METHOD TO SEARCH FOR A USER GENERATED CONTENT WEB PAGE

20-

PREPARE LEXICAL UNIT DATABASE

SELECT SET OF LEXICAL UNITS

22-

INPUT SET OF LEXICAL UNITS TO SEARCH ENGIIC

24-

RESULTS

CONSOLIDATED LISTS

FIG. 2.

Abstract: A method to search for an user generated content web page comprises: preparing (22) a set of lexical units, inputting (24) said set of lexical units into an internet search engine, and consolidating (28) results of said internet search engine. The method further comprises a preliminary step of creating (20) a database of lexical units comprising at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions, and the preparation of the set of lexical units comprises the assembly of at least one expression of each subset.
METHOD TO SEARCH FOR A USER GENERATED CONTENT WEB PAGE

Field of the invention

The present invention concerns a method to search for a user generated content web page and a software to practice the same.

Background of the invention

Nowadays, to search for a web page, the usual methods consists of choosing some words suited to the object of the search. These words are inputted into the query page of an internet engine such as those proposed by Google Inc. or Yahoo Inc.

As a result, the search engine lists a set of web pages by their title and a automatically generated short abstract. A link gives access to the web page.

Search engines contain some internal, and often secret, algorithms to sort the list of web pages and to show to the user the most pertinent, hopefully, web pages at the beginning of the list.

However, this method is not very efficient when the search concerns user's opinion about a product or a service.

Indeed, when a user would like to buy a product or a service it is nowadays a common practice to search for the opinions of the prior buyers or users. This information may be found on blogs, wikis, forums and any other web site where a "standard" user may post a message. These opinions around a product or service generate a buzz which has a positive, or negative, impact on the success of the product/service.

It is therefore important for marketing department as well as for users to dispose of a method which is efficient to find the web pages containing opinions on a
defined product or service while leaving aside "classical" web pages concerned by the product/service such as pages of merchant web site, of price comparators, etc.

Summary of the invention

To better address one or more concerns, in a first aspect of the invention, a method to search for an user generated content web page comprises

- preparing a set of lexical units,
- inputting said set of lexical units into an internet search engine,
- consolidating results of said internet search engine.

The method further comprises a preliminary step of creating a database of lexical units comprising at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions and the preparation of the set of lexical units comprises the assembly of at least one expression of each subset.

Therefore, the method has the advantage to select preferentially web pages containing opinions about the selected matter.

In particular embodiments:
- the database comprises phonetic transcriptions and misspelled versions of said expressions;
- subsets of feeling expressions and action expressions are arranged as thesaurus in the database of lexical units;
- based on a first set of selected lexical units, a list of lexical units sets to be inputted as prepared,
combining various phonetic transcriptions, misspelled versions and synonyms of each selected expressions;
- each lexical units set of the list is inputted into the internet search engine, and the results for all sets are consolidated into a weighted list of the web pages;
- the weight of each web pages is a combination of the order of appearance of the web page in each result and the number of occurrence of the web page in all results.

Aspects of these embodiments may be combined or modified as appropriate or desired, however.

In a second aspect of the invention a device to search for a web page comprises
- means for storing lexical units,
- means for inputting a set of lexical units into an internet search engine,
- means for consolidating results of said internet search engine,

wherein means for storing lexical units comprises a database of lexical units having at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions and means for inputting said set of lexical units is adapted to input set of lexical units comprising at least one expression of each subset.

In a particular embodiment, means for storing lexical units comprises means for storing phonetic transcription and misspelled versions of said expressions.

In a third aspect of the invention, a computer program product to search for a web page comprises
program instructions to execute the steps of the
hereabove method when the computer program product is
executed on a computer.

These and other aspects of the invention will be
apparent from and elucidated with reference to the
embodiment described hereafter where:

- Fig. 1 is a schematic view of a terminal
  connected to internet to practice an embodiment
  of the invention;

- Fig. 2 is a flowchart of a method according to an
  embodiment of the invention; and

- Fig. 3 is a functional view of a terminal
  practicing an embodiment of the invention.

Detailed description

In reference to Fig. 1, a computer is connected to
internet network 3. Through the network 3, the computer 1
is connected to a server 5 on which a search engine is
running. The man skilled in the art understands that the
server 5 symbolized the infrastructure of search
companies such as Google Inc. or Yahoo Inc. In fact,
these companies use server farms containing hundred of
computers dispatched around the world.

A server 7 is also connected to the internet network
3 and contains a web page which is of interest for the
user of the computer 1 but its address is not known by
the computer 1. The web page contains opinion on a
product/service of interest for the user of the computer
1 and is a user generated content web page such as a
blog, wiki or forum page.

The computer 1 is a classical personal computer. It
comprises interface means such as a display 9, a keyboard
II and a mouse 13 or the like.
It comprises also storage means 15 and processing means 17 such as, for instance, hard disk drives and motherboard.

The storage means 15 contains a computer software product which, when executed by the processing means 17, makes the computer 1 execute the steps of a method to search for a web page according to an embodiment of the invention.

In reference to Fig. 2, the method starts with the creation, step 20, of a database of lexical units to search for.

The database is stored in the storage means 15.

The database comprises at least three subsets of lexical units:
- a subset of feeling expressions;
- a subset of action expressions; and
- a subset of context expressions.

Feeling expressions mean lexical units which are related to the mood or feeling of a human being. For instance, words such as "trouble", "happy/unhappy", "unpleasant/pleasant", etc. define a certain state of mind. Generally, they are the reasons for which a user has posted a message.

Action expressions mean lexical units which define action of a user such as "online reservation", "booking", "sales", "offers", etc.

Context expressions mean lexical units which are used to define the context or the specificity of the search. For instance, if the search concerns the comments of travellers having crossed the Channel, the context expressions includes terms like "ferry (ies)", "Channel crossing", "UK-France", etc.
Advantageously, the subsets of feeling expressions and action expressions are structured in a form of thesaurus to allow a user to grab easily a set of words with similar meanings.

Each lexical unit is preferably stored in the database with all its lexical variations such as singular/plural, and with some misspelled forms. Specifically, the most usual misspelled forms are stored in the database as the searched web pages are edited by normal user with different cultural levels or who practices a foreign language. Therefore, it may be useful to be able to select pages ever if the words are misspelled. A particular misspelled form which is often used by teenagers accustomed to the short messages of the mobile phone is the phonetic form.

After the creation of the database, at least, a set of lexical units is prepared at step 22. The set includes at least one lexical unit of each subset.

In the example of the Channel crossing, a set of lexical units is, for instance, "unhappy experience in channel crossing" where "unhappy" belongs to the feeling subset, "experience" belongs to the action subset and "channel crossing" belongs to the context subset.

The preparation is, advantageously, a guided automatic process in which the user selects some groups of lexical units in each subset and all the sets combining elements of these groups as well as their different forms are automatically generated, forming a list.

At step 24, each set of lexical units is inputted into an internet search engine, such as Google engine (www.google.com) or Yahoo engine (www.yahoo.com). This step is done automatically either by building HTTP
request with the adapted syntax or by using some Application Specific Interface (API) provided by these companies to automatize the internet searches with their engine.

At step 26, the computer 1 receives classically the results of the requests as a list of links to web pages. For each request, i.e. for each set of lexical units, the search engine returns at least one web page containing a list of links. Therefore, the computer 1 may receive a huge number of links.

At step 28, these lists of links are consolidated. Identical links found on different lists are merged but with, advantageously, an increased weight associated to the link as the fact that a page is selected through different search requests may be an indicia of relevance.

As known by the man of skilled in the art, the internet search engines sort the found pages to present to the user on top of the list the "most relevant" page. Each internet engine has its own algorithm to sort the pages and this algorithm is often kept secret.

During the consolidation step 28, the sorted list is also used to give to the "most relevant" pages in the sense of the internet search engine a superior weight.

A combination of weights coming from the internet search engine sort and from the number of occurrences is used to sort the consolidated list by decreasing weight.

The sorted list contains page addresses as hyperlink fields. Therefore, each page can be addressed by the user to be read.

By using the sorted consolidated list, the user has a great chance to read web pages which are relevant to his/her search of user's opinion on a product /service.
Indeed, computer 1 comprises, Fig. 3, from a functional point of view, means 30 for storing lexical units. Typically, means 30 for storing lexical units comprises a database 32 of lexical units having at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions.

Computer 1 comprises also means 34 for inputting a set of lexical units into an internet search engine, the set of lexical units comprising at least one expression of each subset.

And computer 1 comprises means 35 for consolidating results send by the internet search engine.

While the invention has been illustrated and described in details in the drawings and foregoing description, such illustration and description are to be considered illustrative and exemplary and not restrictive, the invention is not limited to the disclosed embodiment.

Other variations to the disclosed embodiment can be understood and effected by these skilled in the art in practising the claimed invention, from a study of the drawings, the disclosure and the appended claims. In the claims, the word "comprising" does not exclude other elements and the indefinite article "a" "or" "an" does not exclude a plurality.
CLAIMS

1. Method to search for a user generated content web page comprising
   • preparing (22) a set of lexical units,
   • inputting (24) said set of lexical units into an internet search engine,
   • consolidating (28) results of said internet search engine

   wherein the method comprises a preliminary step of creating (20) a database of lexical units comprising at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions, and the preparation of the set of lexical units comprises the assembly of at least one expression of each subset.

2. Method according to claim 1, wherein the database comprises phonetic transcriptions and misspelled versions of said expressions.

3. Method according to claims 1, 2, wherein subsets of feeling expressions and action expressions are arranged as thesaurus in the database of lexical units.

4. Method according to claims 2, 3, wherein, based on a first set of selected lexical units, a list of lexical units sets to be inputted as prepared, combining various phonetic transcriptions, misspelled versions and synonyms of each selected expressions.
5. Method according to claim 4, wherein each lexical units set of the list is inputted into the internet search engine, and the results for all sets are consolidated into a weighted list of the web pages.

6. Method according to claim 5, wherein the weight of each web pages is a combination of the order of appearance of the web page in each result and the number of occurrence of the web page in all results.

7. Device to search for a user generated content web page comprising
   • means (30) for storing lexical units,
   • means (34) for inputting a set of lexical units into an internet search engine,
   • means (36) for consolidating results of said internet search engine,
wherein means for storing lexical units comprises a database (32) of lexical units having at least three subsets of lexical units: a subset of feeling expressions, a subset of action expressions and a subset of context expressions and means for inputting said set of lexical units is adapted to input set of lexical units comprising at least one expression of each subset.

8. Device according to claim 7, wherein means for storing lexical units comprises means for storing phonetic transcription and misspelled versions of said expressions.

9. A computer program product to search for an user generated content web page comprising program
instructions to execute the steps of the method according to any one of claims 1 to 6 when said computer program product is executed on a computer.
FIG. 1.

FIG. 2.

PREPARE LEXICAL UNIT DATABASE

SELECT SET OF LEXICAL UNITS

INPUT SET OF LEXICAL UNITS TO SEARCH ENGINE

RESULTS

CONSOLIDATED LISTS
A. CLASSIFICATION OF SUBJECT MATTER

According to International Patent Classification (IPC) or to both national classification and IPC:

INV. G06F17/30

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched:

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>WO 01/59625 A (WORD OF MOUTH TECHNOLOGY LLC. [US]) 16 August 2001 (2001-08-16) page 18, line 20 - page 19, line 25 page 74, line 13 - page 75, line 13 page 93, line 18 - page 98, line 26; figures 22,44-46</td>
<td>1,7,9</td>
</tr>
<tr>
<td>A</td>
<td>WO 2005/059772 A (SWISS REINSURANCE CO [CH]; ANDRIS DANIEL [CH]; KELLER LEO [CH]; RUEF F) 30 June 2005 (2005-06-30) page 12, line 10 - page 20, line 17</td>
<td>1-9</td>
</tr>
</tbody>
</table>

X Further documents are listed in the continuation of Box C.

X See patent family annex.

" Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'S' document member of the same patent family

Date of the actual completion of the international search: 29 August 2008

Date of mailing of the international search report: 05/09/2008

Name and mailing address of the ISA:
European Patent Office, P.B. 5818 Palentlaan 2 NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040, Tx. 31 651 epo nl
Fax: (+31-70) 340-3016

Authorized officer

Bowler, Alyssa
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>EP 0 751 471 A (MASSACHUSETTS INST TECHNOLOGY [US]) 2 January 1997 (1997-01-02) page 2, line 3 - page 5, line 5</td>
<td>1-9</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>WO 0159625</td>
<td>16-08-2001</td>
<td>AU 3689901 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2001236899 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2399641 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1272942 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NZ 520663 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2007100898 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2001047290 A1</td>
</tr>
<tr>
<td>WO 2005059772 A</td>
<td>30-06-2005</td>
<td>AU 2003283172 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2005057426 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2007288447 A1</td>
</tr>
<tr>
<td>EP 0751471</td>
<td>02-01-1997</td>
<td>AU 6282596 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2225790 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 11509019 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UO 9702537 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 6041311 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 6049777 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 6092049 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 6112186 A</td>
</tr>
</tbody>
</table>