Automatic translation of user-entered text

Matthew Sharifi
Jakob Foerster

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

Recommended Citation
http://www.tdcommons.org/dpubs_series/743
Automatic translation of user-entered text

ABSTRACT

This disclosure describes techniques that enable users to interact seamlessly with foreign language webpages. When a user loads a webpage in a foreign language and provides input in a language different from the webpage language, the user input is automatically translated. Based on user selection, the original user input or the translated user input, is sent to the website. The translation is performed in a manner suitable to the context and is not a literal translation of user-entered text. For example, translation is performed such that proper nouns are not translated, certain classes of entities are translated via known mappings, etc. The selective translation techniques can be implemented as part of a web browser, an operating system, a software application, etc. The techniques enable users to interact with foreign language websites as if the websites were in the user’s own language.

KEYWORDS

- web browser
- language detection
- language translation

BACKGROUND

Internet users sometimes view and interact with webpages in a language other than their primary language, e.g., English speaking users access French websites, etc. Some web browsers determine the user’s primary language (e.g., English) and automatically translate a webpage that is in a different language (e.g., French). Such inline translation currently works only for text content on webpages. For example, if a user enters a search query or provides form responses on such a webpage, the user-entered content is not translated prior to submission to the website that
served the webpage. This can limit interactions that users can engage in with websites in a language other than their primary language, henceforth referred to as foreign language.

DESCRIPTION

This disclosure describes techniques that enable users to interact seamlessly with foreign language webpages. Consider, for example, an English-speaking user that loads a French shopping website. In this context, the techniques described herein enable a user to enter a search query (e.g., “humidifier”) in English and when users choose, submit a corresponding search query in French to the website. Similarly, user entered data in form fields on foreign language webpages are also automatically translated when users choose to do so. The translation is performed in a manner suitable to the context and is not a literal translation of user-entered text.

Fig. 1: Selective translation of user input
Fig. 1 illustrates selective translation of user input on foreign language webpages. Selective translation is performed specifically upon user consent. In situations where user has not provided consent, an application (e.g., a browser) that implements selective translation obtains user consent prior to performing selective translation. Selective translation is not performed if the user does not provide consent.

When a user loads a webpage in a browser, the language of the webpage is automatically detected (102), when user provides permission for such detection. For example, a language detector is applied over the webpage content when the webpage is loaded, or when the user submits form content. User inputs, e.g., search queries, entry into form fields, etc. are received, and with user’s permission, the language of user input is determined (104) prior to submission of the user input to the website. The language detector used to determine webpage language and the language of user input can be heuristic based, use neural networks, or other translation techniques.

The webpage language and the language of the user input are compared (106). When the webpage language and the user input language are the same, user inputs are sent to the website (110). When the webpage language and the user input language are different, the user is presented with options to submit the search query or form in the user input language as entered, or to automatically translate the user-entered information into the language of the webpage (108). Translation of the user-entered information can be performed on the client-side, the server-side, or both, depending on user preferences.

When the user chooses to translate the user-entered information, such information is replaced with translated versions prior to form submission, e.g., in POST or GET parameters sent to the website. The user is notified that their query or input was translated into the language of
the webpage. The user can choose to always enable translation on the particular webpage or website, and/or for the given language pair.

The language detector and translation techniques utilized are specialized for the specific context. Classification of types of input field is performed, e.g., to determine whether to employ regular or entity-based translation. For example, the translation is performed such that proper nouns, e.g., the name of a person or brand is not translated. Further, translation models are configured with an understanding of entities such that certain classes of entities are translated not literally but via a known mapping. For example, a user query entered in English “Harry Potter and the Sorcerer’s Stone” is translated to French based on a mapping between book titles in English and French, e.g., as “Harry Potter à l’école des sorciers.”

While the foregoing discussion refers to websites and the automatic translation being performed in a browser, the technology can also be implemented as part of an operating system or applications, e.g., a mobile operating system, mobile apps, etc. For example, the automatic translation can be performed, e.g., by the mobile app itself, by a keyboard app, or by the mobile operating system. In this context, language detection is performed over the on-screen text of the app. When the user provides input text in a different language or explicitly seeks translation, the user input is translated into the language of the app.

Translation of webpages or on-screen content, and of user input text is performed specifically upon user consent for such translation. If the user denies permission, translation is not performed. Further, permission is obtained at different levels of granularity, e.g., for each form field on a webpage, for the entire webpage, for a particular domain, for particular language pairs, etc. Users can choose to translate a portion of user input text, while not translation other portions. Further, webpages or applications can also turn off automatic translation, e.g., if the
webpage requests certain types of user data. User input text is translated specifically for the purpose of submission of user data to the website, and is not stored or used otherwise.

Further to the descriptions above, a user may be 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 social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be 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 may be treated so that no personally identifiable information can be determined for the user, or a user’s geographic location may be 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 may have 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 enable users to interact seamlessly with foreign language webpages. When a user loads a webpage in a foreign language and provides input in a language different from the webpage language, the user input is automatically translated. Based on user selection, the original user input or the translated user input, is sent to the website, e.g., as a search query, as form submission, etc. The translation is performed in a manner suitable to the context and is not a literal translation of user-entered text. For example, translation is performed such that proper nouns are not translated, certain classes of entities are translated via known mappings, etc. The selective translation techniques can be implemented as part of a web
browser, an operating system, a software application, etc. The techniques enable users to interact with foreign language websites as if the websites were in the user’s own language.