PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/30 — A1

(11) International Publication Number: WO 99/12108

(43) International Publication Date: 11 March 1999 (11.03.99)

(21) International Application Number: PCT/GB98/02611

(22) International Filing Date: 28 August 1998 (28.08.98)

(30) Priority Data:
97306878.6 4 September 1997 (04.09.97) EP

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and

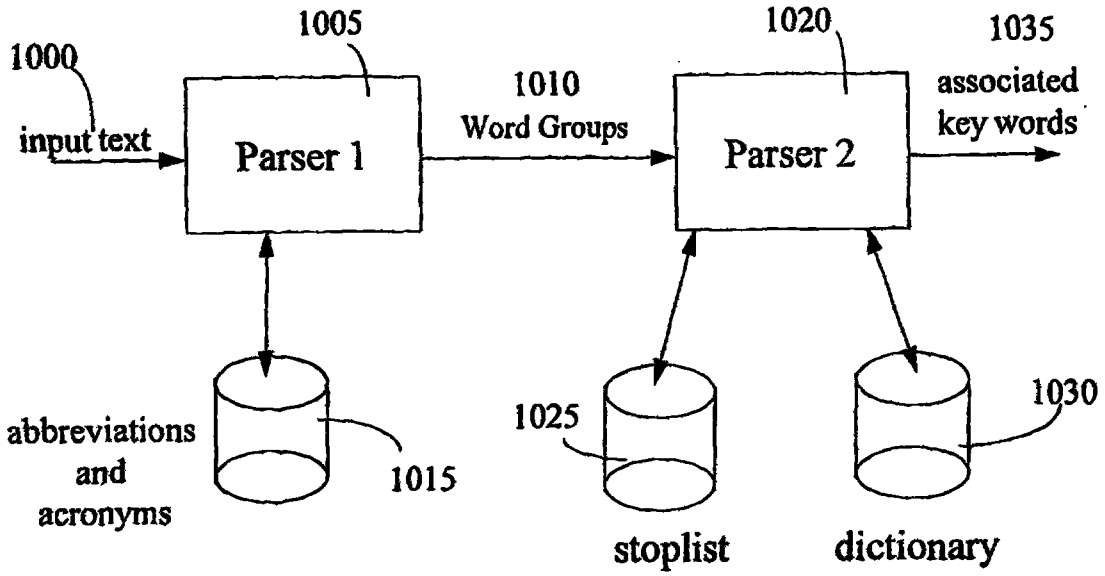
(75) Inventors/Applicants *(for US only)*: DAVIES, Nicholas, John [GB/GB]; 10 Spindle Wood, Colchester, Essex CO4 4SX (GB). WEEKS, Richard [GB/GB]; 44 Glemsford Close, Felixstowe, Suffolk IP11 8UG (GB).

(74) Agent: EVERSHED, Michael; BT Group Legal Services, Intellectual Property Dept., 8th floor, Holborn Centre, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, 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, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, 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
With international search report.

(54) Title: METHODS AND/OR SYSTEMS FOR SELECTING DATA SETS



(57) Abstract

Methods and apparatus for identifying associated key words (1035) in a data set (1000). Associated key words are identified by a parser (1020) which firstly operates to extract key words from a data set (1000). These key words are then analysed by the parser (1020) to identify which key words, if any, have an association as determined by a predefined set of rules. These rules are grammatical and include, for example, two keywords both being nouns that occur one after the other without intervening low value words. A similar rule applies to nouns followed by verbs but does not extend to verbs followed by nouns. These rules allow terms and phrases such as "information technology" and "wide area network" to be identified as associated key words (1035) rather than as individual and unrelated key words.