

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 May 2007 (24.05.2007)

PCT

(10) International Publication Number
WO 2007/058959 A3

(51) International Patent Classification:
G06F 17/27 (2006.01)

Wayne F. [US/US]; 1613 Elmar Way, San Jose, CA 95129 (US). ZAUMEN, William T. [US/US]; 912 Clara Drive, Palo Alto, California 94303 (US).

(21) International Application Number:
PCT/US2006/043837

(74) Agent: VAUGHAN, Daniel; Park, Vaughan & Fleming LLP, 2820 Fifth Street, Davis, CA 95618 (US).

(22) International Filing Date:
8 November 2006 (08.11.2006)

(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, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/272,843 14 November 2005 (14.11.2005) US

(71) Applicant (for all designated States except US): SUN MICROSYSTEMS, INC. [US/US]; MS SCA12-203, 4150 Network Circle, Santa Clara, California 94054 (US).

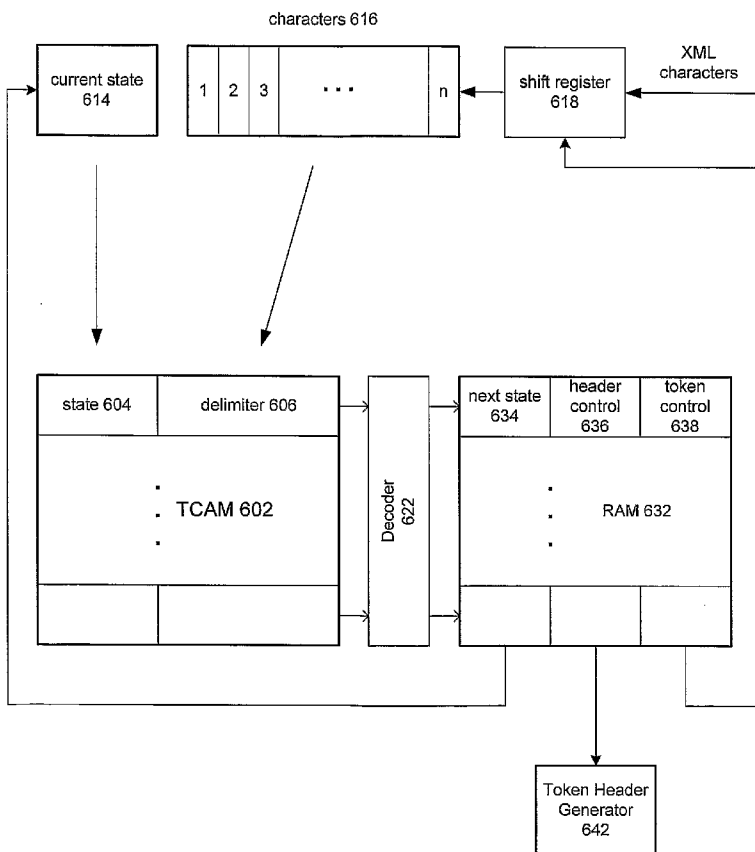
(72) Inventors; and

(75) Inventors/Applicants (for US only): DIGNUM, Marcelino M. [NL/US]; 518 Bay Road, Menlo Park, CA 94025 (US). BEHRENS, Jochen [DE/US]; 224 Arbor Avenue, Santa Cruz, California 95062 (US). SELTZER,

(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: HARDWARE UNIT FOR PARSING AN XML DOCUMENT



(57) Abstract: A hardware unit for parsing an XML document includes embedded logic or circuitry for accessing the document, decoding it to change a character set, validating individual characters of the document, extracting tokens, maintaining a symbol table and generating binary token headers to describe the document's structure and convey the document's data to an application. Tokenization, the process of identifying tokens and generating token headers, may be controlled by a finite state machine that recognizes XML delimiters in the document's markup and activates state transitions based on the current state and the recognized delimiter. The parser unit may be implemented within a hardware XML accelerator that includes a processor, a DMA engine, a cryptographic engine, memory (e.g., for storing a document, maintaining a symbol table) and various interfaces (e.g., network, memory, bus).

WO 2007/058959 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:

16 August 2007

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/043837

A. CLASSIFICATION OF SUBJECT MATTER
INV. 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)
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	US 2003/023633 A1 (ROSS RICHARD [US]) 30 January 2003 (2003-01-30) figure 1 figure 2 figure 6 paragraph [0027] paragraph [0050] paragraph [0077] paragraph [0079] paragraph [0081] paragraph [0100] paragraph [0114]	1-31

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
- *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.
- *Z* document member of the same patent family

Date of the actual completion of the international search

18 June 2007

Date of mailing of the international search report

28/06/2007

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

BOSCH VIVANCOS, P

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/043837

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005/240911 A1 (HUNDLEY DOUGLAS [US] ET AL) 27 October 2005 (2005-10-27) paragraph [0004] - paragraph [0007] paragraph [0010] - paragraph [0011] paragraph [0021] - paragraph [0022] paragraph [0032] paragraph [0042] paragraph [0052]	1-31
A	----- AHO A V ET AL: "Compiler - Principles, Techniques, and Tools" COMPILERS. PRINCIPLES, TECHNIQUES, AND TOOLS, READING, ADDISON-WESLEY PUBLISHING CO, US, 1986, page COMPLETE48, XP002200775 figure 2.29 page 12, line 6 - line 16 page 61, line 18 - page 62, line 15 -----	1-31
A	----- US 2004/083221 A1 (DAPP MICHAEL C [US] ET AL) 29 April 2004 (2004-04-29) abstract figure 3 paragraph [0010] -----	15,29

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2006/043837

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003023633 A1	30-01-2003	WO 03012659 A1	13-02-2003
US 2005240911 A1	27-10-2005	NONE	
US 2004083221 A1	29-04-2004	US 2007016554 A1	18-01-2007