

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14992 A1

(51) International Patent Classification⁷: G06F 15/80,
G06K 9/66

[SG/SG]; Blk 892A, Tampines Avenue 8, #13-30, Singapore 521892 (SG). LAI, Fon, Lin [SG/MY]; 101-G Jervois Road, Singapore 249058 (SG).

(21) International Application Number: PCT/SG99/00089

(74) Agent: GREENE-KELLY, James, Patrick; Lloyd Wise, Tanjong Pagar, P.O. Box 636, Singapore 910816 (SG).

(22) International Filing Date: 25 August 1999 (25.08.1999)

(25) Filing Language: English

(81) Designated States (*national*): SG, US.

(26) Publication Language: English

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

(71) Applicant (*for all designated States except US*): KENT RIDGE DIGITAL LABS [SG/SG]; 21 Heng Mui Keng Terrace, Singapore 119613 (SG).

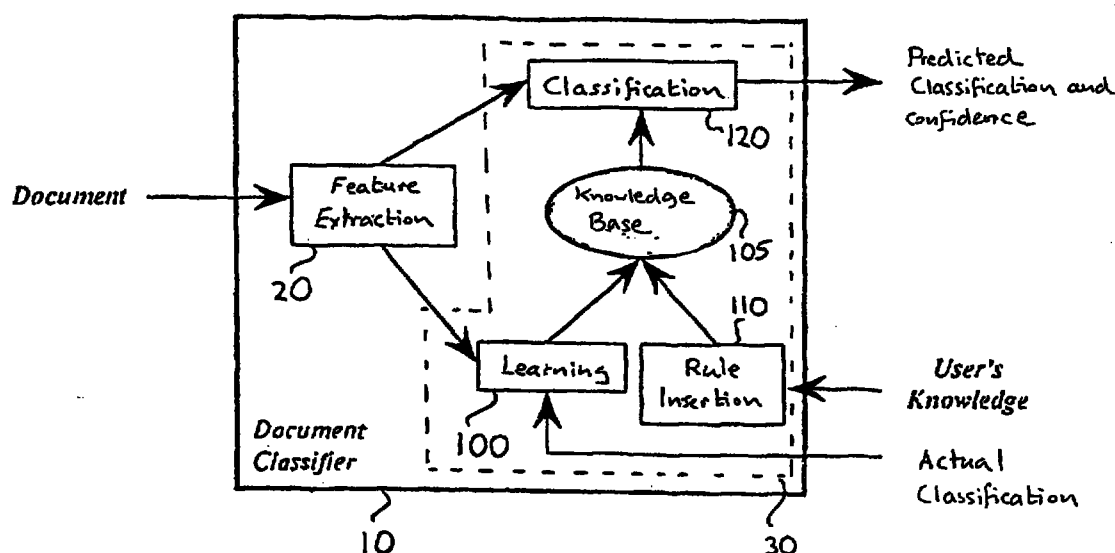
Published:
— With international search report.

(72) Inventors; and

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.

(75) Inventors/Applicants (*for US only*): TAN, Ah, Hwee

(54) Title: DOCUMENT CLASSIFICATION APPARATUS



(57) Abstract: Document classification apparatus is disclosed comprising feature extraction means (20) for extracting a plurality of features from a document and a classifier (30) operable to process the document in a knowledge acquisition mode in which information for classification is acquired from the features of the document and added incrementally to a knowledge base (105) and in a document classification mode in which the classifier (30), using the knowledge base, determines from the features a predicted classification for the document, the classifier (20) being switchable between the modes under user control. The apparatus is further operable to perform rule insertion in the knowledge acquisition mode in which a number of features are input by a user to the classifier together with a classification with which the features are associated.