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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 May 2004 (21.05.2004)

PCT

(10) International Publication Number
WO 2004/042493 A3

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

(21) International Application Number:
PCT/SG2002/000249

(22) International Filing Date: 24 October 2002 (24.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 20 Biopolis Way, #07-01 Centros, Singapore 138668 (SG).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TAN, Ah, Hwee** [SG/SG]; 63 Tampines Avenue 1, #05-01 The Tropica, Singapore 529777 (SG). **KANAGASABAI, Rajaraman** [SG/SG]; Block 331, #06-140, Clementi Ave 2, Singapore 120331 (SG).

(74) Agent: **AXIS INTELLECTUAL CAPITAL PTE LTD**; 21A Duxton Road, Singapore 089487 (SG).

(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, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

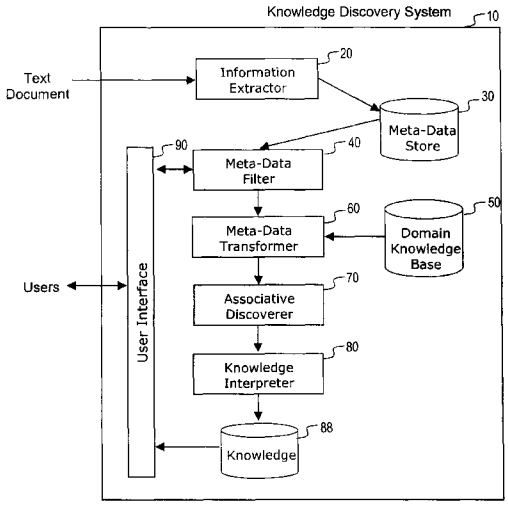
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI,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

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
2 March 2006

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.

(54) Title: METHOD AND SYSTEM FOR DISCOVERING KNOWLEDGE FROM TEXT DOCUMENTS



(57) Abstract: A method and a system for discovering knowledge from text documents are disclosed, which involve extracting from text documents semi-structured meta-data, wherein the semi-structured meta-data includes a plurality of entities and a plurality of relations between the entities; identifying from the semi-structured meta-data a plurality of key entities and a corresponding plurality of key relations; deriving from a domain knowledge base a plurality of attributes relating to each of the plurality of entities relating to one of the plurality of key entities for forming a plurality of pairs of key entity and a plurality of attributes related thereto; formulating a plurality of patterns, each of the plurality of patterns relating to one of the plurality of pairs of key entity and a plurality of attributes related thereto; analyzing the plurality of patterns using an associative discoverer; and interpreting the output of the associative discoverer for discovering knowledge.

WO 2004/042493 A3