

# 第1章 長期的情報検索の提案

## 概要

第3章では、本研究で述べている「長期的な情報」の定義と「長期間」の定義を行う。また、長期的な情報を検索するための手法を提案する。

## 1.1. 対象とする情報

ほぼすべての情報について、その情報が長期的に利用されているかどうかという事実は存在すると考えられる。このような情報の中でも本研究では、特に以下のような情報を取り上げて、研究を行う。

- Web ページ
- 検索キーワード
- ファイル



図 1.1: 対象とする情報

## 1.2. 長期度

本研究では、その情報がどれだけ長期的に利用されてきたかを表す指標として、「長期度」という指標を定義する。長期度が高い程、長期的に利用されている情報、長期度が低い程、短期的にしか利用されていない情報とする。本研究では、長期的に利用されている情報を取得するシステムを実装するため、この長期度が情報を評価するための主な軸となってくる。

## 1.3. 長期的な情報

### 1.3.1. 長期的な情報とは

本研究で述べている長期的な情報とは、以下の 2 種類の意味を含む。

- (1) これまでに長期的に利用されてきた情報
- (2) 情報を取得した人自身が、今後長期的に利用できる情報

2.2 で述べた長期度の高い情報とは、(1) を示す。そして本研究では、(1) の情報は、(2) である可能性が高いという仮説を立てる。この仮説の元、(1) の情報を取得するシステムを開発し、システムの有効性と仮説の検証を行う。

### 1.3.2. 長期的な期間について

本研究では、長期的に利用されている情報という表現をしているが、長期的とはどの程度の期間なのかについて述べる。一般的に、長期的とは絶対的に期間が決められているものではなく、どちらかという相対的な表現として用いられる場合が多い。そのため、こういった情報について述べるかによって、期間が異なってくる。

そのため、1ヶ月以内しか利用されないものが多いタイプの情報ならば、3ヶ月でも長期的となるし、逆に半年ぐらい利用することが当たり前のタイプの情報ならば、半年利用していても長期的とはならない。そこで、本研究では、以下のような平均値と標準偏差を用いて、長期的な情報と短期的な情報を示す。

長期的な情報：長期度  $>$  長期度の平均値  $+$  長期度の標準偏差

短期的な情報：長期度  $<$  長期度の平均値  $-$  長期度の標準偏差

### 1.4. 長期的情報検索の提案

以上を踏まえて、本研究では、長期的な情報を検索するシステムを提案する。これは、検索者が今後も長期的に利用できるような情報を見つけることができるようにすることを目指す。そのために、これまでに長期的に利用されてきた情報に特化して検索可能なシステムを実装する。

長期的に利用されてきた情報に特化して検索するために、長期度の計算手法を提案する。そして、長期度の高い順に順位付けを行い、検索結果を提示する。長期度の具体的な検索手法に関しては、どの情報を取得するかによって、異なってくるため、3章以降で述べる。