6S-4

Fragment Identifier による テキスト資源上のアノテーションの付加

石井 裕介　伊藤 一成　Martin J. DÜRST
青山学院大学理工学部

1 はじめに

コンテンツに対するメタ情報であるアノテーションを表示・付加するにあたり，その対象がファイル全体ではなく，そのコンテンツに対する特定の部分資源である場合，その場所を識別する fragment identifier (以下，素片識別子と呼ぶ) が必要になる．素片識別子は一般的に URI の一部として組み込まれる．

HTML に代表される XML 文書の場合，素片識別子として XPointer が広く用いられている．一方，構造を持たないプレーンテキスト資源に関して汎用的に利用可能な素片識別子が存在しなかったので，アノテーション付加の仕組みを包括的に実現することは難しかった．しかしながら，近年プレーンテキストを対象にした素片識別子の構文が提案され，現在，標準化組織である IETF において策定に向けた活動が活発に進められている [1]．これにより，TEX，e-mail やプログラムソースなど多種多様なコンテンツを対象としたアノテーション利活用技術の創出が見込まれる．それに先駆け，この素片識別子を用いたプレーンテキスト資源に対するアノテーション付加・表示機能を，W3C で開発が進められている Amaya [2] を用いて実現したので報告する．

2 Amaya によるアノテーション付加の仕組み

Amaya は，W3C で開発が行われているオープンソースの Web エディタ兼 Web ブラウザである．図 1 に，Amaya のスクリーンショットの例を示す．一般的なブラウザと異なり，閲覧機能だけでなく編集機能も備わっており両者はシームレスに移行できる．W3C で勧告されている新しいプロトコルやデータフォーマット，規格の実例などを実証・テストするための実験環境としても利用されている．そのため，HTML，MathML，SVG など多くのフォーマットをネイティブでサポートしている．同じく W3C の Annotea Project にて開発されている Web コンテンツにアノテーションをつけるためのフレームワークである Annotea [3] と連携させることで，容易にアノテーションの付加・表示機能も実現出来る．

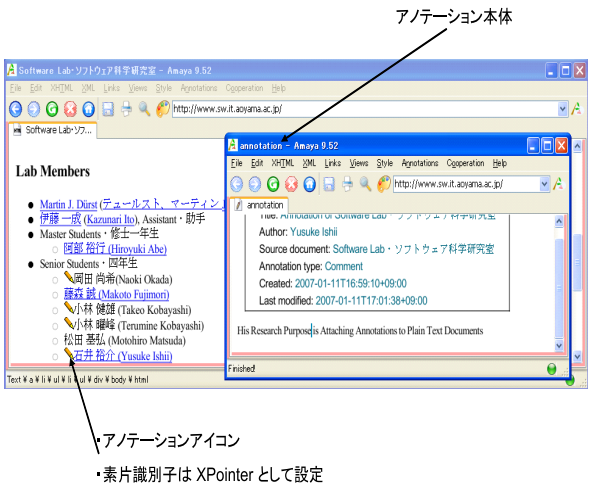

図 1: Amaya を用いたアノテーション表示の例

Amaya はさらに XML ポインタ言語である XPointer をサポートしている．Amaya と Annotea は，アノテーションに関するデータを送受信する際に，対象範囲を指定するための素片識別子の構文として XPointer を用いている．アノテーションを付加する段階でその場所を含む要素までのパスを示す素片識別子が生成される．その識別子が URI の一部に組み込まれた形式が，アノテーションの対象 URI としてやりとりされることとなる．

一般に素片識別子の構文と解釈は content-type (例: text/html , text/plain) によって異なる．プレーンテキスト資源のような文書では，XPointer のように，パスを示す方式は適用出来ない．text/plain 特有の素片識別子をもって対応しなければならない．

Attaching Annotations to Plain Text Documents using Fragment Identifiers
Yusuke ISHII, Kazunari ITO and Martin J. DÜRST
Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University
5–10–1 Fuchinobe, Sagamihara, Kanagawa 229-8558, Japan
yusuke@sw.it.aoyama.ac.jp, {kaz, duerst}@it.aoyama.ac.jp

	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．文字間であるため，実体をもたず，長さゼロの素片とみなす．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	原則として，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．上限が下限よりも大きい場合，長さ 1 以上の実体のある素片を識別する．上限が下限と等しい場合は位置識別と同値である．
`char`	指定する値は文字位置数であり，プレーンテキスト資源の頭 (位置 0) から数える．一つ，または 2 つの数字を使って，文字位置と文字範囲を識別する．
`line`	行位置数を指定し，頭から行単位で数える．`char` と同様に位置と範囲を識別できる．
`match`	正規表現を指定し，文章中に合うもの全ての部分を識別する．
`hash`	識別する機能ではなく，文書の更新を確認でき，それによる素片識別子の消滅の可能性をブラウザに示唆する役割を持つ．

表 1: text/plain 素片識別子の構文一覧

3 text/plain 素片識別子

Wilde によって提案された text/plain 素片識別子では [1]，テキスト資源の特定の部分を文字や行に適用できる位置や範囲の概念を使って参照する．構文の一覧を表 1 に示す．文字位置，文字範囲，行位置，行範囲，正規表現，ハッシュの 6 種類の方式 (scheme) を使った素片識別を規定している．素片識別子の構文は，主に `char`，`line`，`match`，`hash` の 4 種類のスキーマに大別される．XPointer の場合は構造化された文書内の要素を追跡して解釈されるが，この素片識別子は構造の解析に頼らず，テキストの位置をカウントすることで，識別子の示す正確な領域を算出する．

4 Amaya での実装

Amaya 内部のプログラムソースでは，アノテーションの作成，選択領域から素片識別子への変換，アノテーションへのリンクの貼り付けなどの動作はそれぞれ別々のメソッドに分割され，一つのライブラリに定義されている．

はじめに，text/plain 対応の素片識別子を生成するメソッドを新たに追加した．プレーンテキスト内で選択した任意の場所から，その場所に相当する素片識別子を生成することができる．さらに text/plain 対応の識別子の文字列を解釈し，scheme の種類を判別し，scheme が指定した数値を分析することで選択した場所を含む要素を生成するメソッドを新たに加えた．このメソッドはプレーンテキスト上へのリンクの貼り付けに使用した．要素の分割による選択領域の開始点と終了点の算出をプレーンテキストにも対応できるようにし，作成したアノテーションへのリンクを示すアイコンをプレーンテキスト上で表示する．また，解釈に失敗した場合は，素片識別子は無視され，参照先は Web ページ全体でとどまるようにした．

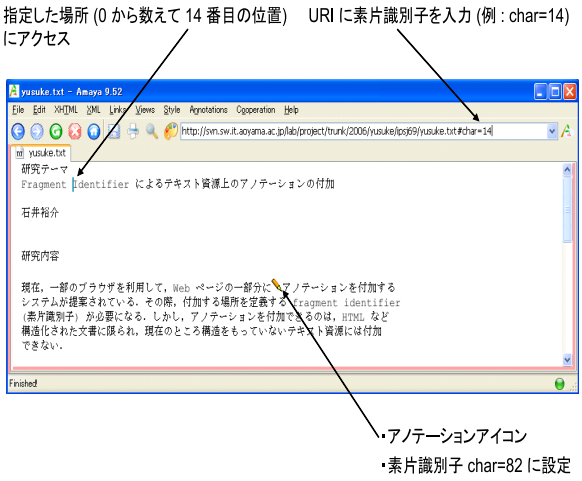

図 2: プレーンテキスト上のアノテーション表示

表示例を図 2 に示す．実際に，Amaya 上で，素片識別子を含む URI を入力することで，指定した位置にアクセスできることを確認した．図 1 と比較してもわかる通り，アノテーションの付加および表示も既存の Amaya と同様のインタフェースで実現される．

5 まとめと今後の予定

本稿では素片識別子を利用することで，プレーンテキスト資源に対するアノテーションを付加する仕組みを Amaya に実装した．さらに，我々の研究グループでは昨年度より，任意のブラウザ上で Annotea クライアント環境を実現できるアノテーションシステム Annoplus も開発している [4]．Amaya だけではなく今後 Annoplus への機能組み込みも予定している．

参考文献

[1] Erik Wilde and Martin J. Dürst: URI Fragment Identifiers for the text/plain Media Type, http://www.ietf.org/internet-drafts/draft-wilde-text-fragment-06.txt (work in progress).

[2] Irène Vatton: Amaya Overview, http://www.w3.org/Amaya/Amaya.html (2006).

[3] Josè Kahan and Marja-Riitta Koivunen and Eric Prud'Hommeaux and Ralph R. Swick: Annotea: An Open RDF Infrastructure for Shared Web Annotations (2001).

[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 Annoplus の拡張, 第 69 回情報処理学会全国大会 (2007).