

ポータルサイトにおける運営支援システムの構築

小林 拓也[†] 岡本東[†] 堀川三好[†] 菅原光政[†][†]岩手県立大学ソフトウェア情報学研究所

1. はじめに

近年、Web 上の情報検索によって多くの情報が収集される。それらの情報は有益無益が混在し、量が多いため、取捨選択が必要とされる場面が増加している。そのため、予め取捨選択が行われ、情報が分類・体系化されているポータルサイトが増加している。

各ポータルサイトは、それぞれ異なる目的を持つが、共通の課題として、利用者を確保する必要がある、そのために運営者は様々な作業を継続的に行うことが重要である。

本研究は、ポータルサイトの運営支援を目的に掲げ、「閲覧数の増加」「コンテンツ数の増加」「運営者の負荷軽減」を行う仕組みを提案し、運営に必要な作業に対して支援するためのポータルサイト運営支援システム（以下、提案システム）を構築する。提案システムを既存のポータルサイトに導入・運用することで有効性を検証する。

2. ポータルサイトの運営に必要な作業

2.1 作業の要求

利用者を確保するために運営者が行う作業¹⁾は「収集」「分類」「提示」「フィードバック」の4つに整理できる。これらの各作業は表1の様な要求が存在している。

2.2 要求に対する支援

各作業の要求に対して表1の支援目的を掲げ、次の手法を提案する。収集の作業に対しては、リンク集の作成支援を目的とし、信頼性が高く²⁾、且つポータルサイトのテーマに一致したWebサイト・ページを自動収集する。テーマとは対象地域と対象分野の2軸で考える。分類の作業に対しては、リンク集の自動分類を目的として、文書分類を用いて運営者の意思を反映させたカテゴリ分類を行う。提示の作業に対しては、利用者の検索目的に合わせた情報提示を目的とす

Construction of a Support System for Portal Site Management

Takuya KOBAYASHI[†], Azuma OKAMOTO[†], Mitsuyoshi HORIKAWA[†], Mitsumasa SUGAWARA[†]
Graduate School of Software and Information Science Studies,
Iwate Prefectural University

表1 運営者が行う作業に対する要求と支援目的

作業	要求	支援目的
収集	テーマに一致した情報収集	公的サイトを起点としたクローリングによるリンク集の作成
	リンク集・独自コンテンツの増加	Webページの内容によるポータルサイトのテーマに一致した選別
	更新頻度の増加	
分類	多くの情報とカテゴリの一致	適合フィードバックを持った文書分類
	利用者が検索しやすいカテゴリの整理	によるWebページの自動分類
提示	利用者の検索目的に合わせた提示	個人プロフィールの作成と利用による情報推薦
フィードバック	利用者のニーズを反映した意思決定	時系列分析による話題の提示

れに基づいて情報推薦を行う。フィードバックの作業に対しては、運営者が作成するコンテンツの増加を目的として、運営者に対して流行している話題などを提示する。

各手法のロジックについては検証実験を行い、各作業の要求を満たしている事を確認した。

3. ポータルサイト運営支援システムの構築

これらの手法をサブシステムに持つ提案システムの構築を行う(図1)。以下に各サブシステムの処理と対象者が行う作業について示す。

(1) 収集サブシステム

運営者は、ポータルサイトのテーマである対象地域を示す地域名、対象分野を示すキーワードを設定する。このサブシステムは設定されたテーマに一致したWebサイト・Webページを自動的に収集し、運営者に対して提示する。運営者はそれらに対して公開の判断を行う。

(2) 分類サブシステム

運営者はカテゴリを設定する。このサブシステムは、設定されたカテゴリの学習文書をWeb上から取得することで、分類ルールの作成を可能とする。この分類ルールを基に収集サブシステムで収集されたWebページを分類し、運営者に対して提示する。運営者はシステムによる分類が適切でない判断すれば再分類を行う。

(3) 提示サブシステム

利用者がリンク集を閲覧することで、プロフィールが作成され、そのプロフィールに基づいて、利用者の検索目的に一致すると思われるWebページを提示する。プロフィールは利用者がリンク集を閲覧するたびに更新される。

(4) フィードバックサブシステム

収集サブシステムで収集されたWebページに対して時系列分析を行い、ある一定期間内に出

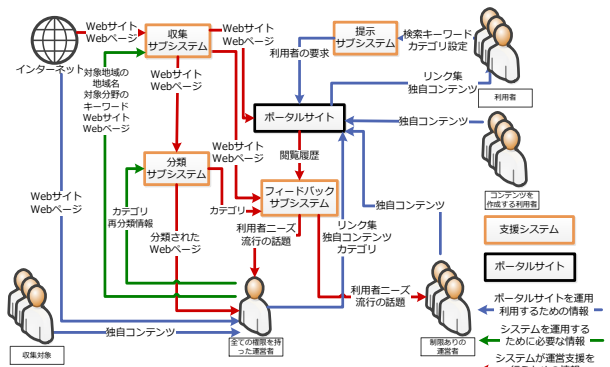


図1 ポータルサイト運営支援システムの構成図

現回数が高い単語を話題のキーワードとして運営者に対して提示する。運営者はそれを基に取材対象の決定や記事作成などの意思決定を行う。

4. ポータルサイトへのシステム導入と運用

提案システムを既存のポータルサイトへ導入し、運用によって得られるデータから有効性を示す。導入を行ったポータルサイトは「岩手県私立幼稚園ポータルサイト」であり、保護者や地域住民を対象とし、岩手県の各幼稚園のおたよりやイベントなどの情報を一元管理し、配信している。また、テーマは対象地域「岩手県」対象分野「子育て」である。提案システムの運用は2012年8月から行なっている。

4.1 ポータルサイトに及ぼす影響

アクセスログを用いてシステムを導入しなかった場合のアクセス数の予測と実績の比較を行う。2009年3月から2012年7月のデータを用いてARIMAモデルによって2012年8～12月の予測を行った。図2に結果を示す。予測値より実績値が増加していることが読み取れ、また、増加傾向があることが分かる。これにより、提案システムを導入したことによって、アクセス数が増加したことが考えられる。

4.2 運営者の負荷とコンテンツの増加

収集サブシステムと分類サブシステムに対して運営者が行う作業の負荷とコンテンツの増加について述べる。

(1) 収集サブシステム

運営者が行う1日の作業時間と収集・更新ページ数の月の平均を以下に示す。

- 1日の作業量(秒) : 109.838
 - 収集・更新ページ数の月平均 : 50.25
- 1日に約2分の作業を行うことで月約50件のページの更新・収集が可能であることがわかった。

(2) 分類サブシステム

分類サブシステムによって自動分類されたWebページに対しての分類精度をF値で示し、

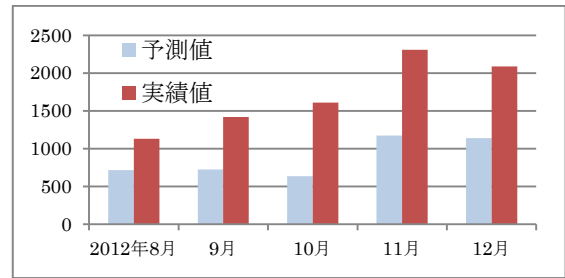


図2 訪問数の予測値と実績値

表2 作業負荷と分類精度

分類数	運用中にカテゴリに分類されたWebページ数	57
自動分類数	システムによる分類数	39
再分類数	運営者による分類数	21
適合率	$= (\text{分類数} - \text{誤分類数}) / \text{分類数}$	0.923
再現率	$= (\text{自動分類数} - \text{誤分類数}) / \text{分類数}$	0.632
F値	$= 2 / (1 / \text{適合率} + 1 / \text{再現率})$	0.750
再分類率	$= \text{再分類数} / \text{分類数}$	0.368

運営者が誤分類を修正した割合を作業負荷として表2に示す。このサブシステムは分類すべきWebページに対して約4割弱の再分類を行うことによって利用が可能である。

5. 汎用性に関する考察

提案システムはテーマの設定やカテゴリの設定を運営者の意思に基づいて行う事で、様々なポータルサイトへ利用可能である。

運用から得られた知見として、リンク集の有無や対象者が明確にされているかなど、提案システムを導入できる条件について明らかになった。また収集されたWebページをポータルサイトが持っている機能と連携するなど、ポータルサイトの特性に合わせた利用方法が考えられる。

6. おわりに

本研究ではポータルサイトの運営に必要な作業について明らかにし、各作業の支援を行う手法について提案した。手法に基づいて提案システムの構築を行い、既存のポータルサイトに導入・運用を行い有効性について明らかにした。

今後の課題として、様々なポータルサイトに導入し、他のポータルサイトへの有効性、汎用性について検証する必要がある。

参考文献

- 1) 田中秀樹, 使われるポータルサイト～ポータルソフトウェア開発を通じて～, 情報の科学と技術, 54巻8号(2004)
- 2) 鷺見克典, 四谷あさみ: 調べる目的で利用する情報源としてのWebサイトに対する評定尺度の作成と信頼性及び妥当性の検討, 情報処理学会論文誌, Vol.45No.3, pp1032-1040(2004)