

要 旨

形容詞を用いた 観光評判情報の抽出に関する研究

山田 忍

近年、インターネットをはじめとする情報ネットワークで提供される情報源は大規模化している。そして、膨大な量の情報の中からユーザにとって役に立つ情報のみを分類することが重要になっている。しかし、膨大な量の Web サイト群の中から、目的の記事を見つけ出すことは困難である。検索結果から記事を絞り込む場合、検索に用いたキーワードに加えて、新たにキーワードを追加する必要がある。しかし、意見や感想を取得するためのキーワードには個人差があり、適切なキーワードを入力することは困難である。そこで、膨大なテキスト群から、掲載されている記事を利用して、情報を分類する手法が提案されている。情報を分類する方法として、文の構造を利用して分類する方法と、出現頻度を利用して分類する方法がある。日本語では、品詞同士の関係によって特定の意味を持つ文が構成される。文の構造を利用する分類方法では、文書中に出現する品詞の組み合わせを利用して分類する。出現頻度を利用した分類方法では、特定の分野を連想させる単語が出現する回数を利用して分類する。本研究では、観光評判情報の文構造における特徴と、文構造の出現頻度を用いて、テキスト群から観光評判情報とその他の情報に分類する手法を提案する。

キーワード 情報配信 検索エンジン データベース

Abstract

A proposal of sightseeing reputation information extraction method

YAMADA, Shinobu

Huge amount of information, such as the Internet, have become accessible in recent year, Then it is important to classify the information that is helpful for a user. However a user needs a lot of time and power when a user takes out useful information from the sources of information. Moreover it is difficult for a user to take out reputation information that is consisting of subjective information. Because many of reputation information are seen at a personal website and weblogs, so in many cases these websites are not reflected in the higher rank of search results. In this paper, we propose the method to classifying reputation information in the sightseeing field using a feature in sentence structure and a frequency appearance.

key words Information retrieval, Classification of information, Reputation information, Sightseeing information,