

(51) International Patent Classification:
G06F 17/27 (2006.01)(21) International Application Number:
PCT/US2012/034871(22) International Filing Date:
25 April 2012 (25.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/097,619 29 April 2011 (29.04.2011) US(71) Applicant (for all designated States except US): **THOMSON REUTERS GLOBAL RESOURCES** [IE/CH];
Neuhofstrasse 1, CH-6304 Baar (CH).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MALIK, Hassan, H.** [US/US]; 402 Pheasant Run, Monmouth Junction, NJ 08852 (US). **BHARDWAJ, Vikas, S.** [IN/US]; 30 River Ct. Apt. 1412, Jersey City, NJ 07310 (US). **FIORLETTA, Huascar** [IT/US]; 510 West 123rd Street, Apt. 34, New York, NY 10027 (US). **RAFAT, Armughan** [PK/US]; 10 Raleigh Way, Franklin Park, NJ 08823 (US).(74) Agent: **DIVITA, Bartholomew J.**; Thomson Reuters, 16th Floor, 3 Times Square, New York, NY 10036 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: REPRESENTING INFORMATION FROM DOCUMENTS

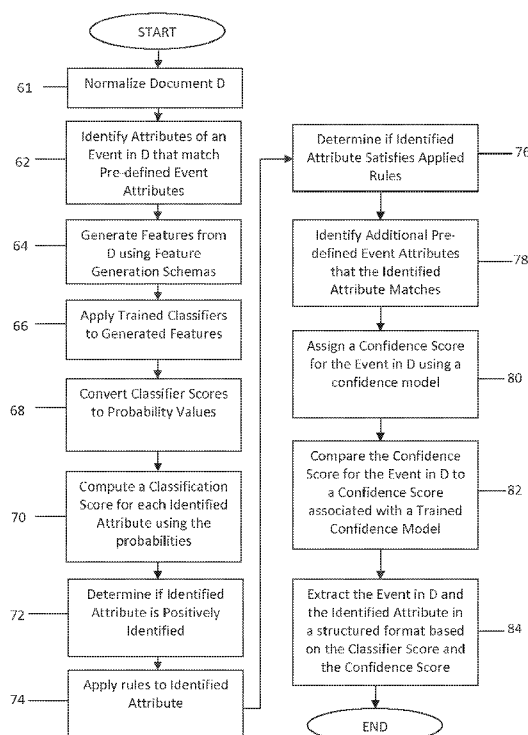


FIG. 3

(57) Abstract: Systems and techniques are disclosed for representing information included in unstructured text documents into a structured format. The systems and techniques identify events and information associated with the events in unstructured documents, classify the identified events and information, and represent the identified events and information in a structured format based on a computed classification score. The systems and techniques may also assign a confidence score to identified events, compare the confidence score associated with events to a confidence score associated with a trained confidence model, and represent the identified events and information associated with the events in a structured format based on the comparison.



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (*Art. 21(3)*)
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (*Rule 48.2(h)*)

(88) Date of publication of the international search report:

20 December 2012

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/034871

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/27
ADD.

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)
G06F

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)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/024769 A1 (FORMAN GEORGE H [US] ET AL) 5 February 2004 (2004-02-05) paragraph [0005] paragraph [0071] - paragraph [0074] ----- -/--	1-88



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search

12 October 2012

Date of mailing of the international search report

19/10/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Abram, Robert

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2012/034871

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JACOBS P S ET AL: "A friendly merger of conceptual expectations and linguistic analysis in a text processing system", PROCEEDINGS OF THE CONFERENCE ON ARTIFICIAL INTELLIGENCE APPLICATIONS. SAN DIEGO, MAR. 16 - 18, 1988; [PROCEEDINGS OF THE CONFERENCE ON ARTIFICIAL INTELLIGENCE APPLICATIONS], WASHINGTON, IEEE COMP. SOC. PRESS, US, vol. CONF. 4, 14 March 1988 (1988-03-14), pages 351-356, XP010011917, DOI: 10.1109/CAIA.1988.196130 ISBN: 978-0-8186-0837-7 the whole document	1-88
A	----- PEGGY M. ANDERSEN ET AL: "Automatic Extraction of Facts from Press Releases to Generate News Stories", ANLC '92 PROCEEDINGS OF THE THIRD CONFERENCE ON APPLIED NATURAL LANGUAGE PROCESSING, 1 January 1992 (1992-01-01), pages 170-177, XP55040218, US DOI: 10.3115/974499.974531 the whole document	1-88
A,P	----- ALEXANDER HOGENBOOM ET AL: "Semantics-based information extraction for detecting economic events", MULTIMEDIA TOOLS AND APPLICATIONS, 1 January 2012 (2012-01-01), XP55040223, ISSN: 1380-7501, DOI: 10.1007/s11042-012-1122-0 the whole document	1-88
A	----- JETHRO BORSJE ET AL: "Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns", INTERNATIONAL JOURNAL OF WEB ENGINEERING AND TECHNOLOGY, vol. 6, no. 2, 1 January 2010 (2010-01-01), pages 115-140, XP55040226, DOI: 10.1504/IJWET.2010.038242 the whole document ----- -/--	1-88

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/034871

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MARTINA NAUGHTON ET AL: "Investigating Statistical Techniques for Sentence-Level Event Classification", PROCEEDINGS OF THE 22ND INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS (COLING'08), 1 August 2008 (2008-08-01), pages 617-624, XP55040229, ISBN: 978-1-90-559344-6 the whole document -----	1-88
A	WANG Y-Y ET AL: "COMBINATION OF STATISTICAL AND RULE-BASED APPROACHES FOR SPOKEN LANGUAGE UNDERSTANDING", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGEPROCESSING, XX, XX, 1 September 2002 (2002-09-01), pages 609-612, XP009035215, section 1 section 4 -----	1-88
X,P	HASSAN H. MALIK ET AL: "Accurate Information Extraction for Quantitative Financial Events", PROCEEDINGS OF THE 20TH ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, CIKM 201, 28 October 2011 (2011-10-28), pages 2497-2500, XP55040212, DOI: 10.1145/2063576.2064001 ISBN: 978-1-45-030717-8 the whole document -----	1-88

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/034871

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004024769	A1	05-02-2004	NONE
