(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]:   G10L 15/00,
G06F 17/30, 7/00

(21) International Application Number: PCT/US01/45599

(22) International Filing Date:
15 November 2001 (15.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2000-353475   20 November 2000 (20.11.2000)   JP

(71) Applicant *(for all designated States except US)*:
HEWLETT-PACKARD COMPANY [US/US]; In-
tellectual Property Administration, P.O. Box 272400, Fort
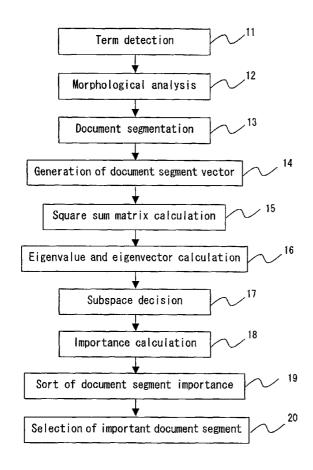Collins, CO 80527-2400 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*:   KAWATANI,
Takahiko   [JP/JP];   1950-21-3-515,   Matsuura-cho,
Kanazawa-ku, Yokohama, Kanagawa (JP).

(74) Agents: GREELEY, Paul, D. et al.; Ohlandt, Greeley,
Ruggiero & Perle, L.L.P., Hewlett-Packard Company, In-
tellectual Property Administration, P.O. Box 272400, Fort
Collins, CO 80527-2400 (US).

(81) Designated States *(national)*: CN, US.

(84) Designated States *(regional)*: European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE, TR).

Published:
— with international search report

[Continued on next page]

(54) Title: METHOD OF VECTOR ANALYSIS FOR A DOCUMENT

(57) Abstract: A method is provided for document represen-
tation and document analysis including extraction of impor-
tant sentences from a given document or determining sim-
ilarity between two documents. The method detects terms
(11) that occur in the document, segments the document into
document segments, each segment being an appropriate sized
chunk and generates document segment vectors (14) where
each vector includes as its element values according to oc-
currence frequencies of the terms occurring in the document
segments. The method calculates the eigenvalues and eigen-
vectors (16) of a square sum matrix (15) in which a rank of
the respective document segment vector is represented by R
and selects from the eigenvectors a plural (L) of eigenvectors
to be used to determine the importance (19). A weighted sum
of the squared projections of the respective selected eigenvec-
tors is calculated.

WO 02/041557 A3

INTERNATIONAL SEARCH REPORT

A.	CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G10L 15/00; G06F 17/30, 7/00
US CL : 702/2,6,911; 704/236, 243, 244, 245; 706/934

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. : 702/2,6,911; 704/236, 243, 244, 245; 706/934

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 Continuation Sheet

C.	DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,619,709 A (CAID et al.) 08 April 1997 (08.04.1997), Abstract, figures 1b-28, Background of the Invention, Summary of the Invention, col. 4, lines 45 et seq.	1-11
X	US 5,794,178 A (CAID et al.) 11 August 1998 (11.08.1998), Abstract, figures 1b-28, Background of the Invention, Summary of the Invention, col. 4, lines 42 et seq.	1-11
X	US 5,634,021 A (ROSENBERG et al.) 27 May 1997 (27.05.1997), Abstract, figures 1-10F, Background of the Invention, Summary of the Invention, col. 3, lines 38 et seq.	1-11
X	US 5,499,366 A (ROSENBERG et al.) 12 March 1996 (12.03.1996), Abstract, figures 1-10F, Background of the Invention, Summary of the Invention, col. 3, lines 44 et seq.	1-11
X	US 5,819,258 A (VAITHYANATHAN et al.) 06 October 1998(06.10.1998), Abstract, figures 1-6, Background of the Invention, Summary of the Invention, col. 4, lines 52 et seq.	1-11
X	US 5,857,179 A (VAITHYANATHAN et al.) 05 January 1999(05.01.1999), Abstract, figures 1-6G, Background of the Invention, Summary of the Invention, col. 4, lines 11 et seq.	1-11
X	US 5,901,244 A (SOUMA et al.) 04 May 1999 (04.05.1999), Abstract, figures 1-4, Background of the Invention, Summary of the Invention, col. 6, lines 46 et seq.	1-11
X	US 5,301,109 A (LANDAUER et al.) 05 April 1994 (05.04.1994), Abstract, figures 1-3, Background of the Invention, Summary of the Invention, col. 3, lines 20 et seq.	1-11

☒ Further documents are listed in the continuation of Box C.　　☐ 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		
"E"	earlier application or patent published on or after the international filing date	"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
"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)	"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
"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	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 June 2002 (12.06.2002)	09 JUL 2002
Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	William D. Thomson
Facsimile No. (703)305-3230	Telephone No. 703-305-3257

Form PCT/ISA/210 (second sheet) (July 1998)

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 6,070,133 A (BREWSTER et al.) 30 May 2000 (30.05.2000), Abstract, figures 1-4c, Background of the Invention, Summary of the Invention, Objects, col. 6, lines 17 et seq.	1-11

Continuation of B. FIELDS SEARCHED Item 3:
EAST
search terms: eigenvectors, eigenvalues, document$, vector$, extract$, rank, matrix, cluster$, chunk$