

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number  
WO 2001/024053 A3

(51) International Patent Classification<sup>7</sup>: G06F 17/30

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, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(21) International Application Number:  
PCT/US2000/025755

(22) International Filing Date:  
20 September 2000 (20.09.2000)

(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), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/156,923 28 September 1999 (28.09.1999) US

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

(71) Applicant: XMLEXPRESS, INC. [US/US]; 8 Central Street, Arlington, MA 02476 (US).

(88) Date of publication of the international search report:  
25 March 2004

(72) Inventors: SOKOLWSKI, Rachael; 79 Bow Street, Arlington, MA 02474 (US). OXENBERG, Philip; 87 Adams Drive, Stow, MA 01775 (US).

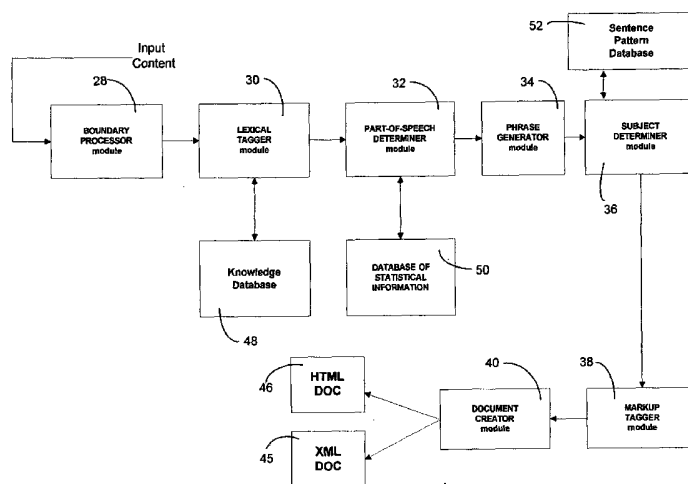
(74) Agents: CANNING, Kevin, J. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(15) Information about Correction:  
Previous Correction:  
see PCT Gazette No. 40/2002 of 3 October 2002, Section II

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,

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: SYSTEM AND METHOD FOR AUTOMATIC CONTEXT CREATION FOR ELECTRONIC DOCUMENTS



(57) Abstract: A system and method for automatically generating a context for information contained in any type of text-based electronic document such as a hypertext markup language (HTML) encoded web page. The contexts generated by the system describe the content or meaning of parts or sections of the electronic document. Additionally, the system automatically generates a hierarchy of how these contexts are organized. The generated contexts do not describe format or appearance such as heading or paragraph; the contexts created are descriptive names that summarize the content. The contexts provided for the electronic document are used to generate descriptive markup of an electronic document, key words, and indices. The system uses a unique combination of sentence and paragraph boundaries, document markup and linguistic information to generate the context and/or keyword. The contexts generated may be used to electronically provide start and end boundaries for the information. The preferred embodiment is the creation of an XML (eXtensible Markup Language) electronic document.

WO 2001/024053 A3

# INTERNATIONAL SEARCH REPORT

Intern:           Application No  
PCT/US 00/25755

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7   G06F17/30

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

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 23584 A (FRIEDMAN MARK M ;IOTA IND LTD (IL); STERN YONATAN PESACH (IL)) 14 May 1999 (1999-05-14) abstract page 1, line 1 -page 3, line 12 page 5, line 20 -page 9, line 22 page 15, line 5 -page 18, line 27 page 23, line 24 -page 32, line 7	1,8-12, 19-22, 25-28
Y	--- ---	2-7, 13-18, 23,24
X	EP 0 802 491 A (SUN MICROSYSTEMS INC) 22 October 1997 (1997-10-22) abstract column 1, line 5 -column 8, line 30 --- --- -/--	1,8-12, 19-22

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

Date of mailing of the international search report

20 January 2004

27/01/2004

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

Lievens, K

## INTERNATIONAL SEARCH REPORT

Intern:           pplication No  
PCT/US 00/25755

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98 26357 A (PRACTICAL APPROACH CORP ;REDFERN DARREN M (CA)) 18 June 1998 (1998-06-18) abstract page 2, line 14 -page 6, line 36 page 10, line 7-13	2-7, 13-18, 23,24
A	----- US 5 488 725 A (TURTLE HOWARD R ET AL) 30 January 1996 (1996-01-30)  abstract column 3, line 29 -column 5, line 37 column 10, line 62 -column 12, line 39 -----	4-6, 15-17, 25-27

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern:      Application No

PCT/US 00/25755

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 9923584	A	14-05-1999	US 6161107 A	12-12-2000
			AU 1371599 A	24-05-1999
			WO 9923584 A2	14-05-1999
EP 0802491	A	22-10-1997	US 6003046 A	14-12-1999
			EP 0802491 A2	22-10-1997
			JP 10116221 A	06-05-1998
WO 9826357	A	18-06-1998	US 6078914 A	20-06-2000
			AU 5471398 A	03-07-1998
			WO 9826357 A1	18-06-1998
US 5488725	A	30-01-1996	US 5265065 A	23-11-1993
			AU 6445094 A	24-10-1994
			WO 9423386 A2	13-10-1994
			AU 2801792 A	03-05-1993
			EP 0607340 A1	27-07-1994
			WO 9307577 A1	15-04-1993
			US 5418948 A	23-05-1995