

XML コンテンツの意味構造および再利用意図の 3Y-01 記述能力を有するマークアップ言語

湯本高行†

灘本明代‡

角谷和俊* 田中克己*

†京都大学工学部 ‡神戸大学大学院自然科学研究科 *京都大学大学院情報学研究科

1 はじめに

Semantic Web に期待されるもののひとつとして、メタ情報 (意味) 記述とそれを利用したコンテンツの自動生成が考えられる。既存のコンテンツの一部を再利用し、そこから新しいコンテンツを生成する場合、ソースとなるコンテンツの再利用に関する作者の意図が尊重されるべきである。しかし、現在の Semantic Web の活動が目指している意味記述では、再利用条件の論理的な記述というものが考えられていない。また、再利用条件を論理的に記述するためには、どの分野の XML 文書にも共通する意味的な構造 (応用独立な意味) を表現する必要がある。しかし、これについても XML 文書に書かれている内容の分野に依存した応用依存な意味記述しか考えられていない。

また、ブロードバンドネットワークの普及に伴い、これからはビデオストリームや SMIL コンテンツなどの同期化コンテンツが増加すると考えられる。しかし、同期化コンテンツについては、応用独立/依存にかかわらず、意味を記述するフレームワークが存在しない。

よって、本研究では、同期化コンテンツを含む Web コンテンツの応用独立な意味構造記述 (意味記述) の手段を提供し、それを利用した自動処理可能な再利用条件の記述 (意図記述) を可能とするマークアップ言語 SaRX を設計する。本言語は、1) 再利用意図記述能力、2) 意味構造記述能力、3) XML との親和性の 3 点に優れている。この 3 点に着目し、本言語の仕様を述べる。

2 言語仕様

2.1 言語構造

意味/意図記述は SaRX ファイルになされ、XPath によって、コンテンツを構成する XML 文書のタグ領域 (以下 region) と関連づけられる。それゆえ、本言語は既存の XML コンテンツを加工することなく、導入することが可能であり、XML との親和性が高い。また、本言語は RDF を拡張する形で実装する。RDF、XTM^[2] との違いを表 1 に示す。

A Markup Language to Describe Semantic Structures
and Reuse Intentions of XML Contents
Takayuki Yumoto[†], Akiyo Nadamoto[‡],
Kazutoshi Sumiya* and Katsumi Tanaka*
[†]Faculty of Engineering, Kyoto University
[‡]Graduate School of Science and Technology, Kobe University
*Graduate School of Informatics, Kyoto University

	SaRX	RDF	XTM
意味記述	意味構造 応用独立	リソースの属性 情報/応用依存	Topic Map 応用依存
意図記述	○	×	×
リソースとの 関連付け	XLink/ XPath (可)	Hyperlink (範囲指定不可)	XLink/ XPointer (可)

表 1: SaRX, RDF, XTM の違い

2.2 region

利用者がコンテンツを再利用する時に、元のコンテンツすべてを再利用するとはかぎらず、元のコンテンツを部分的に再利用することが多い。そこで、SaRX では region というコンテンツを構成する単位を考え、その region に意味記述を行い、その意味記述に基づいた意図記述を行う。region を BNF で記述すると次のようになる。

$$\text{region} ::= (\text{region}, \text{region}, \dots) | \text{タグ領域}$$

つまり、region は単一のタグ領域または region の入れ子である。入れ子の中に入っている region を外側の region に対して、subregion と呼ぶ。また、region は再利用時の引用単位でもあり、再利用条件が記述された region は、再利用の際には、カプセル化されたオブジェクトとみなされ、内部の改変は許されない。

3 意味記述

XML コンテンツ内の意味構造を Author が望むとおりに記述することができる。具体的には、a) region への属性記述、b) region 間の関連記述が可能である。

3.1 属性

Title, Creator, Subject などがある。Creator の属性を region 毎に記述できるので、他の作者が書いたコンテンツを部分的に利用することも可能である。

3.2 内容関連

Web コンテンツではそこでどのような内容が表現されているかが最も重要であるが、それぞれを自然言語処理や画像認識で理解することは困難であり、またそれらの間の関連を理解することはさらに困難である。再利用の際にも内容についての意図は記述できる必要がある。そこで、region 間の内容的な関連を明示的に記述する枠組を用意する。

3.2.1 2 項関連

region の間には 2 項関連を RDF のプロパティとして定義できる。代表的なものを説明する。

region の内容が示す情報の多寡 (informative) は話

題の豊富さ (topic informative) と内容の詳しさ (detail informative) の2つの次元で表現できると考える。

topic informative 次元の比較を wider/narrower (▷/◁) で表現し, detail informative 次元を deeper/shallower (▷/◁) で表現する。各 region の特徴的な Subject のみについて考えたとき, 詳細化/要約については, この関連の組み合わせで

- describe-of (▷, ◁): 一般的な説明/解説
 - specialize (▷, ◁): ある内容に特化する
 - outline-of (◁, ▷): 概要 (specialize の逆関連)
 - summary-of (◁, ◁): 要約 (describe-of の逆関連)
- の4つの関連で表現できる。

3.2.2 多項関連

region 間の多項関連は特殊な region として扱う。

Seq: RDF の標準タグであり, region 間に順番関連があるという関連である。

Part-of: ひとつの概念が複数の subregion の全部によって構成されることを意味する。

Discriminated-union: ひとつの概念が複数の subregion のいずれかによって構成されることを意味する。

3.3 空間関連

表示も一種の意味ととらえられるので, 再利用の際にはレイアウトについても指定できる必要がある。これを簡単な記述で可能とするのが, 空間関連である。Egenhofer の topological relationship^[4] をベースにしており, 特徴的なのは, disjoint の概念を分類し, 並んでいるという関連を neighbor-of, 同じページの中でも離れていることを強調しているものを separate, 異なるページにあるものを isolate とする点である。

4 意図記述

再利用の形態は, さまざまであるが, region をカプセル化されたオブジェクトとして扱い, region の引用という形での再利用を対象とする。また, 再利用先も SaRX コンテンツであることを前提とする。再利用に関するコンテンツ作者の意図としては, 再利用される region 自身の条件 (region の同時使用など) と再利用先の環境についての条件 (他の作者のコンテンツと同じページにしない, ○○について書かれたコンテンツの中ならば許可する, など) に分けられる。これらはすべて, region の属性および関連によって記述可能である。

ここで例を示す。RegionA は subRegionB,C から成り, B は画像である。B と C を同じページ⁽¹⁾ でかつそれしかないページをつくる⁽²⁾ という条件ならば, 使用を認める。また, RegionA より詳しい内容がコンテンツ内にあるはならない。⁽³⁾(図1, 2参照)

5 おわりに

本研究では, 意味構造を記述し, その意味構造を利用して Web コンテンツの再利用意図を記述するフレームワークを提案した。

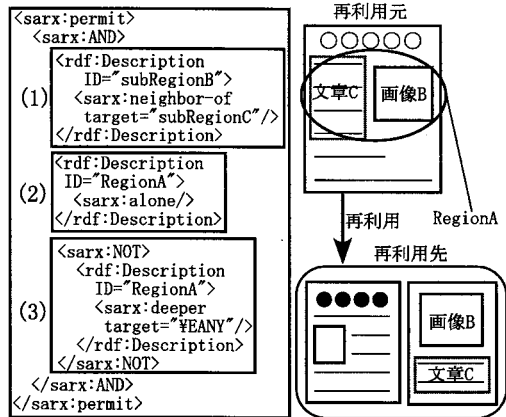


図1: 再利用条件記述例

図2: 再利用の例

また, 応用として意味構造を指定するコンテンツ生成がある。これは, キーワードと内容関連を指定して, 内容関連で記述した意味構造を持つキーワードに関するコンテンツを自動的に生成するものである。生成する元となるコンテンツは検索により決定し, そのコンテンツの意味構造を分析することによって抽出する部分を決定する。コンテンツの生成は再利用意図に基づいて行われる。例えば, 「keyword="マークアップ言語"⁽¹⁾の概要を説明したページ⁽²⁾があり, さらにその中でも Subject="XML", "HTML" について詳しく述べたページ^{(3),(4)}があるコンテンツ。なおかつ "XML" については例がある⁽⁵⁾もの」という指定をした場合のイメージを図3に示す。さらに, ビデオストリームなどの同期化コンテンツへの応用として, キーワードの出現時間を考慮した検索などが可能となる。

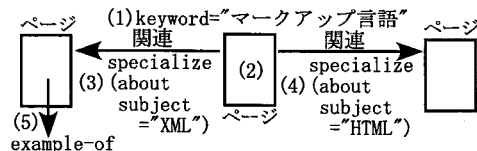


図3: 意味構造に基づくコンテンツ生成のイメージ

謝辞

本研究の一部は, NHK と京都大学との民間共同研究, 文科省科研費 (課題番号 12680416, 13224054), 日本学術振興会未来開拓学術研究推進事業 (JSPS-RFTF97P00501) による。

参考文献

- [1] World Wide Web Consortium: <http://www.w3c.org/>.
- [2] TopicMaps.Org: XML Topic Maps(XTM) 1.0 <http://www.topicmaps.org/xtm/1.0/>.
- [3] Dublin Core Metadata Initiative(DCMI): <http://dublincore.org/>.
- [4] Philippe Rigaux, Michel Scholl, and Agnes Voisard: Spatial Databases with Application to GIS, pp.108-109 (2002).