

Indian Languages on the Internet

By GIRIRAJ AGARWAL

American IT companies help Hindi and Urdu grow on the Web.

Three years ago, Ramprakash Tyagi got very upset when workers from a telecommunications company dug up the road in front of his house in Ghaziabad, Uttar Pradesh, to lay a broadband Internet cable. The 55-year-old teacher did not want any cable laid near his house as he thought he would never need this "hi-fi" technology. This was useful only for the English speaking elite, he felt. Kusum Gupta of Faridabad, Haryana, had similar thoughts when her U.S.-based son bought a computer and Internet connection for her during his visit to India.

But things have completely changed now. Tyagi and Gupta have become regular Internet users and now get upset with Internet service providers whenever they fail to connect to the World Wide Web. Thanks to the rise of Hindi, Urdu and other Indian languages on the Web, millions of non-English speaking Indians are discovering uses of the Internet in their daily lives. They are sending and receiving e-mails, searching for information, reading e-papers, blogging and launching Web sites in their own languages. Two American IT companies, Microsoft and Google, have played a big role in making this possible.

A decade ago, there were many problems involved in using Indian languages on the Internet. "There was mismatch of fonts and keyboard layouts, which made it impossible to read any Hindi document if the user did not have the same fonts," says Balendu Sharma, editor of a Hindi portal (prabhasakshi.com).

"There was chaos, more than 50 fonts and 20 keyboards were being used and if two users were following different styles, there was no way to read the other person's documents," says Sharma, one of the Indian recipients of Microsoft's Most Valuable Professional award in 2007.

But the advent of Unicode support for Hindi and Urdu changed all that. The concept of new character encoding from Unicode Consortium—a nonprofit in California whose members include Google, IBM, Oracle, Microsoft, Sun Microsystems, Yahoo and the Government of India—proved to be a boon for Indian languages.

Microsoft incorporated the Hindi Unicode font, Mangal, in its operating system in 2001. "Since then, the Hindi Unicode support has been a part of all subsequent upgradations of Microsoft's operating systems. We are also providing Input Method Editor facilities which give users the option to use different types of keyboards," says Meghashyam Karanam, product manager, vision and localization, at Microsoft India.





GURINDER OSAN © AP/WIDEWORLD



Courtesy Yahoo India



GAUTAM SINGH © AP/WIDEWORLD

Far left: Rajeev Kaul (left), then-Microsoft India managing director, and Raveesh Gupta, then-Microsoft program manager for localization, at the launch of the Hindi version of Microsoft Office in New Delhi in 2004.

Left: Gopal Krishna, head of audience, emerging markets, at Yahoo.

Left below: Google founders Sergey Brin and Larry Page interact with journalists in Bangalore during their India tour in 2004.

Hindi journalists consider it a “big gift” from Microsoft. “Unicode font was a pathbreaking initiative which gave a big push to Hindi. It provided much-needed freedom of accessibility to Hindi documents,” says Nirendra Nagar, senior editor with navbharattimes.com. The earlier system could incorporate only 127 characters, which is not enough for the Hindi Devnagari script. The Unicode system can incorporate up to 65,000 characters. As most computers in India use Microsoft’s operating system, it ensured that the Hindi font was available to most of them as they upgraded the operating software.

In 2004, the Hindi version of Microsoft Office 2003, which included Word, Excel, PowerPoint and Outlook, was launched. Now the Hindi version of Microsoft Office 2007 is also available. “It includes Hindi language interface packs that allow users to create documents and communicate with others in Hindi. Users can also navigate using the menus and toolbars that are in Hindi. We have received a very good response from the Hindi users,” says Karanam.

Urdu language support is available in Windows Vista and Office 2007. Another Microsoft initiative is Project Bhasha, which was launched in 2003 and now provides support to 13 Indian languages such as Hindi, Tamil, Kannada, Punjabi, Konkani, Oriya.

In 2006, Microsoft, headquartered in Redmond in Washington state, partnered with one of the early Hindi portals, webdunia.com, to launch its MSN Hindi portal. “Webdunia also provided support for the Hindi version of Microsoft Office as well as for language

interface packs,” says Jaideep Karnik, general manager for content and localization at webdunia.com. The Indore, Madhya Pradesh-based company has an office in the United States and helps major software developers localize their products.

If Microsoft built the base for Hindi, Google was ready to put up the superstructure. Realizing the potential of Indian languages, the California-based company has launched various products in the past two years. With the Google Hindi and Urdu search engines, one can search all the Hindi and Urdu Web pages available on the Internet, including those that are not in Unicode font.

“Google offers searching in 13 languages, Hindi, Tamil, Kannada, Malayalam and Telugu to name a few, Gmail in five languages and Google transliteration in Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Punjabi, Tamil, Telugu and Urdu. Urdu is the most recent language that Google has added to its offerings,” says Rahul Roy-Chowdhury, product manager at Google India.

To use the search function, “users can type Hindi words in Roman script and a drop down menu suggests several Hindi phrases. By selecting the appropriate query, users can search for Hindi content without even typing in Hindi,” says Roy-Chowdhury. Google has more useful tools for non-English users. Google News is available in Hindi. With the Google translation engine, one can type English words and get a list of suggested synonyms in Hindi. A transliteration tool allows users to type any word in English, hit the space bar and get the

For more information:
[Microsoft Bhasha](http://www.bhashaindia.com/Community/CommunityHome.aspx)
<http://www.bhashaindia.com/Community/CommunityHome.aspx>
 Google translation
http://translate.google.com/translate_t?sl=en&tl=hi#
 Google transliteration
<http://www.google.com/transliterate/indic>
 Urdu Wikipedia
<http://ur.wikipedia.org/>
 Hindi Wikipedia
<http://hi.wikipedia.org/>

same word in a different language.

Roy-Chowdhury explains the process of adding a new language: “Google offers products first in Google Labs and waits for feedback from users for a couple of months. Then the feedback is collated and the product is updated before introducing the language with its other offerings like Gmail, Search, Blogger, Translate and Orkut, to name a few.”

“Urdu is currently available in Google’s transliteration offering on the Google Labs Web site and the language is soon to be introduced in various other products,” he adds.

The efforts of Microsoft, Google and other developers have begun to produce results. Page views of major Hindi news Web sites are rising fast and most of the popular Hindi newspapers have a Web presence now. “In the last two years, page views of navbharattimes.com have increased significantly and half of them come through Google, as Net users generally search for a specific news item or query,” says Nagar.

Yahoo, with headquarters in California, formed a partnership with *Dainik Jagran* a year and a half ago for the newspaper’s Hindi portal. “The *Jagran* relationship helps us gain significant traction among Indian Internet users. From all the audience measures for this product, this has been a resounding success,” says Gopal Krishna, head of Yahoo’s work for emerging market audiences.

Since Yahoo and *Jagran* started working together, page views have “grown to about 1.4 million from one million a year and a half earlier,” says Upendra Swami, who heads the Internet team at *Jagran*.

Hindi Wikipedia, hosted by the nonprofit Wikimedia Foundation, is also gaining popularity. Started in July 2003, Hindi Wikipedia now has more than 36,000 articles. “It now appears to

be the 52nd largest Wikipedia in size, compared to the over 260 individual language Wikipedias,” says Jay Walsh, head of communications at the California-based Wikimedia Foundation. “Considering there are millions of Hindi speakers, it is certainly an important part of the Wikimedia Foundation’s mission to support the growth of this project,” says Walsh. Urdu Wikipedia, started in January 2004, has more than 10,800 articles.

What are the challenges that still remain in the popularization of Hindi and Urdu on the Internet? “The major challenge is Internet penetration and PC prices. The moment we have better Internet penetration, especially in smaller towns, and PC prices go down, Hindi and Indian languages can flourish on the Net,” says Karnik of webdunia.com. India had more than 49 million Internet users in June 2008 out of which about 9 million used the Internet regularly, according to a study by Juxtconsult India, a research company.

“There is a big opportunity in Indian languages. Our study showed that only 28 percent of Indian Web surfers preferred English on the Web, but as good quality content in Indian languages was not easily available, they did not visit many local language sites,” says Mrutyunjay Mishra, co-founder of Juxtconsult.

IT experts agree. “Localization is the key to success in countries like India. In order to get the widest audience reach, one has to look at Hindi, because in a country of over a billion people, English is spoken by less than 80 million people,” says Krishna of Yahoo. Google’s Roy-Chowdhury agrees with him. “The Web is the democratization of access to information,” he says, adding that the Internet is not a luxury but a powerful tool to improve life.

But is Hindi earning enough revenue to be dubbed successful on the Web? “That’s a tough question. Right now it is not much,” says Mishra. But Roy-Chowdhury thinks revenue is bound to come once Hindi reaches a critical volume.

“If we look at how the Internet developed in the U.S., it may provide a useful analogy. First came content, which was mostly produced by people who had a passion for putting up content they cared about. Traffic and monetization was not the motive. Second came growing readership as people started discovering content. This set off a virtuous cycle in which content eventually became a viable, monetizable business. Third were the application developers who could now focus on moving the online experience beyond passive consumption of information to interactivity, community building, service delivery and a host of other innovations,” Roy-Chowdhury says. “India’s market was stuck in phase one for a long time. And I believe it has recently entered phase two.”

Tahsin Usmani contributed to this article.