

Technical Disclosure Commons

Defensive Publications Series

February 04, 2016

Automatic Detection of User Language

Rachel Blum

Aurimas Liutikas

Alex Ainslie

Rachel Simpson

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

Recommended Citation

Blum, Rachel; Liutikas, Aurimas; Ainslie, Alex; and Simpson, Rachel, "Automatic Detection of User Language", Technical Disclosure Commons, (February 04, 2016)

http://www.tdcommons.org/dpubs_series/157



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

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 of User Language

Authors: Rachel Blum, Rachel Ilan Simpson, Alex Ainslie, Aurimas Liutikas

ABSTRACT

This disclosure describes techniques to automatically detect a user's proficiency in particular languages. Signals from user activity, such as search, browsing activity, location history, mobile usage, etc. are interpreted as indicators of the user's likely language proficiency. Further, user preferences, such as language preferences, spell checking preferences, etc. are analyzed as indicators of the user's language proficiency. Multiple signals are combined to determine languages in which the user is proficient. Detected proficiency is utilized to customize user experience, e.g., to automatically offer translation for content that is in a language in which the user is not proficient.

KEYWORDS

- language detection
- multilingual user
- user experience
- translation

BACKGROUND

Knowledge of a user's language ability permits presentation of content in a manner that is suitable for the user. For example, a browser application can include features for automatic translation of web pages. If the browser application can determine the user's language ability, it can selectively offer translations of web pages. For example, if the browser determines that the user is proficient in Malay, it can present web pages having content in Malay without translation. If the browser determines that the user is not proficient in German, the browser can automatically present a translation (e.g., to Malay) of a web page having content in German.

One indicator of a user's language proficiency is the language that the user selects for a user interface, e.g., for computer software. For example, computer and mobile operating

systems permit users to select a particular language, e.g., English (U.S.), English (U.K.), German, etc. as the user interface language. Application software such as word processors, browsers, spreadsheets, games, etc. also include features allowing users to select a particular language for the user interface.

Multiple factors may influence selection of the user interface language. For example, users that are multilingual may select one of the multiple languages known by the user as the user interface language. In another example, users may select English as the user interface language, e.g., based on a cultural bias towards English rather than based on the user's proficiency in English.

Localization techniques permit implementation of user interfaces that support multiple different languages. In some instances, such localization may be unavailable or may be of poor quality. For example, translations of menu items to the target language may be inaccurate. In such instances, users may select a language in which they are less proficient as the user interface language.

Based on such factors, the selection of a particular language as the user interface language may not be an accurate indicator of the user's proficiency in that language. Further, multilingual users may be proficient in other languages in addition to their selected user interface language. Therefore, an assumption that a user is proficient in a language she has selected for the user interface may not always be accurate or complete.

DESCRIPTION

This disclosure presents techniques that utilize several signals that are indicative of a user's language proficiency. Language proficiency may refer to a user's skill level in a particular language. For example, a user that speaks a language fluently can be determined to be proficient in the language.

Users of computers, phones, tablets, and other devices utilize different software applications to accomplish their everyday tasks. For example, users may utilize a word processor to compose documents, or spreadsheet software to analyze data. Users also access online or server-side services, such as those provided by various online service providers.

Users may use dedicated client software (e.g., an instant messaging client) or an Internet browser to access such services.

User preferences and user data

Users provide their preferences for software applications and online services. For example, users can specify particular settings in application software, e.g., set a particular language as the default language for documents that are composed using a word processor. In another example, users can specify preferences for content, e.g., the preferred language for videos delivered by a video content provider.

Users may also provide consent to software applications and/ or online service providers to collect and utilize data. With user consent, the software applications and/ or online service providers can utilize such data to improve user experience. For example, users may permit a search engine to store search terms that the user enters. In another example, users may permit a mobile operating system and/or a maps provider to store information about the user's location, e.g., as determined from sensors on-board the user's mobile device.

Data and preferences provided by the user, or obtained with the user's permission, are utilized to determine the user's language proficiency. With knowledge of the user's language proficiency, software applications or service providers can provide an enhanced user experience. For example, a browser application can utilize information about the user's language proficiency to determine whether to present a web page in its original language or to present a translated version of the web page in another language that the user is proficient in.

Example environment

Fig. 1 shows an example environment in which techniques of the present disclosure can be implemented. Users, such as User 100 as shown in Fig. 1, utilize user devices such as a computer (160) and/ or a phone (170). User devices are coupled to servers (such as server 180) over a network (185). Servers can provide various services, such as search engine, maps, video delivery, photo storage, translation, word processing etc. User 100 may specify preferences for the user devices and/ or various online services. In implementations where the

user consents to collection and use of user data, user devices (such as computer 160 and phone 170) and online services obtain data about the user.

In the example shown in Fig. 1, computer (160), phone (170), and server (180) include language detection module (190). In various implementations, the techniques of this disclosure can be implemented in user devices, in servers, or in combination of user devices and servers. In different implementations, the language detection module can be implemented as software code, in hardware, or in a combination of software and hardware. In different implementations, the language detection module may be included in an operating system, application software, server software, or a combination of various types of software.

Language proficiency signals

Data about the user include various signals that may be indicative of the user's proficiency in one or more languages. Such signals can be determined by analyzing the user data, e.g., in the language detection module and/or other component of the system. The next several paragraphs describe several examples of analyses that are performed to determine the user's proficiency in a language. Fig. 1 shows several examples of data about the user that can be analyzed to determine various signals.

Browser history

A user utilizes user devices to access various Internet resources, e.g., with a browser application. If the user provides consent, the browser application can maintain a history (105) of the user's visits. Further, an online service can also maintain such history. For example, if the user uses multiple user devices (such as computer 160 and phone 170), the user can benefit from online storage of browser history information by accessing such information across multiple devices. Language detection module 190 can analyze browser history to identify signals indicative of a user's language proficiency. For example, the language detection module determines that the user visited multiple web pages in the Malay language, or visited web pages that belong to websites with top level domain for Malaysia (i.e., .my). The language detection module interprets such browser history as a signal that the user is likely proficient in Malay.

Search history

User 100 utilizes various search tools, such as a desktop or mobile search tool that is integrated into a mobile operating system or application software. User 100 also utilizes various online search engines. With user consent, the operating system, application software, and the online search engine can store the user's search history (110). The language detection module can analyze search history to identify signals indicative of a user's language proficiency. For example, the language detection module analyzes search terms entered by the user and determines that there is a high probability that the language of the search terms is German. Based on this determination, the language detection module interprets the search history as a signal that the user is likely proficient in German. In some examples, the language detection module determines that the user interface language is set to English, while the user's search history has no English search terms. In this example, the language detection module interprets the search history as a signal that the user is likely not proficient in English.

Translation history

A user utilizes translations that are provided by a local software application on a user device or by an online translation provider. For example, translation history (115) includes information indicating that the user requested translation for Czech and Russian. Further, the translation history includes information indicating that the user denied translation for French. The language detection module can analyze such translation history information and interpret such user actions as signals indicating that the user is likely proficient in French, but not in Czech and Russian.

Spell checking settings

A user utilizes spell checking, which may be provided by a local software application or by an online provider. For example, a word processing application installed on the user's computer or accessed over the Internet can provide spell checking functions. The language detection module can inspect the user's spell checking settings (120). For example, the language detection module determines that the user has set spell checking to be active for English and French. The language detection module interprets such a setting as a signal of the user's proficiency in English and French.

Spelling error rate

A user consents to spell checking software tracking the user's rate of errors in spelling. The user's spelling error rate (125) is stored, e.g., locally at the user's computer or at a server. The language detection module can utilize the spelling error rate as a signal of the user's proficiency in a language. For example, the language detection module can detect a relatively higher spelling error rate in English than in French, and interpret this as a signal that the user is more proficient in French than English. In some examples, e.g., when the spelling error rate is available for only one language, the language detection module can assign a relatively low weight to the signal, since the error rate could indicate that the user is a poor speller and not the user's language proficiency.

Keyboard settings

A user sets a preference for keyboards in keyboard settings (130), e.g., a virtual keyboard displayed on a screen of a mobile device such as a phone or a tablet. For example, the user (e.g., a multilingual user) sets the mobile keyboard to a language different from the user interface language. The language detection module interprets such settings as a signal that the user is likely proficient in the language corresponding to the keyboard settings.

Language settings

A user sets a language preference in language settings (135), e.g., in a user profile that is associated with the user. In some cases, the user profile and language settings can be stored on a server, for example. The language detection module interprets such language preference as a signal of the user's proficiency in the preferred language. In some examples, the language detection module assigns a high weight to the signal based on language settings, since the user explicitly indicated a particular language as preferred.

Preferred content language

A user utilizes user devices and servers to access various types of content. For example, the user can access video content from a video sharing website. The user sets a preferred content language (140). The content provider (e.g., the video sharing website) or user account stores the user preference. The language detection module can interpret the user

preference for content in a certain language as a signal of the user's proficiency in the language.

Mobile usage

In some examples, a user utilizes a mobile device such as a phone (170). The mobile device is connected a mobile network operated by a particular service provider. Further, the user communicates with other users using the mobile device (e.g., by placing a phone call). If the user consents to use of such data about the user's mobile usage (145), the language detection module analyzes mobile usage data. For example, the language detection module determines the particular service provider that the user's mobile device is connected to and identifies a location of the service provider and/or the user. The language detection module calculates a signal of the user's likely proficiency in a language, e.g., the local language at the location as determined from the mobile usage.

Maps usage

A user utilizes maps that are provided by an application on the user's device, or provided by a service provider. The maps application or service provider collects data (e.g., user's current location or a start location indicated by the user) if the user consents to such collection. The language detection module can utilize maps usage data (150), e.g., to determine the user's location. The language detection module can utilize the location and other factors determined from the maps usage data (e.g., map searches in a particular language) as signals of the user's language proficiency. For example, the location may indicate proficiency in the local language of the location, and/or user map searches can indicate proficiency in a language used in the searches.

Location

A user device, when connected to a network, is assigned a network address (e.g., an Internet Protocol address). The IP address is indicative of a user's location. In some examples, the user's location can also be determined with other techniques, such as a global positioning system (GPS) sensor built in to the user's device. The language detection module can utilize the user's current and/ or historical location (155) as a signal of the user's language proficiency. For example, if the location information is obtained from the user device IP

address, and indicates that the user device is always in a particular city or country, the language detection module interprets the location as a signal of the user's proficiency in a local language in the city or country.

Combination of signals

The language detection module combines various signals of the user's language proficiency. For example, each signal can be assigned a weight, e.g., based on a likelihood that the signal is accurate. For example, keyboard settings that indicate that the user has set the keyboard to German may be assigned a higher weight than a location signal that the user is in Malaysia, based on the keyboard settings being an explicit user preference and the location signal not being such an explicit user preference. Other techniques of combining the various signals can be used. In different implementations, client-side signals (e.g., signals obtained based on data from user devices) and/ or server-side signals (e.g., signals obtained based on server data about the user) can be used individually or in combination. In some cases, multiple languages of proficiency may be indicated by the various signals. Such multiple determined languages can be ranked and/or scored, based on the weights and/or combinations of the signals, to indicate the likelihood that the user is proficient in each of the determined languages.

Examples of use

The language detection module outputs an indicator of the user's proficiency in one or more languages. In some cases, the output of the language detection module may more accurately indicate a language of proficiency of the user than a selected user interface language. Such output can be used in a variety of ways to improve the user experience. For example, such automatic detection can eliminate the need to ask the user to specify a language preference.

In another example, software that has built-in language translation capability (e.g., a browser application) can automatically translate content to a language that is determined as a preferred language (e.g., spoken language) of the user using described techniques. This example may be particularly useful for users in countries where the user interface language is set to English, but where the users are not proficient in English.

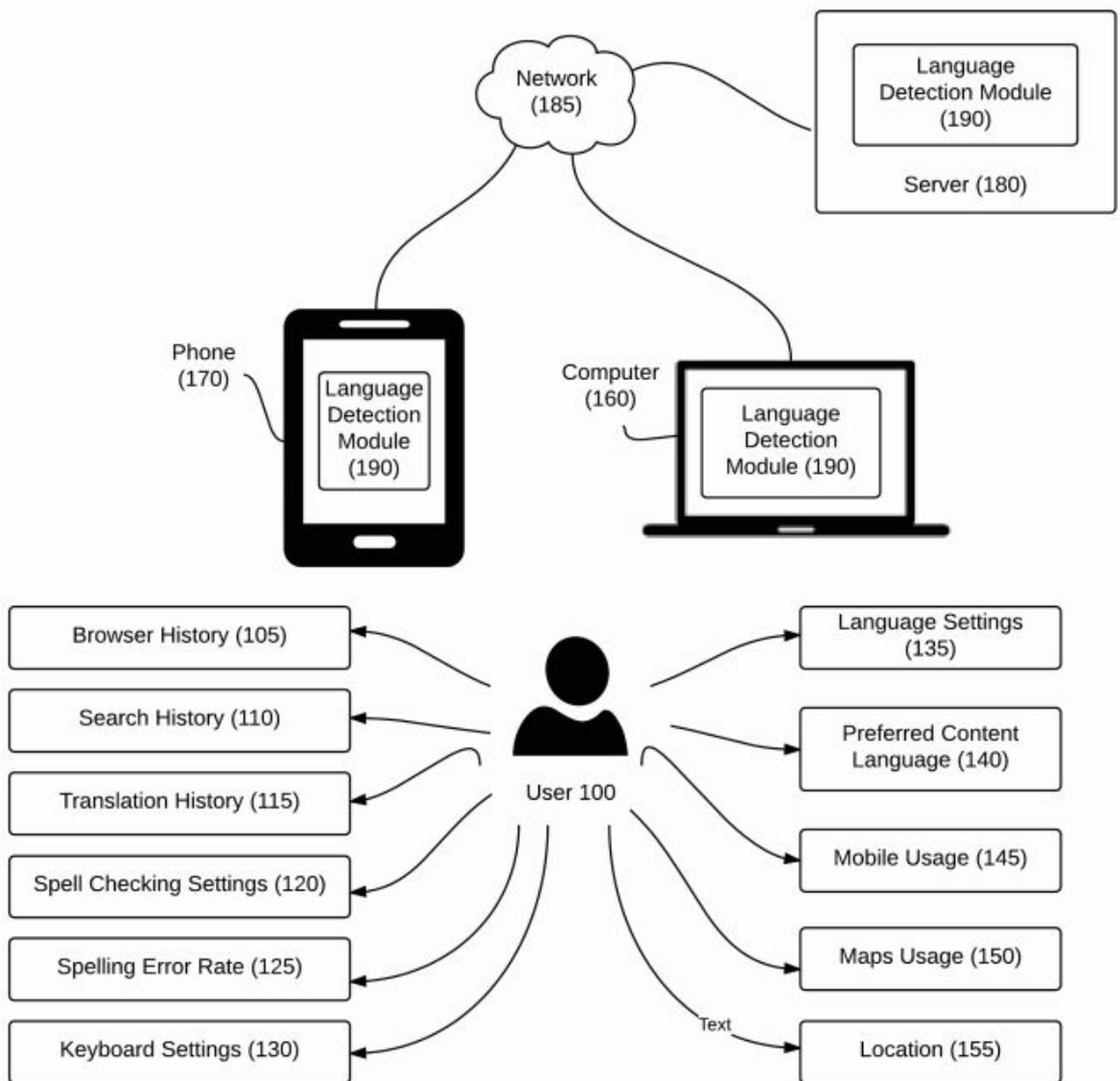


Fig. 1