

文化資源情報における Topic Map の応用と評価

---「源氏物語」画像 DB を使ったモデル作成の試み---

Application and Evaluation of the Topic Map for the Cultural Resource Data ---Experiment for the Graphic Database of the Tale of Genji---

長瀬真理
Mari NAGASE

静岡大学情報学部、浜松市
Shizuoka University, Hamamatsu 432-8011

あらまし：本研究の目的は、現在、文書館・博物館・図書館・研究機関・大学・企業などで構築されている様々な文化資源情報を、情報の構築サイドではなく、個別の利用者の視点を反映して効果的に利用・評価するための新しい機構を調査・研究することにある。具体的には、ISO の標準として採用された Topic Map (ISO-IEC13250:2000) を小規模なデータベースに対し実験的に適用を開始し、逐次実験規模を拡大することによって、Topic Map の効果が期待されているXML型大規模データベースへの適用可能性を検証する。今回は小規模データとして、既に筆者が蓄積した「源氏物語」日英仏ハイパーテキストデータの中から、227 枚の画像データを使用する。画像データは、いくつかの主題に大別することが出来るため、トピックの地図作りのインデックス付加が容易であり、今後ユーザーがこの画像を使うためのモデルを提供することが可能であろう。ただし、本研究は始まったばかりであり、具体的な成果にまでには至っていない。本論文では研究方法の見直し、及びサンプルの DTD 作成までを解説し、読者の批判や助言を仰ぎたい。将来的には、Topic Map を本来の目的である大規模なデータにも適用し、その有効性や問題点も検討するつもりである。更に、それらの成果をデータベース構築にフィードバックし、検討された理論が、これまで蓄積されてきた多様な文化資源情報のマネジメントや評価作業に対して有効であることが証明され、これまで以上にユーザーに使いやすい新たなデータ構築をサポートすることが出来ればと願っている。

Summary: The purpose of this research is to investigate effective methods to use large scale data compiled in various archives, museums and libraries, not from developer's sides but from the standpoint of users. For one example of methods, I would like to deal with the Topic Map (ISO-IEC13250:2000) registered recently as a standard of the ISO and do some researches on its possibility and its validity for large data in the XML format. For the case study, I would apply this method to my multilingual and multi-media data of the Tale of Genji developed in the last ten years and would like to create a model for various usages. Further, if it is possible, I will feed back the results to production sides to promote effective usage of large scale data.

キーワード: トピック・マップ、ユーザー主体、XML、「源氏物語」

Keyword: Topic Map, user-oriented, XML, Tale of Genji

1 序論

伝統的な紙の本には、必ずインデックスがついていた。しかし、電子化されたドキュメントには、利用者側の情報利用に便宜を与えるこの種の配慮がなされていないし、またインデックスを記述するルールも規約も存在していない。確かに現在の高速な検索ソフトウェア（検索エンジン）を使えば、大量の電子文書から自動的に、いわゆる指定された文字列を含むコンテンツを、“正確”にアルファベット順に取り出すことは可能

である。あるいは検索を目的として、yahoo! などを使って、ある特定の項目について膨大な数に及ぶ検索結果集合を導くこともできるであろう。しかし、取り出されたインデックスも検索結果も実際に使うとなると、あまりにも大量であり、無意味な文字列の連鎖にすぎず、かえって我々を途方にくれさせるだけである。

今日、我々は「知や知識の蓄積と共有」を目指して様々なドキュメントをコンピュータに入力・蓄積しているが、それらはいずれも、制作者側の視点から編成されるデータベースであり、個別の利用者からみれば、宝の山ではなく、ゴミの山である。いみじくも Topic Map の提案者の一人である Steve Pepper が「インデックスのない本は、地図 (Map) のない国のような」と言っているように、「インデックスのない電子ドキュメント」は、何ら我々の知識構築には役立たない。こうした状況をユーザーサイドから解決すべく提起されたのが ISO (ISO-IEC13250) の標準として採用された Topic Map なのである。

2 Topic Map の基本概念

「トピック地図: Topic Map」とは、膨大な情報資源と情報操作の間に「橋」をかけようとする試みである。

それは、抽象的なデータ同士の意味をつなぎ、新たな知識を構築し、更にそれを新しい情報資源の地図として描き出す手法なのである。

Topic Map のモデルは①トピック、②トピックのデータ、及び③トピックの連合(association)の三要素から構成される。

具体的な解説の為に、例として、B.LeGrand と M.Soto の実験を紹介する。

まず Topic を定義しておこう。Topic とは、電子ドキュメントにおける検索項目、あるいはデータ項目で(本や印刷物のインデックスに相当する)意味連関の役割を担った情報である。

たとえば、会議を開催するための資料情報を考え、そのインデックスを作ってみよう。そして「この会議がヨーロッパのパリで開かれる XML 会議である」と想定してみよう。

すると、個々の”役割(role)”を担った言葉、即ち、論文、ビデオ、表、会議案内、等々で構成されたインデックスが出来あがる。この場合、個々の言語(インデックス)を、更にメタレベルで、異なった”タイプ(type)”のトピック、例えば、会議、都市、国名等など分けるならば、我々は、これらの個々のインデックス

を結合して新しい情報を作ることができる。例えば「XML ヨーロッパ会議がパリで開催される(take place)。パリはフランスに存在する。(is in)」等々である。即ち、異なったタイプのインデックス同士をメタレベルの意味機能で結びつけることによって、二つの結合、「開催される」と「に存在する」を作り出すことができるのである。

このように、Topic Map は資料情報のメタレベルに作成される地図で、トピックと情報ソースの二つの領域を分離させ、更にそれらのトピックとトピックの結合を、基盤となった情報の資料(データが生じるレベル)を超えたところに、新しい意味連関のネットワークを構築する知識創造の新しい方法なのである。

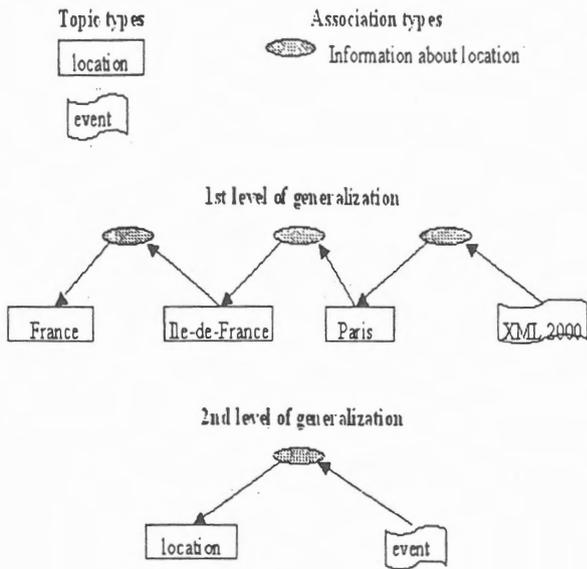
同時に、”地図”という言葉が連想するように、”移動可能な意味ネットワーク”の可視化(visualization)機能も兼ね備える。そこで、次に可視化機能の側面を紹介してみよう。

3. 可視化技術としての Topic Map

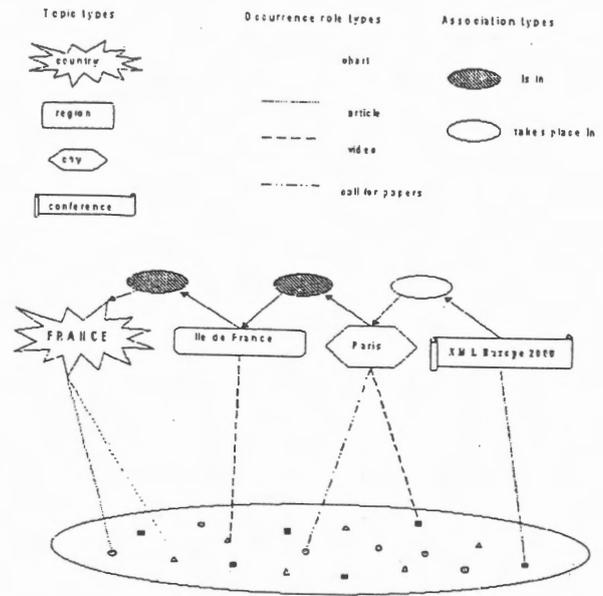
大規模な情報を可視化するために、LeGrand と Soto は開発された MAGE ソフトを利用した。このツールは、もともとは情報に色づけや、三次元表示、動画などの手法を付加して、情報を可視化するために開発された。

これを Topic Map の可視化に利用する。Topic Map は大量のトピックの結合を可能にするが、そのすべてを表示することは困難なため、まずは、関連する情報を選択しなければならぬ。すなわち、関連情報の選択には選別技術が不可欠となる。この選別技術は、XML の属性である、「名前」と「タイプ」によって情報を選別することによって実現される。現在、ネットワーク上に広く用いられている HTML が、情報の表現体裁の制御に限定しているのに比較し、XML は、構造などのタグ情報も提供することから、ネットワーク上での情報共有の新しい機能として注目されつつある。結果出力は次ページを参照願いたい。上段が意味レベルと基本概念、下段が、選択されたインデックスの連合例の可視化である。

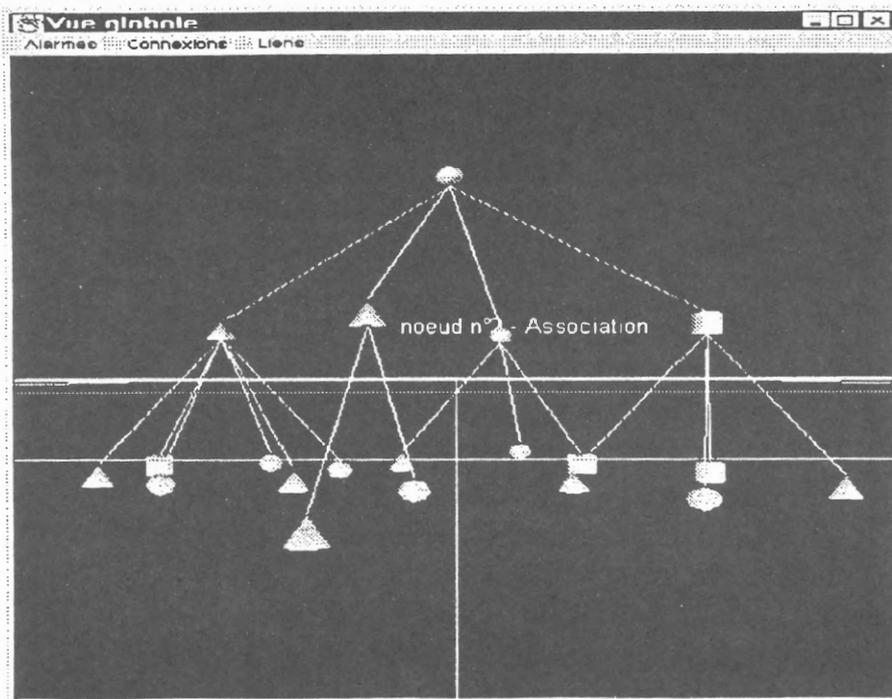
将来は、ユーザーの目的に応じた存在論的な秩序に適合するものへと選別技術が改良されることが期待されている。存在論的知識とは、各分野の基本語からなる語彙で表記される知識を指す。一般に、そこでの概念表記や、当然のことながら、自然言語の記述を基礎付けるものであり、これは、適切な KIF(知識交換形式 knowledge Interchange Format)を指示したものでなけ



Topic Map の意味レベル



Topic Map の基本概念



Topic Map の連合の可視化

ればならない。語彙は新しい概念の挿入がオープン(追記的に可能)である辞書構造と概念構造を持った形で蓄積されるようになる。

こうして、これらの可視化技術や XML 等の基盤技術をもとに、ユーザーは、自己の関心に従った地図(Topic Map)を作成する。今後、トピックと結合は政治地図、経済地図、あるいは文化地図という風に拡大されると推測される。

そこで、もう一つ、Steve Pepper が「オペラの本」電子ドキュメントを題材に行った研究を紹介しよう。

彼は、Topic Map の推進の代表者であり、ヨーロッパを根拠地に活躍している。ヨーロッパは、現在 Topic Map の普及に非常に熱心である。その理由は、もちろん、当地でも大量データの効率的利用が大きな関心事であるからに他ならなし。しかし Topic Map の拡大の裏には、アメリカに遅れをとったヨーロッパの威信のようなものがある。現在、ヨーロッパは、人権問題やデジタル・デバイド(情報格差)に非常に過敏になっている。多くの情報を発信するばかりでなく、受け手が、より適切に、より公平に情報にアクセスすることへの関心が高いのである。また、「知識は力なり」という伝統もあり、情報を、有効な知識の構築に利用しなければ意味がない、という基本的態度を持つ。そこで、一人一人の個人の立場から、大量データをどのように利用し、その知的活動を支援すべきかという問題に関心が高まり、その答えの一つが Topic Map なのである。

本題の Pepper の研究に戻ると、一般に、オペラの本であれば、どんなものであれ、原則的に、作曲家、作品、劇場など、オペラに関連する”トピック”を含んでいるはずである。

図1 トピック

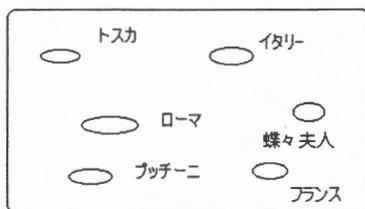


図1は、様々なトピックが、電子ドキュメントの中で並列的に存在する状況を図示したものである。本の裏

のアルファベット順のインデックス同様、このレベルではこれらトピックの役割(role)に差はなく、個々の単語の羅列状態と同じである。

このままでは、個々の語が、同等にテキストに埋没する、いわばプレーン・テキストと同じ状況であり、効果的な意味構築はできない。

そこで、実際の印刷物のインデックス作りでも、言葉の差異化の必要から様々な工夫がなされる。例えば、個々の言葉が特別な意味を持つことを示すために、タイプの違う文字列を作り、それを手がかりに、読者が目的の箇所を見つけやすくするようにしている。例えば、人名だけはイタリック表記の文字にするとか、キーワードはボールド・タイプにするといった具合である。あるいはジャンル分けをして、曲名だけをリストアップすることもあるだろう。

図2 トピック・タイプ

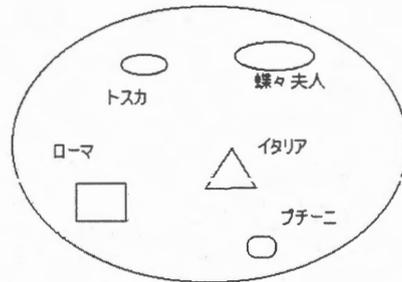


図2は、個々のトピックが、意味レベルで、異なったタイプのトピックであることを図示したものである。

こうすることによって、ユーザーは、個々のトピックの属性を区別することができるようになる。そうすることによって、作曲家と作品のようにトピック同士の結合を行い、有意味な知識の構築を行うこと容易になる。

また、マルチ・インデックス、例えば、名前ごとの、場所ごとの、主題別のインデックスも作成される。

更に、トピックのタイプは、作品名としての「トスカ」と、人物名の「トスカ」も区別できるものにする。あるいは、文中に出現する数字で表記されたデータについては、その数字が、ページ数なのか、脚注の番号を意味しているかなどの区別も可能になる。

トピック的なインデックスとして、ここでも重要なものは、①トピック(名前など一度以上出現する)、②ト

ピック同士の連合、③個々のデータである。

図3トピックの連合

トピックタイプ	出現箇所	連結
△ 国	図	存在
□ 市	論文	開催
○ 人名	ビデオ	
□ 曲名	論文集	

具体的に、先のデータの Topic と連合を行い、データとトピックの間に効果的な橋を架けてみよう。

- * トスカはプッチーニにより作曲された。
- * トスカはローマで演じられた。
- * プッチーニはルッカで生まれた。
- * ルッカはイタリアにある
- * プッチーニはヴェルディの影響を受けた。

といった連合が作り出される。

このようなトピックの結びつきによって表現される Topic Map は、情報資源に結びついた利用者の視点にたった情報資産の再表現を可能にする。すなわち、ここに挙げた情報はそれぞれ、ユーザーが作り出した新しい情報なのである。

更に、情報資源と Topic Map が互いに独立していることが重要とされる。これらが独立していることによって、この地図は他のユーザーにも適用されるし、また別の情報資源にも援用可能となる。

このような考え方は、ユーザーサイドの利用を前提した XML の構想にも共通するものであり、現在、XML と Topic Map の融合実験が精力的に行われている。

次に、ここで筆者の行ったデータベース構築を紹介し、Topic Map への応用の可能性について考察を進めたい。

4.「源氏物語」データベース

1996年から98まで、フランスのシャルル・ド・ゴール＝リール第三大学およびフランス科学技術庁、The Japan Foundationなどの協力で、フランス語版の画像データベースの開発に従事した。

画像(227枚)は、1650年、江戸時代随一の蔭絵師、山本春正の描いた木版画を採用した。実際は、貴重本

刊行会出版の「絵本 源氏物語」をスキャナーで入力した。

画像の原本は鶴見大学図書館と東京大学国文学研究室が所蔵しているが、これらは、大部分紙も傷み、画像もかすれており、公開される機会も少ない。そのため、デジタルカメラのやスキャナーでの入力は不可能であった。

一方、貴重本刊行会で出版された画像は、修正が為されておりスキャナーでの入力がたえるものであった。

更に都合の良いことに、個々の画像には三人の「源氏物語」の専門家の手による5～6行の解説がついており、そこを読むだけでストーリー展開を追うことができるようになっている。

4.1 画像のテーマ

227枚の画像のテーマを調べてみると、大体、以下の6つのカテゴリーに分けることができる。

- 1) 和歌の場面(手紙、手習い、絵画を含む) 38枚
- 2) 音楽の合奏 9枚
- 3) 宴会、遊び(狩り、舟遊び、雛人形、囲碁、双六、舞い) 35枚
- 4) 恋(秘められた恋、覗き) 41枚
- 5) 移動(車、馬、船) 16枚
- 6) その他(自然、語り、お化け、死、祈り、勤行、宗教行事、出産、見舞い、散歩、山歩き、夢、教育)

上記の分類の内 1) 和歌、2) 演奏、3) 宴会、4) 恋の場面みを合計すると、123枚で、全体の54.4%を占める。しかしながら、複数のカテゴリーに属する画像も多く、必ずしもこの通りの数にはならない。

カテゴリーのハイライト・シーンを、いくつか紹介する。

A) 音楽(数字は帖番号と画像通し番号)

音楽がカテゴリーの一つとして選ばれたのは、数の多さによるが、それ以上に、平安文化に占める音楽の役割が非常に大きいからに他ならない。



上記の場面は、有名な「雨夜の品定め」で知られている第2帖の一場で、左馬頭が愛人の家を訪ねた折り、庭先で笛を吹きその到着を知らせると、中から相手がそれに琴の伴奏で答える場面である。(2-8)

楽器が使われるシーンも多様で、宴会やコンサートの場面はもちろん、逢い引き、心の慰め、また踊りの伴奏にも主要な役割を担う。

和琴、唐琴、笛、太鼓、琵琶、笙、太鼓など様々な楽器が登場する。

B)恋

当然のことながら恋の場面は、覗きの場面も入れると41枚もあり、どれを選ぶか選択に困る。これは4帖から取られた画像で(4-18)、光源氏と年上の愛人である六条御息所と霧深い朝の別れのシーンで、女主人は、寝所から腫れぼったい目で源氏を見やっている。

源氏は見送りに出た女房の中将があまりに美しいので、御息所に隠れて、廊下の端の高覧に中将を座らせ、寸暇を惜しんで戯れる。同時に、彼は庭に少女をおろし、愛人へのプレゼント用に美しく咲き誇る萩や朝顔を摘ませる。

浮気っぽい源氏の有様を見事に伝える1枚であろう。



C)宴会

これは右大臣家での藤の花の宴(32-124)の場面である。紅葉や、桜の花見、梅の花見以外に、藤の花見の場面が2回にわたって描かれている。藤の花の色の濃いものは紫色で、蒔絵師は紫式部にふさわしい色と思ったのであろうか。



D) 遊



これは、空蟬と軒端の萩が囲碁に興じる場面(3-14)である。全体で囲碁の場面は3回登場する。2回は遊んでいるところ、もう一つは紫の上が碁盤の上に立って、源氏が髪を切りそろえてもらう場面である。物語全体では、囲碁を打つ場面は沢山あり、姫君達が、自慢の木を賭けて勝負をするエピソードなど様々である。

囲碁以外に、狩り、舟遊び、雛人形、双六、舞いなど遊びの場面は35枚もありカテゴリー第2位を占める。

このように、画像は、テキストだけでは分かりにくい平安時代の生活について、貴重な場面を描き出し、様々な情報を提供する。面白いのは、上記の囲碁の場面も含め、音楽の演奏の場面や、恋の場面で、他人から覗かれている画像が多いことである。こうした版画から、プライバシーが無い世界の有り様が容易に読者に明らかになる。ついでに、妻戸、格子、屏風、几帳、御簾などの仕切りで囲われた世界は、当時の建築や生活様式の理解にも大きな助けとなる。

ところで、こうしたカテゴリー分けは、翻って見れば、実は個人的なものに過ぎない。多数の人々が同意する解釈もあるだろうが、山本春正の画像データを目にしたユーザーはもっと多様な関心を抱くであろうし、新たな発見をすることであろう。だとすれば、データ

の製作者や提供者は、そうしたユーザーサイドにたつて、ユーザーがデータを個人化(パーソナライズ)出来るように、データを再構築することが必要となる。

製作者が出来あがったデータをただ提供するだけではなく、より多くの人々に使いやすいデータベースとして電子ドキュメントを開放することが求められる。

その方法の一つが、今回提起したTopic Mapへの応用なのである。

4.2 画像のテーマからトピックへ

画像データを Topic Map の手法で解析できるようにするための以下のような作業が必要となる。

- 1) 画像データを分類し、トピックと関連項目を洗い出す。
- 2) Topic Map を XML 形式で作成する。
- 3) 個々のデータへのリンクを張る。
- 4) Topic Map Navigator 等のソフトを開発する。

1) の作業については、筆者がテーマ別にカテゴリー分けした以上に、ニュートラルに行わなければならない。即ち存在論的な意味空間を構築できるようなトピックの設計を行うことが必要となる。今回は場所(家、部屋、庭等々)、人物(光源氏、紫の上、六条御息女等々)行為(ここでカテゴリーを生かし、遊ぶ、恋をする、笛や琴などの楽器を演奏する、花見をする、等々を設定)を選択肢とする。

2) の作業で、トピックの定義や結合方式を XML 形式でコーディングする。

3) の作業で、トピックの連合先、いわゆる画像へのリンクをはる。ここで有意義な知識の結合が具体化される。

例えば、

- * 左馬頭が笛を吹く。2-8 図の表示
- * 光源氏が六条御息女と恋をする。4-18 図の表示
- * 右大臣が花見をする。32-124 図の表示
- * 空蟬が囲碁をする。3-14 図の表示

4) のソフト開発に関しては困難が予想されていた。しかし、2000年11月17日に Ontopia Topic Map Navigator が公開され、フリーソフトとしてダウンロードできるという最新アナウンスが出たことから状況は好転している。

このナビゲータを使用するのならば、プログラミングの必要はない。しかしながら、このソフトを利用するためには、Ontopia の定義に添った XML DTD(Document Type Definition)の利用が条件となる。作

成された Topic Map が有効なものであれば、ナビゲータのブラウジングで求める結果を表示できるのである。ただ、残念ながら本論文には間に合わないため、最終的なチェックはできなかった。いずれにしても、Topic Map のソフトがフリーで入手できることから、今後、この手法の他分野への適用実験が一層進むと予想される。

4.3 DTD サンプル

ここで、実験的なものであるが、XML Topic Map の DTD のサンプルの大枠を作成してみた。

XML Topic Map DTD (試案)

```
<?Xml version="1.0" encoding="Shift_JIS"?>
<!ELEMENT topicmap
  (topic | assoc | mergemap)>
<!ELEMENT topic
  (topname | occurs)>
<!ATTLIST topic
  place CDATA #REQUIRED
  person CDATA #REQUIRED>
<!ELEMENT occurs EMPTY>
<!ATTLIST occurs
  instanceOf #PCDATA #IMPLIED>
<!ELEMENT asoc EMPTY>
<!ATTLIST asoc
  id ID #IMPLIED
  xLink #PCDATA #IMPLIED>
```

現在のところ、トピックの指定をどのように行うかは検討中である。また画像データを ENTITY 指定にして外部ファイルとする予定であるが、その場合の ID をどのように記述するかも未定である。

前出の Ontopia Navigator 用の DTD への修正も含め今後の課題としたい。

5. 将来の展望

ネットワークの進展がもたらした大きな変化の潮流の一つに、先にも触れたが、パーソナライゼーション(個人化)が挙げられる。パーソナライゼーションへの傾斜は、例えば One to One マーケティングであり、On Demand Publishing の出現など随所に見いだすことができる。

一方、情報技術の進展とともに、ネットワークを

した情報共有の検討が盛んであり、従来の図書館の機構を発展させた「電子図書館」や「電子博物館」なども実現されつつある。

しかし、上記のパーソナライゼーションの枠組みは、むしろ情報資源そのものを管理対象とするこれらの機構にこそ先んじて適用されるべきと言えよう。すなわち、個々のユーザーの視点に対応した情報提供であり、情報利用者の視点に立ってパーソナライズされた情報の構築が情報の有効利用の鍵を握る。

今後は単にデータの管理者視点からの統制に従った情報蓄積に加えて、個々の利用者の視点に立った電子化ドキュメントの”地図”作りが重要となろう。言い換えれば、Topic Map に代表される情報資源の組織化の研究を通して、情報資源のメタレベルでの再構築・再編成が目指されねばならない。

そのためには、個別分野ごとに、トピックの存在領域に根ざした「トピック地図テンプレート Topic Map Template」を構築し、あわせてその有効性を評価する必要がある。テンプレートの構築には、情報知識、情報工学、知能情報学の研究者の協力を仰ぐ必要がある。また、個別科学を構造的・総合的に構築できる分野ごとの専門化の協力も必要となろうし、ユーザーの利用形態についての実証的研究も忘れてはならない。

今後、Topic Map の研究を通して、情報資源のメタレベルでの再構築の方途を探り、将来的には、各分野のトピックを標準化させ、意味ネットワークとして利用者に流通させることが期待される。本研究は、そのための小さな第一歩である。

7. 謝辞

本論文の執筆にあたっては、(株)シナジーイン・インキュベートの菊田昌弘氏と内藤求氏に様々な助言を頂戴した。紙面を借り感謝の念を表したい。

参考文献:

- 1 Benedicte Le Grand, Michel Soto, "Information management-Topic Map Visualization, Conference Paper at ACM CIKM'99
- 2 Steve Peppe, "The TAO of Topic Maps, Conference Paper at ACM CIKM'99
- 3 ISO/IEC 13250, Information technology-SGML Applications- Topic Maps(ISO, Geneva 2000)
- 4 C. ゴールドバーク「SGMLハンドブック」、アスキー 1993