

[NAME OF DOCUMENT] ABSTRACT

Provided is an orthographical variant detection apparatus which detects orthographical variant candidates with a high precision. The orthographical variant detection apparatus includes a term extraction unit that extracts terms from document data, a similarity computation unit that computes similarity of an arbitrary pair of the extracted terms, an orthographical variant candidate determination unit that determines, based on the similarity, whether or not the terms in the pair of terms are orthographical variant candidates, and a group classification unit that groups the orthographical variant candidates based on a character string commonly included in pair of terms as the orthographical variant candidates.



162162

G00002

