

梅村 武久*1*2、安田 孝美*1*3

通信・放送機構*1 NTT西日本 マルチメディア技術開発センタ*2 名古屋大学 情報文化学部*3

1. はじめに

今後、教育シーンでのインターネット利用は必要不可欠なものとなることが予想される。しかし、限られた授業時間内に欲しい情報の発見や入手が困難であるといった問題がある。

この問題を解決するために、筆者らは情報の検索結果をカテゴリ分類体系に整理して表示することにより、次のアクションを決定するための情報を利用者に提供するWWW検索システムを、通信・放送機構岡崎公共システム開発リサーチセンタに構築し、WWW情報の分類体系化を試みた[1]。本方式で、情報を教育向けにカテゴリライズする自動分類定義ファイルを作成したが、不適切なキーワードの設定や類似カテゴリの判定が困難などの問題により、分類精度は充分と言えず実用に耐えうるものではなかった。

本稿では、その後のフィールド実験による分類精度の向上と利用者アンケートの結果について報告する。

2. 分類精度の向上

WWW情報の分類手法としては、URL名による分類やハイパーリンクの共起による分類[2]など様々な手法が考えられる。本研究では、利用者の環境を教育分野に特化することで、実際の利用者要望や検索ログデータをフィードバックしカテゴリ項目を体系化、更に分類キーワードを繰り返し修正することで、実運用を目指した教育向けカテゴリ分類体系を確立する。

図1の自動分類定義ファイルを用いて、148URLサイト(約 17,000 ページ)を分類した結果、収集ページに対する分類率(分類済みページ/収集総ページ)は約90%と高かったが、目視による分類結果の精度は約40%程であった。精度が低かった原因として、(1)複数

The Classification/Systematization of WWW Information for Education and its Proof

Takehisa Umemura(NTTWEST Multimedia Technology & Development Center), Takami Yasuda(Nagoya University)

/学校の授業: /学校の授業/小学校: /学校の授業/小学校/国語:絵文字,民話,語源,読み方,辞書,ことば,字,片仮名,昔話,語・・・ /学校の授業/小学校/算数:算数,たし算,ひき算,算用,角度,ものさし,定規,数字,数,け・・・ /学校の授業/小学校/理科:ふしぎ,空気,太陽,花粉,摩擦,月,葉っぱ,力,雲,ムシ,葉緑素・・・ /学校の授業/小学校/社会:土器,縄文,曜日,地球儀,古代人,堅穴,恐竜,化石,弥生,ダム,・・・ /学校の授業/小学校/図工:はさみ,のり,カッター,クレヨン,ポンド,ホッチキス,セロテ・・・
↓
/役立つ情報/動物・植物:絶滅,希少,高山,草花,水生,葉緑素,知的,集団,動物,脊椎動物・・・ /その他:乗り物,ビル,プロペラ,おうち,核的,キャベツ,容積,滝,ベルシヤ,物理学,自由・・・

図1 自動分類定義ファイル

の分野に出現するキーワードが数多く定義されていた。例えば、算数カテゴリのキーワードに「円」があるため、値段が表示されているようなWebページが数多く算数カテゴリに出現した。(2)キーワードの頻度を加算して得点の高いカテゴリにWebページを割り当てるが、頻度の均一的な加算に問題があった。(3)分類するカテゴリのないWebページが全く意味のないカテゴリに出現した。(4)類似カテゴリの判定が困難であった。例えば、カテゴリ「小学校の国語」と「中学校の国語」では、差別化できるキーワードが数少なかった。

分類精度の向上を図るため、以下の方法により自動分類定義ファイルを修正した結果、総カテゴリ数は127、総キーワード数は3391となり、分類精度を約70%に向上させることができた。

a. 複数分野に出現するキーワードの削除

該当するカテゴリに特徴的でない(他のカテゴリにも多数出現する)キーワードを削除した。

ex) カテゴリ名: /学校の授業/小学校/国語

字、文、辞書、話、文字、言葉、表現などを削除

ex) カテゴリ名: /学校の授業/小学校/算数

数、時計、時刻、円、面、度、時、分などを削除

b. キーワードの重み付けと追加

最も特徴的と思われるキーワードに対してn倍の重み付けをすると共に、新たなキーワードを追加した。

ex) カテゴリ名: /学校の授業/小学校/算数

重み付け: 算数(n=3)、さいころ(n=2)

追加キーワード: 三角形、四角形、偶数、奇数など

c. 新カテゴリの追加

分類されるべきカテゴリの無いページを抽出し、新たなカテゴリを追加した。

ex) /役立つ情報/行政/郵政省,文部省,etc・・

/役立つ情報/医療 など行政,医療カテゴリの追加

d. カテゴリの統合と削除

分類が困難な類似カテゴリを統合した。

ex) /学校の授業/小学校/国語と/中学校/国語

統合カテゴリ名:/学校の授業/国語

3. 利用者アンケート結果

現在、岡崎市の中学校18校に本検索サービスを試行中であり、各校の教員に対し評価・要望などのアンケートを実施した。回答者は46人で、以下はその主な結果である。

3.1 教育向けカテゴリ分類項目の評価

「分かり易く適格なカテゴリが多い」といった最高評価を5とし5段階評価において、表1の結果となった。ほぼ満足できる評価であったが、2以下の評価が11名あり今後の検討の余地があると言える。

表1 カテゴリ分類項目の評価

評価	5	4	3	2	1	無記
人数	2	14	17	10	1	2

3.2 画面インタフェース(GUI)の評価

カテゴリ分類結果の表示方法は、図2に示すようシンプルかつ直感的な操作が可能となるようフォルダタイプの階層表示となっており、利用者の評価は同じく5段階評価で表2の結果となった。2以下の評価が6名と少なく、操作性の問題は少なかったと言える。

表2 画面インタフェースの評価

評価	5	4	3	2	1	無記
人数	4	11	23	6	0	2

3.3 WWW検索サービスへの要望

今後の改善要望や不具合点の主な内容として、(1)上位カテゴリから想定しにくい下位カテゴリがある、(2)小学生にはカテゴリ名やGUIが難しい、(3)教師 or 生徒向けの区別ができると良い、(4)リンク集を充実して欲しいなどがあった。

3.4 インターネットから入手したい情報

教科毎の素材データや学校通信に利用するスクールカット集を入手したいとの意見が最も多く、次いで(1)教

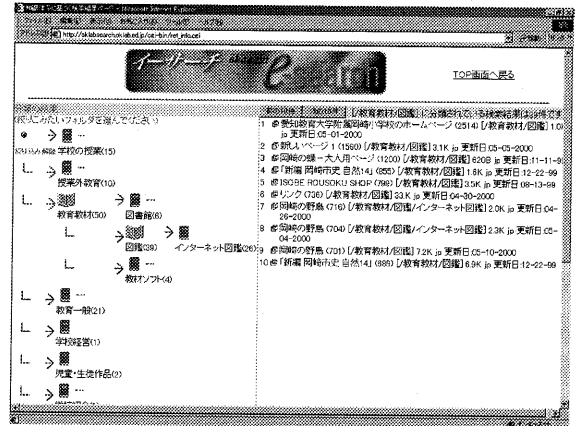


図2 画面インタフェースの例(分類表示)

科の指導案や他校の例、(2)調べ学習や総合学習といった新教科の情報、(3)社会見学や修学旅行など学校行事情報の他、(4)天候・ニュース・路線検索などを希望する声も多かった。

4. 考察

検索対象を教育分野のWWW情報に特化して、カテゴリ項目の検討や分類精度の向上を繰り返したことで、教育分野向けカテゴリ分類体系の基盤を確立した。今後は、検索対象URLの増加やWWW以外の情報も分類の対象とすることで更に汎用性を高める必要がある。

利用者アンケートについては、本論文では一部しか記載できなかったが多くの意見を頂いており、今後のカテゴリ項目検討や検索サービスの改善に向け、有効なデータを得ることができた。しかし、評価や要望には相反する意見も多く、どのように整理するか問題である。

5. おわりに

教育分野におけるWWW情報の分類体系化を中心に、WWW検索サービスのフィールド実験状況を述べた。現在、作成した自動分類定義ファイルを、Oracle やSQLなどのデータベースサーバに対しても適用可能となるシステムを構築中であり、今後、本ファイルの汎用性を更に高めるため実験を続けていきたい。

参考文献

- [1]梅村武久,安田孝美:教育分野におけるWWW情報の分類体系化,第60回情報処大全,pp4-433~434,(2000).
- [2]大久保雅且,杉崎正之,田中一男:リンクの共起関係を用いたWebページ分類法式の検討,第59回情報処大全,pp3-81~82,(1999).