- PDF
- Send to a friend

The Polyglot Internet

Essay by Ethan Zuckerman, January 23, 2009

This essay was originally posted at Ethan Zuckerman's blog and prepared by Ethan Zuckerman for the World Economic Forum Global Agenda Council on the Future of the Internet by Ethan Zuckerman, October 30, 2008

The first wave of the Internet revolution changed expectations about the availability of information. Information that was stored in libraries, locked in government vaults or available only to subscribers was suddenly accessible to anyone with an internet connection. A second wave has changed expectations about who creates information online. Tens of millions of people are contributing content to the modern Internet, publishing photos, videos and blogposts to a global audience.

The globalization of the Internet has brought connectivity to almost 1.3 billion people. The Internet that results from globalization and user-authorship is profoundly polyglot. Wikipedia is now available in more than 210 languages, which implies that there are communities capable of authoring content in those tongues. Weblog search engine Technorati sees at least as many blogposts in Japanese as in English, and some scholars speculate that there may be as much Chinese content created on sites like Sina and QQ as on all English-language blogs combined.

A user who joins the Internet today is far more likely to encounter content in her own language than had she joined ten years ago. But each internet user is able to participate in a smaller percentage of the total interactions and conversations than an English-speaking internet user could in 1997 when English was the dominant language of the net.

There’s a danger of linguistic isolation in today’s internet. In an earlier, English-dominated internet, users were often forced to cross linguistic barriers and interact in a common language to share ideas with a wider audience. In today’s internet, there’s more opportunity for Portuguese, Chinese, or Arabic speakers to interact with one another, and perhaps less incentive to interact with speakers of other languages. This in turn may fulfill some of the predictions put forth by those who see the Internet acting as an echo-chamber for like-minded voices, not as a powerful tool to encourage interaction and understanding across barriers of nation, language and culture.

For the the Internet to fulfill its most ambitious promises, we need to recognize translation as one of the core challenges to an open, shared and collectively governed internet. Many of us share a vision of the Internet as a place where the good ideas of any person in any country can influence thought and opinion around the world. This vision can only be realized if we accept the challenge of a polyglot internet and build tools and systems to bridge and translate between the hundreds of languages represented online.

Machine translation will not solve all our problems. While machine translation systems continue to improve, they are well below the quality threshold necessary to enable readers to participate in conversations and debates with speakers of another languages. The best machine translation systems still have difficulty with colloquial and informal language, and are most reliable in translating between romance languages. The dream of a system that creates fully-automated, high-quality translations in important language pairs like English/Hindi still appears long off.

While there is profound need to continue improving machine translation, we also need to focus on enabling and empowering human translators. Professional translation continues to be the gold standard for the translation of critical documents. But these methods are too expensive to be used by web surfers simply interested in understanding what peers in China or Colombia are discussing and participating in these discussions.

The polyglot internet demands that we explore the possibility and power of distributed human translation. Hundreds of millions of internet users speak multiple languages; some percentage of these users are capable of translating between these. These users could be the backbone of a powerful, distributed peer production system able to tackle the audacious task of translating the internet.

We are at the very early stages of the emergence of a new model for translation of online content - “peer production” models of translation. Yochai Benkler uses the term “peer production” to describe new ways of organizing collaborative projects beyond conventional arrangements like corporate firms. Individuals have a variety of motives for participation in translation projects, sometimes motivated by an explicit interest in building intercultural bridges, sometimes by fiscal reward or personal pride. In the same way that open source software is built by programmers fueled both by personal passion and by support from multinational corporations, we need a model for peer production translation that enables multiple actors and motivations.

To translate the internet, we need both tools and communities. Open source translation memories will allow translators to share work with collaborators around the world; translation marketplaces will let translators and readers find each other through a system like Mechanical Turk enhanced with reputation metrics; browser tools will let readers seamlessly translate pages into the highest-quality version available and request future human translations. Making these tools useful requires building large, passionate communities committed to bridging in a polyglot web, to preserving smaller languages and to making tools and knowledge accessible to a global audience.

If we do not address the problems of the polyglot internet, we introduce another possible way our shared internet can fragment. There are competing - and likely incompatible - visions for future governance of the internet. As the internet becomes less of a global, shared space and more of a Chinese or Arabic or English space, we lose incentives to work together on common, compatible frameworks and protocols. We face the real possibility of the internet becoming multiple internets, divided first by languages, but later by values, norms and protocols.

The internet is the most powerful tool created by humans to allow connection, collaboration and understanding between people of different nations, races and cultures. For the internet to reach its potential in bridging human differences, we need to make the problems of language and translation center to our conversations about the future of the internet.

Ethan Zuckerman became a fellow of the Berkman Center in January, 2003. His work at Berkman focuses on the impact of technology on the developing world. With Rebecca MacKinnon, he founded Global Voices, an international citizen media community, in 2004. His current work includes Media Cloud, a tool for analyzing professional and citizen media, and research on citizen media and cross-cultural understanding. His personal blog is "My Heart's in Accra", located at http://ethanzuckerman.com/blog*.*

Diversity and Global Inclusion

- Understanding our Knowledge Gaps: Or, Do we have an ICT4D field? And do we want one?
  by Michael Best
- Open ICT ecosystems transforming the developing world
  by Matthew Smith and Laurent Elder
- ICT Diffusion: Have we really made any progress?
  by Sabri Saidam

Open Access

- The Need for a "Knowledge Web" for Scholarship
  by Carolina Rossini
- Opening Access in a Networked Science
  by Melanie Dulong de Rosnay
- A Take on Peter Suber's "The Opening of Science and Scholarship"
  by Jean-Claude Guedon

Open Infrastructure

- Metaphors We Regulate By
  by Rikke Frank Jørgensen
- FORWARD WITH FIBER: An Infrastructure Investment Plan for the New Administration
  by Doc Searls

View all thematic areas »

Source URL: http://publius.cc/polyglot_internet/012309#comment-0