

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 December 2000 (14.12.2000)

PCT

(10) International Publication Number
WO 00/075811 A3

- (51) International Patent Classification⁷: G06F 17/30
- (21) International Application Number: PCT/US00/13653
- (22) International Filing Date: 17 May 2000 (17.05.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/328,888 9 June 1999 (09.06.1999) US
- (71) Applicant: THE BOEING COMPANY [US/US]; P.O. Box 3707, M/S 13-08, Seattle, WA 98124-2207 (US).
- (72) Inventors: BILLHEIMER, D., Dean; 2448 - 29th Avenue West, Seattle, WA 98199 (US). BOOKER, Andrew, James; 5726 Keystone Place North, Seattle, WA 98103 (US). CONDLIFF, Michelle, Keim; 10006 - 182nd Avenue S.E., Issaquah, WA 98027 (US). GREAVES, Mark, Thomas; 1434 - 31st Avenue, Seattle, WA 98122 (US). HOLT, Fredrick, Baden; 13244 N.E. 40th, Bellevue, WA 98005 (US). KAO, Anne, Shu-Wan; 17116 N.E. 28th Pl., Bellevue, WA 98008 (US). PIERCE, Daniel, John; 20305 N.E. 133rd Street, Woodinville, WA 98072 (US). POTEET, Stephen, Robert; 17116 N.E. 28th Pl., Bellevue, WA 98008 (US). WU, Yuan-Jye; 24636 S.E. 36th Ct., Issaquah, WA 98029 (US).
- (74) Agent: RICHARDSON, Robert, R.; The Boeing Company, P.O. Box 3707, M/S 13-08, Seattle, WA 98124-2207 (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, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, 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).
- 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:
18 December 2003
- 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.



WO 00/075811 A3

(54) Title: METHOD AND SYSTEM FOR TEXT MINING USING MULTIDIMENSIONAL SUBSPACES

(57) Abstract: A text mining program is provided that allows a user to perform text mining operations, such as: information retrieval, term and document visualization, term and document clustering, term and document classification, summarization of individual documents and groups of documents, and document cross-referencing. This is accomplished by representing the text of a document collection using subspace transformations. This subspace transformation representation is performed by: constructing a term frequency matrix of the term frequencies for each of the documents, transforming the term frequencies for statistical purposes, and projecting the documents or the terms into a lower dimensional subspace. As the document collection is updated, the subspace is dynamically updated to reflect the new document collection.

INTERNATIONAL SEARCH REPORT

International Appl. No. ...
PCT/US 00/13653

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	<p>KOLDA T G ET AL: "A SEMIDISCRETE MATRIX DECOMPOSITION FOR LATENT SEMANTIC INDEXING IN INFORMATION RETRIEVAL" ACM TRANSACTIONS ON INFORMATION SYSTEMS, ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, US, vol. 16, no. 4, October 1998 (1998-10), pages 322-346, XP000831494 ISSN: 1046-8188 abstract page 340 -page 344 page 324 -page 328 page 335</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1-33

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

29 September 2003

Date of mailing of the international search report

22/10/2003

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

Dumitrescu, C

INTERNATIONAL SEARCH REPORT

International App 1 No

PCT/US 00/13653

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>M.W.BERRY & R.D. FIERRO: "Low-Rank Orthogonal Decompositions for Information Retrieval Applications" NUMERICAL LINEAR ALGEBRA WITH APLICATIONS, JOHN WILEY & SONS, LTD, 'Online! vol. 1, no. 1, 10 January 1996 (1996-01-10), pages 1-27, XP002256078</p> <p>Retrieved from the Internet: <URL:http://citeseer.nj.nec.com/cache/papers/cs/143/http:zSzzSzwww.cs.utk.eduzSz{librarySzTechReportszSz1995zSzut-cs-95-284.pdf/berry96lowrank.pdf> 'retrieved on 2003-09-29! page 3 -page 7 page 12 -page 15 page 19 -page 24</p>	1-33
X	<p>US 5 778 362 A (DEERWESTER SCOTT) 7 July 1998 (1998-07-07) abstract column 2, line 17 - line 32 column 3, line 7 -column 5, line 5 column 5, line 63 -column 6, line 33 column 7, line 10 -column 9, line 5 figures 1-4</p>	1-33
A	<p>HETZLER B ET AL: "Multi-faceted insight through interoperable visual information analysis paradigms" INFORMATION VISUALIZATION, 1998. PROCEEDINGS. IEEE SYMPOSIUM ON RESEARCH TRIANGLE, CA, USA 19-20 OCT. 1998, LOS ALMAITOS, CA, USA, IEEE COMPUT. SOC, US, 19 October 1998 (1998-10-19), pages 137-144, 161, XP010313316 ISBN: 0-8186-9093-3 abstract page 138 -page 140</p>	1-33

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 00/13653

Patent document cited in search report	A	Publication date		Patent family member(s)	Publication date
US 5778362	A	07-07-1998	AU	3044497 A	07-01-1998
			EP	0978059 A1	09-02-2000
			WO	9749046 A1	24-12-1997
			ZA	9705477 A	22-12-1997