

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 September 2004 (30.09.2004)

PCT

(10) International Publication Number
WO 2004/084009 A3

(51) International Patent Classification⁷: G06F 15/00

[US/US]; 9 Old Nashua Road, B3, Amherst, NH 03031 (US). SCHOLZ, Carl [US/US]; 332 Goodman's Hill Road, Sudbury, MA 01776 (US). WHITNEY, Ronald [US/US]; 156 Meeting Street, Providence, RI 02906 (US).

(21) International Application Number:
PCT/US2004/006742

(22) International Filing Date: 5 March 2004 (05.03.2004)

(74) Agents: COLEMAN, Roy, J. et al.; Iandiorio & Teska, 260 Bear Hill Road, Waltham, MA 02451-1018 (US).

(25) Filing Language: English

(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.

(26) Publication Language: English

(30) Priority Data:
10/388,685 14 March 2003 (14.03.2003) US

(71) Applicant (for all designated States except US): TEXTERITY, INC. [US/US]; 144 Turnpike Road, Southborough, MA 01772 (US).

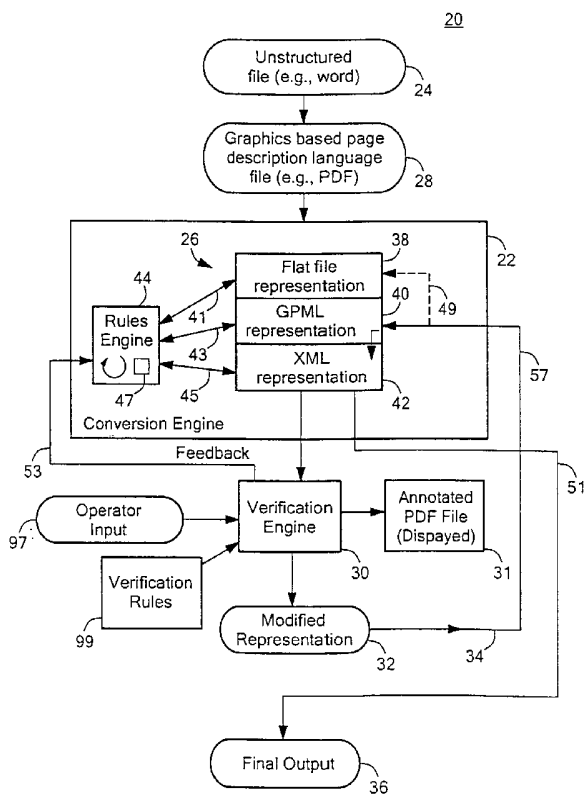
(72) Inventors; and

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

(75) Inventors/Applicants (for US only): MCLURE, Petra

[Continued on next page]

(54) Title: METHOD AND EXPERT SYSTEM FOR DOCUMENT CONVERSION



(57) Abstract: An expert system for more efficiently and accurately deducing document structure from document formatting, the expert system including a conversion engine for converting an unstructured file to a structured file, and a verification engine, responsive to the output of the conversion engine, for generating and displaying a representation of the structured file annotated with a visual depictions of the classified components thereof so that the annotations can be modified and/or classifications can be added and/or classifications can be suggested, and/or rules for classification can be suggested and the structured file reprocessed by the conversion engine.

WO 2004/084009 A3



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, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *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:
10 March 2005

Published:

— *with international search report*

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.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/06742

A. CLASSIFICATION OF SUBJECT MATTER				
IPC(7) : G06F 15/00 US CL : 715/500.000 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. : 715/500.000, 522, 523; 345/764, 809				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM-TDB				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
Y,E	US 6,725,426 B1 (PAVLOV) 20 April 2004 - see col. 2, lines 45-51, col. 4, lines 60-64, col. 6, lines 12-63, col. 4, lines 62-65, col. 4, lines 60-65, col. 5, lines 46, col. 6 lines 35, col. 2 lines 45-51, col. 4 lines 33-35.	1 - 56		
Y,E	US 2004/0107118 A1 (HARNSBERGER et al) 03 June 2004 - see the entire document.	1 - 56		
Y	US 6,360,236 B1 (KHAN et al) 19 March 2002 - see col. 6, line 31, col. 7, line 28.	1 - 56		
Y	US 6,154,213 A (RENNISON et al) 28 November 2000 - see col. 19, line 65, col. 20 line 9.	4, 5 and 50 - 53		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
* Special categories of cited documents: <table border="0" style="width:100%"> <tr> <td style="width:50%"> "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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 </td> <td style="width:50%"> "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 </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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		Date of mailing of the international search report		
13 August 2004 (13.08.2004)		30 DEC 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 Vincent Trans <i>Peggy Harwood</i> Telephone No. 703-305-3900		

INTERNATIONAL SEARCH REPORT

PCT/US04/06742

Continuation of B. FIELDS SEARCHED Item 3:

execut\$4 with button icon, conver\$5 word,text,annotat\$6, classif\$9 component, header,button,time near6
event,graphic\$5,description,database,flat,verificat\$5,engine,intermediate,file,creat\$5,size weight underline strike\$5 color baseline
overlap align,baseline,overlap align,word list,table with content.

INTERNATIONAL SEARCH REPORT

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2002/0154751 A1 (THOMPSON, III et al) 24 October 2002 - see paragraphs 51, 52 and 55.	5 and 53
Y,P	US 2003/0121000 A1 (COOPER et al) 26 June 2003 - see the entire document.	5, 8, 53 and 56
Y	US 2002/0188532 A1 (ROTHSTEIN) 12 December 2002 - page 3, paragraph 40.	8 and 56
Y	US 2002/0015059 A1 (CLARKE) 07 February 2002 - see the abstract page 2, paragraph 24.	9 - 33
Y	US 2002/0165647 A1 (GLENN, III et al) 07 November 2002 -see page 3, paragraph 31-32,	9 - 33
Y,P	US 2003/0167455 A1 (IBORRA et al) 04 September 2003 - see pages 34-35.	35 - 46
A,P	US 2003/0142128 A1 (REULEIN et al) 31 July 2003 - see the entire document.	1 - 56