PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G06F 17/30, 17/27

A1

(11) International Publication Number: WO 96/32686

(43) International Publication Date: 17 October 1996 (17.10.96)

(21) International Application Number: PCT/US96/04945

(22) International Filing Date: 10 April 1996 (10.04.96)

(30) Priority Data:
08/419,126 10 April 1995 (10.04.95) US

(71) Applicant: REBUS TECHNOLOGY, INC. [US/US]; c/o Mind-Works Corporation, 349 Cobalt Way, Sunnyvale, CA 94086 (US).

(72) Inventors: RANGARAJAN, Vijayakumar; c/o Parthasarathy, Aiyaswamy, 5509 Don Pizarro Court, San Jose, CA 95123 (US). RAVICHANDRAN, Natarajan; Flat 15, Site 51, Goodwill Apartment, III Main Road, Attiguppe, Vijayanagar, Bangalore-560040 (IN).

(74) Agents: SUEOKA, Greg, T. et al.; Fenwick & West L.L.P, Suite 600, Two Palo Alto Square, Palo Alto, CA 94306 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
With amended claims.

SG 96122TT-5.

(54) Title: SYSTEM AND METHOD FOR PORTABLE DOCUMENT INDEXING USING N-GRAM WORD DECOMPOSITION

(57) Abstract

A system and method provides for indexing and retrieval of stored documents using a decomposition of words in the documents in n-grams, or linear word subunits. The documents are indexed as pages in a number of banks. For each bank there is a bank index. The individual n-grams are identified for each page and are stored in the bank index. Each bank index further contains an entry map that indicates whether a given n-gram is present in any of the pages of the bank, and then provides an index to a page map that further indicates which page in the bank contains the n-gram. When a search query is input, the query words are decomposed into their n-grams. The query word n-grams are compared first with entry maps to determine if the query word n-grams appear on any page in the bank. If so, the associated page map is traversed to determine which page in the bank contains the query word n-grams. The n-grams on the page are compared with the query word n-grams to determine the presence of a match therebetween. Matching pages are flagged. When all pages in all blanks have been processed, the pages are consolidated with respect to the documents to which they belong, resulting in a list of documents that match the search query. The results are displayed to a user.

