IBM Researchers Lower Language Barrier With Text Translator

November 23 2009

IBM Researchers are helping to break the language barrier with the advent of technology dubbed "n.Fluent" -- smart software that translates text between English and 11 other languages. IBM employees use it to instantaneously translate electronic documents and Web pages -- even live, instant messages exchanged on smartphones.

At the heart of n.Fluent, a remarkably successfully internal IBM crowdsourcing project, is the wisdom of IBM's nearly 400,000 employees in more than 170 countries, where IBM volunteers submit, update and continuously refine word translations. Every time it's used, n.Fluent "learns" and improves its translation engine. To date, the tool has been used by IBMers to translate more than 40 million words.

During a two-week period this past summer, volunteer linguists at IBM crowdsourced approximately 1.3 million words, averaging contributions of approximately 100,000 words per business day. Subsequent campaigns are currently underway across the company to generate and hone more language-related data.

While used exclusively within IBM right now, there is evidence that the technology would be welcomed as a product or service. Companies must enter new markets easily and quickly, and collaborate securely with far-flung employees, partners and suppliers. However, communicating effectively in spite of language differences can be time consuming, awkward, complex and costly.

In addition to providing an alternative to traditional manual translation, n.Fluent offers several key advantages over freely available, Web-based translation programs. In particular, it offers greater security, as the software is hosted behind a company's Internet firewall, and content cannot be seen by non-employees. It also provides translations that are more useful to tech or business-savvy users.

n.Fluent currently can be used for the following languages commonly used in commerce, such as Chinese (Simplified and Traditional), Korean, Japanese, French, Italian, Russian, German, Spanish, Portuguese, Italian and Arabic.

"To become a smarter planet, the world needs a shared vocabulary for collaboration -- particularly the business community," said David Lubensky, an IBM Researcher managing the n.Fluent project. "We see n.Fluent as just such a tool, helping to expand commerce, cement relationships and make the world that much smaller, one word at a time."

IBM volunteers also work on DigiCapE and Agora, two other language-related crowdsourcing projects developed by IBM Researchers. These are technologies that enable people to easily locate, transcribe, synchronize, tag, caption and share audio from a video file. Other examples of crowdscourcing at IBM include Voyage, a wiki where IBMers publish, update, edit and share tips related to business travel.

Real-time language translation was one of 10 emerging business areas in which IBM invested following its global InnovationJam brainstorm in 2006. In September 2009, the Obama-Biden administration identified improved text translation as a technological and investment priority, so as to "greatly [lower] barriers to international commerce and collaboration."

IBM's speech technology research dates back to the early 1960s, when it

developed early speech recognition systems that could understand digits and simple commands. IBM was the first to synthesize text into artificial speech, transcribe the spoken word, speech-enable a personal digital assistant, and translate from one language to the other. Also, IBM was the first to create practical audio-visual speech recognition and speech biometrics using natural language.

Source: IBM