

要 旨

記事の特徴を利用した新聞記事検索手法

友池 貴之

現在，膨大な情報の中から，必要な情報を効率よく取り出すテキスト処理技術に対する関心が高まっている．ユーザの質問文に対して的確な答えを提示する質問応答技術や，重要な部分を損なわずにテキストをコンパクトにまとめる自動要約技術など，様々な観点からの技術研究が進められている．

本論文では，質問応答技術における文書検索手法について新聞記事の特徴を用いた検索手法を提案する．本検索手法では，検索対象に新聞記事データを用いる場合において，新聞記事に見られる特徴を用いることで検索性能の向上をねらうアプローチをとっている．新聞記事の特徴として，本文の1文目には結論が書かれることが多い，各段落の1文目は段落内で重要であることが多い，役職や年齢が添えられた人名は重要であることが多い，等を考えた．

ベースとなる検索手法として，tf-idf法による検索索引語の重み付けを利用しているが，同手法には，検索対象の文書が長いほど優先的に検索されてしまう問題があることが知られている．本論文では，この問題を解決する方法として，重要文抽出法によるテキスト自動要約技術を用いた検索対象の新聞記事データ長の制限手法も提案している．この要約に対してもまた新聞記事の特徴を用いている．

キーワード 質問応答，情報検索，tf-idf法，テキスト自動要約

Abstract

A Method of Article Retrieval Utilizing Characteristics in Newspaper Articles

TOMOIKE Takayuki

The concern about the text processing technology which takes out required information from huge information is increasing now. Technical research is carried out from various viewpoints, such as question answering and text summarization.

This paper describes a document retrieval method which is part of question answering system, utilizing characteristics in newspaper article. The retrieval method aims at retrieving document from newspaper articles. The examples of the characteristics in newspaper article are the first sentence of article has a conclusion in many cases, the first sentence of each paragraph is important in many cases and the name of a person to which an executive and age were attached are important in many cases.

The retrieval method is based on tf-idf weighting. However, it is known that there is a problem in the tf-idf weighting. When there is a long document in newspaper articles, it will be retrieved preferentially as compared with a short one. This paper describes the problem solution method which uses text summarization technique too.

key words Question Answering, Information Retrieval, tf-idf Weighting, Text Summarization