

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 February 2007 (08.02.2007)

PCT

(10) International Publication Number
WO 2007/016232 A3

- (51) International Patent Classification:
G06F 17/30 (2006.01) G06F 17/28 (2006.01)
G06F 17/00 (2006.01)
- (21) International Application Number:
PCT/US2006/029170
- (22) International Filing Date: 26 July 2006 (26.07.2006)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/704,358 1 August 2005 (01.08.2005) US
11/391,889 29 March 2006 (29.03.2006) US
- (71) Applicant (for all designated States except US): BUSINESS OBJECTS AMERICAS [US/US]; 3030 Orchard Parkway, San Jose, CA 95134 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): HAJELA, Swapnil [IN/US]; 920 Lippert Avenue, Fremont, CA 94539 (US). RAJKUMAR, Mareshkumar [IN/US]; 4555 Abbeygate Court, San Jose, CA 95124 (US).

(74) Agents: GALLIANI, William, S. et al.; Cooley Godward Kronish LLP, Attn: Patent Group, 777 6th Street, NW, Washington, DC 20001 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,

[Continued on next page]

(54) Title: PROCESSOR FOR FAST PHRASE SEARCHING

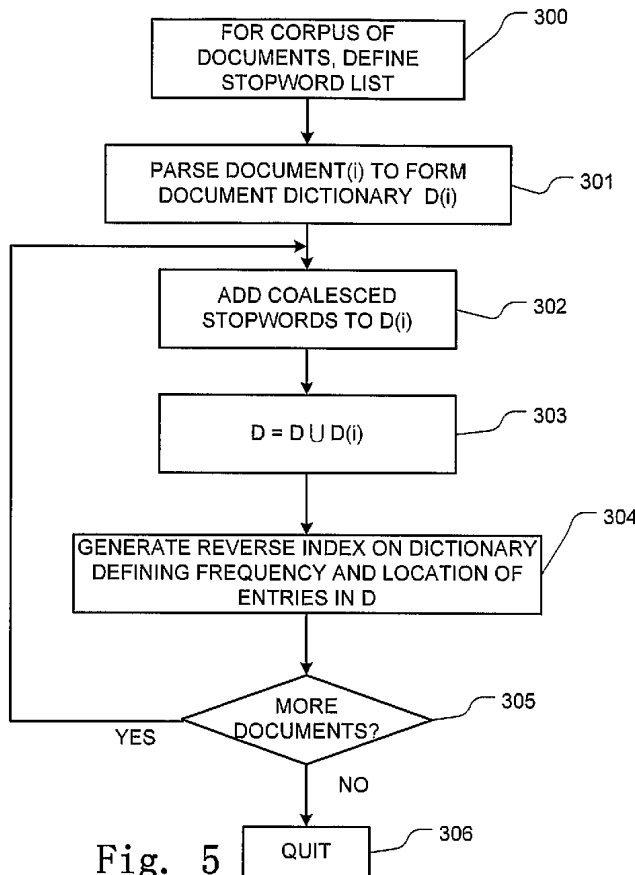


Fig. 5

(57) Abstract: Phrases in a corpus of documents including stopwords are found using a data processor arranged to execute phrase queries. Memory stores an index structure which maps entries in the index structure to documents in the corpus. Entries in the index structure represent words and other entries represent stopwords found in the corpus coalesced with prefixes of respective adjacent words adjacent to the stopwords. The prefixes comprise one or more leading characters of the respective adjacent words. A query processor forms a modified query by substituting a stopword with a search token representing the stopword coalesced with a prefix of the next word in the query. The processor executes the modified query. Also, index structures including coalesced stopwords are created and maintained.

WO 2007/016232 A3



RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

Declaration under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))*

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

30 April 2009

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US06/29170

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **G06F 17/30(2006.01),17/00(2006.01),17/28(2006.01)**

USPC: **707/1,3,5,102;704/9**

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)

U.S. : 707/1,3,5,102; 704/9

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)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,E	US 7,149,748 B1 (Stephan) 12 December 2006 (12.12.2006), whole document	1-35
A	Fast Phrase Querying with Combined Indexes, ACM Transactions on Information Systems. Vol. 22, No. 4, October 2004, pages 573-594, whole document	1-35

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	"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
"E"	earlier application or patent published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
31 May 2008 (31.05.2008)

Date of mailing of the international search report
17 JUN 2008

Name and mailing address of the ISA: US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (571) 273-3201

Authorized officer
Hung Vy
Telephone No. (571)2721650
Deborah A. Thomas
PARALEGAL SPECIALIST
flw