Technical Disclosure Commons

January 19, 2016

DETECTING SEARCH QUERY LANGUAGE

Tal Cohen

DETECTING SEARCH QUERY LANGUAGE

ABSTRACT

A language detection system can be used for detecting language of a search query that is provided at a search engine by a user. The system receives the search query in a first language, that is detected by the system. The search engine determines a first set of search results in response to the search query. The system receives the first set of search results. The system then identifies language of the first set of search results. The identified language is concluded to be the search query language (first language).

PROBLEM STATEMENT

Search engines are used for information retrieval from various sources and databases. A search engine provides a user interface (UI) for a user to input a search query. The user interface of the search engine has an interface language. The user interface and other controls as well as text on the search engine are presented in the interface language. Similarly, the search query has a search query language.

The user interface language may be set based on the location of the device accessing the search engine interface. For example, if the search engine is accessed in China, then the user interface language may be set as Mandarin Chinese. Alternatively, the user may manually set the user interface language based upon user's preferences. The search query language may be same or different from the user interface language.

Presently, a number of methods are used for detecting the search query language. The methods include detecting the input alphabet of the search query, special keywords or letters, or using various language dictionaries to match words to words in a dictionary of a particular

language. Another method may be by detecting the location of the user and the corresponding local language may be considered as the search query language. However, the existing methods are not accurate in detecting search query language. A method is described that accurately detects the search query language.

DETAILED DESCRIPTION

The systems and techniques described in this disclosure relate to a language detection system that detects the language of a search query. The system can be implemented for use in an Internet, an intranet, or another client and server environment. The system can be implemented locally on a client device or implemented across a client device and server environment. The client device can be any electronic device, for example, a laptop, a mobile phone, a computer, or a tablet.

Fig. 1 illustrates an example method 100 that detects the language of a search query from a set of search results. The method 100 can be implemented by a system that detects search query language, for example, the language detection system.

A user provides a search query to the search engine. The system receives (110) the search query. By default, the search engine assumes the search query to be in same language as the user interface language. For example, if the user interface language for the search engine is English, then the engine assumes that the language of the search query is also English. Therefore, the search engine processes the search query in the user interface language. However, sometimes the search query language, might differ from the user interface language. This may happen, for example, when the user accesses the search engine through a global domain (.com) rather than their local/national domain. For example, a user in France who prefers a global domain is routed to the English language user interface but issues a query in French. Alternatively, users travelling

internationally are automatically redirected to a local domain. For example, a U.S. resident travelling to France may be redirected to local French user interface but may issue queries in English. In yet another example, multilingual users, e.g., a user speaking both English and Spanish, may issue queries in both English and Spanish to the search engine.

The search engine determines a first set of search results in response to processing the search query. The system receives (120) the first set of search results corresponding to the search query. The first set of search results may be a subset of the responsive search results. Search engines use ranking algorithms to rank search results according to relevance to the search query. Further, the system identifies (130) a language of the first set of results. The system performs document language detection on the first set of search results. The system applies document language identification algorithms that analyze characteristics of the text, e.g., specific words, relative frequency of letters, to identify the language of the first set of results. The system identifies the language of the first set of results at the same time the system receives the results. Alternatively or additionally, the system may receive the first set of results already annotated with their language, from the search engine.

The first set of results may include one or more results, for example, the ten highest ranked results, the top five ranked results, or the results that fit in a first page of returned search results. The system determines (140) the language of the search query based on the identified language of the first set of results. The system can determine the language of the search query from the language of one search result, e.g., the highest ranked search result. Alternatively, the system can determine the language of the search query from the respective languages of multiple search results, e.g., the highest ranked n number of search results. The system can assign weights to each search result language. The weights can be based on the relative rankings of the

search results. The system can then combine the weights for a given language. The system can determine that the language with the greatest combined weight is the language of the search query.

In offline scenarios, the system stores past submitted search queries and language information associated with the submitted search queries. The language information includes the search engine user interface language and the determined search query language, as described above, for past search queries. For example, when the user queries "tempo mars," the system determines and stores "English" as the search engine user interface language and "English" as the search query language. In another example, when the user queries "tempo jupiter," the system determines and stores "English" as the search engine user interface language and "Portuguese" as the the search query language. When the system receives a new search query at a search engine with a particular user interface language, the system can map the search query and user interface language to the stored information to identify the search query language.

The system may also use a combination of the online and offline methods, e.g., if a query is found in the stored mapping, the system uses the offline method; otherwise, the system identifies the query language using the online method.

The identified language of the first set of results is the determined search query language. The system provides the identified language to the search engine. The search engine returns a second set of results responsive to the search query, in the identified language. The search engine displays the second set of results on the search results page. The system may present results in the form of a "onebox" or "answer box," which provide a direct answer to the search query. In an example, the identified language may be same as the user interface language.

Fig. 2 illustrates example Graphical User Interfaces (GUIs) of a search engine. The search engine's user interface language is "English." The GUI can be displayed on a display screen or another output device associated with the electronic device implementing the language detection system.

As shown in Fig. 2a, the user provides a search query "tempo mars." The search engine processes the search query and provides a first set of search results. The first set of results are displayed at the search results page of the search engine, as shown in Fig. 2a. The system identifies that while the top two results are in "English" language, the third result is in "French" language. The system assigns weights to the results and identifies the language of the first set of results as "English" language. The system provides the determined search query language to the search engine.

Similarly, as shown in Fig. 2b, for the search query "tempo jupiter," the search engine displays a set of search results. The system determines that while the top two results are in "Portuguese" language, the third result is in "English" language. The system identifies the language of the first set of results as "Portuguese" language. The examples of Fig 2a and Fig. 2b illustrate that though the queries appear to be in the same language, however, the two search queries are in different languages and the language detection system determines the search query language accurately.

Fig. 3 is a block diagram of an exemplary environment that shows components of a system for implementing the techniques described in this disclosure. The environment includes client devices 310, servers 330, and network 340. Network 340 connects client devices 310 to servers 330. Client device 310 is an electronic device. Client device 310 may be capable of requesting and receiving data/communications over network 340. Example client devices 310 are

personal computers (e.g., laptops), mobile communication devices, (e.g. smartphones, tablet computing devices), set-top boxes, game-consoles, embedded systems. Client device 310 may execute an application, such as a web browser 312 or 314 or a native application 316. Web applications 313 and 315 may be displayed via a web browser 312 or 314. Server 330 may be a web server capable of sending, receiving and storing web pages 332. Web page(s) 332 may be stored on or accessible via server 330. Web page(s) 332 may be associated with web application 313 or 315 and accessed using a web browser, e.g., 312. When accessed, webpage(s) 332 may be transmitted and displayed on a client device, e.g., 310. Resources 318 and 318' are resources available to the client device 310 and/or applications thereon, or server(s) 330 and/or web pages(s) accessible therefrom, respectively. Resources 318' may be, for example, memory or storage resources; a text, image, video, audio, JavaScript, CSS, or other file or object; or other relevant resources. Network 340 may be any network or combination of networks that can carry data communication.

The subject matter described in this disclosure can be implemented in software and/or hardware (for example, computers, circuits, or processors). The subject matter can be implemented on a single device or across multiple devices (for example, a client device and a server device). Devices implementing the subject matter can be connected through a wired and/or wireless network. Such devices can receive inputs from a user (for example, from a mouse, keyboard, or touchscreen) and produce an output to a user (for example, through a display). Specific examples disclosed are provided for illustrative purposes and do not limit the scope of the disclosure.
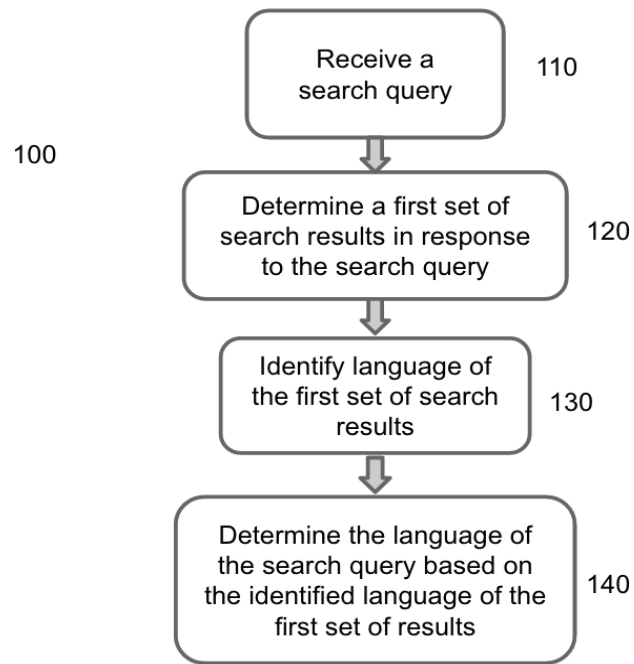
DRAWINGS

100

Receive a
search query   110

Determine a first set of
search results in response   120
to the search query

Identify language of
the first set of search   130
results

Determine the language of
the search query based on
the identified language of the   140
first set of results

Fig. 1

QUERY: | tempo mars

RESULTS: Tempo mars
www.example.com/
Buy A vast collection of men's fashion ranging from suits, shirts, Pants and
more.

Temperature at mars- temperature at mars 20 degrees celsius
www.example.com/
Current temperature at mars is 20 degree celsius.

Previsão do tempo para 25 dias em Mars - AccWeather
www.example.org/
Previsão do tempo de hora em hora · Estendido · Fim de semana ·
Panorama do mês · Situação meteorológica atual · Inicial ... Mars Daily
Weather. Siga-nos em.

Fig. 2a

QUERY: | tempo jupiter

RESULTS: O Tempo para 14 dias jupiter
www.example.com/
O tempo em Jupiter e em todo o mundo para 14 dias, informa-te gratuitamente
das previsões do tempo por hora e condições atuais.

otempo.pt : Weather
Follow this link
Previsão do tempo para 14 dias de Jupiter e do mundo. Mais de 200.000
cidades e locais. Incluindo condições meteorológicas do momento.

Temperature at jupiter- temperature at jupiter 28 degrees celsius
Follow this link
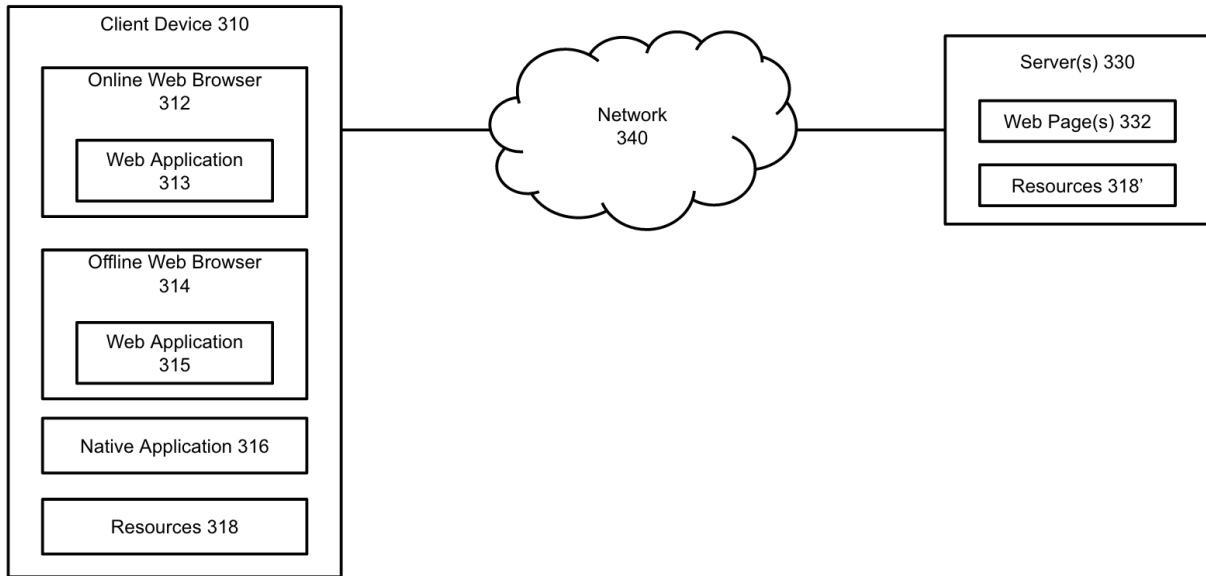Current temperature at jupiter is 20 degree celsius.

Fig. 2b

Client Device 310

Online Web Browser
312

Web Application
313

Offline Web Browser
314

Web Application
315

Native Application 316

Resources 318

Network
340

Server(s) 330

Web Page(s) 332

Resources 318'

Fig. 3