

引用の構造化による マルチメディアコンテンツの意味的統合支援システム

石戸谷 顕太郎[†] 増田 智樹[†] 山本 大介[†] 長尾 確^{††}
[†]名古屋大学 大学院情報科学研究科 ^{††}名古屋大学 情報メディア教育センター

1 はじめに

近年、インターネットの普及・ブロードバンド化を背景に、様々なサービスが多種多様な Web コンテンツを提供するようになった。それに伴って、情報量は急激に増加し、人間が処理可能な限界を超えつつある。そのような状況を考慮し、機械によって人間の情報処理を支援するために、検索・推薦・統合など、マルチメディアコンテンツの高度処理に関する研究 [1] が行われている。

特に、異なる複数のコンテンツの統合は、検索・推薦などを利用して関連性のあるコンテンツを選出し、提示することで人間の情報処理を効率的に支援する方法として注目されている。

マルチメディアコンテンツを効率よく処理するためには掘らの開発したメタコンテンツ基盤システム Annphony [2] のように、コンテンツに関する情報を、そのフォーマットに依存しないメタ情報として取り扱い、またコンテンツの任意の部分要素に関する情報も、コンテンツ全体と同様に処理可能にすることが重要である。

しかし、最近のコンテンツ共有システム、例えば画像共有システム Flickr[†]や、動画共有システム Synvie^{††}では、コンテンツに対する部分定義を不特定多数のユーザーに任せている。この方法では、コンテンツの構造化にかかるコストは軽減できるが、コンテンツの検索・統合・推薦など意味内容を考慮した処理を行う際に、まず意味内容に関するメタ情報の指し先であるコンテンツの部分について、同一性・類似性の推定など、高度で複雑な処理を行う必要性がある。

そこで、本稿では、メタ情報の指し先であるコンテンツの部分についての曖昧性をなくし、メタ情報間の処理を簡便に行えるように、意味単位という新しい概念を Annphony に導入した。

また、意味単位を利用したシステムとして、部分に関する再利用の中で特に重要な行為である引用に注目し、引用を構造化して扱い、引用の意味情報に基づいて異なる複数のコンテンツを統合して提示する仕組みを開発した。

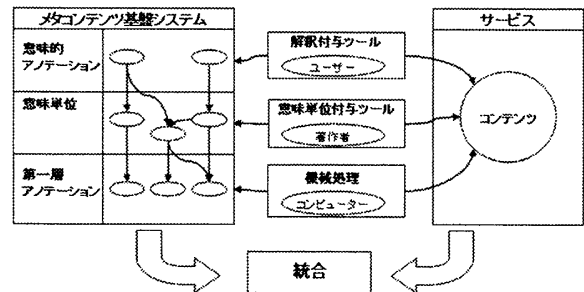


図 1: 意味単位の位置づけとシステム全体図

2 統合支援システム

図1に示したように、本システムは、1) 意味単位を保存するために拡張されたメタコンテンツ基盤システム、2) 著作権者がコンテンツに対して意味単位を設定するためのツール、3) ユーザーが著作権者の付与した意味単位を選択して利用するためのツール、そして、4) 意味的アノテーションを利用してコンテンツを統合するアプリケーションによって構成される。

2.1 意味単位とは

意味単位とは、コンテンツの任意の部分とその部分に対する意味を合わせて定義したものをいう。意味単位の作成は、コンテンツに対して責任のある者、主に著作権者が、コンテンツの制作時に機械処理によって抽出された構造を外部データ化した、第一層アノテーション [3] と、すでに作成された意味単位の中から任意の部分を選出し、その部分に対して意味単位の定義に必要なメタ情報を合わせて保存することで行う。定義に必要なメタ情報は、WordNet[‡]及びSUMO^{‡‡}を利用した内容を表す複数のキーワード、他の意味単位を包含する場合はその意味単位との関連、及び自然言語でのコメントとした。

第一層アノテーション、意味単位、ユーザーの解釈などを表現した意味的アノテーションは、Annphonyで提供されるコンテンツの任意の部分の指示対象とするためのElementPointerを用いて、RDFとして表現し、Annphonyのデータベースに保存する。

図1のように意味単位は、第一層アノテーションと意味的アノテーションの間に位置し、第一層アノテーションを取りまとめ、意味的アノテーションの指示対象としての役割を果たす。

著作権者によって意味単位が作成されることで二つのメリットが考えられる。まず、意味的アノテーション同士の比較などの処理を行う際には、指示対象の同一性・類似性を推定する前処理が必要だったが、指示対象を意味単位に限定することで複雑な前処理が必要なくなる。また、意味単位の定義時のメタ情報は、複数

Semantic Integration of Multimedia Contents Based on Structured Quotation

[†] ISHITOYA, Kentaro
 (ishitoya@nagao.nuie.nagoya-u.ac.jp)

[†] MASUDA, Tomoki
 (masuda@nagao.nuie.nagoya-u.ac.jp)

[†] YAMAMOTO, Daisuke
 (yamamoto@nagao.nuie.nagoya-u.ac.jp)

^{††} NAGAO, Katashi (nagao@nuie.nagoya-u.ac.jp)
 Graduate School of Information Science, Nagoya Univ. (†)
 Center for Information Media Studies, Nagoya Univ. (††)
 Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

[‡] <http://www.flickr.com/>

^{††} <http://video.nagao.nuie.nagoya-u.ac.jp/>

[‡] <http://wordnet.princeton.edu/>

^{‡‡} <http://www.ontologyportal.org/>

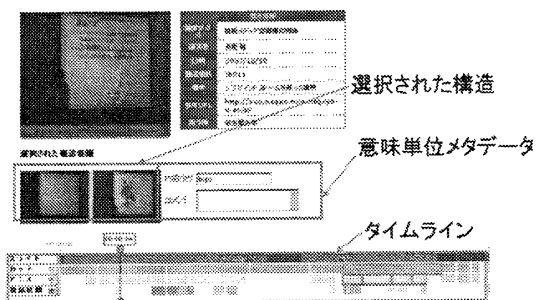


図 2: 意味単位編集ツール

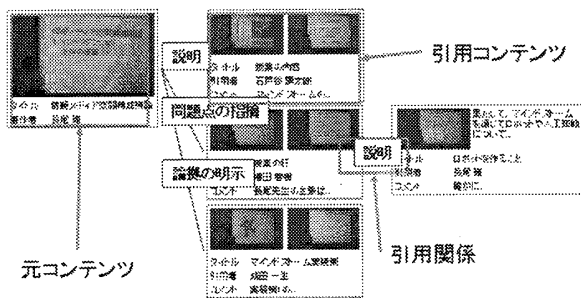


図 3: 引用構造閲覧システム

のユーザーの付与した意味的アノテーションを比較する際の判断基準として利用できる。

2.2 意味単位編集ツール

意味単位の編集は図2に示すようなツールを用いて行う。第一層アノテーションとすでに定義された意味単位がタイムライン上に提示される。著作者は任意の部分を選択し、前述した意味単位の定義に必要なメタ情報を付与することができる。

意味単位の編集作業は、ユーザーによって意味づけを行うオンラインアノテーションとは違い、人的コストの高い作業になる。そこで意味単位の編集を簡便に行えるように、機械処理で取得する第一層アノテーションを可能な限り多数取得し、著作者が簡便に意味単位を付与できる必要がある。たとえば、講義コンテンツでは、プレゼンテーション時にスライドとアニメーションの切り替えタイミング、及び講演者の発話区間を取得し、後日ビデオのカット検出を行っている。

2.3 意味単位に基づく引用ツール

意味単位に基づく引用ツールでは、コンテンツの閲覧時に、著作者によって定義された意味単位を提示し、ユーザーはその中から任意の数の意味単位を選択して、引用情報をメタ情報として設定する。保存された意味情報は候補として Annphony に保存される。引用する際には、引用候補の中から必要なものを選択し利用する。引用が行われると Annphony にコンテンツの情報とともに保存される。

ユーザーが付与するメタ情報は、Weinstock らが定義した引用の理由 [4]、IBIS モデル [5] に基づいて表現した引用対象の意味単位との関係、及び自然言語によるコメントとする。このメタ情報は意味単位に対する意味的アノテーションとして Annphony に保存される。

3 引用構造閲覧システム

引用構造閲覧システムは、図3に示すように、前述したツールを用いて定義された意味単位及び、意味単位に対する意味的アノテーションを利用して、引用被引用関係を可視化して閲覧可能にするシステムである。

あるコンテンツに対する情報取得の要求をシステムが受け取ると、Annphony に対して、コンテンツに付与されている意味単位の検索を行い、その意味単位を対象とした引用情報を列挙する。コンテンツ内に引用情報が含まれる場合には、引用元コンテンツを検索して表示する。また、意味単位及び引用情報に付与されたメタ情報を利用して、コンテンツ間の関連を表現する。これらの処理を、再帰的に行ってグラフ構造として可視化する。ノードを選択すると、そのコンテンツに関する詳細情報と関連するノードを閲覧することができ、さらにそのノードを引用して新たなコンテンツを作成できる。

最近のコンテンツ共有システムでは、ユーザーがそれぞれの判断で解釈の指示対象を選択していたため、複数のユーザーの解釈が、同一の事柄に対して行われているのかを推定することが困難であったが、本システムでは、意味単位に基づいて解釈が行われるために、指示対象を明確に判断することが容易になった。また、著作者が意味単位に対してメタ情報を付与することで、ユーザーの解釈を、著作者の意図と比較することが可能になった。

このように、従来のシステムに意味単位の概念を導入した本システムを利用することで、これまでは高精度で行うことが困難であった、マルチメディアコンテンツの意味情報に基づく検索・推薦・統合などの高度な処理を比較的容易に行えるようになる。

4 おわりに

本稿では、マルチメディアコンテンツの検索・推薦・統合などの処理を容易にするため、意味単位の概念をメタコンテンツ基盤システム Annphony に導入し、意味単位に関連する複数のツールの開発を行った。今後は、意味単位を利用したアプリケーションの開発、及び多くのコンテンツに対して意味単位を付与できるようにし、様々な用途で意味に関する処理を行えるように環境を整備していく予定である。

参考文献

- [1] Katashi Nagao. Digital content annotation and transcoding. Artech House Publishers, London. 2003.
- [2] 梶谷彦, 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.
- [3] 長尾確. SemCode2: オントロジーに基づくアノテーションとトランスコーディング. 情報処理学会第67回全国大会. 2005.
- [4] Melvin WeinStock Citation Indexes. Encyclopedia of Library and Information Science vol.5, pp.16-41 1971.
- [5] Werner Kunz, Horst W. J. Rittel. ISSUES AS ELEMENTS OF INFORMATION SYSTEMS. Working paper 131, Institut für Grundlagen der Planung I.A. University of Stuttgart. 1970.