

# メタデータの豊饒化にむけた公共的アノテーションシステム Web アノテーション付与・管理・再利用・変換のための基盤システム

## 1．背景

広義のアノテーションとは、古くから人間がおこなってきた原文へ二次情報を付加する行為である。情報が文書の形態で流通し始めた時代からそれほど下ることなく、この行為は広く行われてきた。原文の解説、図解、翻訳、推敲、意見など目的は様々であるが、原文と付加情報との関係性を空間的近さが担保するという表現上の制約と、原文に対して二次的な情報を持つことがアノテーションに共通する特徴と言える。

過去に開発者らは、アノテーションの共起頻度にもとづくコミュニティ発見と情報共有システムとして SmartCourier を開発した。SmartCourier では、電子文書へ行われる手書きアノテーションの位置や形状、対象キーワードの共起頻度から各ユーザ間の関心類似度を算出し、潜在的な興味関心、コミュニティの顕在化や適応的な情報共有、論文推薦などの機能を実装した。しかし、SmartCourier は事前に特殊形式で登録した文書のみを対象としている。これは、文書の資料性や正統性とアノテーションの二次性というバランスを前提としたためであるが、原本の原著者を含む流動性の高い情報コミュニティを考えたとき、原著者がアノテータの指摘をすぐに反映するといった、情報共有媒体としての随意性に欠ける。一方、Web においては、その膨大なリソースの二次的利用を目指し、リソースに対するメタデータ付与とフィードの気運が高まりつつある。その原動力となっているのはメタデータの標準化と CMS(Content Management System)の普及である。しかし、メタデータとして記述される内容は、例えば著者名や投稿日時といった CMS が扱う範囲に限定されているのが現状である。なお、メタデータもリソース(原文)に対する二次情報であることからアノテーションと呼ばれることがあるが、ここでのアノテーションとは、原文の著者(以後、原著者と呼ぶ)を含むすべてのアノテーション行為をおこなう人間(以後、アノテータと呼ぶ)が様々な目的で人為的に付与する二次情報を指し、機械的に付与される類の二次情報とは分けて考える。

## 2．目的

本プロジェクトは、Web 利用者の膨大なマンパワーを背景としたアノテーション付きメタデータの流通を目標として掲げ、そのための基盤となる公共的アノテーションシステムの開発を目的とした。アノテータの視点から記されたメタデータは、既存メタデータだけでは叶えられない新しいサービスやコミュニケーション形態を生み出す可能性を秘めており、そのプラットフォームとなる本システムの社会的意義は大きい。この社会的意義を増大させるため、本プロジェクトは、膨大な量のアノテーションの流通、つまり相当数の利用者獲得を目指した設計・開発としている。具体的には、Web 上で任意に更新される HTML コンテンツへのアノテーション付与機能、アノテーション管理機能を実装した。これにより、コンテンツ作成者とアノテータ間で情報の相互参照が可能になり、Web における一次・二次情報の豊饒化へ繋がると考えた。

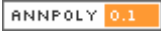
### 3 . 開発の内容

本開発プロジェクトにおいて作成した公共的アノテーションシステムを「annpoly」と名づけた。本システムは、Web サーバに設置され、リソース原著者がHTML に埋め込んだリンクから呼び出される。システムは、呼び出し元のHTML リソースを読み込み、アノテーション入出力インタフェースを実現する JavaScript を埋め込んで再発行する。蓄積されたアノテーションは、入出力インタフェースによるアノテーション表示以外に応用システムからも参照され、要約表示などの二次利用に用いられる。

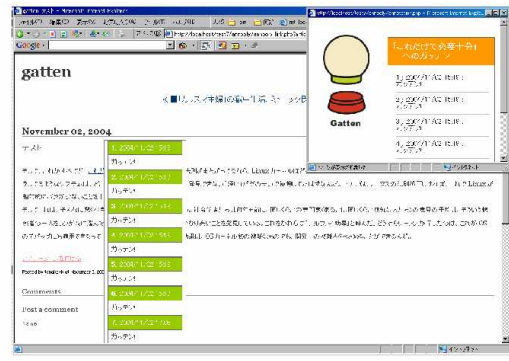
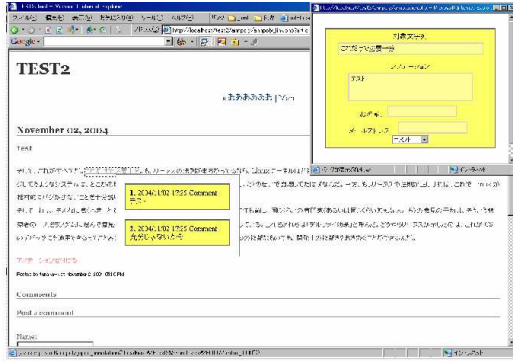
#### 3 . 1 アノテーションの入出力インタフェース

本機能は、コンテンツを公開している原著者の意思にもとづき、アノテーションを付記したいHTML ファイル、もしくはCMS に組み込む形で設置される。原著者は、HTML もしくはCMS のテンプレートにアノテーション管理 CGI を呼び出すタグを一行挿入するという簡単な作業だけでシステムの利用を開始することができる。

まず、入出力の表現様式や意味定義をテンプレートで任意に設定する機能について説明する。この機能は、原著者がアノテーションの表現形式を付箋形式や投票形式等状況に応じて入力項目を容易に設定可能とさせるものである。図1 にテンプレートによるアノテーション入力および表現の変更例を示す。図 1a は、いわゆる付箋形式で通常のアノテーションを実現したものである。入力者は、自由なコメントの入力やプロパティを設定したアノテーションの入力ができる。図 1b は、投票形式による入力と表示を実現したものである。アイコンで示されたボタンを押すだけで、リソースへの賛意を示すことが可能である。このように汎用的に利用可能な形式や、投票のような比較的特定の目的で効果を発揮する形式をテンプレートの記述により実現できるのは本システムの大きな特徴の一つである。

アノテータは、利用に際して専用ソフトやブラウザ用プラグインなどを要求されない。一般的な Web ブラウザで CMS ページにアクセスし、原著者により埋め込まれたリンクアンカ(例 : )をクリックすることで、本システムを利用することができる。新規アノテーションの付与は、対象文字列の選択とアノテーション付与 JavaScript の実行によって行われる。

選択文字列の位置特定は一般的に DOM2 の Range オブジェクトが使われるが、この方法で記録した場合、リソースの内容が1文字でも変更されると、アノテーション対象位置がずれてしまう問題があった。本システムでは、親ノードまでの XPath、選択文字列、文字列の登場順位、それぞれの情報を用いて対象位置を示す。これにより、リソースの変更に対するロバストネスが高められている。また、他のアノテーションシステムとの互換性を想定し、この方法で取得した文字列位置は、DOM2 の Range オブジェクトに変換可能とした。



```

<body onload="self.focus()">
<div class="gComments">
<h2 class="gHeader"><annpoly_target></h2><span class="gJisage">へのアノテーション</span></h2>
<div class="gJisage">
<<ANNOTATIONS>
<<annpoly_num></b><annpoly_name></b><annpoly_date>:<annpoly_annotationTypes></div>
<div class="gForm">
<form id="form1" name="form1" method="post" action="annotation.php">
<div>
<input type="text" value="書き込む"/><input type="hidden" id="annpoly_annotation_id" name="annpoly_annotation_id" value="<annpoly_annotation_id">"/>
</div>
</form>
</div></div>
</body>

```

```

<body onload="self.focus()">
<div class="gComments">
<h2 class="gHeader"><annpoly_target></h2><span class="gJisage">へのガッテン</span></h2>
<div class="gJisage">
<<ANNOTATIONS>
<<annpoly_num></b><annpoly_name></b><annpoly_date>:<annpoly_annotationTypes></div>
<div class="gForm">
<form action="annotation.php" method="post" id="gatten" name="gatten">
<div>
<input type="hidden" id="annpoly_annotation_id" name="annpoly_annotation_id" value="ガッテン"/>
<input type="hidden" id="annpoly_annotation_id" name="annpoly_annotation_id" value="<annpoly_annotation_id">"/>
</div>
</form>
</div></div>
</body>

```

A. アノテーションプロパティ付き付箋形式

B. 投票形式

図1. アノテーション入力インタフェースとテンプレートの例

3.2 アノテーションの二次的利用

アノテーション管理サーバのアノテーション群とアノテーション付与先の原著メタデータを利用して原著をトランスコーディングするためのAPI機能を実装した。また、トランスコーディングのサービスプログラムの例として、以下のサービスを試作した。

- a. 文への民意的な DOI (Degree Of Interest) に基づく要約表示機能
- b. 投票形式インタフェースを用いた、投票結果に応じた文の強調表示機能
- c. アノテーションプロパティを用いた記事のヒトコマ漫画表現
- d. アノテーションの入力時系列に沿った段階的表示機能

機能 a の表示例を図 2 に示す。アノテーション行為を興味対象の表明と捉えて、その件数にもとづき、一般的に「。」や「.」で区切られる一文を強調表示する。これに

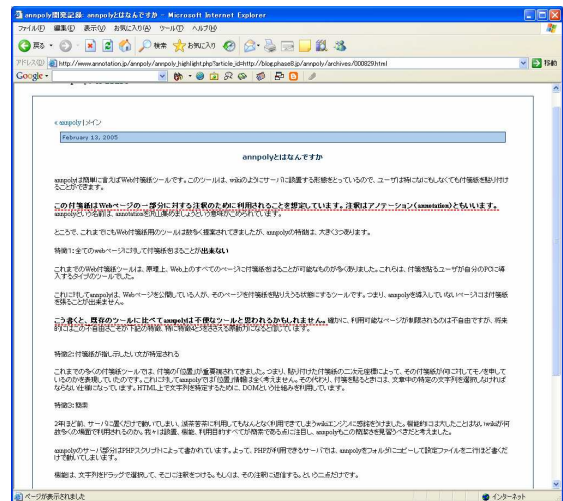


図2. 飛ばし読み用強調表示の例

より、長文の中から興味が集中する箇所をとばし読みするといった要求に応えることができる。

#### 4．従来の技術（または機能）との相違

Web アノテーションシステムは数多く提案されているが、既存システムでは、サーバ、クライアントの両者において専用ソフトや複雑なプラグインのインストールを必要とすることが多く、原著者とアノテータ双方にとって導入の敷居が高い。本提案では、クライアントとして通常のブラウザを利用可能であり、サーバ側も CGI プログラムであるため導入・利用までが非常に簡便である。また、既存システムはアノテーションの入力項目を利用シーンに応じて変更することが想定されていない場合が多いが、本システムではテンプレートによって利用シーンに応じたアノテーション項目の設定が可能である。さらに、本プロジェクトは、メタデータの流通・再利用・変換を各利用者のメリットとすることに徹する点が他のアノテーションシステムと大きく異なる点である。この中でも特に、再利用・変換に関する応用システムの例を示した点を強調したい。

#### 5．期待される効果

Web アノテーションシステムは古くから存在するのに対して、本格的に普及したシステムはなく、現段階ではアノテーションは Web リソースとして十分に流通していない。本システムが普及すれば、コンテンツ所有者のサーバにアノテーションが分散的に蓄積されていくことになり、それらは 3.2 で示したような二次利用や変換を経て Web リソースとして流通することになる。アノテーションが流通することによって展開可能な新規サービスは数多く考案可能なため、このシステムが Web 上のサービスやビジネスに及ぼす効果は大きいと考えている。

#### 6．活用の見通し

開発者らが試験的に blog 上で運用しており、現在のところ問題なく機能している。今後は様々なシーンに応じた利用例を増やしていき、有用性の確認と改良を重ねていく予定である。

#### 7．開発者名（所属）

坂本竜基（ATR 知能ロボティクス研究所）  
伊藤禎宣（ATR メディア情報科学研究所）  
開発パートナー：有限会社 Phase8

（参考）<http://www.annoataion.jp/>