#### (19) World Intellectual Property **Organization**

International Bureau





(43) International Publication Date 29 December 2004 (29.12.2004)

PCT

#### (10) International Publication Number WO 2004/114163 A3

(51) International Patent Classification<sup>7</sup>:

G06F 17/30

(21) International Application Number:

PCT/US2004/004099

(22) International Filing Date: 12 February 2004 (12.02.2004)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10/371,399

19 February 2003 (19.02.2003)

- (71) Applicant (for all designated States except US): IN-SIGHTFUL CORPORATION [US/US]; 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109-3044 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): MARCHISIO, Giovanni, B. [US/US]; 9815 Northeast 130th Place, Unit 303, Kirkland, WA 98034 (US). KOPERSKI, Krzysztof [CA/US]; 2311 Yale Avenue East, Apt. D., Seattle, WA 98102 (US). LIANG, Jisheng [CN/US]; 6343 - 114th Avenue Southeast, Bellevue, WA 98006 (US). MURUA, Alejandro [CL/US]; 1310 East Thomas Street, Apt. 302, Seattle, WA 98102 (US). NGUYEN, Thien [US/US]; 22220 - 98th Avenue West, Edmonds, WA 98020 (US). TUSK, Carsten [DE/US]; 20912 4th Avenue South, Seattle, WA 98198 (US). DHILLON, Navdeep, S. [US/US]; 8011 29th Avenue NW, Seattle, WA 98117 (US). POCHMAN, Lubos [US/US]; P.O. Box 3807, Breckenridge, CO 80424 (US).

- (74) Agents: BIERMAN, Ellen, M. et al.; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, 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, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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 (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: 17 February 2005

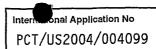
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND SYSTEM FOR ENHANCED DATA SEARCHING

(57) Abstract: Methods and systems for syntactically indexing and searching data sets to achieve more accurate search results and for indexing and searching data sets using entity tags alone or in combination therewith are provided. Example embodiments provide a Syntactic Query Engine ("SQE") that parses, indexes, and stores a data set, as well as processes natural language queries subsequently submitted against the data set. The SQE comprises a Query Preprocessor, a Data Set Preprocessor, a Query Builder, a Data Set Indexer, an Enhanced Natural Language Parser ("ENLP"), a data set repository, and, in some embodiments, a user interface. After preprocessing the data set, the SQE parses the data set according to a variety of levels of parsing and determines as appropriate the entity tags and syntactic and grammatical roles of each term to generate enhanced data representations for each object in the data set. The SQE indexes and stores these enhanced data representations in the data set repository. Upon subsequently receiving a query, the SQE parses the query also using a variety of parsing levels and searches the indexed stored data set to locate data that contains similar terms used in similar grammatical roles and/or with similar entity tag types as indicated by the query. In this manner, the SQE is able to achieve more contextually accurate search results more frequently than using traditional search engines.



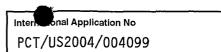
# INTERNATIONAL SEARCH REPORT



A. CLASSIF IPC 7	GO6F17/30			
According to	International Patent Classification (IPC) or to both national classification	ion and IPC		
B. FIELDS				
Minimum do	cumentation searched (classification system followed by classification $G06F$	n symbols)		
]	ion searched other than minimum documentation to the extent that suc			
Electronic da	ata base consulted during the international search (name of data base	e and, where practical, search terms used		
EPO-In	ternal, WPI Data, PAJ, INSPEC			
C. DOCUME	ENTS CONSIDERED TO BE RELEVANT			
Category °	Citation of document, with indication, where appropriate, of the relevant	vant passages	Relevant to claim No.	
X	US 5 933 822 A (CORSTON SIMON H 13 August 1999 (1999-08-03) column 10, line 62 - column 11, lifigure 4 column 13, line 11 - line 52; figurables 2-4 column 20, line 62 - column 21, lifigures 10A,10B column 22, line 29 - line 48; figures 12A,12B,13A	ine 29; ure 5A; ine 23; ure 11	1-121	
X Furt	her documents are listed in the continuation of box C.	Patent family members are listed	in annex.	
"A" docume consider a	ent defining the general state of the art which is not dered to be of particular relevance document but published on or after the international date ent which may throw doubts on priority claim(s) or is cited to establish the publication date of another in or other special reason (as specified) ent referring to an oral disclosure, use, exhibition or means ent published prior to the international filing date but than the priority date claimed	"T" later document published after the inte or priority date and not in conflict with cited to understand the principle or the invention "X" document of particular relevance; the cannot be considered novel or cannot involve an inventive step when the decannot be considered to involve an indocument is combined with one or ments, such combination being obvious the art.  "&" document member of the same patent Date of mailing of the international sea	the application but eory underlying the claimed invention to be considered to comment is taken alone claimed invention wentive step when the ore other such docunus to a person skilled	
1	6 December 2004	23/12/2004		
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 Polzer, A		

2

# INTERNATIONAL SEARCH REPORT



C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	<del></del>
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/14651 A (INVENTION MACHINE CORP) 16 March 2000 (2000-03-16) page 8, line 4 - line 13; figure 2 page 9, line 8 - page 10, line 1; figures 5,8-10,12,13 page 11, line 4 - page 12, line 15; figures 3,7,11	1-121
Α	US 2002/156763 A1 (MARCHISIO GIOVANNI B) 24 October 2002 (2002-10-24) cited in the application  the whole document	1,25,49, 73,86, 87, 89-91, 93,98, 103,108, 113,116,
A	NAGAO K ET AL: "SEMANTIC ANNOTATION AND TRANSCODING: MAKING WEB CONTENT MORE ACCESSIBLE"  IEEE MULTIMEDIA, IEEE COMPUTER SOCIETY, US, vol. 8, no. 2, April 2001 (2001-04), pages 69-81, XP001132036  ISSN: 1070-986X page 72, right-hand column, last paragraph - page 73, right-hand column, paragraph 2	1,25,49, 73,86, 87, 89-91, 93,98, 103,108, 113,116, 119

# INTERNATIONAL SEARCH REPORT

Information on patent family members

# International Application No PCT/US2004/004099

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 5933822	A	03-08-1999	CN CN EP EP JP JP WO	1265209 T 1302412 T 0996899 A1 0998714 A1 2001511564 T 2001511565 T 9905618 A1 9905621 A1	30-08-2000 04-07-2001 03-05-2000 10-05-2000 14-08-2001 14-08-2001 04-02-1999 04-02-1999
WO 0014651	A	16-03-2000	US AU CA CN EP JP NO WO US	6167370 A 5790399 A 2341583 A1 1325513 T 1112541 A1 2002524799 T 20011194 A 0014651 A1 2001014852 A1 2002087497 A1	26-12-2000 27-03-2000 16-03-2000 05-12-2001 04-07-2001 06-08-2002 03-05-2001 16-03-2000 16-08-2001 04-07-2002
US 2002156763	A1	24-10-2002	US US AU CA EP WO	6510406 B1 2003217047 A1 9630401 A 2423476 A1 1323067 A1 0227536 A1	21-01-2003 20-11-2003 08-04-2002 04-04-2002 02-07-2003 04-04-2002