

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 August 2000 (24.08.2000)

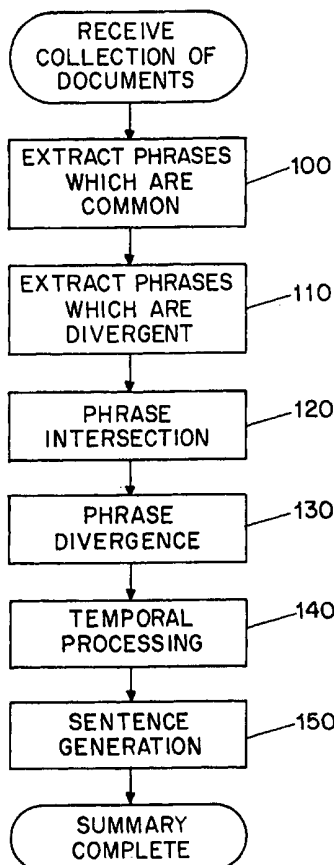
PCT

(10) International Publication Number
WO 00/49517 A3

- (51) International Patent Classification⁷: G06F 17/10, 17/27, 15/00
- (21) International Application Number: PCT/US00/04118
- (22) International Filing Date: 18 February 2000 (18.02.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/120,659 19 February 1999 (19.02.1999) US
- (71) Applicant (for all designated States except US): THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; 116th Street and Broadway, New York, NY 10027 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): MCKEOWN, Kathleen, R. [US/US]; 20 Prospect Road, Wayne, NJ 07470 (US). BARZILAY, Regina [US/US]; 548 Riverside Drive, Apt. 4B, New York, NY 10027 (US).
- (74) Agents: TANG, Henry et al.; Baker Botts, LLP, 30 Rockefeller Plaza, New York, NY 10112-0228 (US).
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, 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, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent

[Continued on next page]

(54) Title: MULTI-DOCUMENT SUMMARIZATION SYSTEM AND METHOD



(57) Abstract: A summary for a collection of related documents can be generated by extracting phrases (100) from the documents which include common focus elements. Phrase intersection analysis (120) is then performed on the extracted phrases (100) to generate a phrase intersection table, where identical or equivalent phrases are identified. Temporal processing (140) on the phrases in the phrase intersection table is performed to remove ambiguous time references and to sort the phrases in a temporal sequence. Sentence generation is then used to combine the phrases in the phrase intersection table into a coherent summary.



WO 00/49517 A3



(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).

(88) Date of publication of the international search report:
30 November 2000

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/US00/04118

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : GO6F 17/10, 17/27, 15/00 US CL : 704/1, 9, 10; 707/530, 531, 532 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. : 704/1, 9, 10; 707/530, 531, 532 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) Please See Extra Sheet.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A,P	US 5,924,108 A (FEIN ET AL) 13 JULY 1999, ABSTRACT.	1-22
A	US 5,638,543 A (PEDERSEN ET AL) 10 JUNE 1997.	1-22
Y	US 5,077,668 A (DOI) 31 DECEMBER 1991, ABSTRACT.	1, 8 AND 16
A	US 5,848,191 A (CHEN ET AL) 08 DECEMBER 1998	1-22
A	US 4,965,763 A (ZAMORA) 23 OCTOBER 1990.	1-22
Y	US 5,838,323 A (ROSE ET AL) 17 NOVEMBER 1998, ABSTRACT.	1, 8 AND 16
A	US 5,297,027 A (MORIMOTO ET AL) 22 MARCH 1994.	1-22
<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:	*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
A document defining the general state of the art which is not considered to be of particular relevance	*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
B earlier document published on or after the international filing date	*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
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)	*G*	document member of the same patent family
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		
Date of the actual completion of the international search 03 JULY 2000	Date of mailing of the international search report 11 AUG 2000	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Patrick N. Edouard <i>Joni Hill</i> Telephone No. (703) 308-6725	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/04118

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 5,689,716 A (CHEN ET AL) 18 NOVEMBER 1997, ABSTRACT.	1, 8 AND 16
A	US 5,778,397 A (KUPIEC ET AL) 07 JULY 1998.	1-22
A	US 5,384,703 A (WITHGOTT ET AL) 24 JANUARY 1995.	1-22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04118

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

WEST/EAST

search term: (summar\$ or abstract\$ or conden\$) same (document or text or information)