



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

(21) International Application Number:  
PCT/US2011/056963

(22) International Filing Date:  
19 October 2011 (19.10.2011)

(25) Filing Language:  
English

(26) Publication Language:  
English

(30) Priority Data:  
12/913,517 27 October 2010 (27.10.2010) US

(71) Applicant (for all designated States except US): **APPLE INC.** [US/US]; 1 Infinite Loop, Cupertino, CA 95014 (US).

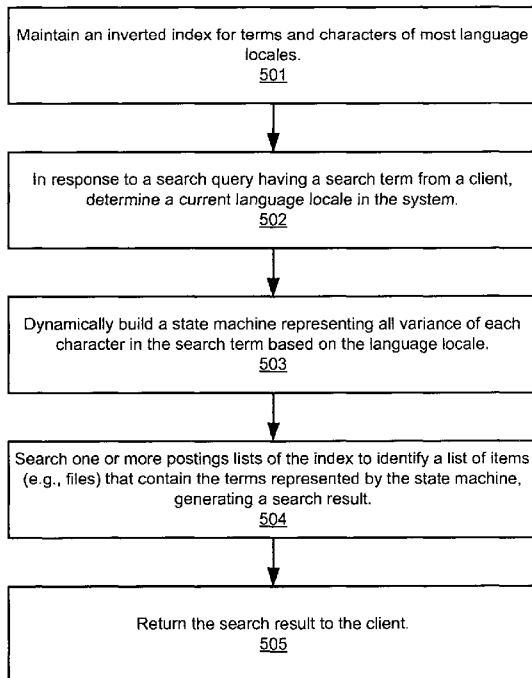
(72) Inventors; and  
(75) Inventors/Applicants (for US only): **HÖRNKVIST, John, M.** [SE/US]; 1 Infinite Loop, M/S 301-2K, Cupertino, CA 95014 (US). **KOEBLER, Eric, R.** [US/US]; 1 Infinite Loop, M/S 303-3KT, Cupertino, CA 95014 (US).

(74) Agents: **VINCENT, Lester J.** et al.; Blakely, Sokoloff, Taylor & Zafman LLP, 1279 Oakmead Parkway, Sunnyvale, CA 94085-4040 (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, 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, QA, RO, RS, RU, RW, 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.

[Continued on next page]

(54) Title: METHODS FOR INDEXING AND SEARCHING BASED ON LANGUAGE LOCALE



500

(57) Abstract: In response to a search query having a search term received from a client, a current language locale is determined. A state machine is built based on the current language locale, where the state machine includes one or more nodes to represent variance of the search term having identical meaning of the search term. Each node of the state machine is traversed to identify one or more postings lists of an inverted index corresponding to each node of the state machine. One or more item identifiers obtained from the one or more postings list are returned to the client, where the item identifiers identify one or more files that contain the variance of the search term represented by the state machine.

FIG. 5

WO 2012/058077 A3

**(84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

**(88) Date of publication of the international search report:**

5 July 2012

**INTERNATIONAL SEARCH REPORT**

International application No  
PCT/US2011/056963

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 practicable, search terms used)  
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GONZALO NAVARRO ED - BAEZA-YATES R ET AL: "MODERN INFORMATION RETRIEVAL, Chapter 8: Indexing and Searching", 1 January 1999 (1999-01-01), MODERN INFORMATION RETRIEVAL, ACM PRESS, NEW YORK, PAGE(S) 191 - 228, XP002457291, ISBN: 978-0-201-39829-8 section 8.2	1-17
A	DAVIS M and WHISTLER K: "UNICODE COLLATION ALGORITHM", Unicode Technical Standard #10, 8 October 2010 (2010-10-08), XP002665206, Retrieved from the Internet: URL:http://www.unicode.org/reports/tr10/tr 10-22.html [retrieved on 2011-12-06] Sections 1 and 8	1-20

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 application or patent 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  3 May 2012	Date of mailing of the international search report  09/05/2012
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/US2011/056963

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2007/203929 A1 (BOLIVAR ALVARO [US]) 30 August 2007 (2007-08-30) abstract, figs. 1 and 2, paras. 29-32, 42 -----	1-17
X	GB 2 331 816 A (IMARKET INC [US]) 2 June 1999 (1999-06-02) the whole document -----	18-20

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2011/056963

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2.  As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
  
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-17

A computer-implemented method for indexing, comprising identifying files containing the variance of a search term by traversing a dynamically built state machine representing variance of the search term having identical meaning of the search term to identify one or more postings lists of an inverted index.

---

2. claims: 18-20

A computer-implemented method for indexing, comprising receiving a term having characters to be indexed and indexing the term in an inverted index using collation keys.

---

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2011/056963

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2007203929	A1	CN 101432685 A	13-05-2009
		EP 1999561 A2	10-12-2008
		US 2007203929 A1	30-08-2007
		WO 2007100812 A2	07-09-2007
-----			
GB 2331816	A	GB 2331816 A	02-06-1999
		US 6026398 A	15-02-2000
-----			