

部分引用の管理に基づく Web コンテンツのマッシュアップ*

梶 克彦

名古屋大学 大学院情報科学研究科

kaji@nagao.nuie.nagoya-u.ac.jp

長尾 確

名古屋大学 情報メディア教育センター

nagao@nuie.nagoya-u.ac.jp

1 はじめに

近年、Web 上に存在するビデオや音楽などのコンテンツの増加に伴い、それらの検索・推薦・変換を行う研究 [1] が進められている。また複数の技術やコンテンツを複合させて新しいサービスやコンテンツを形成するマッシュアップは、既存のコンテンツを再利用し、コンテンツの価値を高める手法として注目されてきている。

様々な種類のコンテンツを組み合わせるためには、関連性の認められるコンテンツ同士を選出することが必要となる。そこで我々は、Nelson[2] もその重要性を指摘する引用情報に着目した。あるコンテンツ内で同時に引用される複数コンテンツ間には明らかに意味的な関連性を認めることができることから、引用・被引用情報をコンテンツマッシュアップに利用することが有効であると考えられる。またコンテンツの部分引用の情報は、コンテンツの意味的な構造把握に有用な情報となるだろう。

また近年ブログやビデオ投稿システムなど、消費者がコンテンツを作成し投稿する CGM(Consumer Generated Media) サービスが盛んになってきており、今後 Web 上のコンテンツの多くは CGM サービスによって管理されることになると予想される。しかし現在の CGM サービスはコンテンツの部分的な引用・被引用を行うために十分な機能を備えていない。

そこで本稿では、CGM サービスが、コンテンツの任意の箇所の部分引用を可能にし、その引用・被引用情報を扱うことを支援するために、任意のコンテンツに対する引用情報の管理を行うプラットフォーム Annphony を構築した。また引用先の CGM サービスとしてブログに、引用元の CGM サービスとしてビデオ共有システム Synvie と認知科学辞典システムに本プラットフォームを適用し、ブログで同時引用されるビデオと辞典を組み合わせたマッシュアップコンテンツを自動生成する仕組みを開発した。

2 Annphony

コンテンツ間の引用情報を扱うためのプラットフォームとして、Annphony[3] を構築した。

本プラットフォームは、コンテンツの部分引用を扱うために、任意のコンテンツの内部を指し示す形式である ElementPointer を採用した。ElementPointer では、コンテンツの種類に応じてそのコンテンツの内部を指し示すための ElementPointer 定義を別途用意し、その定義に基づきコンテンツの部分を URI として記述する。ElementPointer は以下の形式をとる。

```
[C]#epointer([S]([P1],arg1),([P2],arg2)...))
```

コンテンツの URI である [C] に続き、以降にフラグメント識別子としてコンテンツのどの部分を指し示すかを記述する。本手法では、コンテンツの内部は、[S] で表される ElementPointer 定義の URI 以降に、[P] で表される有限個のプロパティの URI とその値を列挙することにより指し示される。実際にはフラグメント識別子は URI としての妥当性を保つため、URL エンコーディングを行う。ElementPointer 定義は RDF(Resource Description Framework)[4] のスキーマ言語である RDFS(RDF Schema)[5] によって記述される。

本形式により、「MIDI のチャンネル 1 における、楽曲開始後 10 秒から 20 秒までの時間範囲」や、「あるビデオ映像の特定オブジェクトの写っているシーン」など、様々なコンテンツの部分を指し示すことが可能になる。また本プラットフォームは、ElementPointer 定義を管理し、さらにどのコンテンツが、どのコンテンツを引用しているかという情報を RDF 形式で保存し、検索することができる機能を備える。

3 CGM サービスの拡張

今後 Web 上のコンテンツの多くは CGM サービスによって管理されることになると予想されるため、Annphony を CGM サービスに適用することで、様々なコンテンツ同士の引用と、その引用情報の共有を可能にした。

引用元の CGM サービスとしてビデオ共有システム Synvie[6] とオンライン辞典システムである認知科学辞典に Annphony を適用した。コンテンツの種類は CGM サービスによって異なるため、コンテンツ内部を指し示す ElementPointer の定義は CGM サービスごとに用意した。ビデオ共有システムではビデオの時間区分を指し示すためには、開始・終了時間が必要であるため、その旨を ElementPointer の定義に記述した。またオンライン辞典システムの各専門用語は、文書の言語構造

*Web Content Mashup Based on Partial Quotation by Katsuhiko Kaji (Graduate School of Information Science, Nagoya University) and Katashi Nagao (Center for Information Media Studies, Nagoya University)

を表現するための言語である GDA(Global Document Annotation)[7] 形式で保持されているため、単語・文節・文など、文書の論理構造に基づく部分を指し示すための定義を記述した。

また引用先となる CGM サービスとして、ブログを採用した。ブログシステムに Annphony を適用し、ブログにビデオや辞典などの他のコンテンツを部分引用した記事を投稿する際に、その部分引用の情報を Annphony に保存する。図 1 はブログにおいてビデオ・辞典を引用した記事を投稿した画面例である。図では、ある分野の詳細な解説として、参考となるビデオと辞典の一部を引用している。このように同時に引用されるビデオと辞典には、明らかな関連性を認めることができる。

一般的にユーザは、あるコンテンツと他の様々なコンテンツとの多項関係を完全に把握してはいないため、一人のユーザによって網羅的な関連性を記述することが困難である。そのため、一般ユーザが記述可能な同時引用の情報は二項、三項関係といった少数の関連性にとどまると考えられる。しかし CGM サービスを通じて同時引用の情報を広く収集することにより、コンテンツ間の多項関係に拡張することができる。

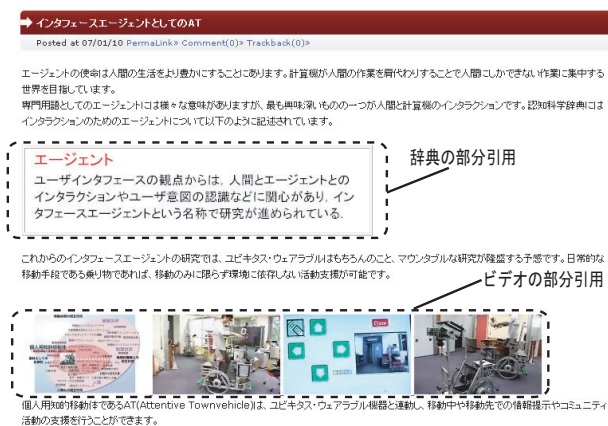


図 1: ブログでのビデオ・辞典の同時引用

4 ビデオ用例付き辞典システム

Web コンテンツを用いてマッシュアップコンテンツを自動生成する例として、オンライン辞典システムを拡張し、ビデオ用例付き辞典システムを構築した。本システムでは、各専門用語の項に、ブログにおいて同時引用された、用例の一つと見なされるビデオを提示することで、その専門用語の理解を深めることができる。

キーワード検索の要求を本システムが受け取ると、オンライン辞典システム・ビデオ共有システムに対して、検索要求のキーワードが存在する辞典の項目・ブログエントリを検索する。次にそれらがブログにおいて過去に同時引用されているかを検索し、同時に引用されたことのあるビデオの一部と辞典項目の一部の組を列挙する。またブログに対してもキーワード検索を行い、検出され

たブログエントリ内で同時引用されているビデオの一部と辞典項目の一部の組を列挙する。

こうして列挙されたビデオの一部と辞典の項目の一部の組を用い、辞典の項目中の同時引用された文以降に、ビデオへのリンクを埋め込む。図 2 はビデオが埋め込まれた辞典の項目の例である。辞典の項目中に具体的な例であるビデオを埋め込むことで、その項目をより深く理解することができる。



図 2: ビデオ用例付き辞典システム

このように同時引用の情報をを用いることにより、ビデオ、オンライン辞典の項目といった異種類のコンテンツ同士を結びつけてマッシュアップを行うことで、付加価値を持ったコンテンツを作成することが可能である。

5 おわりに

本稿では、Web コンテンツのマッシュアップを行うために、任意のコンテンツに対する部分引用と、その引用・被引用情報の管理のためのプラットフォーム Annphony を複数の CGM サービスに適用し、ビデオ用例付き辞典システムを開発した。今後はさらに多くの CGM に Annphony を適用することで、様々なコンテンツを組み合わせることでできる環境を整備する予定である。

参考文献

- [1] K. Nagao. *Digital content annotation and transcoding*. Artech House Publishers, London, 2003.
- [2] T. Nelson. *Computer Lib/Dream Machines*. マイクロソフト出版, 1987.
- [3] 梶克彦, 長尾確. Annphony: メタコンテンツ処理のためのプラットフォーム. *情報科学技術レターズ* vol.5, pp. 381-384, 2006.
- [4] W3C. Resource Description Framework. <http://www.w3.org/RDF/>, 1999.
- [5] W3C. RDF Schema. <http://www.w3.org/TR/rdf-schema/>, 2004.
- [6] 山本大介, 清水敏之, 大平茂輝, 長尾確. Synvie: ブログの仕組みを利用したマルチメディアコンテンツ配信システム. *情報処理学会第 68 回全国大会*, 2006.
- [7] 橋田浩一. GDA:意味的修飾に基づく多用途の知的コンテンツ. *人工知能学会誌*, Vol.13, No.4, pp. 528-535, 1998.