

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number
WO 02/035376 A3

(51) International Patent Classification⁷: G06F 17/27

(21) International Application Number: PCT/US01/32636

(22) International Filing Date: 26 October 2001 (26.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/697,676 27 October 2000 (27.10.2000) US

(71) Applicant: SCIENCE APPLICATIONS INTERNATIONAL CORPORATION [US/US]; 10260 Campus Point Drive, San Diego, CA 92121 (US).

(72) Inventors: BUSCH, Justin, Eliot; 2790 Kelvin Avenue, #1157 Irvine, CA 92614 (US). LIN, Albert, Deirchow; 17128 Patina Street, San Diego, CA 92127 (US). GRAYDON, Patrick, John; 515 Miller Lane, Windston, CA 95492-8068 (US). CAUDILL, Maureen; 11450 Grassy Trail Drive, San Diego, CA 92127 (US).

(74) Agents: ROBINSON, Douglas, W. et al.; Banner & Witcoff, Ltd., Eleventh Floor, 1001 G Street, N.W., Washington, DC 20001-4597 (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, 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, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

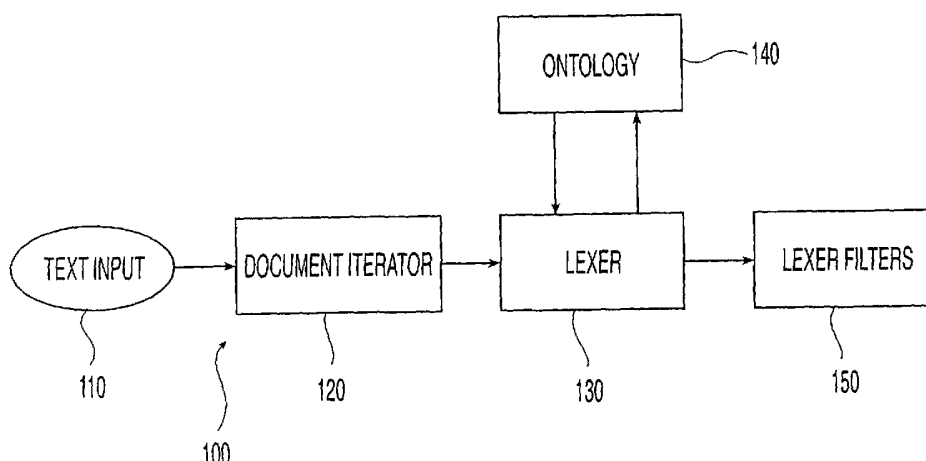
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,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

Published:
— with international search report

(88) Date of publication of the international search report:
28 August 2003

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: ONTOLOGY-BASED PARSER FOR NATURAL LANGUAGE PROCESSING



(57) Abstract: An ontology-based parser incorporates both a system and method for converting natural-language text into predicate-argument format that can be easily used by a variety of applications, including search engines, summarization applications, categorization applications, and word processors. The ontology-based parser contains functional components for receiving documents in a plurality of formats, tokenizing them into instances of concepts from an ontology, and assembling the resulting concepts into predicates. The ontological parser has two major functional elements, a sentence lexer and a parser. The sentence lexer takes a sentence and converts it into a sequence of ontological entities that are tagged with part-of-speech information. The parser converts the sequence of ontological entities into predicate structures using a two-stage process that analyzes the grammatical structure of the sentence, and then applies rules to it that bind arguments into predicates.



WO 02/035376 A3

INTERNATIONAL SEARCH REPORT

Int: 1al Application No
PCT/US 01/32636

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/27

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)
IPC 7 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 practical, search terms used)
EPO-Internal, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 870 701 A (WACHTEL THOMAS JULIUSZ) 9 February 1999 (1999-02-09) column 3 -column 5	1,27
X	WO 00 49517 A (UNIV COLUMBIA ;BARZILAY REGINA (US); MCKEOWN KATHLEEN R (US)) 24 August 2000 (2000-08-24) page 4, line 22 -page 5	2,12,29
A	EP 0 413 132 A (IBM) 20 February 1991 (1991-02-20) abstract	1,27
A	US 5 930 746 A (TING HIAN ANN) 27 July 1999 (1999-07-27) abstract	1,27

Further documents are listed in the continuation of box C. Patent family members are listed in 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 document 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 21 May 2003	Date of mailing of the international search report 28/05/2003
---	---

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Triest, J
--	--

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte | Application No

PCT/US 01/32636

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5870701	A	09-02-1999	EP	0585079 A1	02-03-1994
			WO	9404978 A1	03-03-1994
			GB	2269923 A ,B	23-02-1994
			JP	8502374 T	12-03-1996
WO 0049517	A	24-08-2000	AU	4002600 A	04-09-2000
			CA	2363017 A1	24-08-2000
			EP	1190343 A2	27-03-2002
			WO	0049517 A2	24-08-2000
EP 0413132	A	20-02-1991	US	5146406 A	08-09-1992
			EP	0413132 A2	20-02-1991
			JP	3083167 A	09-04-1991
US 5930746	A	27-07-1999	SG	49804 A1	15-06-1998