

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 December 2007 (27.12.2007)

PCT

(10) International Publication Number
WO 2007/149623 A3

(51) International Patent Classification:
G06F 7/00 (2006.01) *G06F 17/30* (2006.01)

(21) International Application Number:
PCT/US2007/067439

(22) International Filing Date: 25 April 2007 (25.04.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/745,604 25 April 2006 (25.04.2006) US
60/745,605 25 April 2006 (25.04.2006) US

(71) Applicant (for all designated States except US): **IN-FOVELL, INC.** [US/US]; 4600 Bohannon Drive, Suite 220, Menlo Park, CA 94025 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANG, Yuanhua, Tom** [US/US]; 1147 Huntingdon Drive, San Jose, CA 95129 (US). **HU, Qianjin** [US/US]; 20499 Crow Creek Drive, Castro Valley, CA 94552 (US). **YANG, Yonghong, Grace** [US/US]; 1147 Huntingdon Drive, San Jose, CA 95129 (US). **CHEN, Chunnuan** [CN/US]; 1089 W. Olive Avenue, Apt. 4, Sunnyvale, CA 94086 (US). **MEI, Minghua** [CN/CN]; Longjiangqiao Village, Zhenze Town, Wujiang City, Suzhou, Jiangsu (CN).

(74) Agents: **WOLFELD, Warren, S.** et al.; Haynes Beffel & Wolfeld LLP, P.O. Box 366, Half Moon Bay, CA 94019 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, 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

[Continued on next page]

(54) Title: FULL TEXT QUERY AND SEARCH SYSTEMS AND METHOD OF USE

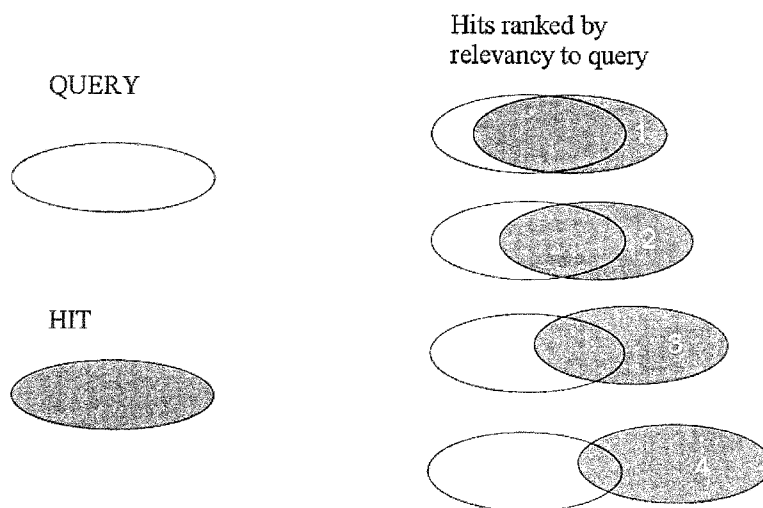


Figure 1

(57) Abstract: Roughly described, a database searching method for searching a database, in which hits are ranked in dependence upon an information measure of items shared by both the hit and the query. The information measure can be a Shannon information score, or another measure which indicates the information value of the shared items. An item can be a word or other token, or a multi-word phrase, and can overlap with each other. Synonyms can be substituted for items in the query, with the information measure of substituted items being derated in accordance with a predetermined measure of the synonyms' similarity. Indirect searching methods are described in which hit from other search engines are re-ranked in dependence upon the information measures of shared items. Structured and completely unstructured databases may be searched, with hits being demarcated dynamically. Hits may be clustered based upon distances in an information- measure- weighted distance space.

WO 2007/149623 A3



(88) Date of publication of the international search report:
12 February 2009

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 07/67439

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 7/00 (2008.04), G06F 17/30 (2008.04)
USPC - 707/3

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 (8) -G06F 7/00 (2008.04)
USPC - 707/3Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 707/1, 4, 7 (See Keywords Below)

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

Pub WEST (USPT, PGPB, JPAB, EPAB), Google Scholar

Search Terms Used: search engine, text, document, database, sub-database, frequency, cumulative information score, shannon score, boolean search, preliminary search, external engine, unstructured text, gap word, term, keyword, result, hit, rank results

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2002/0111941 A1 (ROUX et al.), 15 August 2002 (15.08.2002), entire document, especially para [0019]-[0021]; [0043]-[0044]; [0049]-[0051]; [0053]-[0057] and [0088]-[0091].	1, 12-27, 38, 40-50, 56-58, 67-68, 99-100, 114-116, 120, 123-124 and 127 2-11, 28-37, 39, 51-55, 59-66, 69, 85-96, 118 and 125
Y	US 2004/0024583 A1 (FREEMAN), 05 February 2004 (05.02.2004), entire document, especially para [0127]-[0130].	2
Y	US 5,761,497 A1 (HOLT et al.), 02 June 1998 (02.06.1998), entire document, especially col 5, ln 47 to col 6, ln 60; col 7, ln 40 col 8, ln 44; col 8, ln 40 to col 9, ln 31; col 9, ln 20-64.	3-11
Y	US 2006/0026147 A1 (CONE et al.), 02 February 2006 (02.02.2006), entire document, especially para [0023]-[0026]; [0044]-[0046]; [0076]-[0079]; [0133]-[0135]; [0153]-[0155] and [0194]-[0197].	28-37, 39, and 69
Y	US 5,812,998 A (TSUTSUMI et al.), 22 September 1998 (22.09.1998), entire document, especially Abstract; col 2, ln 25 to col 3, ln 55; col 10, ln 3-57; col 14, ln 16-65; col 15, ln 25 to col 16, ln 13	51-55
Y	US 6,633,817 B1 (WALKER et al.), 14 October 2003 (14.10.2003), entire document, especially col 1, ln 62 to col 2, ln 40; col 5, ln 8-40; col 5, ln 8-40; col 6, ln 37 to col 7, ln 21.	59-66, 85-96, 118 and 125

 Further documents are listed in the continuation of Box C.

* 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 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

12 November 2008 (12.11.2008)

Date of mailing of the international search report

24 NOV 2008

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 07/67439

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

---See Additiona Sheet---

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-69, 85-96, 99-100, 114-116, 118, 120, 123-125, 127

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/US 07/67439

Box No. III

Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-69, 85-96, 99-100, 114-116, 118, 120, 123-125, 127 drawn to a method for searching a database, comprising the steps of: initiating a first search on the database in response to a first query, the first search generating a plurality of hits; ranking the hits in dependence upon an information measure of items shared by both the hit and the first query.

Group II, claims 70-84, 117, drawn to a method for preparing a first database for searching, comprising the steps of: developing a list of items present in the first database, at least one of the items containing more than one token, and at least one of the items constituting less than a full cell in the first database; associating with each of the items in the list an information measure of the respective item; and associating with each of the items in the list an indication of the location of each occurrence of the respective item in the first database.

Group III, claims 97-98, 119, 126, drawn to a method for searching a database, comprising the steps of: developing a synonym list for a first item in a first query, each synonym in the synonym list having an associated similarity score; initiating a search on the database in response to the first query, allowing for substitutions of synonyms from the synonym list for the first item in the first query, the first search generating a plurality of hits; ranking the hits in dependence upon the similarity score of any synonym substituted for the first item in the step of initiating a search.

Group IV, claims 101-107, 121, 128, drawn to a method for searching a database, comprising the steps of: developing a plurality of preliminary queries in dependence upon a provided first query, at least two of the preliminary queries differing from each other; forwarding the preliminary queries to a set of at least one external search engine, each combination of a preliminary search query and an external search engine yielding a respective set of preliminary hits.

Group V, claims 108-113, 122, 129, drawn to a method for organizing segments in a subject database for presentation to a user, comprising the steps of: clustering the segments in dependence upon an information measure of items in each of the segments; and identifying the segments to the user in clusters determined in the step of clustering.

The inventions listed as Groups I-V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The special technical feature of the Group I invention is a method for searching a database, comprising the steps of: initiating a first search on the database in response to a first query, the first search generating a plurality of hits; ranking the hits in dependence upon an information measure of items shared by both the hit and the first query. The special technical feature of the Group II invention is a method for preparing a first database for searching, comprising the steps of: developing a list of items present in the first database, at least one of the items containing more than one token, and at least one of the items constituting less than a full cell in the first database; associating with each of the items in the list an information measure of the respective item; and associating with each of the items in the list an indication of the location of each occurrence of the respective item in the first database. The special technical feature of the Group III invention is a method for searching a database, comprising the steps of: developing a synonym list for a first item in a first query, each synonym in the synonym list having an associated similarity score; initiating a search on the database in response to the first query, allowing for substitutions of synonyms from the synonym list for the first item in the first query, the first search generating a plurality of hits; ranking the hits in dependence upon the similarity score of any synonym substituted for the first item in the step of initiating a search. The special technical feature of Group IV is a method for searching a database, comprising the steps of: developing a plurality of preliminary queries in dependence upon a provided first query, at least two of the preliminary queries differing from each other; forwarding the preliminary queries to a set of at least one external search engine, each combination of a preliminary search query and an external search engine yielding a respective set of preliminary hits. The special technical feature of Group V is a method for organizing segments in a subject database for presentation to a user, comprising the steps of: clustering the segments in dependence upon an information measure of items in each of the segments; and identifying the segments to the user in clusters determined in the step of clustering.

None of these special technical features are common to the other groups, nor do they correspond to a special technical feature in the other groups. Therefore, unity of invention is lacking.