

リンク分離型 Web の実現に向けたハイパーリンクの類型調査

畑佐 賢志[†] 遠山 元道[†]

[†] 慶應義塾大学理工学部情報工学科 〒223-8522 神奈川県横浜市港北区日吉 3-14-1

E-mail: [†]kenji@db.ics.keio.ac.jp, ^{††}toyama@ics.keio.ac.jp

あらまし 著者らは HTML からハイパーリンクを分離する簡潔な手法の一つとして Web Index の応用を試みている。リンク分離型 Web とは、現在の Web で実装されている CSS によるデザインとコンテンツの分離のように、Web ページ中のアンカーテキストとリンク先 URL とをペアにしたファイルを利用してコンテンツとハイパーリンクとの分離を実現する Web である。これにより、Web 文書の更なる情報純度向上や Web 文書作成者とハイパーリンク管理者との分担を実現し作業、管理効率向上を目指すものである。このリンク分離型 Web の実現に向けて、実在の Web ページにリンク分離の手法をどの程度実装可能か判断するために、専用のクロールとスクレイピングを行うツールを設計、開発し、それを利用してハイパーリンクの分析調査を行った。

キーワード 情報抽出, Web クローリング, Web スクレイピング

1. はじめに

近年インターネットの普及とともにユーザは情報を取得するために、検索エンジンを利用して情報検索を行うようになった。また、検索エンジンだけでなく、ハイパーリンクを利用することで、異なる Web ページからさらなる情報を得ることも可能である。しかし、現在の一般的な Web 形態では、Web ページ作成者による特定のページへのリンクしか持つことができない構造となっている。つまり、そのリンクが必ずしも Web ページ閲覧者の欲する情報であるとは限らず、Web ページ作成者の意図したものが提供されているにすぎない。

そこで著者らは Web における利用者主導による情報資源結合を実現するために、Web Index (以下 WIX) と呼ぶ情報資源表現形式の提案、開発を行っている。WIX システムではキーワードと URL の組み合わせであるエントリの集合を XML 形式で記述した WIX ファイルを用い、閲覧中の Web ページに結合 (アタッチ) をすることで、Web ページ内の文章に出現するキーワードをそれに対応する URL へのハイパーリンクに変換する。[1] [2]

現在の Web 形態では文章コンテンツ要素である HTML と文章スタイル要素である CSS を分離することが一般化されている。しかしこの WIX システムを応用することでハイパーリンクに関してもキーワードと URL の組み合わせの集合として分離することでリンクの再利用などにより、Web 作成者の負担を軽減し、リンクの管理効率向上を図ることができると考えられる。ただし現在の Web ページにおいて多種多様なハイパーリンクの使用用途があり、全てのリンクをキーワードと URL の組み合わせとして分離することができるか不明である。

本研究ではリンク分離型 Web の実現に向けて、実在の Web ページのハイパーリンクの利用特性を把握し、どの程度 WIX システムを使用してリンクの分離が実現可能であるのかを判断するため、専用の Web Crawling システムを設計、開発し、収集したハイパーリンク情報を用いてリンクの分析調査を行った。

2. 研究背景

リンク分離型 Web の実現にあたり、現在の Web コンテンツには多種多様なハイパーリンクが用いられており、全てのハイパーリンクを分離すべきかどうかは実際のところは不明である。

現在の Web 形態では文章コンテンツ要素である HTML と文章スタイル要素である CSS を分離することが一般化されている。しかしこの WIX システムを応用することでハイパーリンクに関してもキーワードと URL の組み合わせの集合として分離することでリンクの再利用などにより、Web 作成者の負担を軽減し、リンクの管理効率向上を図ることができると考えられる。

まずはじめに、本研究で用いる一意性に関する定義を以下のように定義する。

[定義 1] ハイパーリンク $l[a, u]$ に関して (a : アンカーテキスト, u : リンク URL)

あるハイパーリンク集合 L において、任意の a がただ 1 つ存在し、 $FD: a \rightarrow u$ ならば、その a は L に置いて一意性を持つと言え、そのときの l を一意性リンクとする。

[定義 2] ある Web ページ集合 P における各 Web ページ p において、一意性リンク l が存在するとき、そのリンク l をページ内一意性リンクとする。

また URL に関しても <https://www.keio.ac.jp/about/campus/yagami.html> で表されるように、ピリオドによって構成されるドメイン、/パス名によって構成されるフォルダ、URL という 3 つの要素で構成されている。各要素において Web ページ集合も異なるため、それぞれの要素における一意性についても定義を行う。

まず、ドメイン階層に関する一意性を定義する。図 1 に示している慶應義塾大学ホームページのドメイン階層を例としてあ

げる。

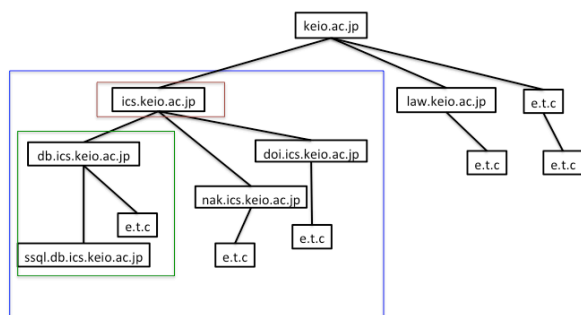


図 1 ドメイン階層

以上の図 1 から、茶臼の範囲内のような、あるドメイン内における Web ページ集合においてページ内一意性リンクが存在するとき、そのリンクのことをドメイン内非再帰的一意性リンクと呼ぶ。

また、青枠や緑枠の範囲内のような、あるドメイン以下における Web ページ集合においてページ内一意性リンクが存在するとき、そのリンクのことをドメイン内再帰的一意性リンクと呼ぶ。

フォルダ階層についてもドメイン階層と同様に定義を行うことができる。

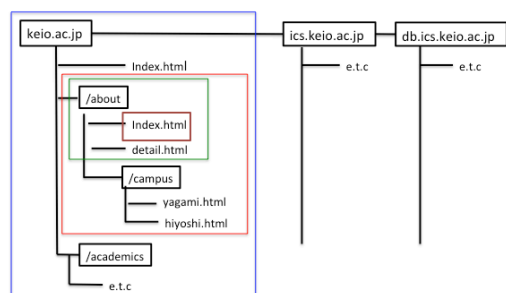


図 2 フォルダ階層

以上の図 2 から, 青枠の範囲内における一意性リンクを先ほど定義したドメイン内非再帰の一意性リンクとする.

そして、茶臼の範囲内のような、あるフォルダ内における特定のファイル内においてページ内一意性リンクが存在するとき、そのリンクのことをファイル内一意性リンクと呼ぶ。

また、緑枠の範囲内のような、あるフォルダ内における Web ページ集合においてページ内一意性リンクが存在するとき、そのリンクのことをフォルダ内非再帰的一意性リンクと呼ぶ。

また、赤枠の範囲内のような、あるフォルダ内における Web ページ集合においてページ内一意性リンクが存在するとき、そのリンクのことをフォルダ内非再帰的一意性リンクと呼ぶ。

本研究において、最も広い範囲であるドメイン内再帰的一意性リンクに関する調査のみを行ったため、このドメイン内再帰的一意性リンクを一意性リンクと呼ぶ。

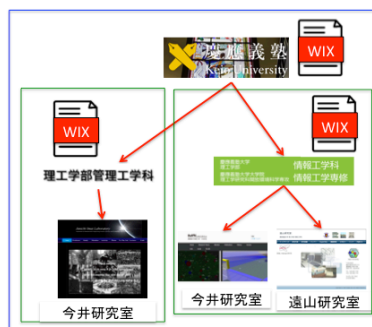


図 3 現在の WIX システムにおける WIX ファイル生成階層の例

さまざまなハイパーリンクの使用用途が考えられるが、現在、図3のように、リンク分離型 Web の実現にあたり WIX ファイルをドメイン毎に配置し、Web Index システムを適用させようと考えている。このとき、あるキーワードと URL のペアの一意性が保たれる最上位のドメイン階層の WIX ファイルにキーワードと URL のペアを配置している。その際に「今井研究室」というアンカーテキストは keio.ac.jp ドメイン以下の全ページにおいては一意性がないが、ae.keio.ac.jp や ics.keio.ac.jp のドメイン以下の全ページではそれぞれ一意性リンクとなる。このようにドメイン階層によってリンクの一意性が変化する。よってリンク分離型 Web を実現するにあたって、ドメイン階層において一意性リンクがどの階層にどの程度存在し、ドメイン以下においてリンクが一意に定まるのか、またリンクが一意性をもたない場合、どの階層において一意性リンクが存在し、リンクが一意性をもつようになるのかを調査することで、現段階でどの程度ドメイン階層に基づく WIX ファイルを生成でき、WIX システムを用いてリンク分離型 Web の実現が可能であるかを把握できると考えている。そこで、リンク分離型 Web の実現を目指し、ドメイン階層における WIX ファイル作成の実現に向け、ドメイン階層に基づくハイパーリンクの利用方法を調査が必要だと考えられる。ハイパーリンクに関する研究は文章の分類 [3] や検索エンジンの向上のための使用 [4] などさまざまな角度から行われている。

3. 関連研究

3.1 WIX ファイル作成支援ツール

当研究室の先行研究により、WIX ファイル作成を支援するためにアンカーテキストと URL の組み合わせをディレクトリ構造を保持したまま抽出し、解析を行うことで、アンカーテキストに対するリンク先の一意度を割合で示すことのできるツールを開発済みである.[5]

3.2 XLink

XML 文章において, XML Linking Language (以下 XLink) という規格が存在する. XLink は, XML 文書の要素にリンクの機能を与えるための規格を定めているものであり, これは, HTML のようなシンプルなリンクだけでなく, 複数リンクや間接リンクなど, さまざまな高度なリンク機能を提供する. XLink は特別な要素タイプを導入するのではなく, XLink で定めるグ

ローバル属性を任意の要素に適用し、その要素にリンクの役割を持たせる.[6][7]

4. Web Index システム

4.1 概要

WIX システムとは、Web における結合可能な情報資源である。現在の Web では、特定のアンカーテキストから特定のページへのリンクがされるという構造が一般的であるが、これに対し WIX では、アンカーテキストとリンクを Web ページから独立した「キーワードとリンク先の集合」(WIX ファイル)として扱い、任意のドキュメントに対しユーザ主導で「結合」を行うことでドキュメント内のキーワードを対応する URL へのハイパーリンクに変換する。

4.2 WIX ファイル

WIX ファイルとは XML 形式で記述されたキーワードと URL の組み合わせであるエントリの集合であり、エントリ内にはキーワードとなる見出し語を keyword 要素に格納し、それに対応する詳細情報を示す参照先の URL を target 要素に格納する。また、WIX ファイルには header 要素にファイル概要、作者コメントなど、その WIX ファイル全体についての meta データを格納することも可能である。また 1 つの entry 要素内に複数の keyword 要素や複数の target 要素を格納することも可能となっている。このような記述がなされた WIX ファイルの entry 情報を格納する場合には、キーワードとターゲットは 1 対 1 の組み合わせとなるように複数のエントリに分けて扱われる。

```
<?xml version="1.0" encoding="UTF-8"?>
<WIX>
  <header />
  <body>
    <entry>
      <keyword>錦織圭</keyword>
      <target>http://ja.wikipedia.org/wiki/錦織圭</target>
    </entry>
    <entry>
      <keyword>ノバク・ジョコビッチ</keyword>
      <target>http://ja.wikipedia.org/wiki/ノバク・ジョコビッチ</target>
    </entry>
    <entry>
      <keyword>ロジャー・フェデラー</keyword>
      <target>http://ja.wikipedia.org/wiki/ロジャー・フェデラー</target>
    </entry>
  </body>
</WIX>
```

図 4 WIX ファイルの記述例

4.3 アーキテクチャ

4.3.1 WIX ライブラリ

WIX ライブラリでは、全ての WIX ファイルを過去のバージョンも含めて保存しており、ファイル単位での情報管理を行っている。アタッチの際には、アタッチエンジンが全ての WIX ファイルのエントリに対して辞書式マッチングを行うため、WIX ファイルをエントリ単位に分解し、WIX DB に格納する。

4.3.2 WIX DB

WIX DB では、WIX ファイルをエントリ単位に分割し、RDB にて管理する。WIX ファイルの情報は、entry テーブルに格納され、管理される。entry テーブルには、各エントリが所属する

WIX ファイルの id (wid), エントリの id (eid), 辞書語となる keyword とそれに対応する URL の target 等が属性として存在する。entry テーブルの例を表 1 に示す。

表 1 entry テーブル

wid	eid	keyword	target
1	1	錦織圭	http://ja.wikipedia.org/wiki/錦織圭
1	1	フェデラー	http://ja.wikipedia.org/wiki/ロジャー・フェデラー
2	3	田中将大	http://ameblo.jp/tanaka-masahiro/
3	6	本田圭佑	http://keisuke-honda.com/voice/detail.php?id=203

4.4 ハイパーリンクの生成

WIX システムのクライアントサイドは、Firefox add-on や Chrome Extension などによって実装されている。図 5 は Chrome Extension の例である。

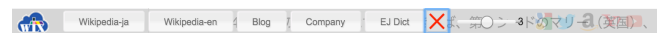


図 5 Chrome Extension(WIX ツールバー)

ユーザは、WIX ファイルの集合をブックマークとしてをクリックすることで、Web ページ中の単語をハイパーリンクに変換することができる。この結合操作をアタッチと呼ぶ。図の記事を閲覧中に、アタッチボタンをクリックすることで、出現しているキーワードに対応する WIX ファイル内の target タグに記述されている URL へのハイパーリンクが図 6 の記事に生成される。



(a) アタッチ前

(b) アタッチ後

図 6 Web ページの例

5. リンク分離型 Web

5.1 背景

現在の Web において一般的に文章デザインの指定に Cascading Style Sheet (以下 CSS) を用いる。CSS により、Web ページからコンテンツ (Web 文書) とデザインの分離を可能とする。

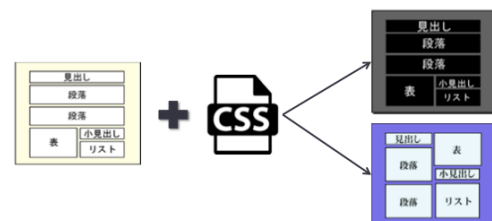


図 7 リンク分離型 Web の概念図

企業や学校等の組織における Web サイトに関して、Web ページにはさまざまなハイパーリンクが張られているが、そのハイパーリンクは企業の製品ページや、イベントの詳細ページなどその Web サイト内での内部リンクである可能性が高いと想定できる。また、企業や学校等の組織では、その組織特有のキーワードに対して、特定のリンク先があり、多くのページでそのリンクが出現することも多い。図 7 において慶應義塾大学情報工学科の研究室紹介ページ (<http://www.st.keio.ac.jp/campus/laboratory/>) を例に挙げる。このページでは各研究室というキーワードに対して情報工学科の各研究室のホームページへのリンクが貼られている。大学の Web ページにおいて、研究室名や教授名に対してリンク先がその教授の研究室のホームページであるリンクは、多くのページで出現する。企業に関しても、製品名などの特有のキーワードに対して、製品詳細ページなどの組み合わせが考えられる。そのため、企業や学校等の組織の Web サイトにおいてこのように頻出するリンクを再利用することができれば、Web ページ作成者の負担を軽減させることもできると考えられる。そこで WIX システムを応用することによるリンク分離型 Web を提案する。



図 8 情報工学科 研究室紹介ページの一部

5.2 概要

リンク分離型 Web とは Web 文章から CSS によるデザイン分離だけでなく、文章中のハイパーリンクの分離を行う Web 形態を指している。この Web 形態実現に向けて HTML からハイパーリンクを分離する簡潔な手法の一つとして WIX の応用を試みている。つまり、各研究室のキーワードとそのリンク先のペアを保存し、他の Web ページにこのリンクをアタッチすることで同じハイパーリンクを生成することが可能となる。リンク分離型 Web のメリットとしては以下の 3 点を挙げることができる。

- (1) リンクの一括管理による文章のメンテナンス性向上
- (2) リンクの再利用によるリンク切れ、リンク忘れ防止
- (3) Web 文章情報の純度向上

6. Web Crawling と Link Analyzer

6.1 Web Crawling システム

6.1.1 概要

Web Crawling システムは、リンク分離型 Web を実現するための手法の 1 つとして考えている WIX システムの応用において、実在の Web ページにリンク分離の手法をどの程度実装可能か判断するために、Web ページに存在している大量のハイパーリンクやアンカーテキストの情報を収集することを目的として、設計、開発したものである。本システムはシステムのユーザが解析したい複数の Web ページの URL (以下 Top ページ) と収集したい階層を自由に指定することでハイパーリンクとアンカーテキストに加え、ハイパーリンクにおけるタグ情報、ハイパーリンクが張られている Web ページの情報を同時に収集することが可能である。また、第 5 章で述べる Link Analyzer システムにおいて様々な条件下でのハイパーリンクの類型調査を行うために、Web ページの情報に関しては、ページのバイト数を含んだ Web ページの meta 情報を収集し、どのページがどの階層のものなのか判別できるようディレクトリ構造を保持したまま収集を行うことができる。

また、ハイパーリンクを収集する際に 1 度収集したページのハイパーリンクを再利用してクロールを行う設計のため、2 度同じページの情報を収集することはない仕様となっている。このシステムはスクリプトになっており、実行コマンドを入力するだけで、収集を行うことができ、収集したハイパーリンク情報は JSON 形式で整形されデータベースに挿入される。

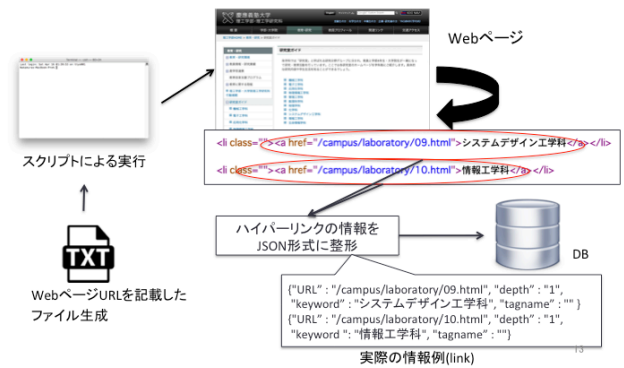


図 9 Web Crawling システムのアーキテクチャ

6.1.2 link 情報と page 情報

本システムで収集した情報は関係データベースに格納する。ハイパーリンクに関する情報は link テーブル、リンクの記載されている Web ページに関する情報は page テーブルにそれぞれ格納する。各テーブルの定義は図 10 に、各テーブルにおいて収集する情報を表 2 に示す。

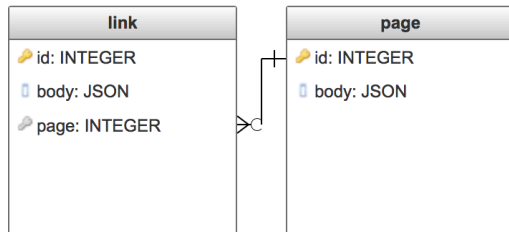


図 10 link,page テーブル

表 2 link,page テーブル情報 (body 要素内)

link		page	
"URL"	ハイパーリンク URL	"URL"	ハイパーリンク記載ページ
"keyword"	アンカーテキスト	"byte"	ページバイト数 (byte)
"depth"	階層数	"depth"	階層数
"tagname"	リンクにおけるタグ情報 (img)	"toppage"	最初に指定した Web ページ URL

6.2 Link Analyzer システム

6.2.1 概 要

Link Analyzer システムでは、Web Crawling システムによって収集されたデータをもとに分析を行う。例えば、ある 1 つのリンクに注目し、それに対してどのようなページにリンクが生成されているか、また同じアンカーテキストを持つリンクにおいて、どのページではどのリンクに対応しているのかなどを調査することができる。システム全体をスクリプト化しているため、何度でもコマンド一つで実行ができるようになっており、長期的に一意性リンクの有無、リンクやページの特徴を調査することができる。例えば、慶應義塾大学理工学部のホームページを例に挙げると、そのページへのリンクがどのページに張られているのか、また、ページ中に含まれている「今井研究室」というアンカーテキストを持つハイパーリンクがどの階層で一意性を持ち、どのようなリンク先に関連付けられているのか、などということを調査することができる。分析結果は txt ファイル形式で保存することで長期的な分析を可能にしている。

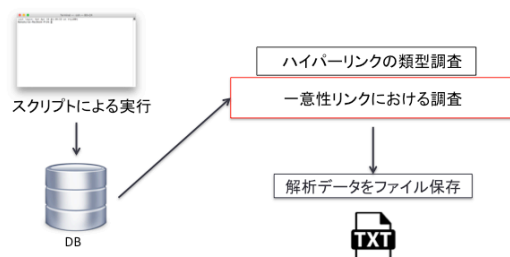


図 11 Link Analyzer システムのアーキテクチャ

7. Link Analyzer によるハイパーリンクの調査

7.1 一意性リンクに関する調査

この調査では、Crawling した DB からアンカーテキスト : a, ハイパーリンク URL : l, リンク記載ページ URL : p を取り出

し、p を用いて、DB 内の全てのドメイン名を取得し、Top ページのドメインに順ずる各ドメイン毎における a と l のリストを作成し、各ドメイン以下で一意性リンクであるリンクをカウントする。その結果をファイル保存し、各ドメインにおける全アンカーテキスト、一意性リンク数と割合が出力される。システム概要が以下の図 12 のとおりです。

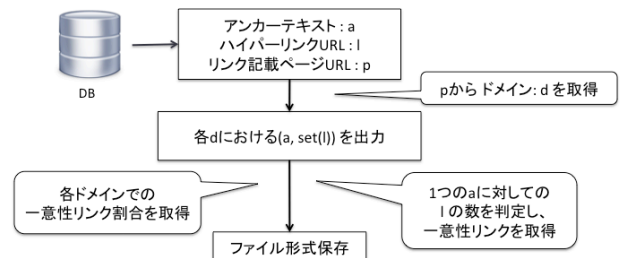


図 12 一意性リンクに関する調査に関するシステム概要

今回の調査において使用した Web ページ情報を表 3 に示す。この Web ページを Top ページとして指定し、Web Crawling システムを用いて link , page 情報を収集した。

表 3 使用した Web ページ情報一覧 (一意性リンク調査)

Web ページ	Top ページ URL	リンク数	ページ数
慶應義塾大学情報工学科 HP	www.ics.keio.ac.jp	19010	434
yahoo ニュース HP	www.news.yahoo.co.jp	246874	2174
NTT 東日本 HP	www.ntt-east.co.jp	225863	2094
ameba HP	www.ameba.jp	1014549	4612

この場合のリンク数、ページ数はそれぞれ Web Crawling システムによって収集された情報数のことを指す。また今回は Top ページから 3 階層分のリンクを対象としている。

この調査では、Top ページのドメインに順ずる各ドメイン毎におけるアンカーテキストとリンク URL のリストを作成し、各ドメイン以下で一意性リンクであるリンクをカウントする。その結果をファイル保存し、各ドメインにおける全アンカーテキスト、一意性リンク数と割合が出力される。指定されたコマンドを実行すると、指定した Top ページ毎にデータファイルが生成し、保存されるようになっている。具体的に取得できる情報を表 4 に示す。

表 4 一意性リンクに関する調査の出力結果 (慶應義塾大学情報工学科 HP)

ドメイン名	全キーワード数	一意性リンク数	一意性リンク割合
yamanaka.ics.keio.ac.jp	206	196	95.15%
doi.ics.keio.ac.jp	23	23	100.0%
arc.ics.keio.ac.jp	65	59	90.77%
sasase.ics.keio.ac.jp	48	44	91.67%
fj.ics.keio.ac.jp	41	34	82.93%
ohtsuki.ics.keio.ac.jp	48	46	95.83%
am.ics.keio.ac.jp	80	75	93.75%
ayu.ics.keio.ac.jp	45	43	95.56%
総サブドメイン数	936	831	88.78%
ics.keio.ac.jp	972	825	84.88%

このように、Top ページのドメインとサブドメインに対する全キーワード数、一意性リンク数とその割合を一度に出力できる。表 4 において、Top ページにおけるドメインと Top ページに対するサブドメインを比較すると、この例においてはサブドメイン以下での一意性リンクの割合が大きくなる傾向があることがわかる。その他のリンクにおける一意性リンク割合を図 13 に示す。

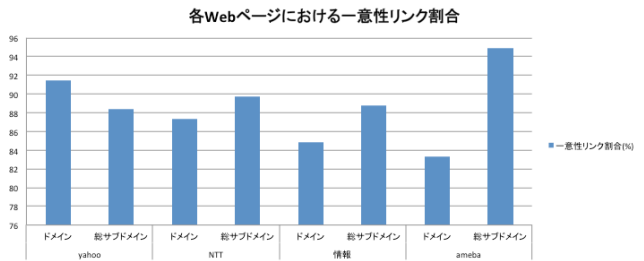


図 13 各 Web ページにおける一意性リンク割合の比較

図 13 の結果より、ドメイン階層における一意性リンク割合は、下の階層になるほど、大きくなる傾向がある。ただし、例外的に上位の階層での一意性リンク割合が大きくなることもある (yahoo サイト)。このようなサイトは、フォルダ階層におけるページ分類を行っている可能性が高いため、ドメイン階層での調査だけでなく、より狭い範囲であるフォルダ階層における調査も必要である。



図 14 一意性リンクの具体例

図 14 は情報工学科 HP での実際の調査で出力された結果である。このように「研究業績」というアンカーテキストに対し Top ページのドメイン以下では、4 つのリンク遷移先を持つ。しかしサブドメイン以下ではリンク遷移先が 1 つに定まり、一意性を持つため、サブドメイン以下での一意性リンク割合が大きくなると考えられる。

ドメイン名	全キーワード数	一意性リンク数	一意性リンク数割合
doi.ics.keio.ac.jp	23	23	100.0%
ayu.ics.keio.ac.jp	45	43	95.56%

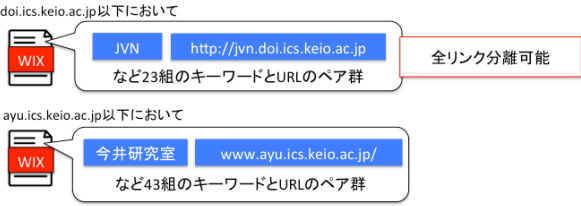


図 15 ドメイン階層における WIX ファイル生成

このような調査結果より、情報工学科 HP ドメイン階層において、図 15 のように”doi.ics.keio.ac.jp”以下のドメイン内では全てのリンクが一意性リンクのためこのドメイン以下では WIX ファイルに全てのペア群を使用でき、リンクを分離することができる。また”ayu.ics.keio.ac.jp”のドメイン以下では 43 組のペア群を使用でき、リンクを分離することができることがわかり、各ドメインの WIX ファイルにどの程度のキーワードとリンクを挿入できるのか判定可能である。

7.2 今後の調査

現状では本システムでドメイン階層での調査しかできていないため、より狭い範囲であるフォルダ階層におけるリンクの調査も行い、最終的にはフォルダごとに WIX ファイルを生成し、リンク分離型 Web を実現させたいと考えている。またドメイン階層においても Top ページに指定する Web ページのカテゴリ数や、試行回数をより増やし、よりリンク特性を把握することを考えている。

本システムで調査する Web ページ対象を大幅に増やし、かつ学校・企業・スポーツなどカテゴライズした Web ページを用いて、ハイパーリンクの利用形態の把握に努めたいと考えている。また調査対象に関しては単一の Web ページだけでなく、ページ集合を対象として調査しなければリンクの特性は把握できないと考えている。以下の表 5 に今後調査する Web ページの一覧を表示する。

表 5 今後の調査対象

カテゴリ	ページの詳細
学校 1	各大学の Top ページ集合
学校 2	各大学の理工学部ページ集合 (専門的な用語を取り扱うページ)
企業	東証一部上場企業の Web ページ集合
スポーツ	日本プロ野球に加盟しているチームの Web ページ集合

8. おわりに

本論文では、リンク分離型 Web を実現するために、専用の Web Crawling システムを設計、開発し、分析したい Web ページを Top ページとして階層を保持したまま、ハイパーリンクや類型調査に必要な情報を大量に収集し、Link Analyzer システムによってハイパーリンクに関する類型調査を行った。調査によって、どのページのどのアンカーテキストにどのような形態のリンクが貼られているのか、そしてそのリンクは WIX ファ

イルという形で分離することで再利用の効率化を図ることができるのかを認識することができる。またリンクとリンク記載ページにおける関係性も同時に把握することができる。今後は調査対象を大幅に増やし、各カテゴリの Web ページによるハイパーリンク特性を把握することに努めたいと考えている。

文 献

- [1] 佐草友也, 遠山元道. Web Index におけるオーサ主導アタッチの拡張に関する研究. ユーザインターフェース. データ工学ワークショップ, DEIM2014. 2014.
- [2] 森良介, 戴達也, 朱成敏, 遠山元道. Keio WIX システム (2) サーバーサイド実装. データ工学ワークショップ, DEIM2011. 2011.
- [3] 鈴木祐介, 松原茂樹, and 吉川正俊. アンカーテキストとハイパーリンクに基づく Web 文書の階層的分類. 人工知能学会全国大会論文集 0 : 227-227. 2005.
- [4] Brin, Sergey, and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. Computer networks 56.18 : 3825-3833. 2012.
- [5] 青木陽介, 遠山元道. Web Index における WIX ファイル作成支援のためのリンク解析ツール. ユーザインターフェース. データ工学ワークショップ, DEIM2014. 2014.
- [6] Behrends, Erik, Oliver Fritzen, and Wolfgang May. Querying along XLinks in XPath/XQuery: situation, applications, perspectives. International Conference on Extending Database Technology. Springer Berlin Heidelberg, 2006.
- [7] Wilde, Erik, and David Lowe. XPath, XLink, XPointer, and XML: A practical guide to Web hyperlinking and transclusion. Addison-Wesley Longman Publishing Co., Inc., 2002.