

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 November 2008 (27.11.2008)

PCT

(10) International Publication Number
WO 2008/144457 A3

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

(21) International Application Number:
PCT/US2008/063808

(22) International Filing Date: 15 May 2008 (15.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/804,627 18 May 2007 (18.05.2007) US

(71) Applicant (for all designated States except US): MICROSOFT CORPORATION [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: LIN, Chenxi; One Microsoft Way, Redmond, Washington 98052-6399 (US). JI, Lei; One Microsoft Way, Redmond, Washington 98052-6399 (US). ZENG, Hua-jun; One Microsoft Way, Redmond, Washington 98052-

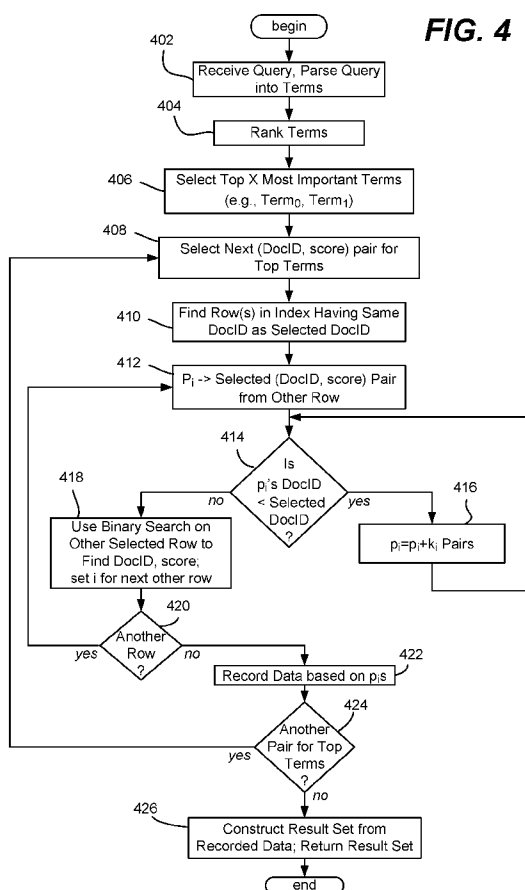
6399 (US). ZHANG, Benyu; One Microsoft Way, Redmond, Washington 98052-6399 (US). CHEN, Zheng; One Microsoft Way, Redmond, Washington 98052-6399 (US). WANG, Jian; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, 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),

[Continued on next page]

(54) Title: EFFICIENT RETRIEVAL ALGORITHM BY QUERY TERM DISCRIMINATION



(57) Abstract: Described is an efficient retrieval mechanism that quickly locates documents (e.g., corresponding to online advertisements) based on query term discrimination. A topmost subset (e.g., two) of search terms is selected according to their ranked importance, e.g., as ranked by inverted document frequency. The topmost terms are then used to narrow the number of rows of an inverted query index that are searched to find document identifiers and associated scores, such as computed offline by a BM25 algorithm. For example, for each document identifier of each important term, a fast search within each of the narrowed subset of rows (that also contain that document identifier) may be performed by comparing document identifiers to jump a pointer within each other row, followed by a binary search to locate a particular document. The scores of the set of particular documents may then be used to rank their relative importance for returning as results.

WO 2008/144457 A3



European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, 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*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

Declarations under Rule 4.17:



- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

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

12 February 2009

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/063808

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06F 17/30(2006.01)i</i>		
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) IPC8 : G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean Utility models and applications for Utility models since 1975 Japanese Utility models and applications for Utility models since 1975		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKIPASS(KIPO internal) "search", "terms", "query", "rank", "index", "invert", "relevant", "document", "row", "reduced", "discrimination"		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STEPHEN ROBERTSON 'Understanding inverse document frequency: on theoretical arguments for IDF.' In Journal of Documentation, 2004, Vol. 60, pp. 503-520.	1-3, 7, 8
A		4-6, 12, 13, 16-20
X	US 5,765,150 A (Michael Burrows) 9 June 1998 see abstract, column 4 line 9 - column 8 line 13, column 19 line 32 - column 21 line 57, column 25 line 28 - column 27 line 63, and figures 2-4, 16, 17, 20-22.	9-11, 14, 15
A		4-6, 12, 13, 16-20
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 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 08 DECEMBER 2008 (08.12.2008)		Date of mailing of the international search report 08 DECEMBER 2008 (08.12.2008)
Name and mailing address of the ISA/KR  Korean Intellectual Property Office Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140		Authorized officer KIM, Yeonho Telephone No. 82-42-481-5758 

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2008/063808

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5765150 A	09.06.1998	None	