



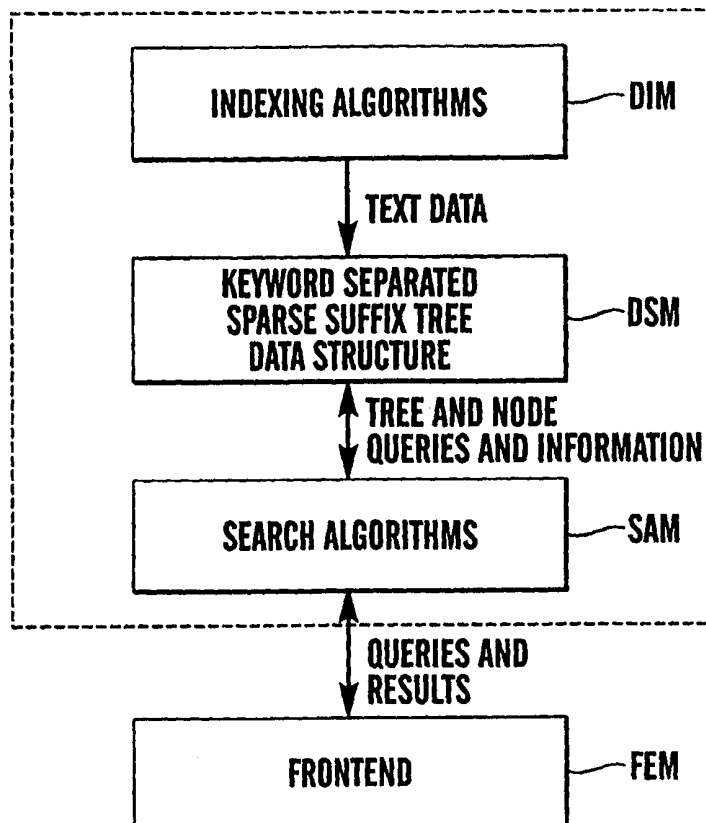
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|--|--|---|---|
| (51) International Patent Classification ⁷ : G06F 17/30 | | A3 | (11) International Publication Number: WO 00/03315 |
| | | | (43) International Publication Date: 20 January 2000 (20.01.00) |
| (21) International Application Number: PCT/NO99/00233 (22) International Filing Date: 9 July 1999 (09.07.99) (30) Priority Data: 19983175 10 July 1998 (10.07.98) NO (71) Applicant (for all designated States except US): FAST SEARCH & TRANSFER ASA [NO/NO]; P.O. Box 1677 Vika, N-0120 Oslo (NO). (72) Inventor; and (75) Inventor/Applicant (for US only): RISVIK, Knut, Magne [NO/NO]; Nygaten 11, N-7014 Trondheim (NO). (74) Agent: GEIRR, I., Leistad; Opticom ASA, P.O. Box 1872 Vika, N-0124 Oslo (NO). | | (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, 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 <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims</i> <i>and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 24 February 2000 (24.02.00) | |

(54) Title: A SEARCH SYSTEM AND METHOD FOR RETRIEVAL OF DATA, AND THE USE THEREOF IN A SEARCH ENGINE

(57) Abstract

A search system for information retrieval comprises a data structure for storing a text T, a combined metric M which includes an edit distance metric for approximate degree of matching between words and/or symbols or sequences thereof in the text T and words and/or symbols in a sequence P, weighting cost functions for edit operations which transform a sequence S of words or symbols into the sequence P, and a search algorithm for determining the degree of matching between words or word sequences in a suffix tree representation of respectively the text T and a query Q. The algorithm searches the data structure with the query Q, retrieving information with specified match to the query. A method in a search system for information retrieval generates a word-spaced sparse suffix tree for storing suffixes of words in a text T as word sequence information, and a word size-dependent edit distance metric for word sequences S, P and including word-weighted cost functions for edit operations, and determines matches between word sequences S_R or retrieved information R and word sequences P_Q of a query Q by calculating the edit distance for all matches. Use in an approximate search engine.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

INTERNATIONAL SEARCH REPORT

International application No.

PCT/NO 99/00233

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: G06F 17/30

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)

IPC7: G06F

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

SE,DK,FI,NO classes as above

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|-----------------------|
| A | WO 9215954 A1 (RED BRICK SYSTEM), 17 Sept 1992 (17.09.92) -- | 1-6 |
| A | US 5627748 A (BRENDA S. BAKER ET AL), 6 May 1997 (06.05.97) -- | 1-6 |
| A | US 5511159 A (BRENDA S. BAKER ET AL), 23 April 1996 (23.04.96) -- | 1-6 |
| A | WO 9600945 A1 (INTERNATIONAL BUSINESS MACHINES CORP.), 11 January 1996 (11.01.96) -- ----- | 1-6 |

☐ Further documents are listed in the continuation of Box C.
 ☒ See patent family annex.

| | |
|---|--|
| * Special categories of cited documents: | "I" 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 | "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 |
| "E" earlier document but published on or after the international filing date | "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 |
| "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) | "&" document member of the same patent family |
| "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 | |

| | |
|--|--|
| Date of the actual completion of the international search | Date of mailing of the international search report |
| 21 December 1999 | 2000 -01- 03 |
| Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86 | Authorized officer Björn Edlund/mj Telephone No. +46 8 782 25 00 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

02/12/99

International application No.

PCT/NO 99/00233

| Patent document cited in search report | | | Publication date | Patent family member(s) | | Publication date |
|---|---------|----|---------------------|----------------------------|-------------|---------------------|
| WO | 9215954 | A1 | 17/09/92 | AU | 1640992 A | 06/10/92 |
| ----- | | | | | | |
| US | 5627748 | A | 06/05/97 | US | 5511159 A | 23/04/96 |
| | | | | CA | 2091503 A,C | 19/09/93 |
| | | | | EP | 0561563 A | 22/09/93 |
| | | | | JP | 6043838 A | 18/02/94 |
| | | | | US | 5953006 A | 14/09/99 |
| ----- | | | | | | |
| US | 5511159 | A | 23/04/96 | US | 5627748 A | 06/05/97 |
| | | | | CA | 2091503 A,C | 19/09/93 |
| | | | | EP | 0561563 A | 22/09/93 |
| | | | | JP | 6043838 A | 18/02/94 |
| | | | | US | 5953006 A | 14/09/99 |
| ----- | | | | | | |
| WO | 9600945 | A1 | 11/01/96 | EP | 0804769 A | 05/11/97 |
| | | | | US | 5787430 A | 28/07/98 |
| ----- | | | | | | |