(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 August 2003 (07.08.2003)

PCT

(10) International Publication Number
WO 2003/065248 A3

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number:
PCT/EP2003/000761

(22) International Filing Date: 24 January 2003 (24.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/066,346 1 February 2002 (01.02.2002) US
10/180,195 26 June 2002 (26.06.2002) US

(71) Applicant: INTERNATIONAL BUSINESS MACHINES CORPORATION [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(71) Applicant (*for LU only*): IBM DEUTSCHLAND GMBH [DE/DE]; Pascalstrasse 100, 70569 Stuttgart (DE).

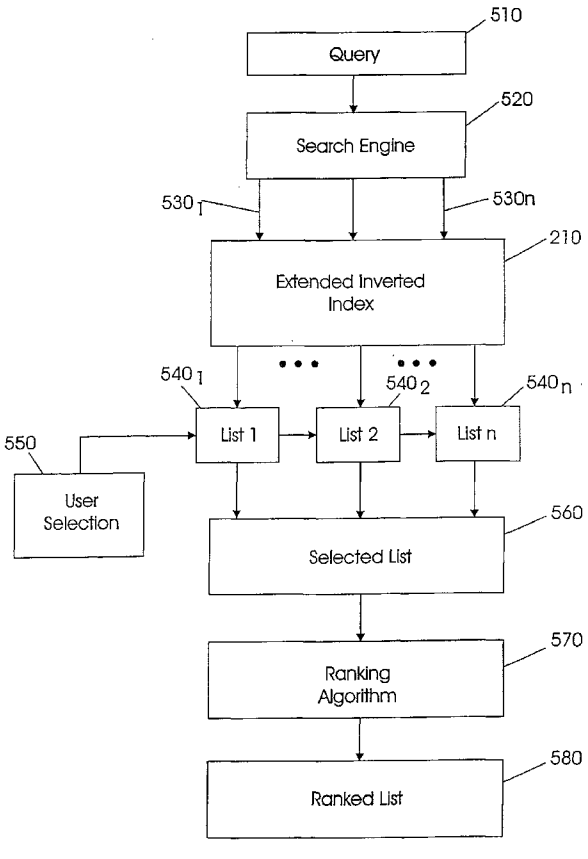
(72) Inventors: BROWN, Gregory, T.; 110 Ohara Drive, Rockmart, GA 30153 (US). DOGANATA, Yurdaer, Nezihi; One Joyce Dv., Chestnut Ridge, Rockland, NY 10977 (US). DRISSI, Youssef; 25 S. Highland Ave., Ossining, NY 10562 (US). FIN, Tong-Haing; 45 Mathews Street, Harrison, NY 10528 (US). KIM, Moon, Ju; 26 Sherwood Heights, Wappingers Falls, NY 12590 (US). KOZAKOV, Lev; 101 Summer Streat, Apt. 721, Stamford, CT 06901 (US). LEON-RODRIGUEZ, Juan; 27 Park Drive, Apt. 103, Pleasant Valley, NY 12569 (US).

(74) Agent: DUSCHER, Reinhard; IBM Deutschland GmbH, Intellectual Property, 70548 Stuttgart (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG,

[Continued on next page]

(54) Title: RETRIEVING MATCHING DOCUMENTS BY QUERIES IN ANY NATIONAL LANGUAGE



(57) Abstract: Search time is reduced with a search engine that includes a bi-directional inverted index facility which can be accessed with a keyword search in one of a number of languages and provide a listing of documents contained in all of those languages. The keywords in all supported languages are preferably stored in an inverted index lookup table cross referenced to documents in those language containing the keywords. Keywords with the same meaning in different languages are accessible together when that keyword in one of the languages is queried. The search engine containing the table can identify pertinent documents either in a selected language, a second language or in all supported languages, as determined by the user. Information about each document can include also information used in ranking the documents such keyword frequency and proximity of keywords.

WO 2003/065248 A3

SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) **Date of publication of the international search report:**
11 March 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.