

第1章 長期的検索の考察と展望

概要

本章では、本研究の考察と展望について述べる。

1.1. 考察

本研究では、長期的に利用される情報に特化して検索する手法である「長期的検索」を提案した。また、本研究では、長期的検索における応用例として、Web ページ、検索キーワード、ファイルの3つの情報に応用した。そして、具体的なアプローチとして、セレクトブックマ、MasteryEye、タームストアの3つのシステムを開発した。本研究のアプローチの特徴は、以下のようにまとめることができる。

- 人の行動から得られる時間情報の分析
- 長期度の提案
- 長期的な情報に着目した検索システムの提案
- 長期的な情報の有用性

1.1.1. 人の行動から得られる時間情報の分析

本研究では、人々の行動から得られる時間情報を分析することによって、この時間情報を検索システムに活かせないか考えた。そこで、ソーシャルブックマークでの人々からブックマークされた時間情報と Google 検索における各検索キーワードの検索された時間情報を取得し、利用した。

最初に、ソーシャルブックマークのブックマーク時間データを分析した。その結果、長い期間新しくブックマークされ続ける Web ページは、長期的に利用される種類の Web ページの割合が多いことが分かった。また、短い期間しかブックマークされていない Web ページは、一時的に利用される種類の Web ページの割合が多いことが分かった。このことから、ソーシャルブックマークという人々のブックマーク行為の時間情報から、長期的に利用される Web ページを取得できることが分かった。

次に、Google 検索での検索回数の時系列データを分析した。この分析では、長期的に多くの人から検索され続けていたキーワードと短期的にのみ多くの人から検索されたキーワードを比較した。その結果、長期的に検索され続けていたキーワードは、短期的にのみ検索されたキーワードと比較して、長期的に利用できるキーワードが多いことが分かった。このことから、人々の検索履歴を利用することによって、長期的に利用できるキーワードを取得できることが分かった。

このように、履歴情報やブックマーク情報など、人の行動から得られる情報の時間情報を利用することによって、長期的に利用できる可能性の高い情報を見つけることができることが分かった。

1.1.2. 長期度の提案

本研究では、その情報がどれだけ長期的に利用されているかの指標である長期度の提案をした。長期度の計算方式として、ブックマーク日数に基づいた計算手法とべき乗則とのズレの大きさに基づいた計算手法を提案した。

セレクトブックマでは、ブックマーク日数に基づき、長期度を計算し、評価実験を行った。その結果、長期度の低い Web ページよりも、長期度が高い Web ページの方が、より今後も利用したいページが多かった（要確認）。

MasteryEye では、べき乗則のズレの大きさに基づいた計算手法により、長期度を計算し、評価実験を行った。その結果、長期度の低い検索キーワードよりも、長期度の高い検索キーワードの方が有意に

1.1.3. 長期的な情報に着目した検索システム

・分析結果に基づき、セレクトブックマを開発した・セレクトブックマは、長期的な Web ページを取得することができた・セレクトブックマの考察をがっと思っていただければいいんじゃない？

・MasteryEye を開発した・MasteryEye は、長期的な検索キーワードを取得することができた・MasteryEye の考察を考えよう（どこが有用なのか、どんなときに使えるのか、何がダメなのか）

1.1.4. 長期的な情報の有用性

・長期的な Web ページ・長期的な Web ページの方が、今後も利用したいページが多かった・長期的に利用されているページは、今後も長期的に利用できるページが多かった・長期的に利用されている情報は、長期的に利用できる可能性が高いことが分かった・また、長期的な情報の有用性が高いと考えられる

・MasteryEye ・評価実験の結果、・長期的な検索キーワードの方が有用だった・長期的な情報の有用性が高いだろう

・このように 2 つの異なった情報に関して、長期的な情報のほうが有用性が高いことが分かった・このため、長期的な情報は有用な可能性が高いだろう

・長期的に利用されるものは長期的に利用できる？

1.2. 議論

1.2.1. 長期的な情報の特徴

1.2.2. 長期的検索の有効性

1.3. 課題

1.4. 展望