

MediaWiki のカテゴリ編集支援システムの開発

瀬高 拓也[†] 矢吹 太朗 佐久田 博司

青山学院大学 理工学部 情報テクノロジー学科[‡]

1 序論

MediaWiki のカテゴリ構造をグラフィカルに編集をするシステムを構築する。

1.1 研究背景

Wiki は Web 上で文書を編集できるシステムである。これを活用することによって、複数人で効率的に文書を作成することができる。しかし、文書の数が増えていくと、Wiki のリンク構造のみでは編集者は文書の整理を行うことが難しくなる。このページ構造化の問題を解決するために、Wiki パッケージの一つである MediaWiki にはカテゴリ機能がある。これはページを関連するテーマでまとめる機能である。編集者は以下の方法でカテゴリの作成を行う。

1. カテゴリに属するページに移動する
2. 編集エリアに `[[category:カテゴリ名]]` と記述する
3. 該当するページ全てに 1 と 2 を行う

しかし、この方法には 2 つの問題がある。

1. MediaWiki 内のカテゴリ構造を把握ににくい
2. ページが所属しているカテゴリを変更するには、そのページ自体を編集しなければならない

この 2 点から編集作業のコストが負担となり、利用者はカテゴリ機能を活用できなくなる恐れがある。これを解決するためには、利用者のカテゴリ機能に対するユーザインタフェースを改善する必要性があると考えられる。そこで以下の機能を持つようなシステムを開発する。

- カテゴリ全体像をツリー形式で表示できる
- ページやカテゴリを GUI で編集できる
- 既存の MediaWiki 上で使用できる

1.2 関連研究

1.2.1 Extension:CategoryTree

MediaWiki には CategoryTree という拡張機能があり、カテゴリの内部をツリー状に表示することができる。しかし、編集者が表示したいカテゴリのタイトルの指定をする必要があり、カテゴリ数が多い場合に全てのカテゴリを把握して記述することは困難になる。この拡張機能で実現されるのはカテゴリの表示だけであり、編集したい場合には 1.1 節で述べた手順を実行しなければならない。

1.2.2 シソーラスの活用

渡邊らはシソーラスに着目し、精度の高いページ分類を行おうという取り組みをしている [2]。シソーラスとは言葉の類似関係によって分類したデータベースである。ページ内、ページ間で分析を行い、シソーラスを作成していくことで、同じテーマを持ったページを分類している。これにより、ユーザのカテゴリライズをサポートできる。しかし、この研究の目的は自動カテゴリライズであるため、編集者がカテゴリを自分で変更したい場合には、1.1 節で述べた手順を実行しなければならない。

1.2.3 Google datawiki

Google labs は構造化データを使いやすくする Wiki として datawiki をリリースしている [1]。この datawiki は構造化データを容易に編集、共有、可視化し、それぞれを相互にリンクさせることができる。しかし、MediaWiki のような既存の Wiki との互換性がないという問題がある。

2 開発したシステム

2.1 システムの利用方法

本システムを利用する編集者は以下の方法でカテゴリライズを行う。

1. 編集者が本システムの GUI にアクセスする
2. ツリー形式で表示されたカテゴリやページをドラッグアンドドロップで編集する
3. 編集結果を保存する

2.2 システム構成

本システムは MediaWiki とカテゴリ編集用 GUI、両者を仲介する中継サーバからなり以下のように動作する (図 1)。

Graphical category editor for MediaWiki.

[†] Takuya SETAKA (a5807043@cc.aoyama.ac.jp)

[‡] Department of Integrated Information and Technology, College of Science and Engineering, Aoyama Gakuin University

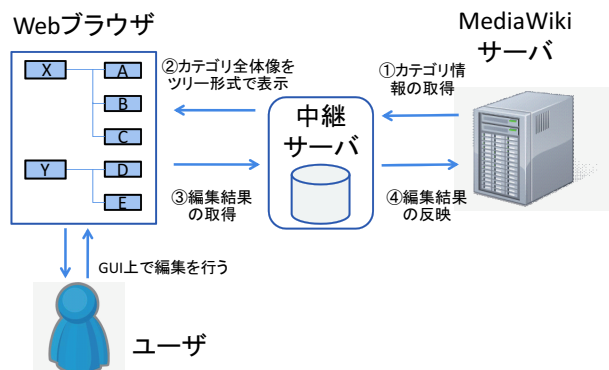


図1 システム構成 (図中の番号は本文のそれと対応している)

1. ユーザが GUI にアクセスした段階で MediaWiki のカテゴリ情報を中継サーバが取得する。
2. その結果をツリー形式の GUI として編集者に表示する。
3. 中継サーバは編集者がボタンを押した段階の編集結果を取得する。
4. 中継サーバが編集結果を MediaWiki に反映する。

2.3 システムの実装

2.3.1 カテゴリ情報

システム間で扱うカテゴリ情報は MediaWiki 内にある以下のデータで構成されている。

- カテゴリの ID と名前
- 属しているページの ID と名前
- ページ (カテゴリ) とカテゴリ間の所属関係

2.3.2 カテゴリ情報の格納

中継サーバは MediaWikiAPI を使ってカテゴリ情報を取得している。API を利用することによって MediaWiki のバージョンアップや仕様変更に対応することができる。ただしカテゴリの ID は API では取得できないため、MediaWiki のデータベースから直接取得している。格納先である中継サーバのデータベースでは、アイテムと関係性のテーブルを扱う。

2.3.3 カテゴリ情報の展開, 編集

編集を行う GUI では中継サーバにあるカテゴリ情報を取得し、ツリー形式の動的なカテゴリ構造を提示する。編集者はその GUI を見ることで、現在の MediaWiki 上のカテゴリとページの関係性を俯瞰できる。その GUI では編集者がドラッグアンドドロップでグラフィカルにカテゴリライズを行うことができる。

2.3.4 編集結果の反映

編集者は編集結果をボタンを押して中継サーバ内のデータベースに編集結果を送ることができる。その結果を中継サーバは取得し、MediaWiki に反映することができる。MediaWiki への反映には MediaWikiAPI を用いている。

2.3.5 実装の詳細

中継サーバは MediaWiki からカテゴリ情報の取得と反映をする PHP のプログラムと、MySQL で構築されたデータベースを持っている。GUI は JavaScript で実装し、PHP で作成したプログラムから情報を取得することで HTML 上にカテゴリツリーを表示している。ドラッグアンドドロップによるアイテム操作は jQuery の Draggable, Droppable プラグインを用いている。

3 結論

本研究では、MediaWiki のカテゴリ機能をグラフィカルに操作するシステムの開発を行った。このシステムによって編集者は複雑になったカテゴリ構造の全体像を把握することができる。これによって Wiki を有効に活用する場面が増えることが期待される。

4 今後の課題と展望

本システムの課題と展望を以下に述べる。

4.1 複数人作業における情報更新の課題

本システムでは編集者がボタンを押すことで編集結果を反映させるリクエストを送信する。現在の方法では、編集者 A が GUI で編集し、反映させるまで (図 1 の (3)) に他の編集者 B の結果が反映されると、変更箇所を把握していない編集者の結果で上書きされる。この状況では上書きされてしまったの編集者 A の変更箇所が消失してしまうという問題がある。その問題を改善するシステムとして以下のようなシステムが考えられる。

- 即時更新性を持ったドラッグアンドドロップ
- 更新履歴などのログを扱ったバージョン管理

しかし、根本的な解決をするためには、複数人における編集のトランザクションについて詳しく検討する必要がある。

4.2 汎用データへの一般化

本システムは多数のデータを持つものならば Wiki に限らず適用できると考えられる。図 1 の (1), (4) に当たるシステムをデータに合わせて作成し、現在のシステムから変更することでカテゴリライズできると考えられる。それにより、より効率的なデータ整理が行えると期待される。

参考文献

- [1] Lawrence Edward Larry Page. Google datawiki. <http://datawiki.googlelabs.com/>.
- [2] 渡邊陽太郎, 浅原正幸, 松本祐治. HTML の木構造を利用した条件付き確立場による固有表現分類: Wikipedia からのシソーラス半自動構築. *Information Processing Society of Japan*, Vol. 47, pp. 73-78, 2007.