

論 文 要 旨

論文題目	時間軸を考慮した Web ニュースの効率的な検索方法の提案
------	-------------------------------

本論文では、ユーザがインターネット検索を行う際の、時間軸を考慮した Web ニュースの効率的な検索方法について提案する。インターネットが家庭に広く普及した近年、Web 上には膨大な情報が存在している。普段我々は、インターネットを利用する際、Web 上の膨大な情報から欲しい情報を得るために、URL を直接指定して情報を得ることはほとんどない。たいていの場合、Yahoo! Japan や Google などの検索エンジンを利用して、欲しい情報を得る。しかし、現在の検索エンジンには、利用する上で一つの問題がある。それは、インターネット検索に不慣れなユーザが検索エンジンを利用して、欲しい情報を容易に得ることができない点である。つまりインターネット検索には、ユーザの経験が必要となってくる。この問題の要因として、1. 検索対象となる情報の範囲、2. 検索キーワードの入力の難しさ、3. ユーザの検索結果に対する処理の負担の三点が考えられる。そのため、インターネット検索に不慣れなユーザでも、欲しい情報を効率的に抽出する技術が求められている。本論文では、検索対象とする情報を Web ニュースに限定する。Web ニュースを配信するサイトは数多く存在する。かつ、記事のリンクも多く存在する。また、記事の数は日々増加し続けているため、インターネット検索の熟練者であっても、Web ニュースの全貌を短時間で把握することは困難である。そのため、インターネット検索に不慣れなユーザが効率的に Web ニュースの全貌を把握できる方法が必要である。そこで本論文では、ユーザのインターネット検索支援を目的とし、Ⅰ. ユーザが与えたキーワードに関する話題、Ⅱ. その話題の注目度、Ⅲ. 注目期間を提示する手法を提案する。はじめに、ユーザが与えたキーワードを元に、関連する記事の集合を抽出する。次に、キーワードで抽出した記事集合を、文書の内容でクラスタリングする。最後に、各クラスタの出現頻度に基づいて、各クラスタの注目度・注目期間を算出し、提示する。提案手法により、ユーザはキーワードに関するニュースの全貌が把握でき、曖昧なキーワードから具体的な話題まで容易にたどり着くことが可能となった。

Abstract

This thesis proposes an efficient search method on Web news that takes into account a time variable. In recent years, the Internet has been widely spread to homes, and there is numerous number of information on the Web. When using Internet, we usually do not input URL directly to obtain information we want. In most cases, we obtain information that we want using search engines. However, there is a problem in using search engines; if an inexperienced user uses a search engine, he cannot obtain the information he is looking for easily. This means that, a certain level of experience is needed to obtain information that the user demands. The following three points are considered as factors of this problem: 1. Range of targeted information, 2. Difficulty of input of search keyword, 3. Load of processing the search result. In view of this need, we must develop a method that allows the inexperienced user to efficiently obtain the information he want.

This thesis aims at a user's information search support. We propose a method for presenting the following three issues; 1. Topics including the given keywords, 2. Degrees of attention of the topics, 3. Attention periods of the topics. First, we extract a document set that includes the given keywords. Next, we generate several clusters based on the appearance frequencies of the words. Finally, we calculate the degree of attention and attention period of each cluster based on the appearance frequency of each cluster, and present it. With the help of the proposed method, the user can take an overview of the whole feature of the news concerning the keywords. Moreover, we can reach from a vague keyword to specific topics easily by the proposal technique.