



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2010/028567

(22) International Filing Date:
25 March 2010 (25.03.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/419,341 7 April 2009 (07.04.2009) US

(71) Applicant (*for all designated States except US*): **MOTOROLA, INC.** [US/US]; 1303 East Algonquin Road, Schaumburg, Illinois 60196 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **MA, Changxue** [US/US]; 4929 Lichfield Drive, Barrington, Illinois 60010 (US).

(74) Agents: **HAAS, Kenneth A.**, et al.; 1303 East Algonquin Road, Schaumburg, Illinois 60196 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

[Continued on next page]

(54) Title: CONTENT ITEM RETRIEVAL BASED ON A FREE TEXT ENTRY

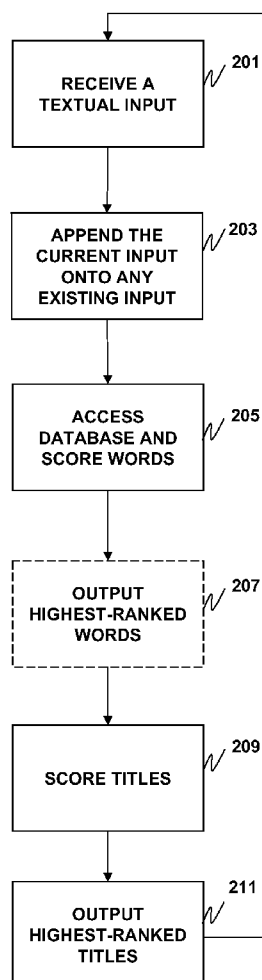


FIG. 2

(57) Abstract: A method and apparatus for textual searching of a database is provided herein. During operation a user will input a letter into a search engine. The search engine will score words based on the letter and display results of the highest-scored words. Another letter will again be received and the process repeated. In situations where titles are returned to the user, additional steps of associating the words with a title and scoring the title take place. The highest-scored titles are provided to the user as the displayed results.



HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,

Published:

— *with international search report (Art. 21(3))*

CONTENT ITEM RETRIEVAL BASED ON A FREE TEXT ENTRY

Field of the Invention

5

The present invention relates generally to textual searching of databases and in particular, to content item retrieval from databases based on a text entry.

10

Background of the Invention

15

20

The use of many electronic devices capable of searching databases has become widespread. For example, a television set-top box used as a multi-media device often has the capability of searching an electronic program guide (EPG) based on a textual input. Often times the searches require large strings of text to be input by the user. The shift of focus across the screen takes time for motor, visual, cognitive processing. This time-consuming process may lead to input error. Therefore a need exists for a method and apparatus for textual searching of databases on electronic devices that is robust to input errors and minimizes the amount of input efforts required from a user.

25

Brief Description of the Drawings

FIG. 1 is a block diagram showing an apparatus used for text entry.

FIG. 2 is a flow chart showing the operation of the apparatus of FIG. 1.

30

35

Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been drawn to scale. For example, the dimensions and/or relative positioning of some of the elements in the figures may be exaggerated relative to other elements to help to improve understanding of various embodiments of the present invention. Also, common but well-understood elements that are useful or necessary in a commercially feasible embodiment are often not depicted in order to facilitate a less obstructed view of these various embodiments of the present invention.

It will further be appreciated that certain actions and/or steps may be described or depicted in a particular order of occurrence while those skilled in the art will understand that such specificity with respect to sequence is not actually required. Those skilled in the art will further recognize that references to specific implementation embodiments such as “circuitry” may equally be accomplished via replacement with software instruction executions either on general purpose computing apparatus (e.g., CPU) or specialized processing apparatus (e.g., DSP). It will also be understood that the terms and expressions used herein have the ordinary technical meaning as is accorded to such terms and expressions by persons skilled in the technical field as set forth above except where different specific meanings have otherwise been set forth herein.

Detailed Description of the Drawings

In order to alleviate the above-mentioned need, a method and apparatus for textual searching of a database is provided herein. During operation a user will input a letter into a search engine. The search engine will score words based on the letter and display results of the highest-scored words. Another letter will again be received and the process repeated. In situations where titles are returned to the user, additional steps of associating the words with a title and scoring the title take place. The highest-scored titles are provided to the user as the displayed results.

When scoring words, a technique is utilized that does not require a user to enter a string of characters in an exact order as a word in the database to identify the word in the database. For example, to find an word consisting of the word “six”, user may just enter “s” and “x”. Words are scored a value in proportion to how many letters they have that are similar to the search string. Additionally, words are scored a higher value if they contain letters in an exact order as contained in the search string. So, for example, even though the word “avalanche” and “avenged” both contain the letters “v” and “a”, the word “avalanche” will be scored higher if the search string is “va”, while the word “avenged” will be scored higher if the search string is “av”.

The above technique provides users with a convenient way to search a database that allows minimal textual input, which allows user to input characters in free format in a very casual manner to reduce attention

demands and while not distracting focus of the attention from entertainment. A user can enter a distinctive letter string as conceived in the user's mind to quickly identify the searched items. Unlike predictive text entry, (which requires the correct spelling to find the items), the above technique may be able to allow users to minimize the keys used for text entry.

Turning now to the drawings, where like numerals designate like components, FIG. 1 is a block diagram showing apparatus 100 used for text entry. Apparatus 100 may be incorporated into any electronic device that is capable of searching databases. Such devices include, but are not limited to TV set-top boxes, cellular telephones, Personal Digital Assistants (PDAs), personal computers, . . . , etc.

As shown, apparatus 100 comprises an electronic processor 101 and database 102. Processor 101 comprises logic circuitry such as a digital signal processor (DSP), general purpose microprocessor, a programmable logic device, or application specific integrated circuit (ASIC) and is utilized to search database 102. Database 102 comprises standard random access memory and is used to store information that can be textually searched. Such information includes, but is not limited to electronic program guides (EPGs), song titles, photo tagged by text, etc.

During operation of apparatus 100, textual input is received by processor 101. The textual input may be from a keyboard, a voice command, a pointing device, or any other means for providing letters to processor 101. In response, processor 101 will score words within database 102. As discussed above, processor 101 will score words based on the letter received and display results of the highest-scored words. Another letter will again be received and the process repeated. In situations where titles are returned to the user, additional steps of associating the words with a title and scoring the title take place. The highest-scored titles are provided to the user as the displayed results.

Scoring Words

As discussed above, when scoring words, a technique is utilized that does not require a user to enter a string of characters in an exact order as contained within a word the database to identify the word in the database. For example, to find an item consisting of the word "six", user may just enter "s" and "x" on the soft keyboard on the screen. Words are scored a value in

proportion to how many letters they have that are similar to the search string (textual input). Additionally, words are scored a higher value if they contain letters in an exact order as contained in the search string. So, for example, even though the word “avalanche” and “avenged” both contain the letters “v” and “a”, the word “avalanche” will be scored higher if the search string is “va”, while the word “avenged” will be scored higher if the search string is “av”.

Other examples are:

- The search string “nn” can return words “cnn” and “thinner”
- The search string “fer” can return the word “winfery”
- The search string “nnoo” can return words “look” and “thinner”
- The search string “kll” can return word “Kelly”
- The search string “ssx” can return words “six”

The scoring of words takes place by initializing a score accumulator (*sword*) to zero. For the current input letter, strings of single letter, bi-letter, tri-letter, etc are formed by pre-pending the previously-supplied inputs. The match of an *M*-letter string at depth *x* produces a score of $\mu_M(x)$. Thus, the match of a single letter in a word at depth *x* produces a score $\mu_1(x)$. The match of a bi-letter in a word at depth *x* produces a bi-letter score $\mu_2(x)$, and so on. With the above in mind:

$$sword = sword * f + (\mu_1(x) + \mu_2(x) + \dots + \mu_n(x)).$$

Where:

f is a forgetting factor to gradually discount the contribution of past inputs. For example, if *f*=1, there is no forgetting factor. If forgetting factor is used we set *f* = 0.97. Alternatively,

$$sword = \mu_1(x) + \mu_2(x) + \dots + \mu_n(x).$$

So, for example, if for a single input letter of “s”, the following score would be returned:

Word	Score
Shoe	$\mu_1(1)$

Basket	$\mu_1(3)$
Sneakers	$\mu_1(1) + \mu_1(8)$
Baseballs	$\mu_1(3) + \mu_1(9)$

For a double letter input of “sc”, the following score would be returned:

Word	Score
soccer	$\mu_1(1) + \mu_1(3) + \mu_1(4)$
sample	$\mu_1(1)$
screen	$\mu_1(1) + \mu_1(2) + \mu_2(1)$
screws	$\mu_1(1) + \mu_1(2) + \mu_2(1) + \mu_1(6)$

5

Considering that a higher score represents a better match of a search string with a particular word, the values for $\mu_M(x)$ are chosen to increase as M increases, and decrease as x increases. Thus, the value for $\mu_M(x)$ is proportional to M and inversely proportional to x . Some sample values for $\mu_M(x)$ are given below with all numerical values being exemplary in nature:

10

M/x	1	2	3	4
1	$\mu_1(1)=3$	$\mu_1(1)=2$	$\mu_1(1)=1.6$	$\mu_1(1)=1.5$
2	$\mu_2(1)=0$	$\mu_2(2)=11$	$\mu_2(2)=10.6$	$\mu_2(2)=10.5$

The score for a particular word can be also estimated from the letter-word occurrences based on the database and heuristics on human behavior modeling. The score can be further processed by a function such as normalization by the number of letters in the word.

15

Scoring Titles

As discussed above, titles may be returned to the user of apparatus 100. When returning titles (e.g., EPG show titles), the score for the words are converted into scores for titles. As an example:

20

- The search string “nn” can return titles “cnn”, “look thinner”
- The search string “fer” can return title “opera winifery”

25

- The search string “nnoo” can return title “look thinner”
- The search string “kll” can return title “Live With Regis and Kelly”
- The search string “ssx” can return title “six sense”

5 A number of highest-scored titles are presented to the user. A score for each title (tscore) comprises a function of individual word scores, and is determined by first setting tscore to zero. Then for each of the N words in the title ($w_1, w_2, \dots, w_N$), Tscore is calculated as:

$$10 \quad \text{Tscore} = \left(\sum_{n=1}^N \text{sword}(w_n) + \beta \cdot \log(Z/i(w_n)) \right)$$

Where $i(w_n)$ is a number of titles having the word w_n , Z is the size of titles in the database and β is determined experimentally to control the reward score given to words appearing less often across the selected titles. For
15 example, $\beta = 10$.

Alternatively, Tscore may simply comprise a summation of the individual word ($w_1, w_2, \dots, w_N$) scores in the title, so that:

$$\text{Tscore} = \sum_{n=1}^N \text{sword}(w_n).$$

20

FIG. 2. is a flow chart showing the operation of the apparatus of FIG. 1. As discussed above, apparatus 100 comprises processor 101 for determining search results based on textual input, and database 102. Database 102 comprises a plurality of words, and optionally comprises a plurality of titles
25 derived from the words.

The logic flow begins at step 201 where processor 101 receives a textual input (search string). Processor 101 appends the current input onto any existing input at step 203 to form an updated search string, and the logic flow continues to step 205. At step 205 database 102 is accessed by
30 processor 101 and all words within database 102 are scored by processor 101 based on the current input and any existing input.

Each word within database 102 receives a score proportional to a number of letters within the search string and the depth of the string within the word. As discussed above, the score of each word comprises a summation of
35 a score for each letter in the textual input, along with a summation of a score

for each combination of letters in the search string. Thus, if a textual input comprises "xyz" each word within database will receive a summation score that is a summation of the score for "x", the score for "y", the score for "z", the score for "xy", the score for "yz", and the score for "xyz". Additionally each word's score for a particular string is proportional to a number of letters within the string (M) and the depth of the string (x) within the word.

At this point in the logic flow (step 207), processor 101 may output the highest scored words as a search result, however, if titles are to be returned as a search result, then the logic flow skips step 207 and proceeds directly to step 209 where each title within the database is scored by processor 101. As discussed above, each title comprise a plurality of scored words, and a score for each title is proportional to a summation of the individual scores of each word in the title. A score for each title (tscore) is determined by summing over the individual word ($w_1, w_2, \dots, w_N$) scores in the title. Thus, the score for each title is proportional to the individual scores of each word in the title. Finally, at step 211 search results comprising titles are output by processor 101. The logic flow then returns to step 201 where processor awaits another letter to append to the search string (textual input).

While the invention has been particularly shown and described with reference to a particular embodiment, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention. It is intended that such changes come within the scope of the following claims:

Claims

1. A method for content item retrieval based on text entry, the method comprising the steps of:
 - 5 receiving by an electronic processor, a textual input (search string);
 - accessing by the electronic processor, a database comprising a plurality of words; and
 - determining by the electronic processor, a score for words from the database based on the textual input, wherein a score for each word is
 - 10 proportional to a number of letters within the search string and the depth of the string within the word.
2. The method of claim 1 further comprising the step of:
 - outputting by the electronic processor, a plurality of words having a
 - 15 highest score.
3. The method of claim 1 further comprising the steps of:
 - determining, by the electronic processor, scores for a plurality of titles within the database, wherein each title comprise a plurality of scored words,
 - 20 and wherein a score for each title is proportional to a summation of the individual scores of each word in the title.
4. The method of claim 3 further comprising the step of:
 - outputting by the electronic processor, a plurality of titles having a
 - 25 highest score.

5. A method comprising the steps of:

receiving by an electronic processor, a textual input;

appending the textual input to a previously supplied textual input to form a search string;

5 accessing by the electronic processor, a database comprising a plurality of words; and

determining by the electronic processor, a score (*sword*) for words in the database based on the search string, wherein

$$10 \quad \textit{sword} = (\mu_1(x) + \mu_2(x) + \dots + \mu_n(x)),$$

where a match of an *M*-letter search string at depth *x* in a word produces a score of $\mu_M(x)$.

15 6. The method of claim 5 wherein $\textit{sword} = \textit{sword} * f + (\mu_1(x) + \mu_2(x) + \dots + \mu_n(x))$, where *f* is a forgetting factor to gradually discount the contribution of past inputs.

7. The method of claim 5 further comprising the steps of:

20 determining, by the electronic processor, scores (*Tscore*) for a plurality of titles within the database, wherein a title within the database comprises a plurality of scored words, and wherein a score for each title is proportional to a summation of individual scores of each word in the title.

25 8. The method of claim 7 wherein for each of *N* words in the title (*w*₁, *w*₂, . . . , *w*_{*N*}), *Tscore* is calculated as:

$$\textit{Tscore} = \left(\sum_{n=1}^N \textit{sword}(w_n) + \beta * \log(Z/i(w_n)) \right), \text{ where}$$

30 *i*(*w*_{*n*}) is a number of titles having the word *w*_{*n*}, *Z* is a size of titles in the database and β is determined experimentally to control a reward score given to words appearing less often across titles.

35 9. The method of claim 7 wherein for each of *N* words in the title (*w*₁, *w*₂, . . . , *w*_{*N*}), *Tscore* is calculated as:

$$\text{Tscore} = \sum_{n=1}^N \text{sword}(w_n).$$

10. An apparatus comprising:

a database comprising a plurality of words;

5 a processor receiving a textual input (search string) and determining a score for words (sword) from database based on the search string, wherein a score for each word is proportional to a number of letters within the search string and the depth of the search string within the word.

10

1/2

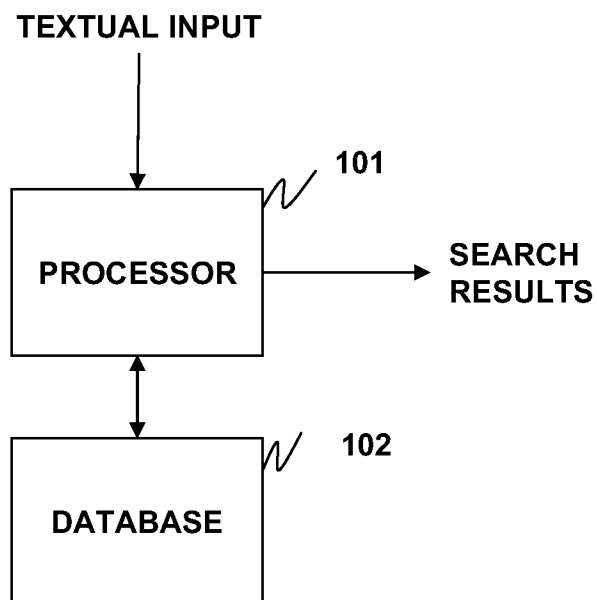


FIG. 1
100

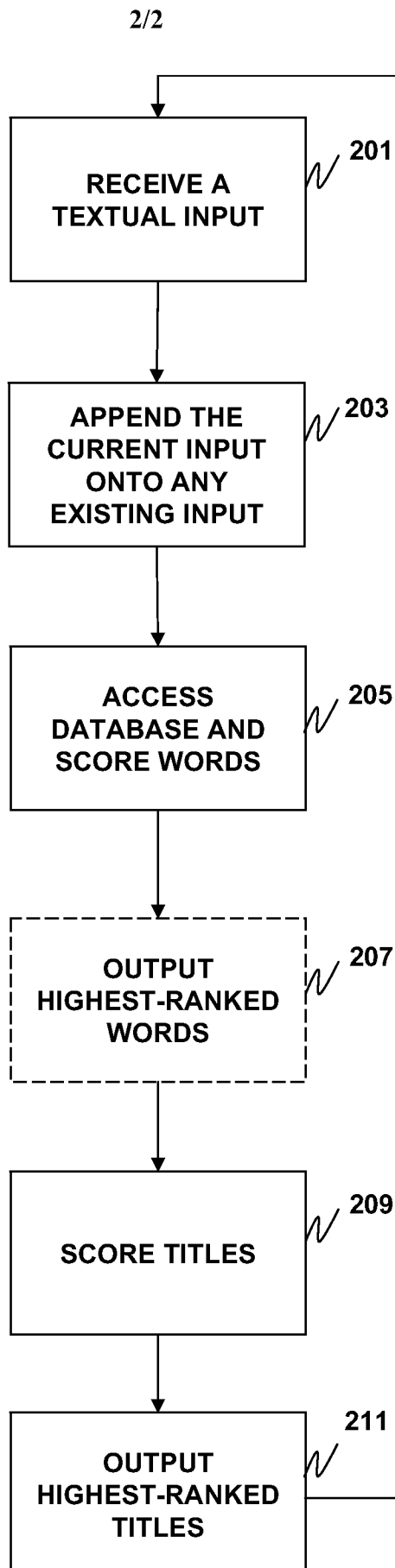


FIG. 2

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2010/028567

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F17/30

ADD.

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

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

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CAVNAR W B: "N-Gram-Based Text Filtering For TREC-2" Presented at the Second Text REtrieval Conference (TREC-2), NIST 30 August 1993 (1993-08-30), 1 September 1993 (1993-09-01), XP002586935 Gaithersburg, MD Retrieved from the Internet: URL: http://cavnar.com/papers/cavnar_1994_n-gram-based_text_filtering_for_trec-2.pdf [retrieved on 2010-06-10] chapters 1 and 2 ----- -/--	1-10



Further documents are listed in the continuation of Box C.



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.

"&" document member of the same patent family

Date of the actual completion of the international search

14 June 2010

Date of mailing of the international search report

02/07/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

Hauck, Rainer

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2010/028567

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FAULK R D: "An Inductive Approach to Language Translation" Communications of the ACM vol. 7, no. 11, November 1964 (1964-11), pages 647-653, XP002586936 New York, NY, USA ISSN: 0001-0782 Retrieved from the Internet: URL: http://portal.acm.org/citation.cfm?id=365067 [retrieved on 2010-06-10] chapter II -----	1-10
X	SHAW R and SOLWAY B and GAIZAUSKAS R and GREENWOOD M A: "Evaluation of automatically reformulated questions in question series" Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering August 2008 (2008-08), pages 58-65, XP002586937 Manchester, UK Retrieved from the Internet: URL: http://portal.acm.org/citation.cfm?id=1641451.1641459 [retrieved on 2010-06-10] chapter 3 -----	1-10