

ディレクトリサービスにおけるカテゴリ再構成作業支援に関する一検討

4Q-8

畑島 隆 元田 敏浩 川崎 隆二 山本 修一郎

日本電信電話株式会社 ソフトウェア研究所

(takashi,motoda,kawasaki)@canary.sl.cae.ntt.co.jp,yamamoto@slab.ntt.co.jp

1.はじめに

内容でカテゴリ分類したディレクトリ構成によって、インターネット上の情報を検索しやすく提供するサービスが行われている。現状のディレクトリサービスではディレクトリの再構成作業は人手により行われている。しかし、カテゴリ分類の追加・削除・変更対象の抽出作業には、以下の問題があった。

- ・全ジャンルのしらみ潰しの検索では作業効率が悪い
- ・担当者の経験に依存するため、偏りが発生する

本稿では、現状のカテゴリ再構成作業について説明し、各カテゴリに属する情報ページの参照履歴に基づいて、再構成すべきカテゴリを自動的に提示する手法を提案する。

2. 現状のカテゴリ再構成作業

2.1 カテゴリの構成

ディレクトリサービスにおけるカテゴリの構成を図1に示す。カテゴリはWeb ページに含まれる情報の内容に従って、大分類・中分類・小分類があり、ある分野の大分類で分類された中でさらに詳細な分類により中分類が存在し、さらに細分化された小分類が存在する。

2.2 カテゴリ再構成の業務フロー

ディレクトリサービスでは、数千のカテゴリが存在する。カテゴリは一度設定ものが永久に利用される訳ではなく、随時変更されており、カテゴリが再構成されている。図2にカテゴリ再構成の業務フローを示す。

現状のカテゴリ再構成案の検討作業においてカテゴリ再構成候補は、表1に示す基準項目により担当者の判断に依存して決定されている。表1のうち(1-1)(1-2)(1-4)は、担当者が着目したカテゴリのみについて検討されている。例えば、表1(1)については担当者が目星をつけたカテゴリについて、日別のアクセス数の一覧をその都度導出し、最近アクセス数が伸びていると判断したカテゴリを再構成の対象としている。この手法では、数千あるカテゴリ数全部について検討すると作業効率が悪くという問題がある。また、この時点で担当者により見落とされているカテゴリが存在し、着目されたカテゴリについても、担当者の判断によって再構成するカテゴリを決定されているため、担当者の偏りに偏る可能性がある。

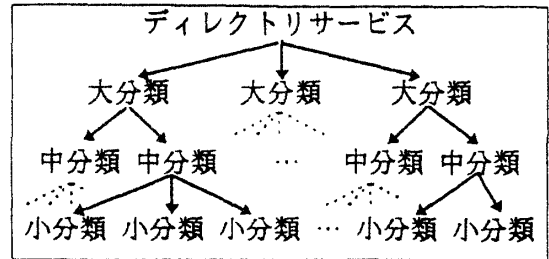


図1:ディレクトリサービスのカテゴリ構成

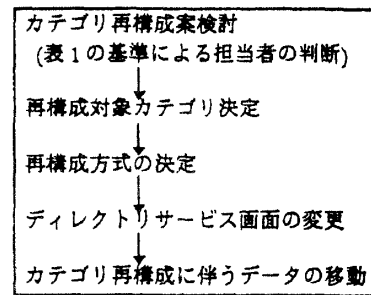


図2:従来のカテゴリ再構成業務フロー

- (1-1)各カテゴリのアクセス状況
- (1-2)各カテゴリの収容ページ数
- (1-3)季節の事項、話題の事項
- (1-4)検索エンジンの投入キーワードログ
- (1-5)利用者からの要望

表1:カテゴリ再構成時の判断基準項目

- (2-1)追加
現在カテゴリが存在していない情報ページを収録し恒常的に使用するカテゴリを新規追加する
- (2-2)削除
アクセス数や登録情報ページ数が少ないカテゴリを削除する
- (2-3)分化
存在するカテゴリのうちアクセス数の増加が認められたり、登録情報ページ数がある程度多くなったものを細分化する
- (2-4)特集
季節的なイベントや社会的ブームにより一時的に多くのアクセスが見込まれる既存のカテゴリを一時的に新規追加する

表2:カテゴリ再構成手法の分類

Reconstructing categories of the Internet Directory Server

Takashi Hatashima, Toshihiro Motoda, Ryuji Kawasaki, Shuichiro Yamamoto

NTT Software Laboratories

¹ NTT DIRECTORY <http://navi.ntt.co.jp/>

Yahoo <http://www.yahoo.com/> など

² NTT DIRECTORY では約2000ジャンル

2.3 カテゴリの再構成方式

カテゴリ再構成の方式は表2に示すように分類される。(2-1)追加は、恒常的に提供するカテゴリを新規作成する。(2-2)削除はアクセスが見込めない等の理由によりカテゴリを廃止するものである。これは現状おほとんど実施されていない。これはアクセスの少ないカテゴリは担当者の目につかないため再構成の対象とならないことや、再構成によるアクセス増加の見込みが少ないカテゴリまで手が回らないことが挙げられる。

3. カテゴリ再構成支援に関する検討

3.1 再構成支援の方針

本方式では、再構成において担当者の負担が大きい「カテゴリ再構成対象の決定」を支援する。図3に示すように、カテゴリ再構成支援の方式では、再構成を推奨するカテゴリの候補を提示する。担当者は提示された候補から実際に再構成するカテゴリを決定し、ディレクトリサーバへの反映作業を行う。本方式は、従来担当者が各ジャンルについて経験に依存した再構成候補カテゴリ決定の自動化を目的とする。本方式を用いたカテゴリ再構成の作業フローを図4に示す。

3.2 再構成候補決定指標

再構成候補の決定に用いる指標の条件を表4に示す。現在アクセス頻度として使われているPageView³⁾に加えて、時系列のアクセス傾向を考慮した指標 [2] を導入し、アクセス履歴を数値化して提示する。また、各ジャンルに含まれる情報ページ数も指標として用いる。

3.3 再構成候補カテゴリの提示方法

担当者の実際の業務をヒアリングした結果をもとに決定した、前述の指標による再構成候補カテゴリの提示条件を表3に示す。ディレクトリサーバのアクセスログを解析し、アクセス頻度の小さいカテゴリは(3-2)削除の対象、アクセス頻度が高いカテゴリは(3-3)分化の対象として提示する。(3-4)特集は終了時期の提示のみを行う。(3-1)追加および特集の開始時期の提示については行わない。これらの事象は、アクセスログを参照する事により得られる情報少なく、むしろ担当者の経験に依存したほうが有効であると考えられるためである。また、各ジャンルの情報ページ数を比較し、特に含まれる情報ページ数が多いページを表3(3)の分化の対象として提示する。

3.3 評価方法

以下の項目を本方式の評価項目として実システムに適用し評価する予定である。

- ・担当者が選定する再構成対象カテゴリの不足ない提示
- ・従来手法と比べた際の作業時間の低減量
- ・カテゴリ再構成前後のアクセス数の変化

4. まとめ

ディレクトリサービスの運営において問題になっているカテゴリの再構成に対して、再構成候補を提示する事により支援

するための手法を提案した。今後はここに述べたシステムを作成し、有効性を評価していく予定である。

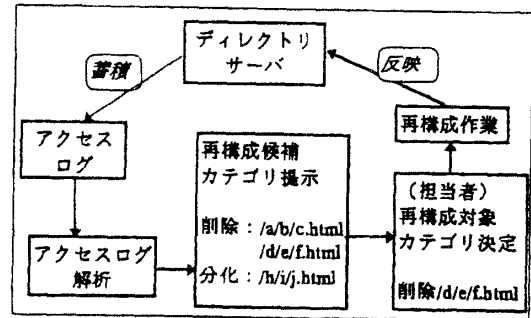


図3：カテゴリ再構成支援の全体図

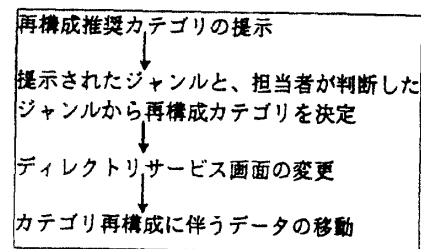


図4：提案方式を用いたカテゴリ再構成作業フロー

- (3-1)追加
キーワード検索のログ解析等が考えられるが、担当者の経験に依存した方が有効であるため考慮しない
- (3-2)削除
アクセス頻度が低い状態が連続しているカテゴリを提示
- (3-3)分化
a)同一分類に属するカテゴリのうち、突出してアクセス頻度が高い小カテゴリを、同一カテゴリ内で複数カテゴリへの分化対象として提示
b)含まれる情報ページ数が多いカテゴリを分化対象として提示
- (3-4)特集
a)特集の設定は(3-1)追加と同様に考慮しない

表3：再構成候補カテゴリの提示条件

- ・再構成候補の決定には以下の指標を用いる
 - 1)PageView
 - 2)時系列によるアクセス傾向を数値化した指標
 - 3)カテゴリに含まれる情報ページ数
- ・全カテゴリに対して指標を求める

表4：再構成候補決定指標の条件

参考文献

[1] NTT DIRECTORY <http://navi.ntt.co.jp>
 [2] 畑島, 元田 "時系列情報を考慮したアクセスログ解析" 第54回情処全大, 分冊3, pp217-218, 1997

³⁾ wwwサーバから引き出された結果、表示されたWebページ数。
 例えばYahoo Japanの広告料金設定で利用されている。
<http://www.dwcj.co.jp/cyberagency/yahoo.htm>