Title (en)

SEGMENTATION OF CHINESE TEXT INTO WORDS

Title (de)

SEGMENTIERUNG CHINESISCHER TEXT IN WÖRTERN

Title (fr)

SEGMENTATION DE MOTS DANS UN TEXTE CHINOIS

Publication

EP 1055182 A2 20001129 (EN)

Application

EP 99902779 A 19990113

Priority

- IB 9900320 W 19990113
- US 2358698 A 19980213

Abstract (en)

[origin: WO9941680A2] The present invention provides a facility for selecting from a sequence of natural language characters combinations of characters that may be words. The facility uses indications, for each of a plurality of characters, of (a) the characters that occur in the second position of words that begin with the character and (b) the positions in which the character occurs in words. For each of a plurality of contiguous combinations of characters occurring in the sequence, the facility determines whether the character occurring in the second position of the combination is indicated to occur in words that begin with the character occurring in the first position of the combination. If so, the facility determines whether every character of the combination is indicated to occur in words in a position in which it occurs in the combination. If so, the facility determines that the combination of characters may be a word. In some embodiments, the facility proceeds to compare the combination of characters to a list of valid words to determine whether the combination of characters is a word.

IPC 1-7

G06F 17/28

IPC 8 full level

G06F 17/27 (2006.01); **G06F 17/28** (2006.01)

CPC (source: EP)

G06F 40/211 (2020.01); **G06F 40/216** (2020.01); **G06F 40/268** (2020.01); **G06F 40/284** (2020.01); **G06F 40/289** (2020.01); **G06F 40/53** (2020.01)

Citation (search report)

See references of WO 9941680A2

Designated contracting state (EPC)

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

DOCDB simple family (publication)

WO 9941680 A2 19990819; **WO 9941680 A3 19991125**; CN 1114165 C 20030709; CN 1290371 A 20010404; EP 1055182 A2 20001129; JP 2002503849 A 20020205; JP 2010157260 A 20100715; JP 4573432 B2 20101104; JP 5100770 B2 20121219

DOCDB simple family (application)

IB 9900320 W 19990113; CN 99802944 A 19990113; EP 99902779 A 19990113; JP 2000531795 A 19990113; JP 2010037953 A 20100223