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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 May 2003 (22.05.2003)

PCT

(10) International Publication Number
WO 03/042872 A1

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

(21) International Application Number: PCT/US02/35650

(22) International Filing Date:
7 November 2002 (07.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/038,197 9 November 2001 (09.11.2001) US

(71) Applicant *(for all designated States except US)*: **INGENUITY SYSTEMS, INC.** [US/US]; 1565 Charleston Road, Mountain View, CA 94043 (US).

(72) Inventors; and

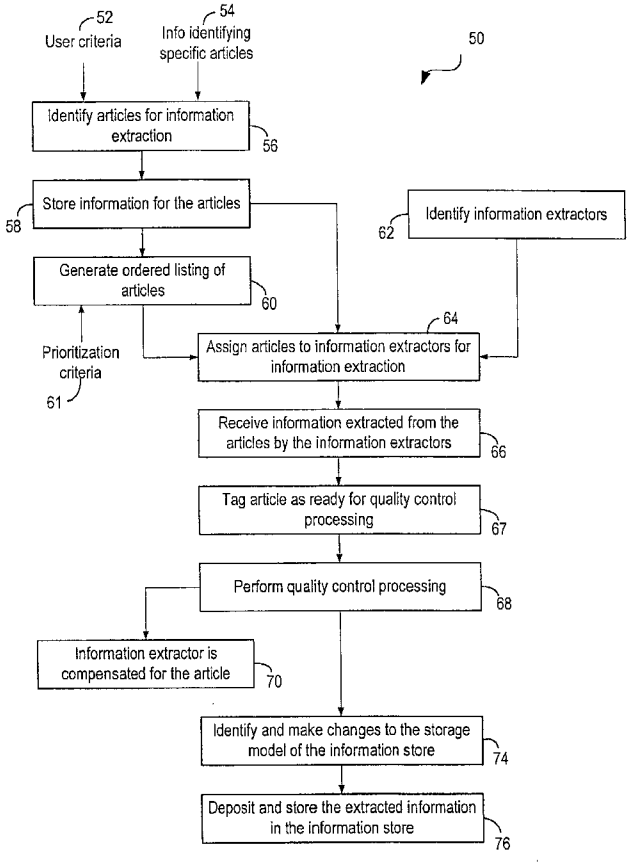
(75) Inventors/Applicants *(for US only)*: **CHO, Raymond, J.** [US/US]; 1063 Morse Avenue, Apt. 13-202, Sunnyvale, CA 94089 (US). **CHEN, Richard, O.** [US/US]; 658 Sierra Vista Avenue, Apt. 9, Mountain View, CA 94043 (US). **FELCIANO, Ramon, M.** [US/US]; 3337 Brittan Avenue, Apt. 7, San Carlos, CA 94070 (US). **RICHARDS, Daniel, R.** [US/US]; 1661 Alma Street, Palo Alto, CA 94301 (US). **NORMAN, Philippa** [GB/US]; 50 West Edith Avenue, Apt. 6, Los Altos, CA 94022 (US).

(74) Agent: **REED, James L.**; MORGAN, LEWIS & BOCKIUS LLP, 1111 Pennsylvania Avenue, N.W., Washington, DC 20004 (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,

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR PERFORMING INFORMATION EXTRACTION AND QUALITY CONTROL FOR A KNOWLEDGE BASE



(57) Abstract: The present invention (fig. 3) relates to the field of information extraction and storage and more specifically to techniques for extracting information from a plurality of articles in a distributed manner and for storing the extracted information in an information store. An embodiment of the present invention identifies a plurality of articles from which information is to be extracted and a plurality of information extractors for extracting the information from the articles (56). A database is provided for storing information related to the plurality of articles and the plurality of information extractors (58). The plurality of articles are assigned to the plurality of information extractors for information extraction. Information extracted by information extractors from the articles is stored in the information store (64).

WO 03/042872 A1

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, 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, IE, IT, LU, MC, NL, PT, SE, 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*

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.