

O-018

## サムネイル配置情報からの RDF 及び RDF スキーマ抽出方式の提案

A proposal of a method for extracting RDF and RDF-Schema from positions of thumbnails

飯塚 京士† 向垣内 岳弥† 村山 隆彦†

Kyoji Iiduka Takeya Mukaigaito Takahiko Murayama

## 1. はじめに

今日のように多種多様なドキュメントがやりとりされる状況では、複数のドキュメント間の意味的な関係性を把握することが著しく困難になる。

上記問題を解決するため我々は、ドキュメントをサムネイル表示し、ドキュメント間の関係性を抽出する方式<sup>[1]</sup>を提案し、Windmill システム<sup>[2]</sup>を実装した。しかし、個人に閉じた GUI 上での情報整理には限界がある。

そこで、GUI 上での情報整理を他者と共有し、ドキュメント間の類似性を伝えることが有効であると、我々は考えている。情報整理を共有する最も簡単な方式は、GUI をそのまま共有することである。しかしこの方式では、他者の異なる観点などに対応できない。そこで、GUI から類似性を表現する構造化データを抽出し、これを他者に伝え、その他者の観点でドキュメント間の類似性を再構成することにする。

本稿では、Windmill から抽出可能な類似性情報を W3C 標準フォーマットである RDF・RDF スキーマ(RDFS)<sup>[3]</sup>に変換し、抽出・共有する方式を提案する。

## 2. 課題

## 2.1. 既存のアプローチ

複数のドキュメントから類似性を抽出し、かつ共有するためには、以下の 2 手段を提供する必要がある。

- 1) 複数ドキュメントを直観的に把握する手段
- 2) 類似性を通信・共有する手段

1)としては、画像データの閲覧を専門とする PhotoMesa<sup>[4]</sup>、iPhoto<sup>[5]</sup>、ドキュメントの管理を目的とした DocuWorks<sup>[6]</sup>など、ドキュメントをサムネイル表示するフォトビューアが挙げられる。しかし、ドキュメント間の類似性を抽出するためには別途手段が必要となる点が問題である。

また 2)としては、セマンティック Web 技術を用いたメタデータ・オントロジの共有方式がある<sup>[7]</sup>。しかし、メタデータやオントロジの構築の難しさ、閲覧性の悪さがあり、エンドユーザの使い勝手を考慮した方式は無い。

これらに対しユーザがドキュメントへアクセスする時系列情報からクラスタリングを行う研究<sup>[8]</sup>などがある。しかし、視覚的かつ直観的なサムネ

イル配置位置から類似性を抽出する方式は、まだ存在していない。

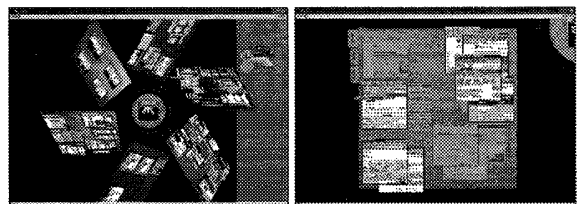
## 2.2. 課題の抽出

サムネイル表示 GUI をベースに、既存方式を分析すると、以下の課題が抽出される。

- 1) サムネイル位置からの類似性の抽出
- 2) 抽出した類似性の共有

1)に関して我々は、サムネイル位置から類似性を抽出する方式を検討し、Windmill システムを実装した(図 1)。Windmill はドラッグ&ドロップ操作によってドキュメント間の類似性を表現できるようにデザインし、2 種の階層関係を GUI の構造に入れた。

Windmill では、ユーザは GUI 上に自由にサムネイルを配置することが可能である。GUI 上のサムネイルの位置はユーザのドキュメントに対する明示されていないクラスタリング情報が反映されると考えられる。そこで、このクラスタリング情報を類似性と定義すると、サムネイル位置から類似性の抽出が可能となる。



GUI 全体像

Wing(3D 空間上の平面)

図 1 ドキュメント類似性抽出 GUI Windmill

本稿では 2)に相当するドキュメント間の類似性の共有方式を検討する。類似性を共有する場合、以下の点が問題となる。

- ・ 登録ユーザと、閲覧ユーザの観点のずれ
- ・ 複数のユーザが登録した情報の閲覧

上記問題に対し、抽出した類似性データを構造化し、閲覧ユーザの観点に合わせてカスタマイズ可能とする必要がある。構造化データの記述方式として RDF・RDFS を用いることとすると、上記問題は以下の 2 つに置き換えられる。

1. RDF・RDFS の抽出
2. RDF・RDFS の共有

## 3. アプローチ

## 3.1. RDF・RDFS の抽出方式

Windmill では、ユーザが GUI 上にドキュメントサムネイルを自由に配置できる。そこで Windmill

†日本電信電話株式会社 NTT 情報流通プラットフォーム研究所  
NTT Information Sharing Platform Laboratories, NTT Corporation

の機能を用い、サムネイルの配置位置が近いものを類似性が高いとみなして検討を進める。

ドキュメントの類似クラスを抽出する方式は、Windmill の機能を考慮すると以下の方式が考えられる。

- 1) GUI の構造上の位置
- 2) GUI 上の座標位置

Windmill では 2 種類の GUI の構造がある。1 つは、3 次元空間上にある平面(Wing)、もう 1 つは Wing 上に重なるように配置されているレイヤである。このように GUI の構造が階層をなす場合、そのまま RDFS のクラス階層関係に置き換えることが可能である。つまり、Wing をクラス、レイヤを下位クラスと割り当てることが可能となる。

また 2) により、座標位置が近いドキュメントをクラスとして抽出することができる。このクラス情報は、Wing により生成されたクラスの下位クラスと位置づける。

個々のドキュメントに関しては、抽出したクラスのインスタンスとして RDF を作成する。また GUI を再構成するため、以下の情報も RDF 化する。

- ・上記で抽出した所属クラス
- ・GUI 上の座標情報

### 3.2. RDF・RDFS の共有方式

RDF・RDFS の共有を行うために、本研究では P2P 型 RDF データベースであるコンテキストビューロ<sup>[9]</sup>を用いる(図 2)。

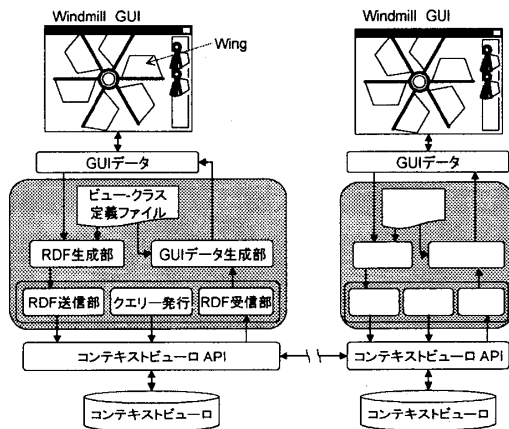


図 2 RDF・RDFS 抽出共有システム

コンテキストビューロには RDF・RDFS を格納するための API が用意されている。本システムでは上記方式で抽出した RDF・RDFS を API 経由で送信、格納する。また、コンテキストビューロは P2P 型の検索機能を有しており、ネットワーク上にある他のコンテキストビューロ上の RDF・RDFS を共有・検索することが可能となる。

また、ユーザ間の観点のずれに対応するために、GUI のカスタマイズ機構が必要となる。本方式では、ビュークラス定義ファイルを用意し、RDFS

で記述されているクラスと、GUI に表示する配置位置とのマッピングをカスタマイズ可能とした。

ビュークラス定義ファイルとして、以下のマッピング情報を利用する。

- ・Wing とクラス/ユーザの関係
- ・レイヤとクラス/ユーザとの関係
- ・配置レイアウト規則

これにより、各ドキュメントの所属クラスにより Wing/レイヤの配置を決定する。また、登録ユーザの観点を眺める場合は、RDF の座標情報を元に GUI に配置することが可能となる(図 3)。

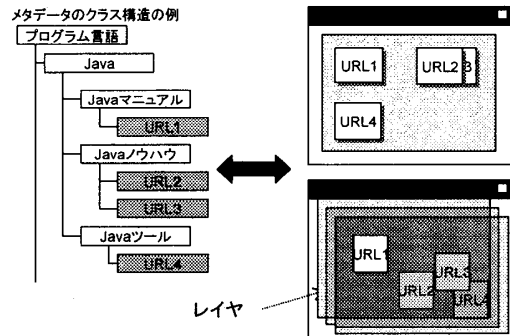


図 3 クラス・配置位置マッピング例

### 4. まとめ

複数のドキュメントの類似性を直観的に把握可能な共有環境を目指し、GUI からの RDF・RDFS 抽出方式及び RDF・RDFS 共有方式を提案した。これにより、GUI から抽出したドキュメント間の類似性のカスタマイズ表示することが可能となり、他者の整理情報を個々人の観点到に合わせた閲覧が可能となる。

今後は、本方式で抽出した類似性の検証を行う予定である。また、類似性以外の関係性が抽出可能であるか、検討を進めたいと考えている。

### 参考文献

- [1] 飯塚,中村,新井,花木,「流動的かつ非構造情報群の処理支援環境の提案」,FIT2003,pp.397-398,2003.
- [2] 飯塚,佐藤,中村,「Windmill コンテキスト抽出 3D デスクトップ」,INTAP セマンティック Web コンファレンス 2003,
- [3] <http://www.w3.org/RDF/>
- [4] <http://www.cs.umd.edu/hcil/photomesa/>
- [5] <http://www.apple.co.jp/ilife/iphoto/>
- [6] <http://www.fujixerox.co.jp/soft/docuworks/>
- [7] 佐藤,「「セマンティック Web」入門第 5 回」,日経インターネットソリューション 2003.8,pp.111-118,2003.
- [8] A. Graham, H.Garcia-Molina, A.Paepcke, T.Winograd, "Time as Essence for Photo Browsing Through Personal Digital Libraries," ACM JCCL'02 July,pp.13-17,2002.
- [9] 佐藤,境,中村,「コンテキストを流通するプラットフォーム技術とアプリケーション」,INTAP セマンティック Web コンファレンス 2003.