

# XML を用いた歴史文書のタグ付け支援ツール

安孫子一敏<sup>†</sup> 中平勝子<sup>†</sup> 三上喜貴<sup>†</sup>

長岡技術科学大学<sup>†</sup>

## 1. はじめに

歴史を電子的に記述するという試みは複数行われており、それぞれ特徴ある仕様や記述項目が提案されているが、標準となるフレームワークはまだ存在しない。筆者は、より簡便な仕様を目指して松井ら [1] と共に XML を用いた歴史記述の枠組みの提案と記述項目の定義を行っており、提案された枠組みを用いて多角的な視点から歴史を分析できるようにしたいと考えている。その実現には歴史記述の枠組みを用いたデータベースによる情報共有、歴史データのタグ付け支援、歴史データの可視化などの仕組みが必要となる。これまで筆者らは、歴史文書へタグ付けしていく際のユーザのタギング作業補助のためのタグ付け支援ツールの実装を試験的に行った [2]。本稿では、試験的に実装したタグ付け支援ツールの改良を行う。

## 2. 研究の全体像

ツールはタグ付け支援ツールの部分と可視化ツール部分から構成され、データは DB によって共有される。これらのツールは Web アプリケーションとし、ユーザは Web ブラウザから操作を行うことができる。

タグ付け支援ツールは歴史データを入力・編集・保存する機能を持っており、データ入力時にユーザの作業を補助する。この機能はユーザのデータ入力を助け、入力されたデータの整合性を保つ目的がある。

## 3. 歴史事象の記述方法

表 1 概念モデル

| 要素名   | 内容                 |
|-------|--------------------|
| イベント名 | イベントを端的に表すラベル      |
| 時間    | イベントが起きた時間         |
| 場所    | イベントが発生した場所        |
| 人物    | イベントの登場人物          |
| 原因    | イベントが起きる原因となったイベント |
| 証拠    | イベントの存在を裏付ける証拠     |

歴史事象記述の枠組みは松井ら [1] によって提案されたものを採用する。ドキュメントは人間が理解

しやすいう microformats で文書を記述した後に、歴史の資料情報で汎用形式となっている RDF へ変換する。歴史事象記述の枠組みの概念モデルを表 1 に示す。本研究では、表 1 の定義に従って歴史事象を記述していく。

## 4. タグ付け支援ツール処理の流れ

タグ付け支援ツールのフローを図 1 に示す。ツールのメイン画面はエディタ編集画面である。エディタに入力された文章や単語をドラッグし、エディタのボタンクリックによって要素のマークアップする。マークアップに際して文章や単語の修正が必要な場合は、上記の作業中に行われる。マークアップの完了した文書は歴史ドキュメント保存 DB へ保存される。

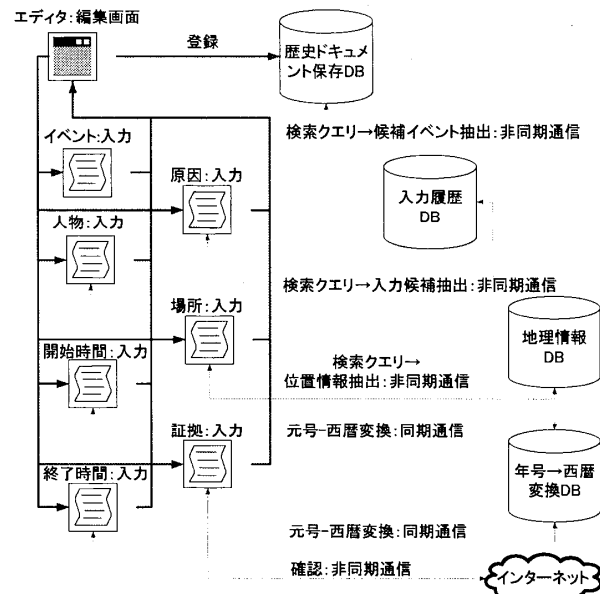


図 1 タグ付け支援ツールのフロー

タグ付け支援機能は人物名補完・場所名補完・年号名変換を用意した。人物名補完機能は文章中で人物名の省略が行われた場合にユーザを補助するため、入力履歴 DB との非同期通信を基に入力候補をユーザに示す。場所名補完機能は文章中の地理情報入力を補助するために、地理情報 DB との非同期通信を基に地理情報をユーザへ提示する。年号名変換機能は複数の形式で記述された暦を西暦に変換するため、年号変換 DB との同期通信を基に年号の変換を行う。

Tagging Support Tool for History Text Using XML  
<sup>†</sup> Kazutoshi Abiko, Katsuko T. Nakahira, Yoshiki Mikami  
 Nagaoka University of Technology

## 5. タグ付け支援機能の改良

### 5.1. 暦の統一

西暦にはユリウス暦やグレゴリオ暦が存在し、国ごとにそれぞれの暦を採用した年が異なっている。従って、暦を統一せずに歴史事象を記述すると、データの整合性がとれなくなる。そこで時間軸の基本となる暦を決定するため、他フレームワークの調査を行った。その結果を表2に示す。

表2 フレームワーク採用暦の比較

| フレームワーク名            | 年代記述の暦指定             |
|---------------------|----------------------|
| CIDOC CRM[3]        | 規定無し(運用側が規定)         |
| CDWA[4]             | グレゴリオ暦               |
| ミュージアム資料情報構造化モデル[5] | グレゴリオ暦               |
| HEML[6]             | グレゴリオ暦<br>(但し、他の暦も可) |

歴史に関するフレームワークである CIDOC CRM や CDWA やミュージアム資料情報構造化モデルや HEML を比較した結果、現在最も使用されている暦であるグレゴリオ暦が多数を占めた。また現時点で記述される歴史文書はグレゴリオ暦で記述されるため、入力のし易さからも年号名変換の西暦はグレゴリオ暦を採用し、各年号変換 DB はグレゴリオ暦への1対1対応を準備することとした。

### 5.2. ユーザ管理機能

多くの歴史文書を扱うためには、それぞれの歴史に対して専門家が記入することが望ましい。その意味では複数のユーザが本ツールを使用することを想定している。そのため、ユーザ毎のアクセス/編集制限を実装する必要がある。本稿では、ユーザ管理機能を強化した。

ユーザ管理機能強化は実装環境に汎用 CMS (Contents Management System) の XOOPS (eXtensible Object Oriented Portal System) [7] を採用することで対応した。XOOPS はモジュールと呼ばれるプラグイン型プログラムで機能を拡張でき、ユーザ管理の機能が初期状態から組み込まれている。これに静的コンテンツ用モジュールの pico [8] と JavaScript で動作するエディタの FCKeditor [9] を組み合わせて実装した。pico はタグ付けされた歴史文書の分類と整理を担う。FCKeditor へは従来のタグ付け支援ツールで動作していたマークアップ機能、タグ付け支援機能の移植を行った。結果、複数ユーザの管理・コンテンツへのユーザ毎アクセス制御が行えるようになった。実装画面を図2に示す。



図1 改良後の動作画面

## 6. おわりに

本稿ではタグ付け支援ツールの改良を行い、ツールに採用する暦の決定とツールの複数ユーザへの対応を行った。これにより、複数ユーザが同時に使用できるようになる。今後は、多くのユーザに試用してもらい評価と改良を進め、実際の運用を行いたい。

## 参考文献

- [1] 松井正志, 中平勝子, 三上喜貴「XML を用いた技術史の知識基盤表現」『FIT2007 第6回情報科学技術フォーラム』, pp. 425-427, 2007年
- [2] 安孫子一敏, 中平勝子, 三上喜貴「歴史事象記述の枠組みを用いた歴史教材入力支援ツールの実装」『FIT2008 第7回情報科学技術フォーラム』, pp. 307-308, 2008年
- [3] CIDOC CRM, <http://cidoc.ics.forth.gr/>
- [4] CDWA, [http://www.getty.edu/research/conducting\\_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)
- [5] ミュージアム資料情報構造化モデル, <http://webarchives.tnm.jp/docs/informatics/smmoi/>
- [6] HEML Project, <http://heml.mta.ca/heml-cocoon/>
- [7] XOOPS Cube, <http://jp.xoops.org/>
- [8] PEAK XOOPS サポート&実験室, <http://xoops.peak.ne.jp/>
- [9] FCKeditor, <http://www.fckeditor.net/>