

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
17 December 2009 (17.12.2009)

(10) International Publication Number  
**WO 2009/151861 A3**

- (51) **International Patent Classification:**  
*G06F 17/30* (2006.01)      *G06F 17/21* (2006.01)
- (21) **International Application Number:**  
PCT/US2009/043816
- (22) **International Filing Date:**  
13 May 2009 (13.05.2009)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
12/139,213      13 June 2008 (13.06.2008)      US
- (71) **Applicant (for all designated States except US):** MICROSOFT CORPORATION [US/US]; Attn: Tuan Ngo, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US).
- (72) **Inventors:** KHASSANOV, Raif; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US). MERRIGAN, Chadd, Creighton; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US). PETRIUC, Mi-

hai; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US). **KOKHAN, Artsiom Ivanovich;** c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US).

- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, 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, LY, MA, MD, ME, 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, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (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, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

[Continued on next page]

(54) **Title:** SEARCH INDEX FORMAT OPTIMIZATIONS

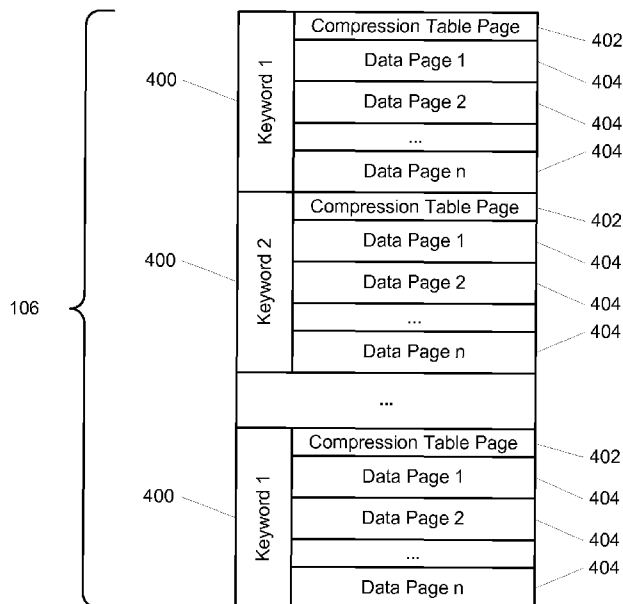


FIG. 4

(57) **Abstract:** A search index structure which extends a typical composite index by incorporating an index which is optimized for fast retrieval from storage and which eliminates data which is specific to phrase searching. Other data is represented in a manner which allows it to be calculated rather than stored. Associating variable length entries with logical categories allows their length to be inferred from the category rather than stored. Using delta values between document IDs rather than the ID itself generates a compact, dense symbol set which is efficiently compressed by Huffman encoding or a similar compression method. Using an upper threshold to remove large, and thus rare, delta values from the symbol set prior to encoding further improves the encoding performance.

WO 2009/151861 A3



MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR),  
OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,  
MR, NE, SN, TD, TG).

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

**Declarations under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**(88) Date of publication of the international search report:**

18 February 2010

**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/30(2006.01)i, G06F 17/21(2006.01)i**

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 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models since 1975

Japanese Utility models and applications for Utility models since 1975

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

KOMPASS (The internal search system of Korean Intellectual Property Office)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7028045 B2 (FRANZ, M. et al.) 11 APRIL 2006 See abstract, figures 5-9, column 2 line 60 - column 3 line 67, column 6 lines 15-57, column 7 lines 3-8, claims 1-10.	1, 10, 19
A	US 6055526 A (AMBROZIAK J.) 25 APRIL 2000 See abstract, figures 3, 4, 7, 8, 11, column 1 line 52 - column 2 line 10, column 7 line 49 - column 8 line 14, column 11 line 66 - column 12 line 10, column 15 lines 6-25, claims 1-3, 5, 8-13, 20, 21.	1, 10, 19
A	US 5649183 A (BERKOWITZ, B. T. et al.) 15 JULY 1997 See abstract, figures 1-3, column 1 line 45 - column 2 line 41, column 3 lines 1-20, column 4 lines 19-56, claim 1.	1, 10, 19

 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 application or patent 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 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

29 DECEMBER 2009 (29.12.2009)

Date of mailing of the international search report

**30 DECEMBER 2009 (30.12.2009)**

Name and mailing address of the ISA/KR

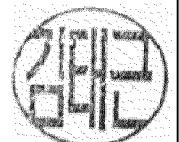
Korean Intellectual Property Office  
Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Tae-Geun KIM

Telephone No. 82-42-481-8120



**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/US2009/043816**

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 07028045 B2	11.04.2006	US 2003-144995 A1	31.07.2003
US 06055526 A	25.04.2000	AU 1999-34596 A1	31.03.1999
		US 06460047 B1	01.10.2002
		WO 1999-052046 A2	14.10.1999
US 05649183 A	15.07.1997	EP 0601569 A1	15.06.1994
		EP 0601569 B1	24.10.2001
		JP 06-243009 A	02.09.1994
		JP 03-550173 B2	04.08.2004
		US 05832479 A	03.11.1998