

# ブラウザに依存しないプレーンテキスト上の Web アノテーション機能

石井 裕介<sup>†</sup> 松本 章代<sup>††</sup> Martin J. DÜRST<sup>††</sup>

<sup>†</sup> 青山学院大学大学院 理工学研究科 理工学専攻

<sup>††</sup> 青山学院大学 理工学部 情報テクノロジー学科

## 1 はじめに

現在までセマンティック Web の実現に向けて様々なサービスが提案されている。その一つに、Web コンテンツへの注釈付与を行うアノテーションがある。そのフレームワークとして W3C が開発した Annotea [1] が代表的であり、Amaya [2] をはじめとしたいくつかのブラウザに導入されている。

本研究では特定のブラウザに依存せずに Annotea を利用できる Annoplus の開発を進めている [3][4]。これによりアノテーションのユーザ増加、さらにアノテーションの普及に貢献していきたい。その一方で、プログラムソース等のプレーンテキストには付与できない。したがってアノテーションの利用範囲拡大が急務である。

アノテーションの付与には、文書内の正確な場所を示す fragment identifier (以下、素片識別子と呼ぶ)が必要である。近年プレーンテキスト (text/plain) 用の素片識別子が IETF で標準化された [5]。一昨年に text/plain 素片識別子の仕様を元に Amaya で実装し、プレーンテキスト用アノテーション機能を追加した [6]。E-Learning 教育や論文作成を活性化し、アノテーションをさらに普及させたい。そのために、プログラムソースや TeX ソースなど対象コンテンツを多種多様にする。その実現に向けて、Annoplus にプレーンテキスト上のアノテーション付与・表示、さらに選択したコンテンツの範囲の表示機能を追加したので報告する。

## 2 Annoplus の開発

Annoplus は任意のブラウザでの Annotea 環境の確立を目指している。Annotea ではアノテーションは RDF で記述され、専用サーバで保持される。コンテン



図 1: クライアント側の付加作業の処理

ツ上の正確な位置は XPointer で定義される。付与対象が HTML の場合でも XPointer で対応する仕組みになっている。図 1 に、Annoplus のスクリーンショットの例を示す。Annoplus はアノテーション専用フォームよりアノテーションの表示・非表示を選択できる。さらにこれまでの開発でアノテーションの表示スタイル選択による視覚的理解の助長、アノテーション付与を実現した。

昨年まで Annoplus はサーバ側で Perl によって実装された。しかし、Perl は可読性が低い。そこで本年度に、ソース全体をより読みやすい Ruby に一新した。結果としてソースの拡張が容易になった。クライアント側は JavaScript で実装されている。アノテーション付与の際、JavaScript で XPointer が生成される。しかし、これまでは一部の場所では XPointer が生成されない問題があった。そこで JavaScript を改良し、任意の場所からのアノテーション付与を可能にした。

さらにプレーンテキスト用のアノテーション付与機能をサーバ側とクライアント側に実装した。

### Browser-Independent Web Annotation Tool for Plain Text

Yusuke ISHII<sup>†</sup>, Akiyo MATSUMOTO<sup>††</sup> and Martin J. DÜRST<sup>††</sup>

<sup>†</sup> Graduate School of Science and Engineering, Aoyama Gakuin University

<sup>††</sup> Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University  
5-10-1 Fuchinobe, Sagami-hara, Kanagawa 229-8558, Japan  
yusuke@sw.it.aoyama.ac.jp, {akiyo, duerst}@it.aoyama.ac.jp

表 1: text/plain 素片識別子の概念, 構文一覧

	説明
位置	文字自体ではなく, 2 つの文字間の位置を識別する. 行頭や行末も一つの位置とみなす. 位置は 1 ではなく, 0 から数える.
範囲	上限と下限の両方のパラメータを指定する. 2 つの位置の間に囲まれる部分を識別する.
char	char=3 先頭より 3 番目と 4 番目の文字間を示す. char=4,10 4,5 番目の文字間と 10,11 番目の文字間の間全体を示す.
line	line=4 先頭より 3 行目の行頭を示す. line=4,6 4 行目と 6 行目の行頭の間全体を示す.

### 3 プレーンテキスト用のアノテーション機能

Wilde らによって提案された text/plain 素片識別子では, プレーンテキストの特定の部分を文字間や行間を起点に位置や範囲を参照する. XPointer と異なり, 構造解析に依存せず, 純粋なテキストのみに準拠して場所を示し, その位置はコンテンツの先頭にあたる 0 から数える. さらに, 2 つの数値の指定により範囲も識別できる. 位置, 範囲の概念と識別子構文の一覧を表 1 に示す. 素片識別子の構文は char, line の 2 種類のスキーマで記述される.

以上の構文を元にした実装により, プレーンテキスト上のアノテーション付与を実現した. その際に, Anno-plus の Javascript に char, line による text/plain 素片識別子生成機能を追加した. その際に, プレーンテキストを簡易な XHTML に変換するようにした.

### 4 対象範囲の表示機能の実装

ほとんどのアノテーションサービスはアノテーション表示の際, 元の文書のツリー構造を維持するため対象範囲の開始点のみをアイコンで表示する. 対象範囲が把握できず, 結果としてアノテーションの視覚的理解が難航する.

そこで本研究ではリンクだけでなく対象範囲も示す機能を提案し, 実装した. これにより, コンテンツの視覚的理解度を向上させたい.

手法としては対象範囲を span タグで囲み, スタイルシート生成によりコンテンツを色分けする. その処理手順は以下の通りである.

1. アノテーションの内部データからアノテーション ID と 素片識別子を含む Context を取得する.
2. Context からアノテーションの開始位置, 終了位置を示す数値を取得し, span 要素の開始タグと終了タグの位置として設定する.
3. ソーティング処理, クラス値の追加処理により, 個々のタグの位置データから, 他の span タグと重複せず小刻みな配置データに変換する.

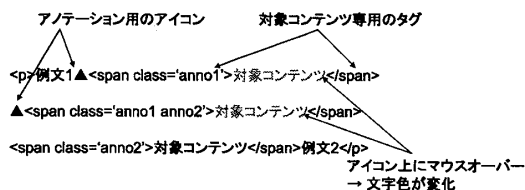


図 2: 対象範囲の表示例

4. ロードしたコンテンツと span タグの位置情報によりアノテーション付きコンテンツを作成する.

結果として, 他のアノテーションや元のマークアップとの対象範囲が重複した場合でも個別に範囲を表示できるようになった. 対象範囲の表示例は図 2 に示す. 表示の際に, 指定したクラス値を持つ span タグ内のテキストの文字色を変化させる機能を JavaScript に追加した. コンテンツのロード後, アノテーションのアイコンにマウスを当てると, アイコンのクラス値に対応する span 内のテキストの文字色を変化できるようにした.

### 5 まとめと今後の予定

本稿では text/plain 素片識別子と span タグを利用することで, プレーンテキスト上のアノテーション付与と対象範囲の識別を Anno-plus に実装した. さらに, 同じプレーンテキストに分類される TeX や プログラムソースへのアノテーション付与や大学の e-Learning への導入も予定している.

### 参考文献

- [1] Kahan, J., Koivunen, M.-R., Prud'Hommeaux, E. and Swick, R. R.: Annotea: An Open RDF Infrastructure for Shared Web Annotations, *WWW10 International-Conference*, pp. 623-632 (2001).
- [2] Vatton, I.: Amaya Overview, <http://www.w3.org/Amaya/Amaya.html> (2006).
- [3] 村田菜々子, 伊藤一成, Dürst, M. J.: 効果的な視覚表現が可能な Web アノテーションシステム Anno-plus, 第 68 回情報処理学会全国大会 (2006).
- [4] 川上建一郎, 伊藤一成, Dürst, M. J.: Web アノテーションシステム Anno-plus の拡張, 第 69 回情報処理学会全国大会 (2007).
- [5] Wilde, E. and Dürst, M. J.: URI Fragment Identifiers for the text/plain Media Type, RFC 5147 (2008).
- [6] 石井裕介, 伊藤一成, Dürst, M. J.: Fragment Identifier によるテキスト資源上のアノテーションの付加, 第 69 回情報処理学会全国大会 (2007).