(A1)

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

(19) World Intellectual Property Organization
International Bureau

G00002

(43) International Publication Date
11 July 2002 (11.07.2002)

PCT

(10) International Publication Number
WO 02/054289 A1

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

(21) International Application Number: PCT/US01/49260

(22) International Filing Date:
19 December 2001 (19.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/258,991 29 December 2000 (29.12.2000) US

(71) Applicant *(for all designated States except US)*: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CARMEL, David** [IL/IL]; 42 Alexander Yanai St., 34816 Haifa (IL). **COHEN, Doron** [IL/IL]; Mitzpe Gilon, D.N. Misgav, 20103 Leshem (IL). **FAGIN, Ronald** [US/US]; 112 Cherrystone Court, Los Gatos, CA 95032 (US). **FARCHI, Eitan** [IL/IL]; Hashmonaim 49, 37052 Pardes Hanna (IL). **HERSCOVICI, Michael** [IL/IL]; 1 Kabirim St., 32922 Haifa (IL). **MAAREK, Yoelle** [IL/IL]; 57 Albert Schweitzer St., 34995 Haifa (IL). **SOFFER, Aya** [IL/IL]; 33 Disraeli St., 34333 Haifa (IL).

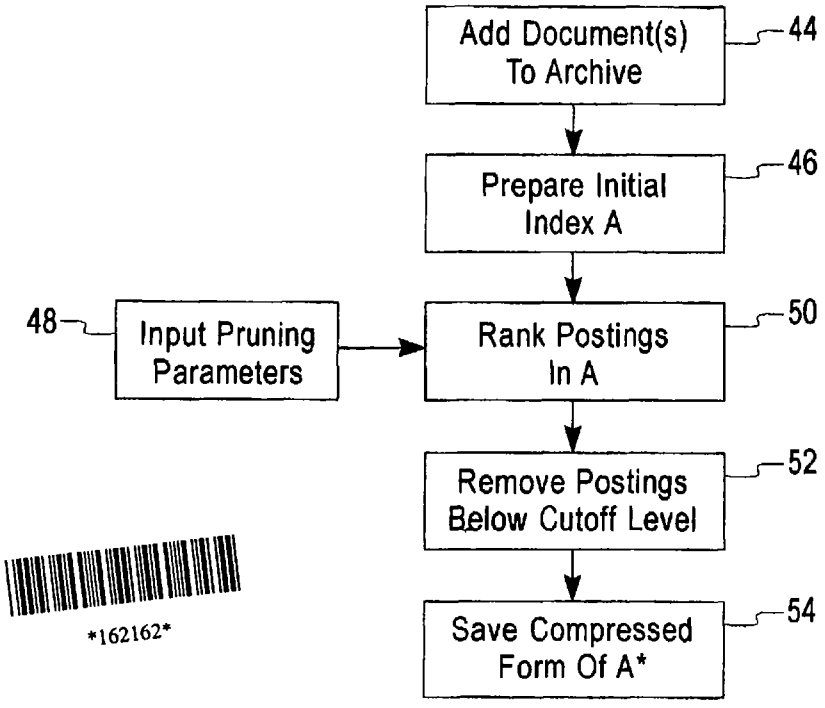
(74) Agent: **ZARICK, Gail**; IBM Corporation, Intellectual Property Law Department, P.O. Box 218, Yorktown Heights, NY 10598 (US).

(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, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: LOSSY INDEX COMPRESSION



(57) Abstract: An apparatus is provided for performing a method (Fig. 2) for pruning an index of a corpus of text documents, wherein the method includes steps for ranking (50) the postings in the index and pruning (48) from the index the postings below a given level in the ranking. The pruning methods of the invention are lossy, since some document postings are removed from the full index; however, the user cannot differentiate the lossy index from the full index.

WO 02/054289 A1

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (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*

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.