

# 第1章 序論

## 概要

本章では，本研究の背景・目的および論文の構成について示す．

## 1.1. 研究の背景・動機

近年、情報の短命化が進んできている[?]. その理由として、Web2.0の一般化やネットワーク速度の向上、電子機器の記憶容量の増大などが考えられる。Web2.0的なサービスの一般化と記憶容量の増大により、Web上には膨大な情報が記録されるようになってきた。このように大量に記録されたWeb上の情報は、玉石混交であると言われており[?], あまり利用されない情報も大量に存在する。また、例えばマイクロブログのような情報伝播速度が早いサービスも一般的に利用されるようになってきた。このようなサービスの一般化やネットワーク速度の向上などによって、情報が流行する速度、さらには情報が廃れる速度も速くなってきた[?]. こういったことから、寿命の短い情報やデジタルコンテンツがだんだんと増加してきてると考えられる。

このように寿命が短く短期的にしか利用しないような情報が増えてきたため、これまでの情報検索システムや情報提示システムでは、長期的に利用可能な情報を見つけることが難しくなってきたと考えられる。これに対して、長期的に利用する情報に特化して取得するようなシステムはほとんど存在しない。そこで、本研究では、長期的に利用可能な情報を見つけようと思ったときに、そういった情報を見つけることが容易でないという課題があると考えた。

ところで、本や商品に関しては、ベストセラーとロングセラー[?]という指標がある。ロングセラーと長期的に利用されるということは、「売れる」と「利用される」の違いはあるものの、類似した指標であると考えられる。そこでロングセラーの有用性を例として、長期的に利用される情報の有用性について考える。本や商品に関して言えば、ベストセラーを好む人もいれば、ロングセラーを好む人もいる。以上のような特徴がある一方、Web上の情報はベストセラーといったような観点か、他人からの推薦といったような観点からの評価が主で、ロングセラーという観点での評価はほとんどされていない。Webの歴史はまだ浅いし、商品の売れ方とWebでは異なるところも多いが、Web上の情報についてもベストセラーのような人気の高い情報を好む人もいれば、ロングセラーのように長期的に利用される情報を好む人もいると考えられる。また、どちらの情報が欲しいかは状況によって変わってくる。そのため、Web上の情報に関しても、長期的に利用されているかどうかという指標は、有用である可能性が高い。

そこで、本研究では、長期的に利用される情報という視点に着目した。そして、こういった情報に特化して取得するためには、今までよく利用されていたリンク数、アクセス数、評価数、関連度などといったような指標だけではなく、長期的な時間情報を利用した新しい情報の評価指標が必要であると考えた。

## 1.2. 研究の目的

本研究では、主に以下の3点を目的とする。

- 自分が長期的に利用可能な情報を手軽に見つけられるようにすること
- 長期的に利用されている情報の有用性が高いことを示すこと
- 本研究で実装した長期的な情報を検索するシステムの有効性を示すこと

本研究の一番の目的は、自分が長期的に利用可能な情報を手軽に見つけられるようにすることである。その理由は、長期的に利用可能な情報は有用である可能性が高いにもかかわらず、見つけるための手法が確立されておらず、発見することが容易ではないからである。

次に、本研究では、長期的に利用されている情報が有用である可能性が高いことを示す必要がある。実際に本研究では、情報の利用状況に基づいて、ユーザが長期的に利用可能な情報を発見し

ようとしている。ある情報が、長期的に利用可能であるということは、その情報が有用である可能性が高いということが言える。そこで、長期的に利用されている情報が、長期的に利用可能であることを示すことを目的とする。

さらに、実際に提案して実装したシステムが本当に有効であるかどうかを示す必要もある。

### 1.3. 用語定義

本論文で使用するいくつかの用語に対して、用語の意味を説明する。以下に示すのは、本論文で利用する際の用語の意味である。

**\*1 長期的に利用されている情報**

長期的にアクセスされており、これまで長期的に利用されてきた情報

**\*2 長期的に利用可能な情報**

今後も自分が長期的に利用できる情報

**\*3 長期間**

長期間とは、その情報がどれだけ長期的に利用されているかを表す指標とする

**\*4 情報フィルタリング [?]**

大量の情報の中から、ユーザーにとって必要な情報を取り出し、不要な情報を除外する処理を自動的に行う技術のこと。本論文では、広義の意味として情報フィルタリングという用語を利用しており、情報の収集と排除双方の意味を含む。

**\*5 情報レコメンデーション [?]**

ユーザの興味や嗜好に応じて、お勧めの情報を提供すること。

**\*6 folksonomy**

Web 上のデータにおいて、ユーザ自らが情報の分類・収集を行うこと。「人々」(folks)と「分類」(taxonomy)とを掛け合わせた造語である。例えば、ソーシャルブックマークなどでは、ユーザがブックマークしたページに自由にタグを付与できるが、このユーザ自身が自由にタグをつける行為も folksonomy の一つである。タグをつけることにより、タグを付けた情報にアクセスするための検索や分類に役立てることを目的としている。

**\*7 体系だった知識を得られる Web ページ**

ある分野に対して、基礎的な内容から応用的な内容までの幅広い知識を得られる Web ページ。また、ある分野に関する重要な知識がまとめて記述されているような Web ページ。

**\*8 RSS リーダー**

指定した Web サイトの更新情報を一定時間ごとに自動的にダウンロードし、更新があると記事へのリンクを表示してユーザに知らせるツールのこと。

**\*9 ソーシャルブックマーク**

インターネット上で自分のブックマークを不特定多数のユーザに公開し、有益な Web ページを共有する Web サービスのこと。

#### \*10 マイクロブログ

主に 140 文字程度の短い文章を書いて Web に公開する短いブログ。ブログとの違いは、手軽である点とユーザ同士のコミュニケーションを支援するソーシャルネットワーク的な側面が強いところである。

#### \*11 適合率

検索結果として得られた文書中にどれだけ検索に適合した文書（適合文書）を含んでいるかという正確性の指標である。情報検索システムの評価を行う際に一般的に利用される指標であり、これとともに再現率<sup>\*9</sup>が利用される場合が多い。適合率は以下の数式で表される。

$$\text{適合率} = \frac{R}{N}$$

$R$  : 検索された適合文書の数

$N$  : 検索結果の文書の数

#### \*12 再現率

検索対象としている文書の中で検索結果として適合している文書（適合文書）のうちで、どれだけの文書を検索できているかという網羅性の指標である。情報検索システムの評価を行う際に一般的に利用される指標であり、これとともに適合率<sup>\*8</sup>が利用される場合が多い。再現率は以下の数式で表される。

$$\text{再現率} = \frac{R}{C}$$

$R$  : 検索された適合文書の数

$C$  : 全対象文書中の適合文書の数

### 1.4. 論文の構成

図 1.1 に本論文の構成図を示す。

## 第 2 章 背景と関連領域

第 2 章では、本研究の背景と関連領域について述べる。

## 第 3 章 長期的情報検索の提案

第 3 章では、本研究で述べている「長期的な情報」の定義と「長期度」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。

## 第 4 章 長期的な Web ページの検索

第 4 章では、長期的な Web ページを検索するシステム「セレクトブックマ」について述べる。

## 第 5 章 長期的な検索キーワードの提示

第 5 章では、長期的な検索キーワードを提示するシステム「MasteryEye」について述べる。

## 第6章 長期的なファイルと Web 履歴へのアクセス

第6章では、自分が長期的に利用してきたファイルと Web ページへ手軽にアクセスするためのシステム「タームストア」について述べる。

## 第7章 関連研究

第7章では、本研究に関連する研究領域について整理し、本研究の特徴や位置づけについて述べる。

## 第8章 考察

第8章では、本研究の考察と展望について述べる。

## 第9章 結論

第9章では、本研究の成果をまとめ、本論文を総括する。

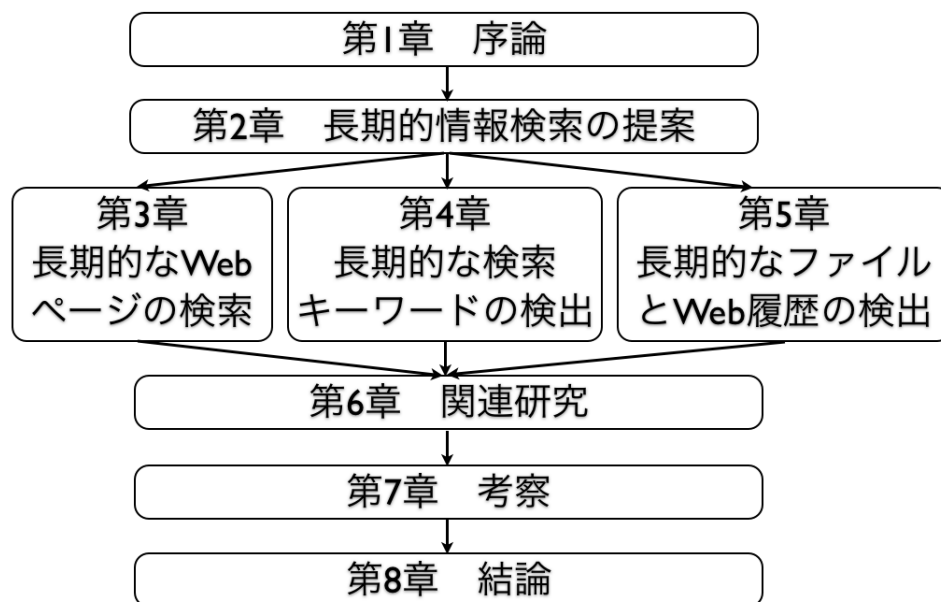


図 1.1: 本論文の構成