

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
14 April 2005 (14.04.2005)

PCT

(10) International Publication Number  
**WO 2005/033967 A3**

(51) International Patent Classification<sup>7</sup>: **G06F 17/27**

(21) International Application Number:  
PCT/US2004/029772

(22) International Filing Date:  
13 September 2004 (13.09.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
10/676,724 30 September 2003 (30.09.2003) US

(71) Applicant (for all designated States except US):  
**GOOGLE INC.** [US/US]; 1600 Amphitheatre Park-  
way, Mountain View, CA 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MITTAL, Vibhu**

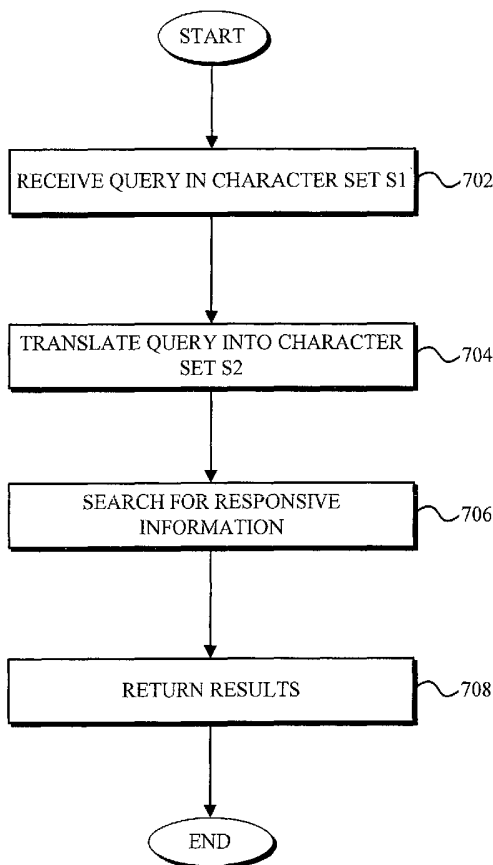
[US/US]; 1327 Elsona Drive, Sunnyvale, CA 94087 (US).  
**PONTE, Jay, M.** [US/US]; 2439 Marjorie Court, Moun-  
tain View, CA 94043 (US). **SAHAMI, Mehran** [US/US];  
3238 Hoover Street, Redwood City, CA 94063 (US). **GHE-**  
**MAWAT, Sanjay** [US/US]; 111 North Renfstorff Avenue,  
#184, Mountain View, CA 94043 (US). **BAUER, John, A.**  
[US/US]; 415 Del Medio Avenue, #8, Mountain View, CA  
94040 (US).

(74) Agent: **KUO, Jung-Hua**; P.O. Box 3275, Los Altos, CA  
94024 (US).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, 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, 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, MZ, NA, NI, NO, NZ, OM, PG,  
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,

[Continued on next page]

(54) Title: SYSTEMS AND METHODS FOR SEARCHING USING QUERIES WRITTEN IN A DIFFERENT CHARACTER-SET AND/OR LANGUAGE FROM THE TARGET PAGES



(57) Abstract: Methods and apparatus consistent with the inven-  
tion allow a user to submit an ambiguous search query and to re-  
ceive relevant search results. Queries can be expressed using char-  
acter sets and/or languages that are different from the character set  
and/or language of at least some of the data that is to be searched.  
A translation between these character sets and/or languages can be  
performed by examining the use of terms in aligned text. Probabil-  
ities can be associated with each possible translation. Refinements  
can be made to these probabilities by examining user interactions  
with the search results.

WO 2005/033967 A3



TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, IT, LU, MC, 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*
- *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:**

13 October 2005

*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/US2004/029772

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G06F17/27

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, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|----------|--|-----------------------|
| A        | EP 0 597 611 A (CANON RESEARCH CENTRE EUROPE LIMITED; CANON EUROPA N.V; CANON RESEARCH) 18 May 1994 (1994-05-18) abstract  | 1,30                  |
| A        | -----<br>RESNIK P ET AL: "The Web as a parallel corpus"<br>COMPUTATIONAL LINGUISTICS MIT PRESS FOR ASSOC. COMPUT. LINGUISTICS USA, vol. 29, no. 3, July 2002 (2002-07), pages 1-29, XP002327039<br>ISSN: 0891-2017<br>paragraph [2.1.1] - paragraph [2.1.2]<br>----- | 1,30                  |

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

19 May 2005

Date of mailing of the international search report

17. 08. 2005

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

Triest, J

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2004/029772

## Box 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 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 additional 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 an additional fee, this Authority did not invite payment of any additional fee.
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.:

1-13, 30-33

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-13,30-33

method and computer program product for aligning text using  
hyperlinks

---

2. claims: 14-21, 39-43, 44-45

method for performing a search based on a translated query

---

3. claims: 22-29,34-38

method for creating a probablistic dictionary using aligned  
data

---

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US2004/029772

| Patent document cited in search report |   | Publication date | Patent family member(s) | Publication date |
|--|---|------------------|-------------------------|------------------|
| EP 0597611                             | A | 18-05-1994       | GB 2272091 A            | 04-05-1994       |
|  |   |                  | DE 69322741 D1          | 04-02-1999       |
|  |   |                  | DE 69322741 T2          | 02-06-1999       |
|  |   |                  | EP 0597611 A2           | 18-05-1994       |
|  |   |                  | ES 2128395 T3           | 16-05-1999       |
|  |   |                  | US 5893134 A            | 06-04-1999       |
| -----                                  |   |                  |                         |                  |