Automatic Detection and Use of Content Available in Multiple Languages

Matthew Sharifi

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Sharifi, Matthew, "Automatic Detection and Use of Content Available in Multiple Languages", Technical Disclosure Commons, (March 06, 2020)
https://www.tdcommons.org/dpubs_series/2988

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
Automatic Detection and Use of Content Available in Multiple Languages

ABSTRACT

Content producers often create and serve versions of their pages and websites in multiple languages. The versions in each language produced by the content provider typically incorporate the provider’s in-depth knowledge of the content and associated nuances. When available, users are likely to prefer content in their preferred language that originates from the content producer to that generated by automated translation of the content from another language, e.g., as provided by a web browser or translation service. This disclosure describes techniques that, with permission from the user, detect whether the content provider of the web page or website has made content available in multiple languages. When content is available in a language that matches the user’s preferred language, the user is automatically redirected to such content.

KEYWORDS

- Automated translation
- Multilingual website
- Multilingual content
- Multiple language support
- Web browser
- Preferred language
- Language detection

BACKGROUND

Many web browsers include an automatic page translation feature that can translate a web page from its original language into another language that is preferred by the user. For instance, a Spanish-speaking user visiting a web page in the German language can avail of such a feature to
view the web page with the German content automatically translated by the browser to its Spanish equivalent. Such a feature is important to enhance and extend the accessibility of web content for all users regardless of the languages they understand. The feature may be provided by a server-side or on-device translation system.

In many cases, the content producers may themselves have created versions of their pages and sites in multiple languages. For instance, content that is expected to be accessed by native speakers of a number of different languages is often made available separately in each of the languages. For example, content creators in Switzerland may provide their content in each of the main languages in the country, German, French, Italian, and English. Given the contextual connection of the content creator to the delivered content, the versions of the content that originate from the creator are likely to be more accurate than versions of the content produced by automatically translating a version from another language. Therefore, users are likely to prefer to view versions of the content that originate from the content producer when such versions are available for a given piece of web content.

DESCRIPTION

This disclosure describes techniques to enhance the user experience for users who prefer to view web content in their preferred language regardless of the language of the content originally accessed. The techniques include detecting whether the content provider of the web page or web site makes the content available in multiple languages. If content provider serves the content in multiple languages and one of those languages matches the user’s preferred language (determined with user permission) for accessing content, the user is automatically redirected to the version of the content in that language as made available by the content provider.
Detection of languages in which the currently accessed content is available can be performed on the client side within the web browser. To that end, with user permission, the browser can include functionality to search content for appropriate markers that signify the availability of the content in additional languages beyond the currently access language. Such markers can include: links with text such as “language_name” (e.g., “Spanish”), icons commonly associated with specific languages (e.g., flags of countries in which the language is predominant such as the flag of China used to point to Chinese language version of the site), etc.

Alternatively, or in addition, the availability of content in multiple languages can be determined on the server side. For instance, a process running periodically on a server can be used to crawl web content in advance and analyze the crawled content to detect versions of the same content provided in multiple languages. Apart from techniques similar to the client based operations described above, the server side processes can employ automated translation and matching techniques to mark the same content provided in different languages. For instance, such automated approaches can involve automated translation of the content to a common language, such as English, followed by fuzzy matching comparisons to determine whether the translations are close enough to indicate that they correspond to the same content delivered in different languages.

If the user permits, operational implementation of the techniques described above within a web browser can be performed automatically based on the user’s preferred language as specified by the user and/or the relevant browser setting such as language id. In such a case, if the content accessed by the user is not in the user’s preferred language, the user is shown the version of the content in the user’s preferred language as made available by the content producer, if such content is determined to be available. Alternatively, if the user prefers, the user can be
offered a choice between the content producer’s version and automated translation of the content to the user’s preferred language.

**Fig. 1: Redirecting a user to content provider’s website in the user’s preferred language**

Fig. 1 shows a user (102) that prefers to access content in Italian navigating to a URL (106) of a site within the web browser (104). The URL points to the default version of the site in English. Upon detecting that the site creator provides an Italian version of the site (108), the user is automatically and transparently redirected (110) to the URL of the Italian version of the originally accessed English language page.

If the provider of the accessed content does not make the content available in the user’s preferred language, the browser falls back to the automated translation functionality. As such, for the user’s preferred language, the techniques described in this disclosure prioritize content originating from the content provider over automated translation from content in another language. Since content generated directly by the creator is likely to be of much higher quality
compared to automated techniques that lack the same level of nuance and contextual understanding.

The techniques described herein can be implemented within any web browser or other application and can work seamlessly with any web content made available in multiple languages by content providers. Further, if the users permit, the techniques can be employed and extended to support any language in which web content is available. Implementation of the techniques with user permission can enhance the user experience for accessing the content in the user’s preferred language and improve the content translation features within web browsers.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user’s language preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data is treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user’s identity is treated so that no personally identifiable information can be determined for the user, or a user’s geographic location is generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user has control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that, with permission from the user, detect whether the content provider of the web page or website has made content available in multiple languages. When content is available in a language that matches the user’s preferred language,
the user is automatically redirected to such content. Alternatively, if the user prefers, the user is offered a choice between the content producer’s version and automated translation of the content to the user’s preferred language. The detection of languages in which the currently accessed content is available can be performed on the client and/or server side. When content is not available in the user’s preferred language, built-in automated translation functionality is used to provide a translated version. The described techniques can be implemented within any web browser or other application and support any content that is made available in multiple languages by content providers. Selective serving of source content in the user’s preferred language can enhance the user experience and provide an improvement over automated content translations.