	9.8
14	Czech	991,075	12,000	12.1
15	Basque	36,321	588	16.2
16	Slovenian	134,454	2,218	16.5
17	Korean	4,046,530	75,000	18.5
18	Latvian	60,959	1,550	25.4
19	Russian	5,900,956	170,000	28.8
20	Hungarian	498,625	14,500	29.1
21	Portuguese	4,291,237	170,000	39.6
22	Greek	287,980	12,000	41.7
23	Spanish	7,573,064	332,000	43.8
24	Lithuanian	82,829	4,000	48.3
25	Polish	848,672	44,000	51.8
26	Hebrew	198,030	12,000	60.6
27	Chinese	12,113,803	885,000	73.1
28	Turkish	430,996	59,000	136.9
29	Bulgarian	51,336	9,000	175.3
30	Romanian	141,587	26,000	183.6
31	Arabic	127,565,000	202,000	1,583.5

Source: Adapted from Carvin (2001).

and the Internet, in its early iterations, functioned best in the ASCII code, which is very difficult to read and write with non-Roman alphabets.

These short-term reasons have already started to fade in significance and impact. As discussed earlier, Internet access is starting to reach saturation point in the United States but it is just taking off in many other countries around the world. As a critical mass of users gets online in a particular language, more people and businesses create Web sites in that language, and speakers of the language also have a greater number of potential partners for computer-mediated communication. This trend is also accelerated by the expansion of operating systems and Web page authoring software in non-Roman scripts, which allows people to communicate more easily in non-alphabetic languages such as Japanese, Chinese, and Hebrew. Because of these trends, the proportion of Web sites in English is expected to drop to 40% in the next decade (Graddol 1997).

However, even as English's short-term advantages decrease, it will still maintain a strong position over other languages on the Internet because of its long-term advantages. Principal among these is the historical fact that English was already the *de facto* global language at the time the Internet was created, and remains so today. The Internet, by enabling global communication, requires a global standard, and English's default advantage thus remains and is in fact strengthened. A mutually reinforcing cycle takes place, by which the existence of English as a global language motivates (or forces) people to use it on the Internet, and the expansion of the Internet (and online English communication) thus reinforces English's role as a global language. This cycle can occur even when more and more people are using the Internet in their own languages. They may use the Internet in their own language for local or regional communication, but they will continue to use the Internet in English for global communication.

This trend is illustrated in a study conducted by the Organization for Economic Cooperation and Development ("The Default Language" 1999). The OECD study found that while only 78% of the regular Internet sites surveyed were in English, some 91% of the sites on what are called secure servers were in English and 96% of the sites on secure servers in the .com domain were in English. This is significant because

secure servers, especially in the .com domain, are most frequently used for e-commerce. This means that even as people increasingly use languages other than English for local communication, they can be expected to use English for many international transactions. Of course, this latter trend may not be permanent as more companies may localize their e-commerce in the languages of the consumers.

Furthermore, long-term advantages do not necessarily mean permanent advantages. It is entirely possible that a century from now English will no longer be the dominant language on the Internet (or on whatever has replaced the Internet), either because of the weakening of English as a global language (because of demographic or economic changes) or because of the development of improved machine translation techniques (thus allowing everyone to communicate in the local language). Machine translation already exists online, but it is of such poor quality (based on word-by-word translation) that it does not now mitigate the need for a *lingua franca*, nor is it predicted to be of sufficient quality to mitigate such a need for a long time. For the foreseeable future, then, a disproportionate percentage of the world's Web sites, especially those necessary for international exchange, will be in English, and that is an important factor limiting access to Internet content.

The role of English vis-à-vis other languages online can be illustrated through analyses from Egypt (where English is spoken as a foreign language), India (where it is spoken as a second language), and Hawai'i (where it is spoken as a first language, though in a diverse multilingual setting).

Language Online in Egypt

Egypt is an excellent example of a highly stratified country in which English plays a dual role. On the one hand, English helps connect Egypt to the world by facilitating international commerce, tourism, and exchange. On the other hand, unequal access to English within the country serves to heighten the nation's already substantial social and economic disparities.

Arabic is the official language of Egypt and virtually the entire population speaks a dialect of this, referred to as Egyptian Arabic. Those that can read and write also know Classical Arabic, the main written variety

of the language. Other languages in the country include ancient Coptic (used in Coptic Christian church services), a variety of African languages spoken by refugees, and European languages used in business and tourism. The use of European languages in Egypt has a long history dating back to periods of French and British colonialism, and at the time the Egyptian elite often preferred to be educated in French or English rather than Arabic (Haeri 1997). Most recently, though, the use of English has far surpassed that of French and other foreign languages by Egypt's elite. English is used not only in communication with foreigners—for example, in international commerce—but also for internal communications in a number of privileged occupations, especially in the fields of information technology, engineering, medicine, dentistry, and sciences. It is not unusual for Egyptian professionals in these areas to hold their conferences or produce their publications in English, even if the intended audience is other Egyptians.

English is essential for participation in elite professions, yet it is spoken by only a small minority of the population, estimated at some 3% (Warschauer 2001b). English is a mandatory foreign language taught in all schools beginning in the fourth year of elementary school, but it is learned very poorly, if at all, because of huge class sizes, poorly trained teachers (many of whom themselves barely know the language), and the country's high dropout rate (with half the adult population completing less than five years of schooling) (Fergany 1998). The elite, many of whom are bilingual in English and Arabic, usually learn English in private schools (a large number of which offer English medium instruction), private tutoring, English-medium private universities in Egypt, and study abroad in England or the United States (see discussion in Schaub 2000).

Even though English is spoken by just a small percentage of the population in Egypt, it is a dominant language of the Internet in that country (Warschauer, Refaat, and Zohry 2000). Many Egyptian Web sites, including those targeted exclusively for use inside the country, are only in English (see, for example, figure 4.4, showing Otlob.com, a popular site for ordering food delivery from restaurants in Cairo and Alexandria). Some 70% of young professionals I surveyed use English



Figure 4.4

Ordering food in Egypt.

Source: <http://otlob.com>. Used with permission.

exclusively in formal e-mail communication (Warschauer, Refaat, and Zohry 2000). Arabic, when added in e-mail, is most often written in Roman characters and used principally for religious or highly emotive expressions.

The reasons for English's dominant role in Egyptian online communications are multiple. First, no single standard of Arabic-language computing has emerged yet, so Web producers are often forced to convert Arabic-language content into slow-loading images if they want to guarantee that their content can be read in Arabic. This lack of a common standard also discourages Arabic-language e-mail. In addition, the Internet first arose in Egypt in the very sectors that operate in English, such as the information technology industry and international businesses. Finally, the early adopters of the Internet in Egypt were mostly people who—owing to their schooling and work experience—write, compute, type, and keyboard better in English than they do in Arabic, and using English online thus comes naturally to them.

Some of these conditions are bound to change over time, especially with the emergence of common Arabic-language standards for computing. However, in the meantime, the 97% of the Egyptian people who do not know English are excluded from full access to Egyptian online content written in English, let alone international English-language Web sites.

Multilingual Computing in India

India is another country where English plays a stratifying, if also a unifying, role. People in India speak some 850 local and regional languages (Todd and Hancock 1987) of which 58 are taught in schools, 87 are used in newspapers, 71 in radio programs, and 15 in films (“Indian Languages” 2001). A total of 18 of these are considered official languages.

English is spoken by an elite throughout the country and is used in scientific, technological, and business communications. However, despite the national prominence of English in India and India’s reputation as an English-speaking country, only about 5% of the people speak English (Crystal 1997). In contrast, nearly half the people of the country speak Hindi and almost another quarter speak dialects of Tamil, Bengali, Kannada, or Marathi (“India” 2001).

Indian-language computing thus is complicated by many factors, including the fact that each major Indian language has its own script. Not surprisingly, English has emerged as the dominant language of the information technology industry in India, which serves the industry’s booming export business well. However, English is much less useful as a language of communication for national development purposes, especially for projects that target India’s poor. In short, the potential of ICT for aiding rural development will not be reached without adequate Indian-language software and content.

For this reason, a number of Indian organizations are trying to develop software solutions to promote Indian language computing. One of the more promising is being produced by a group called Chennai Kavigal, which is collaborating with the Institute of Indian Technology in Madras to develop low-cost Indian-language software solutions for both Windows and Linux platforms. One of their products, a complete office

suite with a word processor, spreadsheet, database, and presentation software, is being made available to development projects for only \$6 USD. Other products include e-mail, paint, browser, and programming software. The products are being developed with special attention to the needs of Indian users. For example, all products have complete Indian-languages interfaces, including navigation menus. The e-mail software comes with separate password-protected folders, ideal for a situation in which many users share one machine. A compiler of the C and C++ computer languages enables programming to be done in English (essential for getting a job) but allows comments and error messages to be written in Indian languages, thus providing important support to limited-English speakers who are in the process of becoming software programmers. Finally, to encourage international communication between India's different regions, the word processing and communication software products are programmed to automatically convert from one language script to another. This is especially helpful for the many millions of people in India who can read and write one Indian language but speak others. For example, a writer of Tamil (but speaker of Hindi) can write a Hindi-language message in the Tamil script and have it automatically converted to Hindi script to be read by someone in a Hindi-speaking region of the country. These conversions can even be performed instantaneously using synchronous communication software so that one's own script appears on one's own screen while the other script appears instantly on the correspondent's screen.

These software solutions are helping rural development projects in India to develop Internet content in local languages. Much useful content in these projects is not necessarily original material but rather public information already available elsewhere on the Web that is carefully selected, translated into Indian languages, and presented in a simplified form in easy-to-navigate pages.

Language and Identity in Hawai'i

Multilingual software solutions are only the first step. To help promote technology for social inclusion, it is also necessary to use the software to develop and make available relevant content. This can be done by translating material and, more important, by developing original

material in the minority language itself. One of the more interesting examples of minority language content development is from Hawai'i.

The situation in Hawai'i differs markedly from those in India and Egypt. In Hawai'i, almost everyone except some recent immigrants speaks English. Why then is Hawaiian-language computing and Internet content even necessary? To understand this, it is necessary to examine briefly from a sociopolitical angle the history of Hawaiian language use.

Hawaiian was the national language of the sovereign Hawaiian monarchy of the nineteenth century. At the end of that century, however, wealthy American landowners backed by the U.S. government overthrew the Hawaiian kingdom and forcefully incorporated Hawai'i as a U.S. territory. Laws forbidding the use of the Hawaiian language were passed and vigorously enforced through beating children who dared speak their native tongue in school (Wilson 1998). By the time Hawai'i became a U.S. state in 1959, Hawaiian was spoken by only a few thousand elders and the language was seriously endangered. The suppression of the Hawaiian language went hand-in-hand with the subjugation of the native Hawaiian people, whose numbers gradually diminished and who found themselves at the bottom of all social and economic indicators with little room to move in terms of rescuing their dying language.

Nonetheless, a strong Hawaiian resistance movement emerged in the 1970s to fight for Native Hawaiian rights, including revitalization of the Hawaiian language. As a result of this effort, the Hawaiian language was legalized and a number of Hawaiian immersion schools were established by the state government. New undergraduate and graduate programs in Hawaiian studies and Hawaiian language were launched in the state's public universities. Ever since, defense of the Hawaiian language has continued to be a central element for defense of the Hawaiian people. This is especially so because of the complex nature of Hawaiian identity. There are almost no "pure" Hawaiians left, although some 20% of the people in the islands have part-Hawaiian ancestry. In this context, many Hawaiians believe that revitalization of their language is critical to their survival as a people.

The Hawaiian language is no longer facing extermination; nevertheless, it still faces an uphill battle to be reestablished as a stable living lan-

guage used regularly in daily life. There are no daily Hawaiian-language newspapers or full-time Hawaiian-language television stations. The handful of elders who learned to speak Hawaiian as a first language are dying. The Hawaiian immersion schools are trying to teach Hawaiian but are confronted by a lack of curricular content and a dispersion of its speakers in small pockets spread out over several islands.

It was within this context that Hawaiian educators from the University of Hawai'i launched Leokī (powerful voice), believed to be the first online bulletin board system that functioned completely in an indigenous language (Warschauer and Donaghy 1997). The provision of Hawaiian-language content and communication media was intended to boost the Hawaiian revitalization effort, and particularly the Hawaiian-language programs in the immersion schools. As Keiki Kawai'ae'a, the director of Hawaiian-language curriculum materials, explained,

Without changing the language and having the programs in Hawaiian, they wouldn't be able to have computer education *through* Hawaiian, which is really a major hook for kids in our program. They get the traditional content like science and math, and now they are able to utilize this 'ono (really delicious) media called computers! Computer education is just so exciting for our children. In order for Hawaiian to feel like a real living language, like English, it needs to be seen, heard and utilized everywhere, and that includes the use of computers. (Quoted in Warschauer 1998, 147–148)

All the interfaces, menus, and content for Leokī are written in Hawaiian. The board also contains an extensive array of features (figure 4.5):

- *Leka Uila* (electronic mail). Each user has a private mailbox for sending and receiving mail to and from other users on Leokī as well as via the Internet.
- *Laina Kolekole* (chat line). An online chat area for real-time interaction. Users can also create their own private chat rooms.
- *Ha'ina Uluwale* (open forum). Public synchronous computer conferences for discussion, debate, and surveys.
- *Ku'i ka Lono* (newsline). Advertisements, announcements, and information about Hawaiian-language classes and important upcoming events.
- *Hale Kū'ai* (marketplace). Announcements and online order forms for the purchase of Hawaiian-language books and materials.

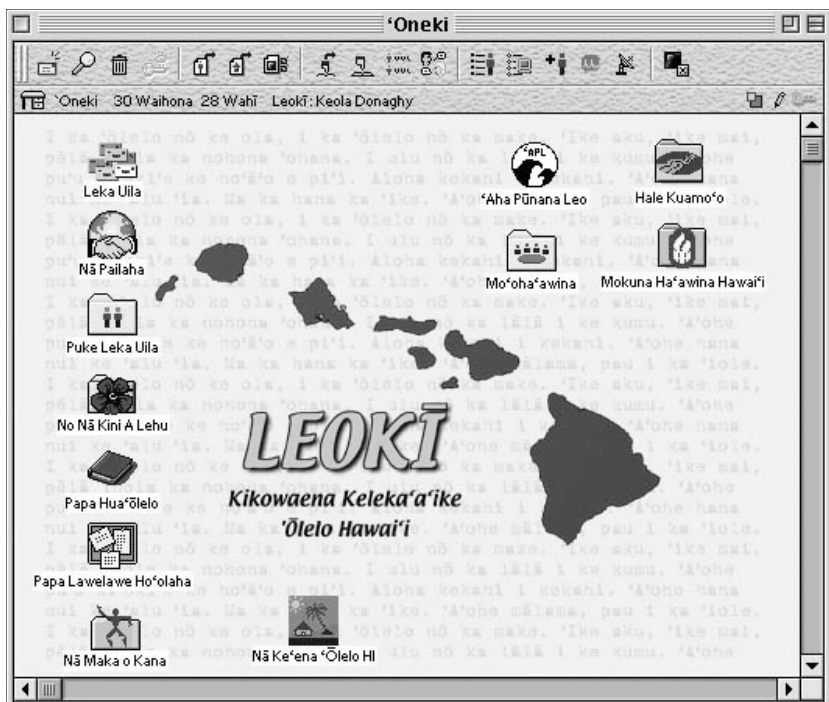


Figure 4.5

Hawaiian language bulletin board system.

Source: Kualono (2001). Used with permission from Kualono.

- *Papa Hua'ōlelo* (vocabulary list). Dissemination of Hawaiian words being coined by the Hawaiian-language lexical committee. Users can suggest new words and offer input on terms being considered or search vocabulary databases.
- *Nā Maka o Kana* (The Eyes of Kana). The current and all back issues of the *Nā Maka o Kana* newspaper, published by and for the Hawaiian immersion program.
- *Noi'i Nowelo* (Search for Knowledge). Shared resource area for old and new Hawaiian-language materials. Posting of stories, articles, and songs.
- *Nā Ke'ena 'Ōlelo Hawai'i* (Hawaiian language offices). An information section about the various agencies that provide educational support

for Hawaiian studies coursework and Hawaiian-medium programs throughout the state.

Leokī has been installed on the computers in the Hawaiian immersion schools and preschools, Hawaiian-language departments of colleges and universities, and in Hawaiian-language support organizations. Most recently, a public version has been made available to any Hawaiian-language speaker.⁸

In addition, Hawaiian educators have actively mobilized the Hawaiian-language community to create content for Leokī and for the broader Internet. In 1997 I conducted an ethnographic study of a Hawaiian-language class at the University of Hawai'i that worked to develop a Web site in the Hawaiian language. The students in the course chose topics related to Hawaiian history and culture, including Hawaiian chanting and music, Hawaiian leaders, and the geography and nature of Hawai'i. The Web pages that they developed were highly sophisticated in language, content, and design, and the experience proved highly motivating for the students. As one of the students explained,

It's like a double advantage for us, we're learning how to use new tools, like new technology and new tools, at the same time we're doing it in Hawaiian language, and so we get to learn two things at once. . . . It looks almost as if it's a thing of the future for Hawaiian, because if you think about it, maybe there's [only] a few Hawaiian-language papers. But instead of maybe having a Hawaiian-language newspaper, you have something that might be just a little bit better, like the World Wide Web, it's like building things for all the kids who are now in immersion and even for us, someplace to go and get information, and so that's kind of neat what we're doing, we're doing research and then finding out all that we can about a topic and then actually putting it on the World Wide Web, and then having that be useful to somebody else in the future. (Warschauer 1999)

Nancy Hornberger (1997) once said that "Language revitalization is not about bringing a language back, it's about bringing it forward." For students in Hawai'i, indigenous-language content creation is bridging a very special divide, that between a recent past in which their language and culture faced persecution and near elimination, and a future in which their words, chants, songs, and stories will thrive with the assistance of new media.

Conclusion

Physical resources such as computers and connectivity mean little without sufficient digital content that is relevant to people and in the language of their communities. As many of these examples have shown, the most important content production is often done by people in the targeted communities themselves. This, in turn, demands literacy and education. It is to those issues that I turn now.

This is a section of [doi:10.7551/mitpress/6699.001.0001](https://doi.org/10.7551/mitpress/6699.001.0001)

Technology and Social Inclusion

Rethinking the Digital Divide

By: Mark Warschauer

Citation:

Technology and Social Inclusion: Rethinking the Digital Divide

By: Mark Warschauer

DOI: 10.7551/mitpress/6699.001.0001

ISBN (electronic): 9780262257176

Publisher: The MIT Press

Published: 2004

The open access edition of this book was made possible by generous funding and support from the author



The MIT Press

First MIT Press paperback edition, 2004
© 2003 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Sabon by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Warschauer, Mark.

Technology and social inclusion : rethinking the digital divide /
Mark Warschauer.

p. cm.

Includes bibliographical references and index.

ISBN 0-262-23224-3 (hc. : alk. paper), 0-262-73173-8 (pb.)

1. Digital divide. 2. Marginality, Social. I. Title.

HN49.I56 W37 2003

303.48'33—dc21

2002075130

10 9 8 7 6 5 4 3 2