

# 要 旨

## Weblog に適した観光評判情報抽出システムの開発

畠山 英之

近年，旅行計画のために，観光評判情報を含む Weblog を参考にする人が増加している [1]．しかし，Weblog は，インターネット上に多数存在する．多くの情報から観光評判情報を含む Weblog を絞り込むのは難しい．そこで，観光評判情報を含む Weblog の絞り込み手法を提案する．提案手法は，文章を用いて分類する．分類手法は，係り受けと話題性のある単語を用いる．係り受けとは，単語の被修飾する回数である．話題性のある単語とは，目的の文章に多く含まれる単語である．本研究は，抽出率と除去率の 2 点から評価する．抽出率とは，観光について記された Weblog を絞り込む確率である．除去率とは，観光評判情報以外の記事を取り除く確率である．

キーワード 観光評判情報，情報フィルタリング，係り受け，Weblog

# Abstract

## A proposal of sightseeing reputation information extraction method for weblog

Hatakeyama, Hideyuki

In late years, the users who use traveler's diary (Sightseeing reputation Information) written down in the Weblog for the reference of travel plans increase[1]. The retrieval of the diary uses the search engines such as Google and Yahoo. But, there is a problem in the search engine. Because news and the advertisement are included, the retrieved number becomes huge. It is difficult to excerpt the Sightseeing reputation Information from huge information. Then, the solution that excerpts the Sightseeing reputation Information from huge information is proposed. The proposal technique uses, and classifies the sentence. There is a technique of two classifications. One uses the structure of the sentence. Another one uses the occurrence rate of the word. The proposal technique uses Lexical Analysis for the structure analysis on sentences. Lexical Analysis is frequency to which the word is modified. The technique of the classification that uses the occurrence rate of the word uses the frequency of a feature word. The proposal technique classifies the article into sightseeing reputation information and the remainder. The technique of the classification uses Lexical Analysis and Occurrence rate of feature word. The proposal technique is evaluated by the sampling fraction and the elimination factor. The sampling fraction is accuracy by which sightseeing reputation information is extracted. The elimination factor is accuracy from which the article that is not sightseeing reputation information is removed.

*key words* information filtering, Sightseeing reputation Information, Lexical Analysis, Internet