20000935-1 IPOS.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F — A2

(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:
1) 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

Without international search report and to be republished upon receipt of that report.

This Sheet Has Been Replaced

1 PD. IPC: G06F 17/30
1 appl^n. IPER
1 Inv. 2/2/2000.

(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.

