



Multilingual, Federated Search Solution Provides Global Access to Scientific Research

Overview

Customer: WorldWideScience Alliance

Customer Website: worldwidescience.org

Customer Size: Members include 14 national libraries and information centers from around the world

Country or Region: Worldwide

Industry: Public Sector – Library and Museums

Customer Profile

The WorldWideScience Alliance is a multinational partnership that provides access to scientific and technical information sources around the world.

Partner Profile

Deep Web Technologies specializes in multilingual, federated search solutions for multiple industries.

Business Situation

The WorldWideScience Alliance wanted to improve global access to scientific and technological content.

Solution

Create a solution based on Microsoft Translator that searches across multiple data sources worldwide to deliver results in the reader's preferred language.

Benefits

- Increases access to scientific information worldwide
- Facilitates international collaboration
- Expands research opportunities

“Instead of having to go to 100 different sources in ten languages to find global science research information, WorldWideScience.org, powered by Deep Web Technologies and Microsoft Translator, offers the ability to search all of them with a single query.”

Lorrie Johnson, Project Manager, WorldWideScience.org

Deep Web Technologies, a provider of innovative search solutions across multiple industries, was asked to improve access to scientific research for the WorldWideScience Alliance. Millions of articles are published each year in multiple languages, with only a fraction accessible through conventional search engines. To improve access, the company took advantage of Microsoft Translator to create WorldWideScience.org, a portal that can search 100 different databases across 70 countries, and then rank and translate the results into the user's preferred language. Available to anyone with a web browser, including mobile users, the solution increases access to scientific and technological research worldwide, facilitates international collaboration, and provides new opportunities for research in multiple industries.

Situation

Located in Santa Fe, New Mexico, Deep Web Technologies creates innovative solutions for retrieving and analyzing data that would be inaccessible to standard search engines. Its software is used by a variety of corporate, government, and academic organizations for federated search solutions that simultaneously query and rank data from multiple sources to provide relevant results. Its projects include the Science.gov gateway, which provides a portal to information from multiple United States federal science agencies. In 2006, at the International Council for Technical and Scientific Information (ICSTI), representatives from Science.gov, and others familiar with Deep Web Technologies and its federated search capabilities, suggested creating a new portal that would search scientific information worldwide.

The idea kicked off the WorldWideScience Alliance, a multinational governance organization that oversees the project. The organization's mission is to eliminate barriers to finding and sharing research across national boundaries. Lorrie Johnson, Operating Agent for the WorldWideScience Alliance and Product Manager for WorldWideScience.org, says, "Our goal is to make scientific and technical information available to anyone, whether they are scientists, government workers, or interested private citizens. So the portal needed to be a web-based solution—we didn't want unique logins, restrictions, or fees for any users, whatever their location."

Scientific articles and other published data are typically found in what is called the Deep Web, which consists of documents, images, and records located in disparate databases, instead of information provided through webpages. The Deep Web includes billions of pages of research, and the volume is growing rapidly. For example, in 2008 alone, Chinese scholars published approximately 110,000 papers in international journals and 470,000 papers

in domestic publications. And only a fraction of that material is published in English and easily accessible online.

Even if the material could be located, retrieving it typically required filling out request forms, and then translating the article once it was found. The WorldWideScience Alliance turned to Abe Lederman, Chief Executive Officer and Chief Technology Officer of Deep Web Technologies, to realize its vision of a better, more automated solution with multilingual support. "We wanted to create an application that would make scholarly material more accessible worldwide to both English and non-English speakers," he says. "For instance, we wanted a French-speaking user to be able to type in a query and find documents written in any language."

Solution

Deep Web Technologies was already using its Explorit Everywhere! software for federated searches in other projects, including Science.gov. The next step was to extend its capabilities with multilingual search features. The company and the Alliance chose Microsoft Translator, a cloud service that automates translation across multiple languages. "Microsoft Translator was very attractive because we could use the service to do more than just translate from English to one other language. We were able to implement an approach which allowed users to translate between multiple languages; for example, a German-speaking user could obtain results from databases in French, Japanese, Korean, Russian, etc.," says Johnson, "Plus, Microsoft was really supportive and saw the value in being able to provide a scientific and technical solution worldwide."

In addition to providing the high levels of scalability and reliability required for the project, the translation service could be easily integrated with a wide variety of data sources and would work with virtually any operating system and device. By using the Microsoft Translator application

“Many countries are starting to be more open with their research, and it’s exciting that we can provide access through WorldWideScience.org using Microsoft Translator and the Deep Web Technologies Explorit Everywhere! software.”

Lorrie Johnson, Project Manager,
WorldWideScience.org

programming interface (API), Deep Web Technologies would be able to create almost any type of interface and work with dozens of languages without building a complicated integration stream of its own.

In June 2010, Deep Web Technologies and the Alliance launched multilingual search and translation capabilities with WorldWideScience.org, which today searches across more than 100 databases in more than 70 countries. Users worldwide can search databases and translate results in 10 languages: Arabic, Chinese, English, French, German, Japanese, Korean, Portuguese, Russian, and Spanish. The solution also takes advantage of the Microsoft Audio Video Indexing Service (MAVIS). In 2011, multimedia search capabilities were added so that users could retrieve speech-indexed content as well as text.

The site handles approximately 70,000 queries and 1 million page views each month, and all traffic, including that from automated crawlers and search engines, amounts to approximately 70 million transactions per year. When a user enters a search term, WorldWideScience.org instantly provides results clustered by topic, country, author, date, and more. Results are ranked by relevance, and users can choose to look at papers, multimedia, or research data. Divided into tabs for easy usability, the interface also provides details about

each result, including a summary, date, author, location, and whether the full text is available. Users can print the search results or attach them to an email. They can also set up an alert that notifies them when new material is available.

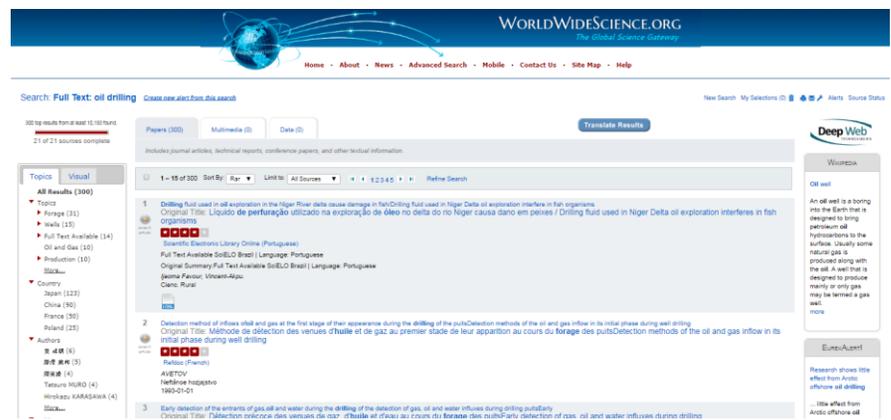
To provide better accessibility, WorldWideScience.org also offers a mobile interface. Deep Web Technologies is launching a streamlined HTML5 version that will work with virtually any device, whether PC, phone, or tablet. Other future enhancements include a localization feature that will provide search portals in the user’s native language.

Benefits

Using Microsoft Translator, Deep Web Technologies created a federated, multilingual search solution for WorldWideScience.org that improves global access to scientific research, encourages international collaboration, and provides new opportunities to share data.

Increases Access to Scientific Information Worldwide

WorldWideScience.org is the result of years of research and innovation. Although the underlying technology itself is exciting, Deep Web Technologies and the WorldWideScience Alliance are most interested in what it enables for users. “This solution increases access to worldwide information, which is the biggest benefit,”



For More Information

For more information about Microsoft products and services, call the Microsoft Sales Information Center at (800) 426-9400. In Canada, call the Microsoft Canada Information Centre at (877) 568-2495. Customers in the United States and Canada who are deaf or hard-of-hearing can reach Microsoft text telephone (TTY/TDD) services at (800) 892-5234. Outside the 50 United States and Canada, please contact your local Microsoft subsidiary. To access information using the World Wide Web, go to: www.microsoft.com/translator

For more information about Deep Web Technologies products and services, call (505) 820-0301 or visit the website at: www.deepwebtech.com

For more information about WorldWideScience.org, visit the website at: <http://www.worldwidescience.org>

explains Johnson. "We search approximately 100 repositories that we estimate include more than 500 million pages of science and technology information. So instead of having to go to 100 different sources to find content, WorldWideScience.org using Microsoft Translator offers the ability to search all of them with a single query."

Facilitates International Collaboration

While WorldWideScience.org can aid research virtually anywhere in the world, the Alliance believes that the solution has particularly strong potential in developing nations that previously lacked sizeable localization budgets and easy access to international research. The portal could encourage international collaboration, too. "For example, a researcher in Korea might be working on a project, do a search with WorldWideScience.org, and realize that someone in the United States is working on a similar problem," Johnson says. "Collaboration among peers is critical to the scientific research process. We believe WorldWideScience.org can connect people, as well as provide information."

Expands Research Opportunities

Deep Web Technologies looks forward to broadening access to information. "Using Microsoft Translator, WorldWideScience.org has gained a lot of attention and visibility in government circles," says Lederman, who sees opportunities to use his company's Explorit Everywhere! application for multiple purposes. "The intelligence community, multinational companies, and risk-analysis firms are all examples of organizations that could really benefit from federated, multilingual searches."

To support the global open-data trend, Deep Web Technologies and the WorldWideScience Alliance are exploring possibilities for integrating new information from publicly funded research, including articles, conference records, and even unstructured raw data. Johnson says, "Many

countries are starting to be more open with their research, and it's exciting that we can provide access through WorldWideScience.org using Microsoft Translator and the Deep Web Technologies Explorit Everywhere! Software."

Software and Services

- Microsoft Translator
- Microsoft Audio Video Service