

## ページタイプと閲覧目的による WEB ブラウジング支援

小田 寿則<sup>†</sup> 石井 恵<sup>†</sup> 片岡 良治<sup>†</sup>

NTT サイバーソリューション研究所<sup>†</sup>

### 1. はじめに

本稿では、閲覧中 WEB ページの内容に関する周辺ページを先行して検索しておき、それらを閲覧する目的別に分類・提示することにより、目的指向にブラウジングを支援する WEB 情報ナビゲーション方式を提案する。提案方式は、WEB ページのページタイプに関連度の高い閲覧目的が存在し、次に見たい WEB ページについても特定の目的を持って閲覧されるという仮定に基づいているが、本稿では、この仮定の正当性を検証するために実施した実験結果を示すと共に、プロトタイプ化したブラウジング支援 AP について述べる。

### 2. 関連研究

本稿にける WEB ページの情報ナビゲーションとは、現在閲覧している WEB ページの次に見たい WEB ページの候補を、システム側が何らかの方向性を示した上で、自動的に提示するブラウジング支援である。従来の研究においては、閲覧中の WEB ページと文書類似度の高い WEB ページを次の閲覧候補として、閲覧中ページの周辺に提示するブラウジング支援 [1] の提案が行われてきた。また、ブラウジング支援とは異なるが、検索結果を WEB ページのページタイプ (コンテンツ形式) で分類・提示することによる検索効率化の提案 [2] も行われてきた。

### 3. 提案方式の概要

提案方式は、閲覧中の WEB ページの内容によって、次の閲覧候補ページを変化させるという基本戦略においては [1] と同様であるが、ページ候補を類似度に基づいて提示するのではなく、閲覧目的のクラスに分類して提示することにより、目的から WEB ページを絞り込んでブラウジングを支援できる点で [1] と異なる。閲覧目的は WEB ページの内容から容易に推定できないため、本稿では WEB ページのページタイプと閲覧目的の関係を利用する。図 1 に示すように、閲覧中のページタイプと次に見る WEB ページの閲覧目的は、経験的にカタログ WEB ページの次には掲載されている商品を販売する店の WEB ページを閲覧したいなどの関連性があり、また、閲覧目的を達成するために特定のページタイプの WEB ページを閲覧することも行われているなど、ページタイプを利用すればおおよその閲覧目的が推定できると考えた。以降に提案方式の処理概要を示す。

### 提案方式の処理概要

#### < 事前処理 >

- 1) 閲覧するページを収集する
- 2) 収集したページをページタイプに分類する
- 3) ページタイプと閲覧目的の関係を計算する

#### < ブラウジング支援処理 >

- 1) ページが閲覧されると、事前に求めたページタイプをブラウザ側に伝える
- 2) 周辺ページ群を検索しページタイプを付けてブラウザ側に伝える
- 3) ブラウザ側でページタイプと次の閲覧目的の関連度の高い順に、閲覧目的を提示する
- 4) 閲覧目的の下に関連度の高いページタイプのページを提示する

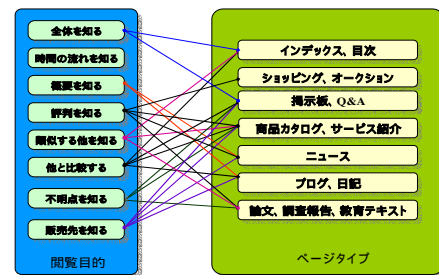


図1. ページタイプと閲覧目的

### 4. 検証実験の結果

提案方式の正当性を検証するために以下の実験を行った。

#### 4.1 ページ収集・分類

閲覧中ページの周辺ページとしては、構造的周辺である同ドメインページやリンク/被リンクページ、また、意味的周辺ページとして、語ベクトル類似度が高いページが考えられるが、本方式ではドメインを超えたナビゲーションを対象としているため、後者を採用した。実験では話題性がある数種類の検索語による検索エンジンの検索結果 (合計 1841 ページ) を収集し、収集した WEB ページを事前に各種属性 (ジャンル、ページタイプ、該当のページの閲覧目的、該当のページを閲覧することによって次に見たいページの閲覧目的) の観点で人手により分類し、それらをインデックス化した。

ページタイプとは、“政治”、“経済”、“社会”などのジャンルとは異なり、WEB ページのコンテンツ形式であり、実験では“インデックス”、“掲示板”、“カタログ”、“ニュース”などの11クラスの分類を用いた。

閲覧目的とは WEB ページを何のために WEB ペー

ジを閲覧するかという目的であり、実験では”全体を知る”，”時間の流れを知る”，”評判を知る”，”販売先を知る”などの8クラスとし、それぞれ閲覧中のWEBページの閲覧目的と、次に閲覧したいページの閲覧目的として分類した．  
表1に分類結果を示す．

表1．収集したページの分類結果

ジャンル	件数	ページタイプ	件数	閲覧目的	件数
政治	107	インデックス	296	全体を知る	372
経済	120	ニュース	304	概要を知る	480
生活	319	ブログ	149	詳細を知る	657
スポーツ	133	日記	112	評判を知る	102
社会	121	カタログ	140	時間の流れを知る	40
趣味	251	サービス紹介	118	不明点を知る	109
芸能	179	オークション	68	類似する他を知る	10
旅行	144	Q & A	79	販売店を知る	5
グルメ	89	掲示板	99		
科学	142	調査報告	140		
		論文	79		

#### 4.2 ページタイプ、閲覧目的間の関連性

収集・分類したWEBページから、ページタイプと閲覧目的の関連度を調べた結果を表2に示す．ページタイプによって、特定の閲覧目的との関連度が高い事が判るが、このことは提案方式の正当性を示している．尚、ジャンルと閲覧目的の間にはこのような関連は無かった．

表2．ページタイプと閲覧目的との関連（上位3目的）

ページタイプ	閲覧目的	関連	次の閲覧目的	関連
インデックス	全体を知る	54.4%	詳細を知る	69.6%
	概要を知る	32.4%	概要を知る	13.5%
	詳細を知る	8.1%	目的完了	8.1%
ニュース	詳細を知る	59.2%	目的完了	44.1%
	概要を知る	26.6%	詳細を知る	27.0%
	全体を知る	7.2%	類似する他を知る	20.1%
ブログ	詳細を知る	41.6%	目的完了	75.2%
	評判を知る	34.2%	類似する他を知る	9.4%
	概要を知る	18.8%	詳細を知る	8.7%
日記	詳細を知る	57.1%	詳細を知る	24.1%
	概要を知る	11.6%	時間の流れを知る	21.4%
	全体を知る	11.6%	類似する他を知る	17.9%
カタログ	概要を知る	48.6%	詳細を知る	52.1%
	詳細を知る	25.0%	目的完了	26.4%
	全体を知る	16.4%	販売店を知る	6.4%
サービス紹介	概要を知る	39.8%	詳細を知る	63.6%
	全体を知る	35.6%	目的完了	21.2%
	詳細を知る	20.3%	販売店を知る	3.4%
オークション	全体を知る	45.9%	概要を知る	47.1%
	概要を知る	19.1%	詳細を知る	27.9%
	不明点を知る	17.6%	評判を知る	8.8%
Q & A	不明点を知る	88.6%	詳細を知る	88.6%

	概要を知る	5.1%	目的完了	7.6%
	全体を知る	2.5%	概要を知る	2.5%
掲示板	詳細を知る	35.4%	詳細を知る	38.4%
	評判を知る	19.2%	時間の流れを知る	25.3%
	全体を知る	14.1%	目的完了	25.2%
調査報告	詳細を知る	44.3%	詳細を知る	53.6%
	概要を知る	34.3%	目的完了	32.9%
	全体を知る	15.0%	類似する他を知る	5.0%
論文	詳細を知る	65.8%	目的完了	50.6%
	全体を知る	15.2%	詳細を知る	32.9%
	概要を知る	11.4%	類似する他を知る	6.3%

#### 4.3 ブラウジング支援

提案方式の実現性と、実際にブラウジングの支援が可能か検証するため、プロトタイプAPを作成した．このAPでは、図2に示すように、ユーザがあるページを閲覧すると、閲覧候補を目的指向に分類して、ブラウザの右側に提示し、ユーザが提示されたページを選択することにより、閲覧候補の分類が更新される．これらの操作を繰り返すことにより目的に沿ったブラウジング支援が可能となる．



図2．ブラウジング支援の例

#### 5. まとめと今後の課題

閲覧中WEBページのページタイプによる目的指向WEBブラウジング支援の提案を行い、実験とプロトタイプ実装により方式の正当性と実現性を検証した．今後は、ページタイプの自動分類と提案方式の評価を行う予定である．

#### 参考文献

- [1] 池田新平, 是津耕司, 小山聡, 田中克己: Web コンテンツの周辺情報提示によるナビゲーション支援, DEWS2003 8-P-02
- [2] 川前徳章, 高橋克己: Web コンテンツの機能に着目した検索結果の構造化に関する提案, 人工知能学会 第5回セマンティックWEBとオントロジ研究会 2004