

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 September 2006 (08.09.2006)

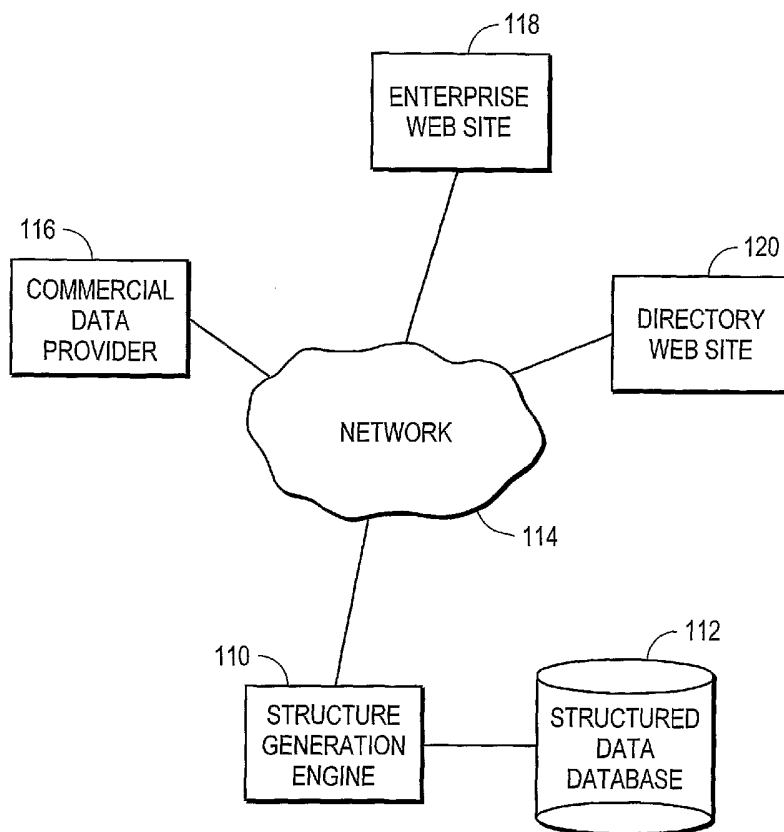
PCT

(10) International Publication Number
WO 2006/094206 A3

- (51) International Patent Classification:
G06F 7/00 (2006.01)
- (21) International Application Number:
PCT/US2006/007639
- (22) International Filing Date: 2 March 2006 (02.03.2006)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/658,214 2 March 2005 (02.03.2005) US
11/366,162 1 March 2006 (01.03.2006) US
- (71) Applicant (for all designated States except US):
GOOGLE, INC. [US/US]; 1600 Amphitheatre Park-
way, Building 41, Mountain View, California 94043 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): PASZTOR, Egon
[US/US]; 777 W. Middlefield Rd., #205, Mountain View,
California 94043 (US). EGNOR, Daniel [US/US]; 139
Rinconada Avenue, Palo Alto, California 94301 (US).
- (74) Agents: HOFFMAN, Brian, M. et al.; FENWICK &
WEST LLP, Silicon Valley Center, 801 California Street,
Mountain View, California 94041 (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, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI,
NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG,
SK, SL, SM, 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, NA, 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,

[Continued on next page]

(54) Title: GENERATING STRUCTURED INFORMATION



(57) Abstract: Structured and/or unstructured data about enterprises are acquired from one or more sources such as commercial data providers, enterprise web sites, and/or directory web sites. Strings are extracted from the unstructured data. The strings contain key, value pairs describing facts about the enterprises. The extracted strings are parsed to normalize the keys and values and place them in a machine-understandable structured representation. Some keys and/or values cannot be normalized. The facts are clustered with the enterprise to which they pertain. Normalized facts from different sources are compared and confidence levels and/or weights are assigned to the facts. These confidence levels and weights are used to select the facts that are displayed on a page for the enterprise in a directory.

WO 2006/094206 A3



FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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.

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:

23 November 2006

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US06/07639

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **G06F 7/00(2006.01)**

USPC: 707/1,10,102

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. : 707/1, 10, 102

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 --- Y	US 2003/0154071 A1 (SHREVE) 14 August 2003 (14.08.2003), paragraphs 15, 59, 78, 92, 114, 132.	1-5, 7, 10-16, 18-19, 21-26, 28-29, 31 ----- 6,8-10,17,20,27,30,32
Y	US 2004/0107125 A1 (GUHEEN et al.) 03 June 2004 (03.06.2004), paragraphs 0113, 0143-0145, 0701-0711, 1474, 1703, 2251, 2445, 2907, 3106, 3138, 3150, 3158.	6, 8-10, 17, 20, 27, 30, 32

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	"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
"E"	earlier application or patent published on or after the international filing date	"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
"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)	"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
"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	"&" document member of the same patent family

Date of the actual completion of the international search
26 June 2006 (26.06.2006)

Date of mailing of the international search report
13 SEP 2006

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. (571) 273-3201

Authorized officer
Linh Black
Linh Black
Telephone No. 571-272-2100