

(19) World Intellectual Property Organization  
International Bureau



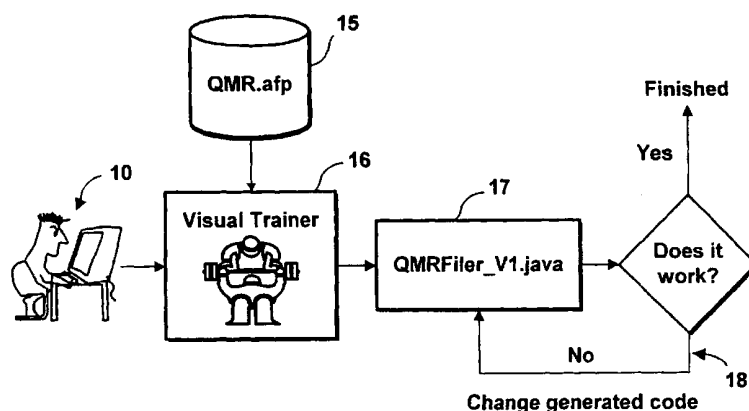
(43) International Publication Date  
17 August 2000 (17.08.2000)

PCT

(10) International Publication Number  
**WO 00/48088 A3**

- (51) International Patent Classification<sup>7</sup>: **G06F 17/60.** (74) Agent: **CHACLAS, Angelo, N.**; Pitney Bowes Inc., Intellectual Property Law Department, 35 Waterview Drive, P.O. Box 3000, Shelton, CT 06484-8000 (US).
- (21) International Application Number: PCT/US00/03734
- (22) International Filing Date: 11 February 2000 (11.02.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/119,777 11 February 1999 (11.02.1999) US
- (71) Applicant: **PITNEY BOWES INC.** [US/US]; 1 Elmcroft Road, P.O. Box 741, Stamford, CT 06926 (US).
- (72) Inventors: **CLARKE, William, D.**; 38 Longfellow Drive, Northampton, MA 01060 (US). **FLYNN, James, P.**; 138 Liberty Road, Tappan, NY 10983 (US).
- (81) Designated States (*national*): AU, CA, IL, JP, ZA.
- (84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
- Published:  
— with international search report
- (88) Date of publication of the international search report:  
6 December 2001
- 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: DATA PARSING SYSTEM FOR USE IN ELECTRONIC COMMERCE



**WO 00/48088 A3** (57) **Abstract:** A parsing technique suitable for use in electronic commerce that avoids the disadvantages of known monolithic parsers. The parsing system generates code specific to each input document and data stream type, which may then be updated as needed to handle document-specific idiosyncrasies without requiring modifications to the core parser code. As a user (10) defines parsing rules for extracting data from a representative sample document (15), a visual trainer (16) automatically generates code in the background, referred to as a filer (17), that is specific to that document and that embodies the rules for extracting data from that particular document. The generated code may be modified manually (18) as needed to account for any idiosyncratic conditions associated with the document. Each representative document has associated with it its own filer. A parsing engine (20, 21) comprises a collection of such individual filers appropriate for the types of documents that arise in any given organization. A mapping or other association is maintained between representative document types and their filers. In regular operation, a user selects a data set to be extracted from certain documents. When the parsing engine receives a document in an input data stream, the associated filer is loaded and parses that document for the selected data set. Then another filer is loaded in response to another input document, and so on. The filers may be especially efficiently generated using an object-oriented approach and then dynamically instantiated at run time as may be conveniently achieved, for example, in the Java programming language.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/03734

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G06F17/60 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)  
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, WPI Data, PAJ, INSPEC, IBM-TDB

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KLEIN E ET AL: "THE PARSER GENERATING SYSTEM PGS" SOFTWARE PRACTICE & EXPERIENCE, GB, JOHN WILEY & SONS LTD. CHICHESTER, vol. 19, no. 11, 1 November 1989 (1989-11-01), pages 1015-1028, XP000095040 ISSN: 0038-0644 abstract	1
A	--- KLEIN, B. ET AL.: "Error Tolerant Document Structure Analysis" PROC. IEEE INT. FORUM ON RESEARCH AND TECHNOLOGY ADVANCES IN DIGITAL LIBRARIES ADL '97, 7 - 9 May 1997, pages 116-127, XP002163487 abstract --- -/-	1

☒ 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.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

21 March 2001

Date of mailing of the international search report

04/04/2001

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

Schmidt, A

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/03734

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

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 631 244 A (XEROX CORP) 28 December 1994 (1994-12-28) abstract	1
A	US 5 675 801 A (LINDSEY ANTHONY HAYDEN) 7 October 1997 (1997-10-07) abstract	1

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 00/03734

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0631244	A	28-12-1994	US 5519608 A	21-05-1996
			JP 7056954 A	03-03-1995
US 5675801	A	07-10-1997	JP 8202545 A	09-08-1996