

平成 24 年度

慶應義塾大学

博士論文

長期的に利用する情報へのアクセスに関する研究

政策・メディア研究科

上野 大樹

2013 年 1 月

長期的に利用する情報へのアクセスに関する研究

論文要旨

本研究では、長期的に利用される情報に特化した情報検索手法および情報アクセス手法の提案を行う。ここでは、このような手法を「長期的検索」と呼ぶ。また、長期的検索を行うために、情報がどれだけ長期的に利用されているかの指標として、「長期度」という指標を提案する。

情報にはさまざまな利用のされ方がある。例えば、ある一定の短い期間でしか利用されないもの、長期的に利用されるもの、周期的に利用されるものなどである。このような利用のされ方によって、情報の種類や今後の有用性などある程度予測できる。ここで筆者は、長期的に利用されている情報は、これまで長期間にわたって利用されてきたという実績があるため、今後も長期的に利用される可能性が高いと考えた。そのため、こういった情報が有用である可能性は高いのではないかと考えた。情報の検索手法として、人気の情報や今流行りの情報を検索する手法は数多く提案され、システムとして実現されているが、長期的に利用される情報に特化して検索する手法やシステムはほとんど存在しない。そのため現状では、長期的に利用されている有用な情報でも、特别人気が高くない場合、見つけることが難しい場合もある。そこで本研究では、長期的検索という新しい概念を提唱し、長期的に利用される情報を見つける手法を提案する。

長期的に利用される情報の具体的な例として、「Web ページ」、「検索キーワード」という 2 つの例に焦点を当て、「セレクトブックマ」、「MasteryEye」という 2 つのシステムを開発した。

セレクトブックマは、長期的に利用できる Web ページに特化して検索するためのシステムである。ここではまず、ソーシャルブックマークのブックマーク時間情報を利用することによって、Web ページがどのように利用され、ブックマークされているかを分析した。そして、その分析結果に基づき、長期的な Web ページを検索できるように設計・実装をおこなった。セレクトブックマでは、画面上でキーワードを入力して検索を行うことによって、入力したキーワードに関連した Web ページの中で、長期的に利用されている Web ページを検索することができる。例えば、体系的だった知識を得られる Web ページ、Web サービス、リファレンスなどに特化して検索することを目的としている。

MasteryEye は、長期的にされ続ける検索キーワードを見つけるためのシステムである。Google 検索における検索回数の時系列データを利用することによって、このような検索キーワードを見つけることができるようにシステムの設計・実装をおこなった。MasteryEye では、画面上で入力したキーワードに関連するものの中で、長期的に検索されている検索キーワードを提示する。ある分野について新しく学びたい場合に、定番の検索キーワードを手軽に取得することを目的としている。

これら 2 つのシステムを試作し、分析および評価実験を行った。本論文では、分析および評価実験の結果を示し、長期的検索の意義や有効性についても議論する。

キーワード

情報検索，情報フィルタリング，情報レコメンデーション，長期的利用，ソーシャルブックマーク，検索キーワード，アクセスパターン

慶應義塾大学 大学院政策・メディア研究科

上野 大樹

Abstract of Doctor's Thesis Academic Year 2012

Research on Access to Long Term Information

Summary

In this research, we proposed information retrieval method and information access method that find information used for a long-term. We call these methods "long-term retrieval". Moreover, we propose "degree of long-term" that is an index of which information was used for a long-term.

The people have used information in variety ways. For example, there are the information used for a short-term, the information used for a long-term and the information used periodically. We can predict information types and values by using these data. Thus, the author consider information that used long-term until in the past is used for a long-term in the future too, because of past records. Moreover, these information is a high possibility of practical value. There are a lot of information retrieval methods that can search popular information and trend information, but there are few information retrieval methods that can search specialized long-term information. Therefore, now it is difficult to find long-term information if it is not popular. In this research, we propose a new concept that is "long-term retrieval", and we propose retrieval methods that can search long-term information.

For concrete examples, we developed two systems, "SelectBukuma" and "MasteryEye" that search long-term "Web pages" and "retrieval keywords".

"SelectBukuma" is a system that can search specialized long-term Web pages. First, we analyzed social bookmark time data, and we developed a retrieval system that can search long-term Web pages. "SelectBukuma" can search long-term Web pages by input keywords in the screen. For example, the system is aimed for search specialized Web pages that can get structured knowledge, Web service and some references.

"MasteryEye" is a system that can search long-term retrieval keywords. To achieve the purpose, "MasteryEye" uses Google's time-series data of search times. In the case when a user wants to acquire new field knowledges, the user can get standard retrieval keywords about that field.

We developed these two systems, and we analyzed and evaluated about systems. In this paper, we indicate the result of analysis and the result of evaluation experiment, and discuss about availability and the signification of "long-term retrieval".

Key Word

Information Retrieval, Information Filtering, Information Recommendation, Long Term Usage, Social Bookmark, Retrieval Keyword, Access Pattern

Keio University Graduate School of Media and Governance

Taiki Ueno