

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
19 February 2004 (19.02.2004)

PCT

(10) International Publication Number
WO 2004/015905 A3

(51) International Patent Classification⁷: **G06F 9/45**

(21) International Application Number:
PCT/US2003/024097

(22) International Filing Date: 1 August 2003 (01.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/217,114 12 August 2002 (12.08.2002) US

(71) Applicant: **REUTERS RESEARCH INC.** [US/US]; 100 William Street, New York, NY 10038 (US).

(72) Inventors: **MAHONEY, John**; 4 Country Squire Lane, Princeton Junction, NJ 08550 (US). **BOROVIKOV, Dmitry**; 62 Elwood Road, Northport, NY 11768 (US). **CURTIS, Kevin**; 91 Linvale Road, Ringoes, NJ 08551 (US). **KOLFMAN, Michael**; 8118 23rd Avenue, Apt. 3B, Brooklyn, NY 11214 (US).

(74) Agent: **BRANDT, Jeffrey**; Axiom Legal Solutions c/o PortfolioIP, P.O. Box 52050, Minneapolis, MN 55402 (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, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, 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, 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
- 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:
12 August 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.

(54) Title: METHODS AND SYSTEMS FOR CATEGORIZING AND INDEXING HUMAN-READABLE DATA

(57) Abstract: Systems and methods (20, 200) for processing content packages such as human-readable documents identify and analyze content type. Structural (300) and logical (500) evaluation of a content package is performed, followed by analysis and indexing of concepts within the package. Analysis and identification of concepts and sub-concepts may be an iterative process. Concepts are indexed (800) in accordance with different rule sets representing different consumer needs and perspectives. Customers can then use the indices to navigate large groups of content packages based on the concepts contained within those packages and also on keywords associated with concepts.



WO 2004/015905 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/24097

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 9/45
US CL : 717/121, 136; 707/3, 5, 102; 715/515

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 717/121, 136; 707/3, 101; 715/500.1, 515, 517, 526

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,202,201 B1 (Dom) 13 March 2001, Abstract, column 2, lines 55-57, col. 3, lines 54-67, col. 4, lines 1-38; col. 9, lines 12 - col. 10, line 25; tables 1-3 and associated texts; Fig(s) 3-14 and associated texts.	1-18
A	US 6,263,335 B1 (Paik et al). 17 July 2001, All /	1-18

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 11 May 2004 (11.05.2004)	Date of mailing of the international search report 02 JUN 2004
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Susan Y Chen Telephone No. 703-305-3900

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/24097

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-18

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US03/24097

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

A telephone interview has been conducted on May 10, 2004 (Please see attached telephone interview for reference) for the restriction of instant claims as following:

Group 1, Claims 1-18, drawn to managing software component for designing document.

Group 2, Claims 19-20, drawn to generate database index structure.

Group 3, Claims 21-34, drawn to compound document processing.

Group 4, Claims 35-37, drawn to database query processing.

Application has elected Group 1 for continuing the instant PCT searching/examination processing.