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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 November 2003 (20.11.2003)

PCT

(10) International Publication Number
WO 03/096220 A1

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

(21) International Application Number: PCT/GB03/01800

(22) International Filing Date: 28 April 2003 (28.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/142,673 8 May 2002 (08.05.2002) US

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

(71) Applicant *(for MG only)*: **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, North Harbour, Portsmouth, Hampshire PO6 3AU (GB).

(72) Inventors: **DENESUK, Matthew**; 1770 Hudson Drive, San Jose, CA 95124 (US). **GRUHL, Daniel, Frederick**; 6675 Dorene Place, San Jose, CA 95120 (US). **MCCURLEY, Kevin, Snow**; 6721 Tannahill Drive, San Jose, CA 95120 (US). **MEYER, Joerg**; 3193 Kilo Avenue, San Jose, CA 95124 (US). **RAJAGOPALAN, Sridhar**; 19206 Vineyard Lane, Saratoga, CA 95070 (US). **TOMKINS, Andrew**; 737 Hary Road, San Jose, CA 95120 (US). **ZIEN, Jason, Yeong**; 928 Wright Avenue No 408, Mountain View, CA 95043 (US).

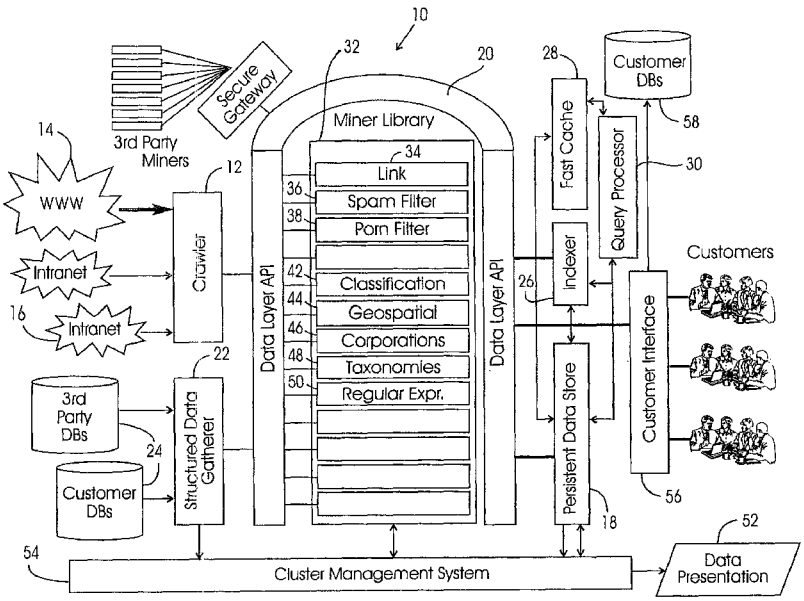
(74) Agent: **LITHERLAND, David, Peter**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester, Hampshire SO21 2JN (GB).

(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, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, 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, RO,

[Continued on next page]

(54) Title: DATA STORE FOR KNOWLEDGE-BASED DATA MINING SYSTEM



(57) Abstract: In a data mining system, data is gathered into a data store using, *e.g.*, a Web crawler. The data is classified into entities and stored into underlying vertical and horizontal tables respectively representing miner outputs and entities that can be the subjects of indexing. Data miners use rules to process the entities and append respective keys to the entities representing characteristics of the entities as derived from rules embodied in the miners, with the keys being associated with the entities in the tables. With these keys, characteristics of entities as defined by disparate expert authors of the data miners are identified for use in responding to complex data requests from customers.

WO 03/096220 A1

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*

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.