Abstract of the Disclosure

METHOD AND APPARATUS FOR PROCESSING CHINESE TELETEXT

A method and apparatus for processing a Chinese teletext, which can automatically extract and display several key phrases from a teletext by using computer technology in order to aid users to rapidly capture the essential meaning of the teletext. In processing a mass of Chinese teletext, the apparatus can enable users to rapidly capture the content of the teletext to expedite the processing of these useful information. In general, phrases containing two or more Chinese characters are often used in a Chinese text. According to statistics, about 88% in a Chinese text are phrases. Among these phrases, those with higher appearing frequencies in a text are highly related to its content. Whenever a phrase comprises more Chinese characters, it will capture more essential meaning of the text and aggrandize more the theme of the text. Most of the other 12% characters are conjunctions. If these characters are used as separation reference for dividing sentences, the text can be divided into respective character strings. These characters are called the first reference character, hereafter. There's another category of characters, which sometimes exists separately in a text and sometimes form a phrase with adjacent characters. In the case where few characters in such category form a phrase, the characters are considered as conjunctions rather than key phrases. This category of characters is referred to the second reference characters. This invention extracts key phrases based on the phrases, their frequencies and the length of the phrase.

Fig. 1