

# 高品質ハイパーメディア実現と その利用に関する研究

原 良 憲

高品質ハイパーメディア実現と  
その利用に関する研究

原 良 憲



Doctorial Dissertation Series of Tanaka Laboratory  
Department of Social Informatics, Graduate School of Informatics,  
Kyoto University

Copyright © 2005 Yoshinori Hara

## 概 要

本論文では、Web 情報活用の優劣による編集・利用格差が増大している状況をふまえ、大規模な情報洪水に埋もれて必要な情報を探しだせない問題（情報空間の迷子問題）、関係づけの複雑化を回避する設計課題（ハイパーメディアの設計問題）、並びに、編集者・利用者間の認識共有による信頼構築の問題（情報のトラスト問題）に焦点をあて、ハイパーメディアの関係づけ（リンク構造）に関する高品質化についての研究開発とその成果・意義について報告する。これらの問題解決に際しては、データベースの設計モデルと、人間を系に含む情報処理サイクルモデルとからの統合アプローチで実現をめざす。

本論文で述べた研究開発による当該技術分野への寄与は、第一に、ハイパーメディア抽象化という考え方を導入し、例外を考慮したリンク構造のクラスタリング方式により、大規模ハイパーメディアに対する見通しのよい全体構造把握と効率的な物理構造設計とを実現した点である。ハイパーメディア抽象化の考え方は、後年の Hub-Authority 等のリンク解析、Web コミュニティ発見による情報分析へと発展拡張される。

第二に、関係づけが複雑になることを回避するため、ハイパーメディアに対するトップダウン的な論理構造設計を提案した。また、言語規約に基づくデータベースからのハイパーメディアのリンク自動生成についても導入した。これらの活動により、編集・利用効率のよいハイパーメディアシステムを作成することができる。現在において、大きな比重を占める深層 Web の有効活用を支援する活動とも位置づけられる。

第三には、ハイパーメディア抽象化の考え方を、情報提示と検索・ナビゲーションに応用し、直感的で、かつ、周辺情報としての文脈を意識した情報活用ツールを提供した点である。今日的には、多次元情報分析ツールの先駆けとも位置づけられる活動である。

また、第四には、情報利用者から編集者へのフィードバック情報を円滑に提供できる仕組みを想定したハイパーメディア(Augmented Hypermedia)、並びにシステム構成を提案し、実際の Web 美術システムに適用を行った点である。このような考え方は、自律的機構を組み込むことによる信頼性向上や、Blog のような双方向リンクをもつハイパーメディアにおける情報活用へと発展している。

今後は、さらに安心して快適な活用が行える情報アクセスナビゲーション基盤実現のため、ハイパーメディアの情報科学的側面からの分析だけでなく、人間の行動様式の分析などを含む社会科学的側面からの分析や統合が重要である。



# 目次

概要	i
目次	iii
図表一覧	vi
第1章 はじめに	1
1. 1 背景	
1. 1. 1 World-Wide Web とハイパーテキスト/ハイパーメディア	
1. 1. 2 情報ナビゲーションアクセス基盤としての Web の利点	
1. 2 課題認識 — 大規模な情報洪水の中での制御・活用 —	
1. 3 高品質ハイパーメディア	
1. 4 アプローチ	
1. 4. 1 データベース分野からのアプローチ	
1. 4. 2 CHI 分野からのアプローチ	
1. 5 本論文におけるハイパーメディア研究開発の貢献と意義	
1. 6 本論文の構成	
第2章 ハイパーメディアの抽象化と物理設計	21
2. 1 ハイパーメディアの抽象化とは	
2. 2 関連研究	
2. 3 大規模ハイパーメディアの問題点	
2. 4 効率的かつ効果的なナビゲーションに対する要求事項	
2. 5 ハイパーテキストアグリゲーション (Hypertext Aggregation)	
2. 5. 1 ハイパーテキストアグリゲーションの例	
2. 5. 2 ACE: 例外を考慮した集約化クラスタリング	
2. 5. 3 ACE クラスタリング手法の定式化	
2. 5. 4 ACE クラスタリングの計算量	
2. 5. 5 ハイパーメディア物理設計の分析モデル	
2. 5. 6 主記憶 DB における最適格納コスト	
2. 5. 7 効率的な検索のための二次情報管理	
2. 5. 8 ACE クラスタリングを実現する一探索的アルゴリズム	
2. 5. 9 実験結果	
2. 5. 10 分析モデルとの比較	
2. 5. 11 今後の検討課題	

## 2. 6 本章のまとめ

### 第3章 抽象化によるハイパーメディアの論理設計

57

- 3. 1 ハイパーメディア論理設計の必要性
- 3. 2 関連研究
- 3. 3 ハイパーメディア論理設計の課題と要件
  - 3. 3. 1 課題
  - 3. 3. 2 ハイパーメディア論理設計の要件
- 3. 4 拡張 E-R モデルによるハイパーメディアデータモデル
  - 3. 4. 1 ハイパーメディアプラットフォーム Himotoki
  - 3. 4. 2 システム評価と課題
- 3. 5 関係抽象化によるハイパーメディア論理設計
  - 3. 5. 1 データベース設計における抽象化の考え方  
(従来のデータモデル設計)
  - 3. 5. 2 Database Abstraction の例
  - 3. 5. 3 関係の抽象化設計 (Relationship Abstraction)
  - 3. 5. 4 関係の抽象化に対する要件
  - 3. 5. 5 ハイパーメディア論理設計手順
  - 3. 5. 6 関係抽象化による論理設計の評価と課題
- 3. 6 言語規約によるデータベースからのハイパーテキスト生成
  - 3. 6. 1 ハイパーメディア DB の設計指針
  - 3. 6. 2 Hypertext Projection
  - 3. 6. 3 Hypertext Projection 変換言語 (HPTL: Hypertext Projection Translating Language)
  - 3. 6. 4 Hypertext Projection の評価と展望
- 3. 7 本章のまとめ

### 第4章 大規模ハイパーメディア構造可視化とブラウジング

97

- 4. 1 ハイパーメディアの構造理解と可視化技法の必要性
- 4. 2 関連研究
- 4. 3 **Focus+Context Views**
  - 4. 3. 1 Landmark Node の発見・規定
  - 4. 3. 2 Focus+Context Views の実現機能
  - 4. 3. 3 評価と考察
- 4. 4 スキーマブラウザ
  - 4. 4. 1 スキーマブラウザ実現の背景

4. 4. 2	集合操作リンクに基づくスキーマブラウザ	
4. 4. 3	集合操作リンクを実現するデータモデル	
4. 4. 4	トラバース (当該の読み方) のパターン	
4. 4. 5	評価と考察	
4. 5	本章のまとめ	
<b>第5章</b>	<b>Augmented Hypermedia</b>	<b>119</b>
	— 編集・利用効率をめざすシステム構成 —	
5. 1	Augmented Hypermedia 構成の背景	
5. 2	関連研究	
5. 3	Augmented Hypermedia	
5. 3. 1	関係情報の共有	
5. 3. 2	関係情報の精緻化	
5. 3. 3	関係情報の抽象化	
5. 4	システム構成	
5. 4. 1	3層構成のシステムアーキテクチャ	
5. 4. 2	AMOREによるシステム構成	
5. 5	Augmented Hypermedia の利用に関する知見	
5. 6	本章のまとめ	
<b>第6章</b>	<b>まとめと今後の展望</b>	<b>133</b>
6. 1	高品質ハイパーメディア研究のまとめ	
6. 1. 1	高品質ハイパーメディアのモデル、機能、プロセス	
6. 1. 2	情報空間の迷子問題に対する対処と知見	
6. 1. 3	ハイパーメディアの設計問題に対する対処と知見	
6. 1. 4	情報のトラスト問題に対する対処と知見	
6. 2	今後の研究展望	
6. 3	結言	
<b>謝 辞</b>		<b>145</b>
<b>主要文献・その他関連文献・参考文献</b>		<b>147</b>



## 図表一覧

- 図 1-1 ハイパーメディア・Web の進展
- 図 1-2 人間を系に含む Web 情報アクセスモデル
- 図 2-1 学生の受講関係の例
- 図 2-2 参照マニュアルの例
- 図 2-3 ACE(Aggregation Clustering with Exceptions)クラスタリング
- 図 2-4 ACEクラスタリングの E-R ダイアグラム
- 図 2-5 完全2部グラフ(Complete Bipartite Graph)における頂点集合
- 図 2-6 ACEクラスタリング例
- 図 2-7 式の図的表現
- 図 2-8 データベースサイズ $S$ とストレージ関数 $\alpha^2$
- 図 2-9 ACEクラスタリングの2次記憶構成
- 図 2-10 “Hypertext on Hypertext”の大局構造
- 図 2-11 ランダムグラフの入力データ
- 図 2-12 ランダムグラフの結果例 (Aggregation Graph + Inclusive Links)
- 図 2-13 ランダムグラフの結果例 (Aggregation Graph)
- 図 3-1 ハイパーメディアプラットフォーム Himotoki
- 図 3-2 Himotoki のナビゲーション機能
- 図 3-3 パリ観光案内ハイパーメディアの拡張 E-R モデル
- 図 3-4 静的リンクを用いたナビゲーション例
- 図 3-5 スキーマブラウザを用いたナビゲーション例
- 図 3-6 Aggregation と Generalization
- 図 3-7 映画と俳優・女優との関係 (多対多)
- 図 3-8 関係の一般的属性
- 図 3-9 Augmentation
- 図 3-10 Globalization
- 図 3-11 関係抽象化の例
- 図 3-12 Hypertext Projection の概念図
- 図 3-13 データベースモデルからハイパーメディアの生成
- 図 4-1 Landmark Node に対する Cut-off 値の変化
- 図 4-2 Focus+Context View の例
- 図 4-3 Focus+Context View のイメージ検索システムへの適用
- 図 4-4 観点を導入したナビゲーション
- 図 4-5 観点及び属性の関係表現の例

- 図4-6 カテゴリーブラウジング
  - 図4-7 制約トラバースの例
  - 図4-8 ズームイン/ズームアウトの例
  - 図4-9 クロスリファレンス
  - 図4-10 トラバースの例
  - 図5-1 Augmented Hypermedia のコンセプト
  - 図5-2 Augmented Hypermedia 機構機能構成
  - 図5-3 Augmented Hypermedia システム構成
  - 図5-4 Augmented Hypermedia 3層構成の詳細
  - 図5-5 AMORE システムアーキテクチャ
  - 図5-6 AMORE システム構成
  - 図5-7 Augmented Hypermedia 応用
  - 図6-1 課題に対する具体的解決施策のまとめ
- 
- 表2-1 実験結果

# 第1章 はじめに

## 1. 1 背景

### 1. 1. 1 World-Wide Web とハイパーテキスト/ハイパーメディア

1989年に Tim Berners-Lee により World-Wide Web (以下、Web と略称) が提唱されて以来、10 数余年の歳月が経過した。その間、インターネットの普及とあいまって、Web 情報活用は劇的に進化した。Web は、その言葉の如く、世界中に張り巡らされた情報のくもの巣である。コンテンツ情報というノードと、関連のあるコンテンツ間を結合したリンクとで構成される。このような直感的にわかりやすいメタファーが、情報作成者や利用者を受け入れられ、結果として大規模な情報ネットワークの構築とその活用とがなされてきた。Web は、日常生活やビジネスにおける情報発信・共有利用メディアとして、今や必要不可欠な存在である。出版、放送、物品売買、情報交換、遠隔教育、企業内活動、企業間連携、コミュニティ活性、公共サービス等々、あらゆる分野での情報活用基盤となっている。

UC Berkeley の”How Much Information?” プロジェクト [HoMI03]によれば、2003年時点で、世界中で数十億ページの Web 情報 (約 170 テラバイト) があると推定されている。このタイプの Web は、表層 Web (Surface Web)と呼ばれ、あらかじめ Web ページとして作成されている静的なもので、かつ一般に公開された情報アクセス可能なものである。実は、その背後に、データベース等から動的に生成される Web 情報や、ファイアウォール等のセキュリティ機構で外部アクセス制限された Web 情報が存在する。これらの情報は、深層 Web (Deep Web)と呼ばれる。深層 Web の量は、表層 Web の量に比べ、約 500 倍多くあると推定され (推定モデルにより、約 7.5~92 ペタバイト)、人間が感覚的に理解できないほどの巨大な情報空間を形成している [Berg01] (図1-1 参照)。

情報コンテンツ利用基盤としての Web の出現は、歴史的に見ても、印刷技術、電信電話の発明や、郵便サービスの普及などと同様に、非常に大きなパラダイムシフトである。情報ネットワーク基盤をもとにした情報蓄積・利用メディア実現に対しては、通信・コンピュータ技術の発展と共に、ネットワーク、データベース、情

## 第1章 はじめに

報検索、ワークフロー等の要素技術に基づく応用システムが、多数実現されてきた。しかしながら、今日の Web のように、日々の生活に浸透した形態で、全世界規模の必要な情報に直接アクセスできる情報作成・蓄積・検索・提示・流通利用メディアは、歴史上初めてである。

このような Web 情報の根幹となる情報表現・操作モデルは、ハイパーメディアという情報表現法と、そのナビゲーションな情報操作法とを基本としている。ハイパーメディアとは、マルチメディアで構成された文書を関連づけた情報構造の表現法であり、対象情報が、テキスト情報のみのものは、ハイパーテキストと呼ばれる。これらは、1945年に Vannevar Bush の提唱した Memex [Bush45]から連綿と精緻化された非線形文書モデルを源泉とし、ノードとリンクとからなるグラフ構造を対象としたグラフ理論などが、その理論的基盤となっている。グラフ理論に基づく Web 関連技術の著名な具体例としては、高精度 Web 検索エンジンである Google があげられる。これは、グラフのリンク構造に基づくページ重要度評価アルゴリズムが基本となっている。また、共通性ある Web サイト群を発見する Web コミュニティ分析など、意味ある情報の自動抽出をめざす Web マイニング技術も、グラフ理論的情報分析を基盤としている。

グラフ構造に基づくハイパーメディアや Web の情報操作法は、データベースアクセスにおける SQL 言語問い合わせのような宣言的なアクセス手法との対比で差別化され、ナビゲーションな情報アクセス法と呼ばれる。これは、利用者とのインタラクションの中から順次関連情報を手続き的（芋づる式）にアクセスする手法である。ナビゲーションな情報アクセスは、ユーザインタフェースとしての工夫も行いやすく、いろいろな理解度の利用者にとっても直感的に認識でき、広範に受け入れられてきている。

このように Web とハイパーテキスト／ハイパーメディアの歴史的発展経緯は厳密には異なるが、Web 情報利用の普及が進んだ 1990 年代後半以降、両者の構築、利用フレームワークの考え方は、ほぼ同義となってきている。以降の記述においては、ハイパーメディアの高品質化という本論文の主題を明確化して焦点をあてることにするため、Web やハイパーメディアを情報ナビゲーションアクセス基盤として位置づけ、両者を統合して説明を行うこととする。

# 第1章 はじめに

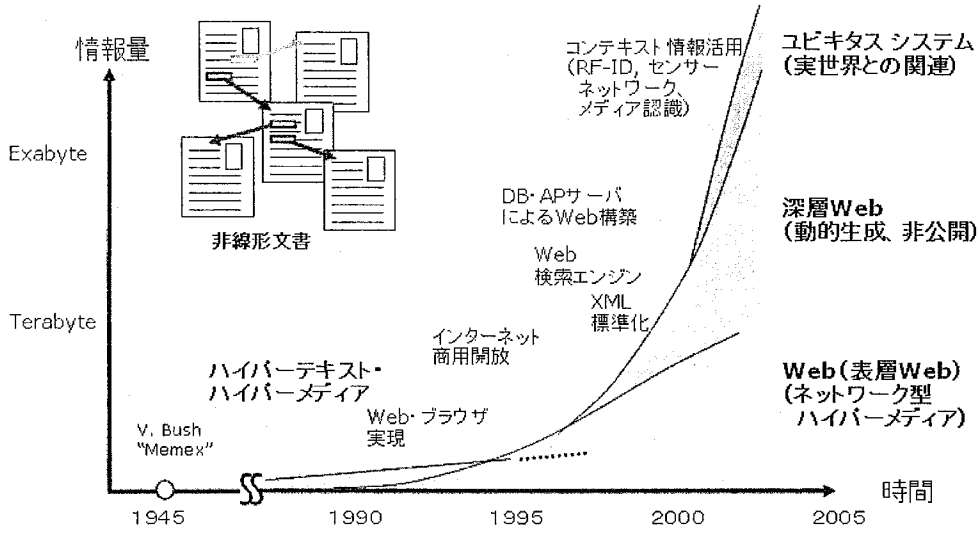


図1-1 ハイパーメディア・Webの進展

## 第1章 はじめに

### 1. 1. 2 情報ナビゲーションアクセス基盤としての Web の利点

Web システムの構築や運用という側面では、使いやすく拡張可能なシステムアーキテクチャとしての寄与が大きい。Tim Berners-Lee らによる HTTP (HyperText Transfer Protocol)と HTML(HyperText Markup Language)の開発と、その数年後に開発された Marc Andreessen らによる MOSAIC ブラウザの開発により、Web システムアーキテクチャの基本が実現された。これに、米国におけるインターネットの商用開放、基本 OS 寡占化に対するブラウザや Java 関連技術の対抗、アプリケーションサーバ実現による電子商取引の促進などがあいまって、Web 情報システムインフラ基盤が実現されてきた。Web 情報の生成・蓄積は、このようなインフラ基盤の進展と共に、相互進化的に増大していった。

一方、Web システムインフラ実現に加えて、情報処理サイクル（編集・蓄積・検索・提示・配信等）において、Web 情報の提供者、享受者双方に、Web 情報利用の効用やその価値が理解されてきたことも重要な要素である。以下では、情報ナビゲーションアクセス基盤としての Web の利点を、整理して概説する。

#### (1) 非同期並行編集による効率的な Web 情報作成

まず、第一には、情報コンテンツ構築の基本となるデータ作成・編集の柔軟性、拡張性に関するものである。Web 情報の作成・編集は、非同期の並行型情報処理である。Web 情報の編集者は、自身のページの作成と関連するリンクの作成という Web 情報空間のごく一部の編集を担うだけで、結果的に大規模な Web 情報空間が形成される。すなわち、Web 情報へのアクセス頻度といった利用者側の便益に対し、非常にコストパフォーマンスのよい情報コンテンツ編集環境を提供している。

#### (2) サーチエンジン等による大規模 Web 情報の分類・検索

第二には、Web コンテンツ検索処理の効率化があげられる。Yahoo, Looksmart をはじめとするディレクトリ型の分類・検索サービス、Google などの大規模サーチエンジンによる検索サービスなどが実現され、Web コンテンツの検索・アクセスが容易になった。また、サーチエンジンによる Web 全体の情報検索については、インデックス情報の効率的実装に加えて、高速 Web クローラ (Web 文書の自動収集) の寄与も大きい。適切な新鮮度での Web 文書のスナップショットが得られ、このスナップショット情報をもとに、巨大な Web 情報空間のインデックス作成が可能となった。また、リンク解析による検索結果の自動ランキング技術や、効率的なトピックフォーカス型情報収集技術は、ノイズが少なく、かつ、もれも少ないという意味での高精度な Web 情報検索の実現を導いた。

## 第1章 はじめに

### (3) Web 情報マイニングに基づく新しい情報分析・活用

第三には、情報メディアとして、いままでにはない新しい価値実現に関するものがあげられる。たとえば、大規模な蓄積情報を基に、全体的な情報のトレンド分析、異常データの検出・把握、関連トピック発見といった新しい価値を提供するものである。このような Web 情報の分析・活用は、Web マイニングと呼ばれているが、分析対象により、Web 内容マイニング、Web 構造マイニング、Web 利用マイニングとに大別される。特に、Web 構造マイニングは、Hub-Authority リンク構造に基づくページランク [BrPa98]、Web コミュニティ [KRRT99, FILG00]など、ハイパーメディア固有の構造を生かした情報分析手法として進展した。近年では、Blog に代表される双方向リンクに基づく情報活用も活発である。また、アクセス頻度などの利用者行動特性を分析し、Web 情報の人気度、評判、推薦情報等を得ることも重要となってきた。

### (4) Web 利用を介した業務プロセス・生活様式の改善

第四は、人間の経済的活動や文化的活動に対応するプロセス効率化に関するものである。たとえば、企業におけるリソース管理・計画 (EPR)、顧客管理 (CRM)、電子物品売買等の改善、効率化である。近年の Web ベースの情報処理システムの継続的成功は、このような具体的業務や生活に密着した情報活用によるところが大である。また、このような情報活用が、新たなビジネスモデルを生み出し、経済的価値のあるプロセス改善・革新を引き出している。たとえば、Web におけるポータルサイト (玄関サイト) の出現は、情報閲覧者 (訪問者) 数を背景にした Web 広告ビジネスモデルを引き出し、新しい情報アクセスプロセスの革新を生み出した。また、電子ショッピングやインターネットオークションに対しても Web 情報検索技術が重要な役割をもち、インターネットならではの、新しいビジネスプロセスの実現がなされている。

上述したように、情報ナビゲーションアクセス基盤としての Web の利点は、広範、多岐にわたる。Web 情報利用環境は、インターネットバブル崩壊による過剰な Web 情報サービスの淘汰の時代を経て、現在では、世界中の人々が価値ある情報を相互活用できる時代へと成熟し、さらに大きな成長の可能性を示している。

## 1. 2 課題認識 — 大規模な情報洪水の中での制御・活用 —

以上概観したように、Web の利点は広範、多岐にわたり、かつ、さらに大きな成長の可能性を示しているが、一方では、表層 Web だけでも数十億ページにわたり、さらにその 500 倍多くあると推定される深層 Web の存在により、人間が感覚的に理解できないほどの巨大な情報空間になってしまっている。このような状況においては、実は、最先端の Web 検索エンジンもあまり有効に機能はしていない。たとえば、[LaGi99]によれば、単一の Web 検索エンジンは、表層 Web のただかたかた十数%程度しかカバーできていない。また、更新頻度の高い Web サイトに対する巡回収集（クローリング）の限界や、全域的な情報解析をする限界などにより、リアルタイム的な情報分析もできていない。また、深層 Web に対しては、全く別のメカニズムで情報の有効活用を考えなければ、分析すら行えない状況である。さらには、利用者層が広がったことにより、Web 情報活用のしかたの優劣による情報利用格差（情報リテラシー格差）がますます増大してきている。

我々は、このような大規模な情報洪水の中で、多大な労力を費やしても関連のない情報ばかりに翻弄されてしまうのか、あるいは、大量の情報を積極的に活用して意味ある情報検索や傾向分析に活用できるのかという岐路に立たされている。本論文においては、Web やハイパーメディア情報の大規模な洪水の制御対処法を大きな究極課題と捉える。課題解決へ向けてのマイルストーン規定に際しては、本論文では、読み手（情報利用者）固有の課題、書き手（情報作成者）固有の課題、ならびに、読み手と書き手との双方に関連する課題とに分けて議論を進める。

### (1) 読み手固有の課題：情報空間の迷子問題

利用者にとって、大規模な情報洪水に埋もれて、最新の役に立つ必要な情報をなかなか探しだせないことは、深刻な問題である。ハイパーメディアの分野では、情報空間の迷子問題 (Lost in Hyperspace、あるいは Disorientation Problem) として位置づけられ、古くから重要な課題の1つとされてきたものである。すなわち、情報利用者が、Web や非線形文書の情報コンテンツ利用の枠組み（情報空間）の中で、いま自分がどこにいるのか、また、どちらの方向に進めば、目標とする情報コンテンツに到達できるのかがわからなくなるというものである。比喩的にいえば、訪問した新しい街で、地図を持たずに路地に入り込んでしまったような状況である。また、Web 情報では、1つの主題が複数のページにまたがって記載されていることもあり、リンク構造の複雑化とあいまって、情報空間の迷子問題がさらに深刻な課題となっている。



### (2) 書き手固有の課題：ハイパーメディアの設計問題

大規模な情報の洪水に対しては、大量の情報が生成されても、氾濫することがないように予防的な対応が必要である。すなわち、書き手側で大量の情報が生成される時点、あるいはそれ以前の時点で、適切な制御対応が重要となる。Web では、複数の編集者が同時にページを作成したり、リンクづけしたりできるため、重要なアプローチとしては、Web やハイパーメディアの設計方法論、たとえば、編集者間でのシナリオ共有やリンクづけ規約の明確化などがあげられる。特に大規模な環境では、ノードの2乗のオーダーで複雑度が増す関係構造（リンク）に対する設計指針が重要となる。ハイパーメディアの利点である簡便性を生かしつつ、トップダウン的な設計で関係構造の制御困難性を打破し、品質向上を行うことが必要である。

### (3) 読み手と書き手とに関連する課題：情報のトラスト問題

本論文で議論する情報のトラスト問題とは、読み手－書き手間での情報に対する信頼が確立できない問題である。これは、書き手側の提供する情報品質（誤った情報、誤った関係づけなど）の程度にも依存するが、情報編集者と情報利用者間での目的意識・意図の相違という問題にも起因する。Web のような大規模ハイパーメディアを対象としたコンテンツの品質を高め、かつ、情報リテラシーの習熟度が異なる利用者也想定して、情報活用効率をあげる仕組みが必要とされる。単に関連情報を繋ぎあわせればよいのではなく、どのような利用者がどんな意図・目的で情報活用しようとするのかという側面を考慮する必要がある。書き手の意図を正確に読み手が認識・理解したり、読み手が安心して情報をアクセス・利用し、書き手にフィードバックできたりする仕組みが求められる。

### 1. 3 高品質ハイパーメディア

本論文では、上述の3つの具体的な課題認識に対し、高品質ハイパーメディアという考え方を提案し、その実現と利用とを通じて、解決するアプローチをとる。「高品質」という言葉は、一般的な用語であり、様々な解釈が可能である。たとえば、内容に誤った情報が記載されていない、安心かつ安全に利用できる、高級感がある、いつでも利用できる、作りやすい、伝えやすい、探しやすい、正しく探せる、間違いなく理解できる等々である。

以下、本論文では、技術的な側面から課題解決に寄与する高品質性に限定し、モデル、機能や、プロセス面での提案とその実証とを行う。また、課題解決に際しては、データベース分野とCHI（コンピュータ・ヒューマン・インタラクション）分野双方からのアプローチを援用し、特に、ハイパーメディアの特長である関係構造の編集と利用双方の向上をはかるという高品質性を実現する。従って、セキュリティや、情報コンテンツそのものの信頼性向上方策等は、本論文の範囲外とする。

高品質を目指す機能を説明する際に、まず、どのような目的に対する誰のための技術なのかを明確にして、対応づけておく必要がある。ここでは、書き手もしくは読み手に対する効用と、書き手もしくは読み手に対する技術との組み合わせという観点から提案方式を整理し、上述の3つの具体的課題（情報空間の迷子問題、ハイパーメディアの設計問題、情報のトラスト問題）の解決を試みる。

#### (1) 大規模関係構造分析に基づく「全体構造の概略整理・把握」による高品質化（書き手への効用を目的とした書き手に適用する技術）

まず、現状のハイパーメディア関係構造を分析し、どのような関係構造が良い構造かの議論を行う。また、大規模ハイパーメディアに対しても見通しの良い解を見出すため、スケーラブルで効率のよいアプローチを検討する。本論文では、ハイパーメディアの抽象化という考え方を提案する。この考え方は、ハイパーメディアの概略構造（粗構造）を抽出し、情報空間の迷子問題や、ハイパーメディアの設計問題（物理構造設計）の解決を目指すものである。本内容は、ハイパーメディアの抽象化として2章に詳細記述する。

#### (2) ハイパーメディアトップダウン設計による「効率的な情報活用」を追求した高品質化（読み手への効用を目的とした書き手に適用する技術）

次に、高品質な関係構造を有するハイパーメディアを設計、構築するにはどのような指針がよいかについて議論する。本論文では、ハイパーメディアのトップ

## 第1章 はじめに

ダウン的論理設計法を提案する。具体的には、データベース的なアプローチの良さを取り入れたハイパーメディア設計、関係構造に着目した概念設計、ハイパーメディア関係構造の自動生成についてである。特に、編集・利用時期の差異や、編集者-利用者間での意識のずれを軽減し、あるいは、差異がある場合には、観点や差異を認識することにより、コンテンツの信頼性向上を図る。この考え方は、ハイパーメディアの設計問題（論理設計）や情報のトラスト問題の解決を目指すものである。本内容は、ハイパーメディアの論理設計法として3章に詳細記述する。

### (3) 全体概略構造を活用した「探しやすい・理解しやすい」情報提示と検索・ナビゲーションによる高品質化（読み手への効用を目的とした読み手に適用する技術）

さらに、対象とする大規模ハイパーメディアに対し、どのような情報検索・ナビゲーション手法や、情報提示手法が有用かについて議論する。本論文では、全体構造と詳細構造とのバランスをとったハイパーメディア情報提示手法（Focus+Context View）や目的に沿った検索・ナビゲーション機能（スキーマブラウザ）を提案する。一般に、限られた表示スペースや紙面においては、大規模大量情報の全体・大局的表示と個別・詳細的表示とは、トレードオフの関係にある。本論文では、魚眼図表示のような考え方を情報空間に応用し、理解しやすい情報提示手法を提案すると共に、データベースモデルを援用した情報検索・ナビゲーション手法について説明する。この考え方は、直接的に、情報空間の迷子問題の解決を目指すものであり、本内容は、ハイパーメディアの情報提示・ナビゲーションとして4章に詳細記述する。

### (4) ハイパーメディア情報の自律的運用を基本とした「フィードバック機構のある」仕組みによる高品質化（書き手への効用を目的とした読み手に適用する技術）

最後に、ハイパーメディア情報を、自立的、発展的に運用させていくためには、どのような機能、システム構成が良いかについて議論する。本論文では、情報利用から得られた情報を新たな編集情報として活用し、ハイパーメディアの編集・利用効率を高める機構(Augmented Hypermedia)を提案する。特に、アドホック的、個別的な関連づけの対応だけではなく、体系的かつスケーラブルな処理が行えることを考慮する。このようなプロセスを循環させるという考え方により、ハイパーメディアの設計問題や情報のトラスト問題の解決に寄与することを目指す。本内容は、Augmented Hypermediaとして5章に詳細記述する。

## 第1章 はじめに

以上述べたように、本論文では、関連づけ構造の編集効率と利用効率の双方の向上をはかるという意味での高品質性（高品質ハイパーメディア）を実現することに焦点をあてる。関連づけは、ハイパーメディアの本質的特徴である。その品質管理を行うことが、1次情報、並びに2次情報としてのハイパーメディア情報やハイパーメディアで構成される知識空間把握を促進させる。また、ひいては、編集者自身、利用者自身の情報活用目的・モデルを明確化させることにも役立つ。課題解決に際しては、次節で述べるようにデータベース分野とCHI（コンピュータ・ヒューマン・インタラクション）分野とからのアプローチを援用し、分析を掘り下げる。このようなアプローチにより、誰もが必要な情報を有効に活用できる環境をめざすことを目標とする。

### 1. 4 アプローチ

本論文では、データベース分野と CHI(Computer-Human Interaction)分野との統合アプローチにより、課題解決に向けての研究開発を規定する。

#### 1. 4. 1 データベース分野からのアプローチ

上述したように Web やハイパーメディアは、ノードとリンクという簡便なデータモデルを基本としているため、非同期の並行編集など、人々が情報作成に簡単に参加し、利用できる特性をもっている。しかしながら、データベース分野における設計指針・モデルのような明確な考え方が提案当初からあったわけではない。ハイパーメディアの情報活用としての良し悪しは、編集者に委ねられていた。本稿では、大規模ハイパーメディアを構築する際の設計指針について、データベース分野における設計技法と対比しつつ、本論文での提案アプローチについて言及する。

まず、Web・ハイパーメディア設計への適用に対し、データベース技術分野における特長を生かせる設計方法論としては、データ独立の考え方にに基づく設計方法論（概念設計・論理設計・物理設計）と、「もの」を抽象化したスキーマ（クラス）の考え方にに基づくオブジェクトデータ設計方法論があげられる。データ独立の考え方を反映することにより、蓄積・格納する情報のデータ構造に影響されずに、アプリケーションの設計や、データを操作する言語（SQL）などが規定されてきた。また、実世界の複雑な「もの」をコンピュータ内部で効率的に表現する際に、オブジェクト設計方法論が有効であり、Web コンテンツ管理や GUI 設計にも用いられてきた。このようなデータベース設計方法論は、ハイパーメディアシステムをデータベース管理の応用システムとして捉えることにより、ハイパーメディア設計手法として活用することができる。

しかしながら、データベース設計をそのままハイパーメディア設計に導入しても、全てが機能する訳ではない。場合によっては、ナビゲーションな情報アクセス手法に制約をきたし、Web やハイパーメディアの特長を失わせてしまう。また、従来のデータベース設計は、基本的に対象物に着目した設計方法論であるため、ハイパーメディアの本質的特長である関係（リンク）に対する設計の見通しをたてることができない。たとえば、リレーショナルモデルに基づく関係データベースにおいては、関係を表現する関係表（対応する実体表のキー属性値の組み

## 第1章 はじめに

合わせ)により実現可能であるが、関係表そのものに対する設計方法論がなく、大規模な情報洪水を制御・活用するという意味での寄与はない。

本論文のアプローチは、データベース分野の技法を援用しつつ、Web やハイパーメディアの大規模な情報洪水を制御・活用する目的での研究課題の規定とその解決を目指すものである。特に、情報空間の迷子問題、ハイパーメディアの設計問題、情報のトラスト問題に焦点をあて、以下に具体的な研究の視点・対象項目をまとめる。

### (1) 関係の抽象化に基づく設計

Web やハイパーメディアでは、多数のコンテンツに対するナビゲーションのためのリンクが付与されているが、このリンク構造の取り扱いの可否が、課題解決の糸口となる。たとえば、枝葉末節なリンク情報は省略し、基幹となる関係情報を浮き彫りにしたり、また、いくつかのリンク情報を束ねてグループ化し、全体構造におけるマクロ的、仮想的な流れ(関係の遷移)を把握したりすることである。このような Web やハイパーメディアの全体概略構造(粗構造)の抽出を、本論文では、関係の抽象化と捉えて、設計方法論やツールへの適用を目指す。

### (2) 間接的關係表現と直接的關係表現との変換、操作

Web やハイパーメディアのリンクは、基本的に、明示された直接的な関係(explicit link)表現である。これに対し、関係モデルによるデータベース設計では、属性間に基づく動的かつ間接的な関係(implicit link)表現に価値がある。大規模なリンク構造を制御するためには、必要に応じ、実行時に関係を規定する動的かつ間接的な関係表現を活用するなどして、見通しよく設計・整理することが有用である。実現のためには、間接的關係表現と直接的關係表現との変換操作などが研究課題である。

### (3) ハイパーメディア情報操作・データ構造(ナビゲーション等)の規定

データベースの操作言語は、一階述語論理に基づく操作であり、この範囲では、完備な操作である。従って、ハイパーメディアにおけるナビゲーションな操作も実現可能である。しかしながら、ナビゲーションな情報操作を、そのつど、データベースの操作言語で規定することは現実的な対応ではない。また、同様に、情報検索・分析・提示等の各フェーズにおいて、使用頻度の高い操作、データ構造の規定等は、操作性向上のために必要である。このようなアウトターDB としてのハイパーメディアデータモデルも研究課題である。

## 第1章 はじめに

### 1. 4. 2 CHI 分野からのアプローチ

また、情報空間の迷子問題、ハイパーメディアの設計問題や、情報のトラスト問題に対する課題解決に向けては、人間を系に含む情報処理モデルからのアプローチも重要である。ハイパーメディアのナビゲーションな情報操作は、利用者との対話操作を基本としているため、CHI (Computer-Human Interaction) 分野における課題として対処する必要がある。

具体的に、情報検索品質の劣化、情報信頼性の低下の問題について考えてみる。CHI 的な視点で情報処理プロセスを概観すると、以下のような暗黙の前提を想定していることがわかる。しかし、実際にはこのような前提のどこかが成り立たなくなった際に、問題が顕在化されるといえる。

- a. 編集者・利用者の情報活用目的、及びモデルが明確である。  
(人間の判断・思考モデルの問題)
- b. 情報コンテンツをアクセスする手がかり情報(2次情報)が明確である。  
(人間と2次情報との親和性・想起容易性の問題)
- c. 手がかり情報に矛盾がない。(2次情報の完備性の問題)
- d. 手がかり情報が情報コンテンツを適切に表現している。  
(2次情報と1次情報との親和性・規定容易性の問題)
- e. 情報コンテンツに矛盾がない。(1次情報の完備性の問題)
- f. 編集者・利用者が情報コンテンツ全体の内容に対して明確に理解している。  
(人間と1次情報との親和性・理解容易性の問題)

これら a~f についての関係を、図1-2に示す。Web 情報が現在に至るまで利用が促進されてきた領域は、定型情報プロセスの効率化における情報コンテンツ利用領域(たとえば、B-to-B における定型業務プロセス、B-to-C における電子購買等)、もしくは、非定型情報プロセスであるが前述の情報アクセス・利用格差が問題にならないようなコンテンツ活用領域(たとえば、趣味のコミュニティサイト)等である。これらの領域は、人間にまつわる部分の前提 a, b, f が比較的成り立ちやすいものであるか、もしくは、仮に前提が成り立たなくてもあまり問題とならない領域である。

情報空間の迷子問題、ハイパーメディアの設計問題や、情報のトラスト問題が顕在化している背景は、単に情報利用格差のある情報利用者が増大するという一面だけではなく、非定型、動的、かつ、大規模複雑化した情報コンテンツの特性にも

## 第1章 はじめに

影響を受けている。また、このようなダイナミズムが要請される大規模複雑化した Web 情報コンテンツの比率が一層高まると、情報空間の迷子問題、ハイパーメディアの設計問題や、情報のトラスト問題は、一層本質的な課題となる。このような状況では、特に、上記前提 a, b, f を満足させるような、人間を系に含む Web 情報処理プロセスからの情報アクセス処理が一層必要となってくる。以下、4つの課題解決の実現を目指すことから、アプローチの具体化を進める。

### (1) 1次情報としての情報・知識空間把握の問題（上記 f に対応）

編集者・利用者によるハイパーメディア 1次情報の理解・認識促進をめざし、これらの全体情報や関連はどのようにになっているのかを明確化させる。

### (2) 2次情報としての適切な手がかり情報規定の問題（上記 b に対応）

編集者・利用者によるハイパーメディアの 2次情報の理解や認識促進をめざし、これらの全体情報や関連はどのようにになっているのかを明確化させる。

### (3) 編集者・利用者の情報活用目的・モデル明確化の問題（上記 a に対応）

編集者・利用者が、何のためにどのようなハイパーメディア 2次情報を活用して、必要な 1次情報をアクセス・利用させようとしているのかを明確化させる。

### (4) 上記問題解決を支援、及び促進するためのツールの整備と評価

対象となる情報編集者・情報利用者に対し、より具体的な解決支援・促進のためのツールの提案と実現を行う。また、問題解決支援に対する客観的な評価基準・方法を導入する。



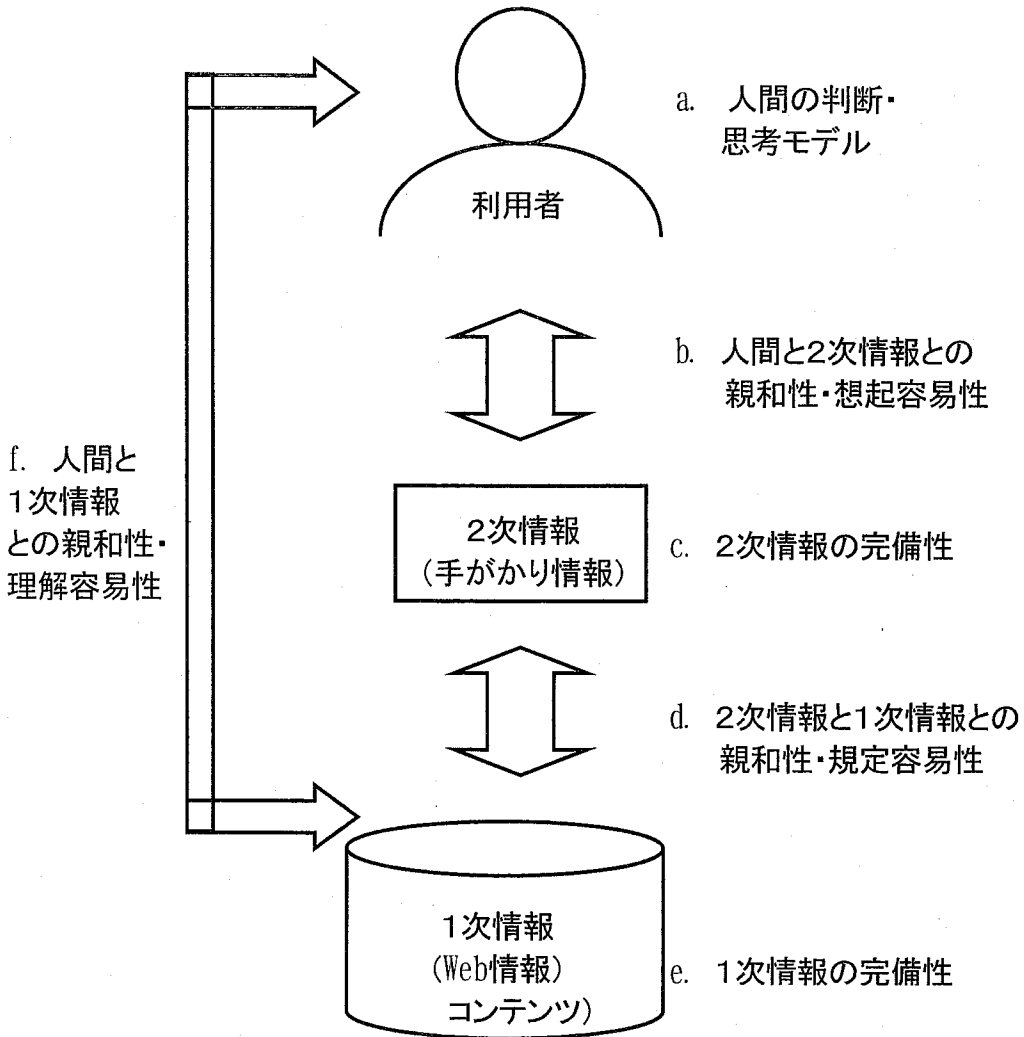


図1-2 人間を系に含む Web 情報アクセスモデル

## 1. 5 本論文におけるハイパーメディア研究開発の貢献と意義

本節では、1. 3に掲げた高品質ハイパーメディアの貢献・意義につき、1. 1. 2で概説した情報ナビゲーションアクセス基盤としての Web の利点や、1. 2でまとめた課題と関連づけて説明する。

### (1) 大規模関係構造分析に基づく「全体構造の概略整理・把握」(ハイパーメディアの抽象化)

ハイパーメディア高品質化のために提案した関係構造に基づく抽象化は、大規模複雑な Web やハイパーメディア情報の中で、本質的な骨組みとなる関係(粗構造)を抽出することである。具体的アイデアとしては、ハイパーメディアのリンク構造のなかから、完全2部グラフ(Complete Bipartite Graph)を発見し、マクロなリンク構造の規定(関係構造の抽象化)を行うものである。ただし、完全2部グラフのみに着目するとまとまった大局的な構造抽出が困難なので、例外を考慮したマクロなリンク構造(欠損リンクのある2部グラフ)の抽出を行う。例外のコストを考慮することにより、一元的な評価基準で、比較的まとまった大きさの大局構造を抽出し、かつ、個別リンク、例外リンクを考慮することにより、オリジナル構造の再生が可能となる。

このような関連情報の抽象化により、大規模な表層 Web に対して、見通しのよい構造把握を行う。また、全体のハイパーメディア構造を分析することにより、編集者間や、編集者と利用者、利用者間での対象情報コンテンツに対する共有理解を促進させ、設計・編集の効率化、検索・利用の効率化、高精度化をめざす。抽出された構造は、表層 Web やハイパーメディアの物理構造設計への適用により、効率的なハイパーメディア設計に寄与する。分析モデルによれば、約25%程度の削減が期待できる。また、全体構造と局所詳細構造とを同時に表示させるオーバビューダイアグラムへの適用により、情報空間の迷子問題解決に寄与する。

また、本論文で提案した関係のクラスタリング方式は、リンク構造のマクロ的パターンを抽出するので、このような情報を活用したコンテンツの重要度判定に適用することもできる。このような考え方は、多数の文書から参照されている文書(Authority)や、多数の文書を参照している文書(Hub)は、結果的に精度の高い重要な文書とみなされるという Hub-Authority モデルのようなグラフ構造分析の先駆けとなるものである。Google等の Web 検索エンジンもこのような Hub-Authority モデルを基本としたランキングアルゴリズムを実現している。

さらに、近年のホットなトピックスである R. Kumar らの Web コミュニティ（共通性ある Web サイト群）の発見[KRRT99]は、本論文の方式提案の 8 年後に提案されているが、基本的考えは、本論文のハイパーメディアの抽象化と同様である。すなわち、Fun と呼ばれる Web サイトの集合から Center と呼ばれる Web サイトの集合への完全 2 部グラフの抽出を基本とした、同じ興味をもつコミュニティの発見を行うものである。これは、例外のない完全 2 部グラフの効率的な逐次型抽出アルゴリズムの提案と、具体的な Web 情報空間への適用である。より大きなクラスター間の大局的な関係把握を行おうとすると、本論文で提案しているような欠損リンクを含む例外を考慮する必要がでてくると想定される。

このように、本論文で提案するハイパーメディアの抽象化、ならびに関係のクラスタリングアルゴリズムは、非同期並行編集による効率的な Web 情報作成、サーチエンジン等による大規模 Web 情報の分類・検索や、Web 情報マイニングに基づく新しい情報分析・活用へ寄与する活動と位置づけられる。

### (2) ハイパーメディアトップダウン設計による「効率的な情報活用」

(1) では、ハイパーメディア抽象化の概念を、物理的なリンク構造の分析アルゴリズムと捉えて説明したが、抽象化の意味するところは、このようなレベルに留まるものではない。本論文では、関係抽象化の概念をハイパーメディアのトップダウン的論理設計にも適用させて、関係構造に着目した抽象化概念の有効性、広汎性を示す。まず、データベース設計時に有用な E-R スキーマを援用した関係構造の抽象化設計について説明し、提案機能に対して、プロトタイプシステム開発による機能検証を行う。次に、設計プロセスへ、関係の抽象化概念を直接適用する場合について説明する。関係の抽象化概念は、大局的な関係情報を表現することであり、例外的な関係情報を区別しつつ、大規模なハイパーメディアの構築指針と情報共有化を行う。たとえば、映画ハイパーメディアを作成する場合、映画と出演俳優との関係づけの個別のリンクをはずし、まず、映画タイプと俳優タイプとの分析を行う。そして、タイプ間での関係で一次的な関連づけを行おうとするものである。

さらに、関係データベースで表現される間接的な関係表現から、ハイパーメディアの直接的な関係構造を自動生成する Hypertext Projection という方式を提案する。これは、事前の言語表現を用いたテンプレートを基にした体系的なアプローチにより、人為的に陥りやすいリンクの張り間違い、変更し忘れ、消去し忘れ等のエラーを防ぎ、意味あるリンクを生成・維持するものである。このような Web、ハイパーメディアコンテンツの信頼性向上方策は、現在では、Web Content Management という製品領域で実現、運用されてきている。

このように、本論文で提案したトップダウン的な関係の論理設計手法は、非同期並行編集による効率的な Web 情報作成へ寄与する活動と位置づけられる。また、深層 Web に対する制御・利用方策としてのハイパーメディアの一設計手法としても位置づけられる。

### (3) 全体概略構造を活用した「探しやすい・理解しやすい」情報提示と検索・ナビゲーション

「探しやすい・理解しやすい」情報提示・検索手法として、まず、Focus+Context View と呼ばれる全体構造把握のための視覚表現(Visualization)への適用を紹介する。この方式では、対象情報の構造的、意味的近傍にある情報を具体表示するだけではなく、情報空間におけるブラウジング指標として意味ある情報(ノード)とその関係も表示させる。このような大局的に意味あるノードを Landmark Node と呼ぶが、効率的な抽出アルゴリズムを提案している。重要度を判断する尺度として、リンク構造、アクセス頻度、Web サイトにおける階層の深さ情報等を組み合わせて活用している。このような視覚表現ツールは、全体構造としての大局的情報を直接表示することも重要であるが、用途によっては、情報解釈の多義性・あいまい性を減少させる表現や、目標とする情報コンテンツへ到達しやすくするガイドラインを提供するような意味付けを持たせることも重要であり、抽象化操作の新たな意味づけとなる。また、特徴量間での重み付けを変更することにより、利用者の特性によって重要度判断を適応的に変更することができる。

加えて、「探しやすい・理解しやすい」情報提示・検索手法として、ナビゲーションの多様性を提供できるブラウジング方式とその実現モデルについても言及する。これは、集合操作リンクと呼ぶ一種の動的リンクモデルを基本としている。1.4.1のデータベース分野からのアプローチのところでも述べた、データベースとの統合処理から派生した間接的なリンクに基づくハイパーメディアナビゲーションである。対話的操作により、編集者の意図と利用者の検索行為との差異を減少させることを目指す。

本論文の方式は、アウターDB としての一機能として位置づけられる。主たる目的は、ハイパーメディアのナビゲーション機能であるが、情報分析ツールとして捉えると、OLAP システムにおける多次元情報操作(スライス、ダイス等)の先駆けとなるものである。また、田中らの質問対リンクなどの一連の研究[TaQi92]は、データベース問い合わせ処理からの動的リンク実現に対する先駆的研究である。本論

## 第1章 はじめに

文における活動は、多次元分類構造（観点）でのナビゲーション操作による動的リンク実現を提案するもので、田中らの活動と相補的な先駆的研究と位置づけられる。

このように、本論文で提案した全体・概略構造に基づく情報提示・検索・ナビゲーション手法は、サーチエンジン等による大規模 Web 情報の分類・検索や、Web 情報マイニングに基づく新しい情報分析・活用へ寄与する活動と位置づけられる。

### （4）ハイパーメディア情報の自律的運用を基本とした「フィードバック機構のある」仕組み

本論文では、情報のフィードバック機構のある仕組みとして、Augmented Hypermedia を提案しているが、それは、利用から得られた情報を新たな編集情報として活用し、ハイパーメディアの編集・利用効率を高める機構である。このようなプロセスを循環させるという考え方により、ハイパーメディアの設計問題や情報のトラスト問題の解決に寄与することを目指す。アドホック的、個別的な関連づけの対応だけではなく、体系的・スケーラブルな処理が行えることを考慮し、3層のアーキテクチャの利点について、具体的プロトタイプを交えて説明する。

情報の利用者から情報の作成者へのフィードバック機構があるものとしては、近年 Blog が脚光を浴びている。これは、利用者側のコメントや引用を簡便に行いやすい双方向リンクのあるハイパーメディアとみなすことができ、情報発信や情報コミュニケーションとしての様式を変革するような仕組みとなってきた。Augmented Hypermedia は、このような Blog へとつながる活動の先行的研究と位置づけられる。

このように、本論文で提案した自律的運用を基本とした「フィードバック機構のある」ハイパーメディアは、Web 利用を介した業務プロセス・生活様式の改善へ寄与する活動と位置づけられる。

### 1. 6 本論文の構成

本章では、Web、ハイパーテキスト／ハイパーメディア技術領域における歴史的経緯と現状の背景・課題について概観した。大規模な情報の洪水を制御・活用することを目指して、関係構造に着目した高品質ハイパーメディアに対する要請と提案について説明を行った。また、情報空間の迷子問題解決、ハイパーメディアの設計問題、並びに、コンテンツのトラスト向上に向けて、データベースとCHIからのアプローチについての言及を行った。さらに、本論文の位置づけとして、ハイパーメディア物理設計、論理設計、構造表現の可視化・検索、自律的ハイパーメディア実現につき、今日的な状況においての貢献や意義を明確にして、概説を行った。

以下、2章では、高品質化のためのハイパーメディア関係構造の分析と把握に対する基本的指針として、関係構造（リンク）に基づく抽象化の提案と導入を行う。ここで、関係構造に基づく抽象化とは、大規模複雑なWeb、ハイパーメディア情報の中で、本質的な骨組みとなる関係をグラフ理論の援用により、抽出することである。また、ハイパーメディアのリンク情報の物理構造設計の効率化にも言及する。

3章では、高品質ハイパーメディア設計に対する指針と構築のため関係抽象化の考え方について説明する。具体的には、拡張E-Rモデルによる設計、関係抽象化の論理設計(Link Relationship)への適用、ハイパーメディアリンク構造の自動設定につき言及する。

4章では、高品質ハイパーメディアを用いた情報提示（可視化）と、検索・ナビゲーションの多様化について説明する。Focus+Context Viewsというオーバビューダイアグラム、並びに表示構成要素としてのLandmark Nodeの抽出について説明する。また、集合操作リンクによるスキーマブラウザについても説明する。

5章では、高品質ハイパーメディアを自律的に発展、運用させるための仕組みについて議論する。体系的・スケーラブルな処理が行えることを考慮した3層のアーキテクチャの利点について、具体的プロトタイプを交えて説明する。

第6章では、本論文のまとめと今後の展望に対して説明する。本論文で提案した高品質ハイパーメディアの考え方の今日的意義、情報空間の迷子問題、ハイパーメディアの設計問題、並びに、コンテンツのトラスト問題における貢献に関してまとめる。また、今後の研究展望について言及し、結言とする。

## 第2章 ハイパーメディアの抽象化と物理設計

### 2.1 ハイパーメディアの抽象化とは

本章では、大規模ハイパーメディアの高品質化をはかるため、関係構造（リンク）に基づく抽象化の提案と導入を行う。関係構造に基づく抽象化とは、大規模複雑な Web、ハイパーメディア情報の中で、関係構造のグルーピングを行うことにより、本質的な骨組みとなる関係を抽出することである。この考え方を、ハイパーメディア設計、提示、アクセス多様化等への対処に対する具体的マッピングを行い、情報空間の迷子問題や、コンテンツのトラスト問題に対する解決を目指す。

具体的には、ハイパーメディアのリンク構造の中から、完全2部グラフ(Complete Bipartite Graph)を発見し、マクロなリンク構造の規定（関係構造の抽象化）を行うものである。ただし、完全2部グラフのみに着目するとまとまった大局的な構造抽出が困難なので、例外を考慮したマクロなリンク構造（欠損リンクのある近似的完全2部グラフ）の抽出を行う。例外のコストを考慮することにより、一元的な評価基準で、比較的まとまった大きさの大局構造を抽出し、かつ、個別リンク、例外リンクを考慮することにより、オリジナル構造を再生することが可能である。

このような関連情報の抽象化の目的は、全体のハイパーメディア構造を分析することにより、編集者間や、編集者と利用者、利用者間での対象情報コンテンツに対する共有理解を促進させ、物理構造設計の効率化、検索の高精度化をめざす。また、直接的には、オーバビューダイアグラムの実現手段として、情報空間の迷子問題解決に寄与する。

大規模複雑なシステムやコンテンツを見通しよくする手段としては、対象の分類やクラスタリングが役に立つ。従来の分類やクラスタリングに基づく対象は、情報コンテンツそのもの（ノード）である。従って、ハイパーメディアに従来手法で適用することを考えた場合、グラフ構造のグラフ分割問題とみなすことができる。すなわち、グループ間での関係構造を最小限に留めて、かつ、同一グループ内の関

## 第2章 ハイパーメディアの抽象化と物理設計

係構造を最大限に引き出してマクロ構造を抽出するものである。このような従来手法は、類似の情報コンテンツ自体をグルーピングするという意味では、見通しのよい概略構造が抽出されるものの、マクロな関係構造が消失される。本論文では、マクロな関係構造自体を抽出する目的として、提案手法としての関係構造に着目したクラスタリング手法を提案する。

近年のホットなトピックスである R. Kumar らの Web コミュニティ（共通性ある Web サイト群の発見）[KRRT99]は、本論文の方式提案の 8 年後に提案されているが、基本的考えは同様である。すなわち、Fun と呼ばれる Web サイトの集合から Center と呼ばれる Web サイトの集合への完全 2 部グラフの抽出を基本とした、同じ興味をもつコミュニティの発見を行うものである。これは、例外のない完全 2 部グラフの効率的な逐次型抽出アルゴリズム(Trawling)の提案と、具体的な Web 情報空間への適用である。より大きなクラスター間の大局的な関係把握を行おうとすると、本論文で提案しているような欠損リンクを含む例外を考慮する必要がでてくると想定される。



### 2. 2 関連研究

ハイパーメディアの抽象化手法は、ハイパーテキストなどグラフ構造と対象としたクラスタリングの一手法である。従来は、グラフ分割問題[GaJo79]のように、クラスター内の関係（リンク）を最大化し、クラスター間の関係（リンク）を最小化する基準のもとに解析されていた。これは、回路設計や種々の最適化問題で、大規模問題を分割化して解析することを目的としていた。しかし、マクロな関係構造を極力少なくする目的なので、本論文で対象とする関係構造の把握とは異なる。

本論文で目的とする全体の関係構造の把握手法としては、大別して2つのアプローチがある。1つは、仮想的なクラスタノードを抽出せず、対象グラフの中から本質的に重要な少数のノードを規定する方法である。初期の代表例は、Botafofo らのクラスタリング分析である[BoSh91, BoRS92, Bota93]。これは、大規模グラフの中で、パス（任意のノード間の最短経路）に着目した場合に、要所となるノードを見出す。大規模ハイパーメディア構造の視覚化で重要な要素である Landmark Node は、この方式の延長上の概念である。この方式では、特定のノード（あるいは、Web ページ）のみで他の情報を代表できる場合はよいクラスタリングが行える。しかし、ノードの集合として意味がある場合には、このような情報がクラスタリングにより欠落してしまう可能性がある。

これに対する2番目のアプローチとしては、仮想的なクラスタノード（ノードのアグリゲーション）を何らかの基準で抽出する方法である。本論文で提案する ACE クラスタリング手法は、このアプローチに該当する。従来のグラフ分割問題も、広義にはこのアプローチに該当するが、本手法は、ノードのグルーピングに評価基準をおくのではなく、例外を考慮したリンクのグルーピングに評価基準をおいている。計算量は、前者のアプローチに比して一般的に多いが、グラフ構造全体に対してより深い解析を行える。

なお、本論文で説明するハイパーメディアクラスタリング方式や、R. Kumar らの提案する Trawling による Web コミュニティ発見は、直感的には、Web 情報間の関係傾向をはかろうとする立場にある。これに対して、Flake らの提案する Web コミュニティ[FILG00]は、コミュニティ内の Web 情報とのリンク関係が、コミュニティ外の Web 上情報とのリンク関係よりも密になる場合と定義されている。Flake らは、このように定義された Web コミュニティを最大流最小カット定理のアルゴリズムを用いて抽出している。この場合は、グラフの分割問題に相当する問題に帰着される。コミュニティ間のマクロな関係抽出を直接目的としている訳ではないが、たと

## 第2章 ハイパーメディアの抽象化と物理設計

えば、倫理上問題とされるコミュニティサイトに対する効率的発見とその情報フィルタリング等に有効な方式である。

R. Kumar [KRRT99]や Flake[FILG00]らが提唱した Web コミュニティは、このような大規模グラフに対する構造の偏在性、相関性を考慮して、Web 情報のビヘイビアを分析している。上記のようなカテゴリーを勘案すれば、Flake らのアプローチ（最大流最小カット定理の適用）は、前者のアプローチ（集約リンクのないケース）に該当し、R. Kumar らのアプローチ（完全2部グラフの抽出）は、後者のアプローチ（集約リンクのあるケース）に該当する。ただし、例外を考慮していないため、抽出クラスターのサイズは小さい。Web コミュニティは、リンク構造に基づく情報フィルタリングとしても、広く活用されている。

このように、現在の Web コミュニティの研究開発活動は活発であり、利用目的により、コミュニティの定義、抽出アルゴリズムを適合させながら進展している。これら関連する研究活動は、Web をグラフ構造とみなし、グラフ理論に基づくアルゴリズムを援用して、効率的な解法を見出している。この点で、本論文で説明するハイパーメディアクラスタリングは、ハイパーメディア初期の時代に問題提起、方式提案した先駆的な研究と位置づけることができる。

### 2. 3 大規模ハイパーメディアの問題点

ハイパーメディア構造の利点は、ユーザにナビゲーション情報アクセスを提供することにある。ナビゲーション情報アクセスとは、何らかの指針もしくはガイダンスをたよりに、次々に情報検索・提示が行えることである。構造が、比較的小規模であったり、疎なグラフ構造であったりする場合には、有用であることが示されている。しかしながら、大規模な構造になるとナビゲーション自体が複雑となり、ハイパーテキストの特徴が失われてくる。以下では、このような大規模な構造になった場合の問題点と、それに対する要求事項を整理する。たとえば、Web 情報をはじめとし、電子百科事典[HaKa88]、医学マニュアル[FrCo89]、協同型文書作成環境[CoBe88]等の大規模ハイパーメディアを扱う際に問題点が顕在化する。

#### (1) 情報空間中の迷子 (Disorientation) の問題

利用者が、情報空間の中で自分が今どこにいるのかわからない。また、次にどこへ進めば良いのかわからない。これは、大規模ハイパーテキスト/ハイパーメディアを取り扱う場合の共通の不満である[Conk87]。この問題は、大規模かつ複雑な構造の空間把握が困難なことに起因する。

この問題は、複数利用者が使用する共有ハイパーテキストにもあてはまる。すなわち、各利用者が各々のビュー (見方) をもっていることや、作成者と利用者においてもビューが異なるためである。標準化による解決方策も議論・提案されている[MoBB90]。しかしながら、柔軟かつ動的なビューの変更には、まだ不十分である。

#### (2) 関連に対する正則化技法の欠如

関連 (ハイパーメディアでは、リンク付け) に対する正則化が、従来検討されてきていない。ここで、正則化とは、概略化、詳細化、階層化などの抽象化により、構造を再編成して、情報利用効率を改善させることにある。

ハイパーメディアには、大別して2種類の関連がある。一方は、構成リンク (organizational Link) であり、もう一方は、参照リンク (associative Link) である。前者は、ドキュメントとそのサブドキュメント間の内的関係を表現しており、階層構造を形成する。後者は、複数ドキュメント間の外的関係を表現しており、ネットワークもしくはループ構造を形成する場合もある。構成リンクに対しては、“is-a” や “is-part-of” のリンクタイプを用いることができる。これは、一種の抽象化

## 第2章 ハイパーメディアの抽象化と物理設計

操作とみなすことができる。属性の継承 (inheritance) により、システム設計者は、コンパクトにデータ構造を設計することが可能である。しかしながら、ハイパーメディア構造の本質的部分である参照リンクの正則化に対しては、まだ適切な対処法が見いだされていない。対処法を見いだすために、まず、理論的かつ分析的検討が必要である。

### (3) 作成のコストと利用の便益とのトレードオフ

もし、システム設計者が、単純かつ疎な関連づけ方策を採用すると、利用者は、非常に簡単なブラウジングしか用いることができない。一方、非常に多様な情報関連づけ並びに検索機能を提供すると、システムの作成／更新コストが増大する。このトレードオフは、いかなるサイズのシステムにおいても存在するが、特に大規模になると、一層顕著である。ハイパーテキスト構造をデータベース諸機能と関連づけることは、このトレードオフを改善する1つの方策である。データに対する柔軟なスキーマの提供が、データ作成／利用の両面において、効果を発揮することができるためである。

## 2. 4 効果的かつ効率的なナビゲーションに対する要求事項

以下が、大規模情報空間におけるナビゲーションを改善するための要求事項である。

### (1) 集合操作表現への変換

密なハイパーテキスト構造を対象とする場合には、関係代数のような集合操作モデリングでの操作が必要である。すなわち、関連の相手先が特定の一つのノードだけでなく、ノード集合も対象とすることである。仮想リンク[RoMN81]や、集合操作リンク[HaKa90]が、集合操作モデリングに基づいている。これらは、一種の問い合わせ言語インタフェースとみなすことができ、対象となる集合ノードは、関係DBモデルにおけるビューに相当する。

しかしながら、静的なハイパーテキスト構造から、集合操作モデルへの変換はまだ行なわれていない。もしこのような変換が実現されると、データベース分野における各種技法が、大規模ハイパーテキストシステムに対しても適用できることになる。

### (2) 高レベルのセマンティック構造の提供

データベース分野における関係モデリングやセマンティックモデリングと比較すると、ハイパーテキストモデルにおける意味的制約は不十分である。その結果、いくつかの冗長な表現が存在することになる。一般のハイパーテキストシステムでは、通常以下の2つの意味的構造しか用いられていない。1つは、構成リンクに対する“is-a”、“is-part-of”関係の利用である。もう1つは、参照リンクに属性を付与することによる意味的構造の追加である。Web[UtYa89]は、参照リンクについた属性により、部分グラフを規定し、ブラウザとして用いている。

それゆえ、データベースの分野で用いられているような属性間の関数従属や、関係間の基数制約(1対1、1対多)などの間連の制約手法をハイパーテキストに適用する余地がある。実際、実システムにおいてノードとリンク間には相関があり、冗長性を低減できる可能性が高い。

### (3) 多対多関連に対する効率的なインデクシング技法

Single multiple-attribute index [ClGa90] が、階層構造形の効率的なインデックス方式として提案されている。しかしながら、ネットワーク構造(親が2つ以上の階

## 第2章 ハイパーメディアの抽象化と物理設計

層構造) や、ループを含む構造に対するインデクシングとしては、まだ不十分であり、一般のグラフ構造に対する効率的なインデクシング技法が必要である。

また、稠密なグラフ構造に対するコンパクトなグラフ構造の表現も要請されている。閉移推包 (transitive closure) に対する圧縮表現はいくつか提案されている [AgBJ89, CrNo89]。しかしながら、閉移推包でない一般の稠密なグラフに対して、コンパクトに表現する方式はまだ提案されていない。

## 2. 5 ハイパーテキストアグリゲーション (Hypertext Aggregation)

本節では、2. 4節で述べた要求事項に基づいて、ハイパーテキスト構造の関連の抽象化・集約化方式（ハイパーテキストアグリゲーション：例外を考慮した関連の概略構造抽出）について提案する。まず、具体的な例を説明した後、集約と例外の双方を考慮した物理構造設計や、関連の概略化モデルを説明する [HaKG91-1]。

### 2. 5. 1 ハイパーテキストアグリゲーションの例

#### 例1：学生の受講関係

最初に、学生と学生が受講する履修科目という簡単な例を考える。明らかに、この関係は、多対多の関連である。図2-1(a)にあるように、関連の数（リンク数）は、学生数と履修科目数の積のオーダーである。学生がたくさん受講すればするほど、グラフ構造が密になる。

しかしながら、実際のシステムにおいては、リンク間に相関があるため、適切なクラスタリングを用いることにより、単純なリンク構造を得ることができる。図2-1(b)は、(a)に対するクラスタリングの一例である。大部分の計算機科学科の授業は、計算機科学科の学生によって受講され、同様に大部分の数学科の授業は、数学科の学生によって受講される。もちろん、例外が存在する。学生は、他学科の授業を履修してもよいし、また、全ての授業が同一学科内の学生によって受講されるわけではない。一般に、このような例外が相対的に少ない場合には、データベースの物理構造設計や、構造の視覚的表示に有効であると想定される。

#### 例2：参照マニュアル

次に、参照マニュアルを考える。図2-2に示したように、各項目は他の項目を参照する。ある種の参照は、階層構造をなし、またある種のもの、相互参照している。適切に集約を行なうことにより、構造は単純化される。これは、関連項目集合間は、同様の参照項目集合を形成しがちであるためである。大局的な参照構造は、図2-2(b)に示すように強調させることができる。

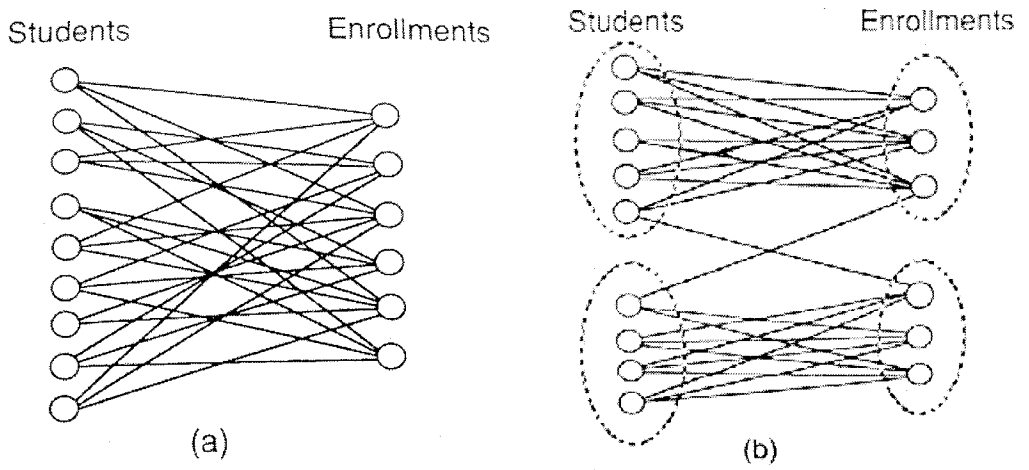


図2-1 学生の受講関係の例

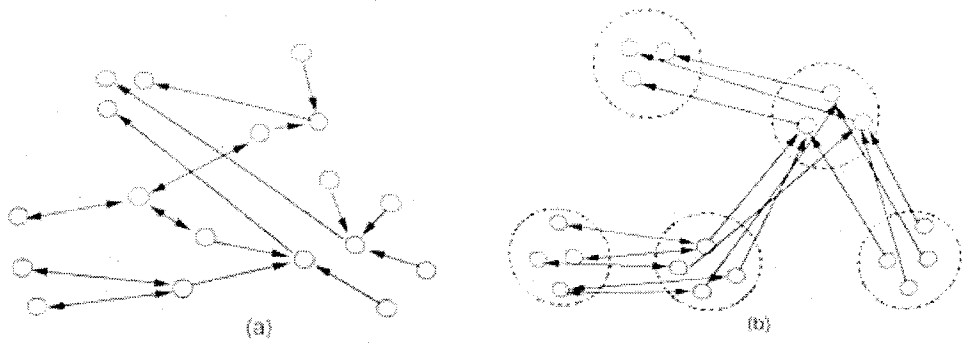


図2-2 参照マニュアルの例



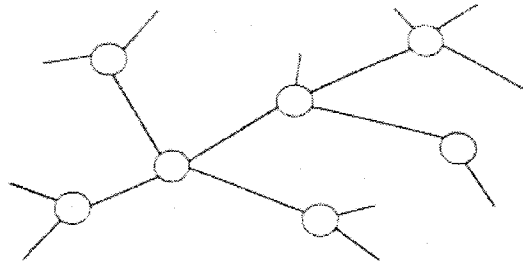
## 2. 5. 2 ACE：例外を考慮した集約化クラスタリング

ここで用いる基本となる考え方は、図2-3に示したような例外を考慮することにより、効果的な集約化クラスタリングを提供することである。ブロック対角化手法をはじめとした同様のクラスタリング手法に比較すると、本手法の特徴は、所与のハイパーテキスト構造から、集約化関係を重視して抽出することにある。一方、従来型クラスタリング手法は、カット数（クラスタ間にもたがるリンク数）を最小にすることが目的である。このため、得られた大局的構造のマクロな関係があいまいになりやすい。以下、提案するクラスタリング手法を **ACE (Aggregation Clustering with Exceptions)** と呼ぶ。

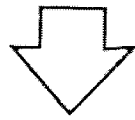
ここでは、集約と例外とを対等に扱うことにより、より自然で、情報理論的に意味のある高レベル構造を抽出することを試みる[Yama89, 90]。すなわち、クラスタに関する情報と、後に残った例外に関する情報との双方をバランスよく考慮する。このようなグラフデータ構造の圧縮は、効率的なデータベース物理構造の表現手法として有効である。また、利用者側においても、ヒューマンインタフェースに関する要求事項を満たし、理解しやすい視覚表現を提供できる可能性がある。

図2-4に、ACEクラスタリングを表わすE-Rダイアグラムを示す。図2-4に示す様に、2種類のACEクラスタリングがある。例1のような場合は、対応ノード集合が明白に2分されているケースであり、E-Rモデルにおける多対多関連の集約化に相当する。また、例2のような場合は、参照元/参照先が同一のノード集合の集約化に相当する。これは、E-Rモデルにおける、自己関連の集約化に相当する。

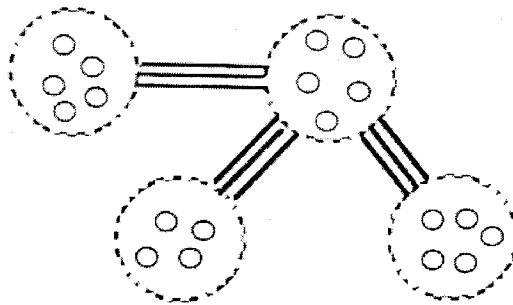
入力グラフ  
 $G = (V, E)$



Input Graph Structure

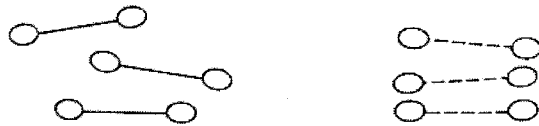


集約グラフ  
 $G_A = (V_A, E_A)$



Global Structure

+



Exceptions

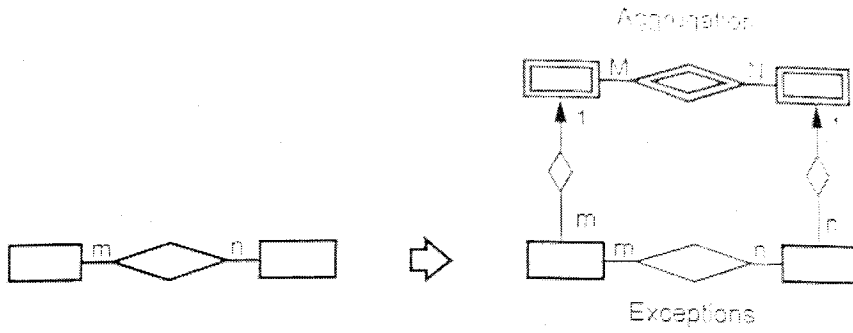
正の例外グラフ

$G_I$

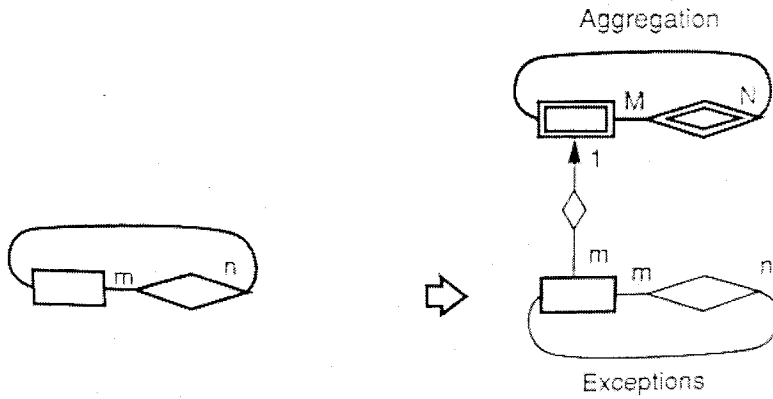
負の例外グラフ

$G_E$

図2-3 ACE(Aggregation Clustering with Exceptions) クラスタリング



(a) Many-to-many relationship aggregation



(b) Self relationship aggregation

図2-4 ACEクラスタリングのE-Rダイアグラム

### 2. 5. 3 ACEクラスタリング手法の定式化

本節では、ACEクラスタリングの定式化と、その分析モデルにつき説明を行なう。以下では、ハイパーテキスト構造のうち参照リンクのみを対象とし、これを一般に有向グラフとみなして議論を進める。実際的なACEアルゴリズムは、2.5.8で説明を行なう。

#### (1) 問題の定式化

ACEクラスタリングに対する入力  $G = (V, E)$  は、ノード  $V$ 、リンク  $E$  で表わされる有向グラフである。対象とする出力は、集約グラフ  $G_A = (V_A, E_A)$  と、例外グラフ  $G_X$  であり、両者の情報を併せて入力グラフの情報となる。

集約グラフ  $G_A$  は、ハイパーノード  $V_A$  と、ハイパーリンク  $E_A$  とを有するハイパーグラフである。各集合  $V_A$  は、入力ノード  $V$  の部分集合に対応する。しかも、 $V$  の各要素は、 $V_A$  の唯一の要素に対応する多対1の関係である。また、各集合  $E_A$  は、 $V_A$  の各要素間の関連を表わすリンク集合である。すなわち、 $E_A \subseteq V_A \times V_A$  である。

例外グラフ  $G_X$  は、入力グラフ  $G$  と、集約グラフ  $G_A$  との補完を表わすグラフである。例外グラフには、2種類のタイプがある。1つは、正の例外リンク (Inclusive Link) であり、ハイパーリンクが存在しないところに存在する小数の入力関連情報である。正の例外リンクで構成されるグラフ  $G_I = (V, E_I)$  を正の例外グラフ (Inclusive Graph) と呼ぶ。またもう1つは、負の例外リンク (Exclusive Link) であり、ハイパーリンクは存在するが、実際の入力関連情報としては存在しないリンクを意味する。負の例外リンクで構成されるグラフ  $G_E = (V, E_E)$  を負の例外グラフ (Exclusive Graph) と呼ぶ。

ここで、集約グラフ中のハイパーリンクが生成されるか否かは、対応するハイパーノード内の個々の関連情報の多少に依存する。一定の閾値以上、たとえば過半数、であれば生成されることとする。このような集約グラフと例外グラフの抽出方針は、いかにグラフを単純化するかに基づく。すなわち、抽出のための評価基準は、たとえば、グラフのノード数とリンク数の総和、あるいは、リンク数のみの総和として設定することができる。入力グラフをこのようなグラフ構造に分解することは、データ圧縮だけでなく、グラフの概略構造を把握する上で重要である。

より一般的には、集約スキーマ構造の抽出に関する問題定義は、以下のよう  
に表現することができる。

## 第2章 ハイパーメディアの抽象化と物理設計

入力： 入力グラフ  $G=(V, E)$

出力： 集約グラフ  $G_A=(V_A, E_A)$ ,

正の例外グラフ  $G_I=(V, E_I)$ , 及び負の例外グラフ  $G_E=(V, E_E)$

入力情報と出力情報とは、下記の制約関係をみたす。

$$\begin{aligned} V &= \cup_i V_{A_i}; \forall i, j, i \neq j: V_{A_i} \cap V_{A_j} = \emptyset \\ E &= (E'_A - E_E) \cup E_I \\ E'_A &= \{(v_i, v_j) \mid v_i, v_j \in V, \exists V_{A_i}, V_{A_j} \in V_A, \\ &\quad v_i \in V_{A_i}, v_j \in V_{A_j}, (V_{A_i}, V_{A_j}) \in E_A\} \end{aligned}$$

図2-5に示しているように、 $E'_A$ は、集約グラフ  $G_A$  で構成される完全2部グラフ (Complete Bipartite Graph) 中のリンク集合である。

ここでは、集約の過程において、下記のコスト関数を最小化することが最適であるとみなす。

$$\text{Cost\_function}(E_A, E_I, E_E) = |E_A| + |E_I| + |E_E| \rightarrow \text{Min.}$$

図2-6(a)に例を示す。入力データは、一般に隣接行列として表される。行列のサイズは、入力ノード数に等しい。また、各要素において、 $x_{ij} = 1$ は、ノード  $i$  から、ノード  $j$  へリンクがあることを意味し、逆に、 $0$ は、リンクがないことを意味する。

コスト関数を最小化するという条件のもとで、適切なクラスタリングを行なうことにより、たとえば、図2-6(b)のような行列を得ることができる。ここで、注意すべき点は、 $0^\#$ と $1^*$ とが、各々、正の例外リンク、負の例外リンクを表している点である。これらは、各クラスターにおける小数の例外関係を示している。

次に、ACEクラスタリングの意味を、関係代数の観点から規定する。下記の条件が、関係代数を用いることにより、入出力データ間を満足する条件である。

$$\begin{aligned} (R - R_I) \cup R_E &= \prod_{R_{M1}.Node, R_{M2}.Node} (R_{M1} \mid X \mid R_{M1}.Anode = R_A.Snode \ R_A \\ &\quad \mid X \mid R_A.Enode = R_{M2}.Anode \ R_{M2}) \end{aligned}$$

ただし、 $|R_A| + |R_I| + |R_E| \rightarrow \text{Min}$ .

ここで、

$R$ ; 入力リレーション

$R_I$ ; 正の例外リレーション

$R_E$ ; 負の例外リレーション

$R_M \equiv R_{M1} = R_{M2}$ ; メンバーシップリレーション

$R_A$ ; 集約リレーション

図2-7 に示すように、ACE クラスタリングは、差と和の代数操作を用いて入力リレーションを整形した後、要素の集約を行なうクラスタリング手法である。メンバーシップリレーションは、入力グラフのノード属性から構成されているので、ACE クラスタリングは、個々の関連の情報を、ノード属性値に変換する方法とみなすことも可能である。

また、ACE クラスタリングを、グラフ分割問題と対比すると、後者は、隣接行列のブロック対角化する指針（グループ間のリンクを最小化）であるのに対し、前者（ACE クラスタリング）は、非ブロック対角成分も含めて、全体的の最適な構造を抽出しようとするものである。

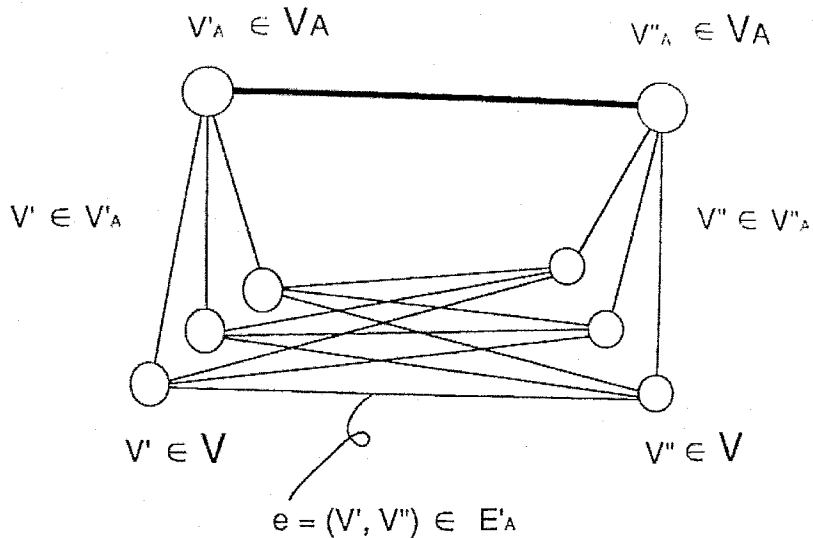


図2-5 完全2部グラフ(Complete Bipartite Graph)における頂点集合

第2章 ハイパーメディアの抽象化と物理設計

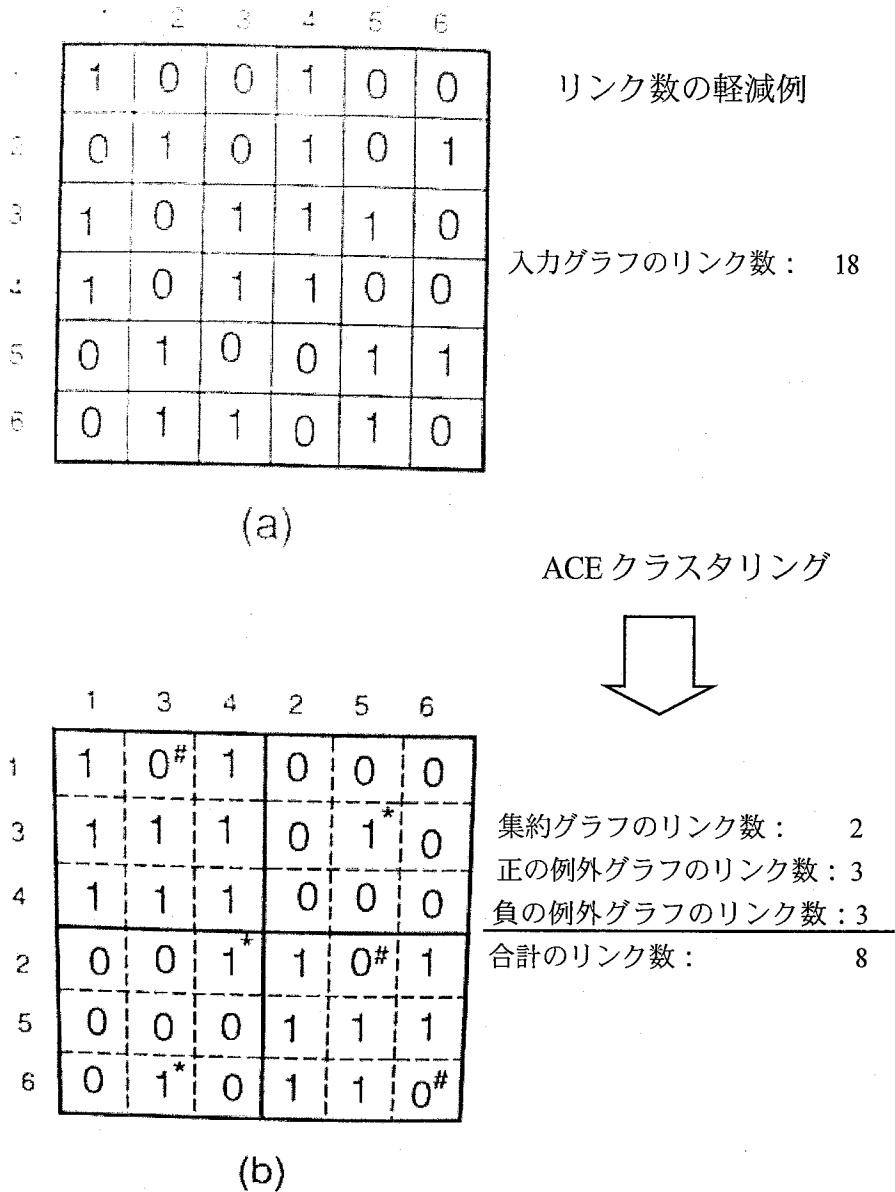


図2-6 ACEクラスタリング例

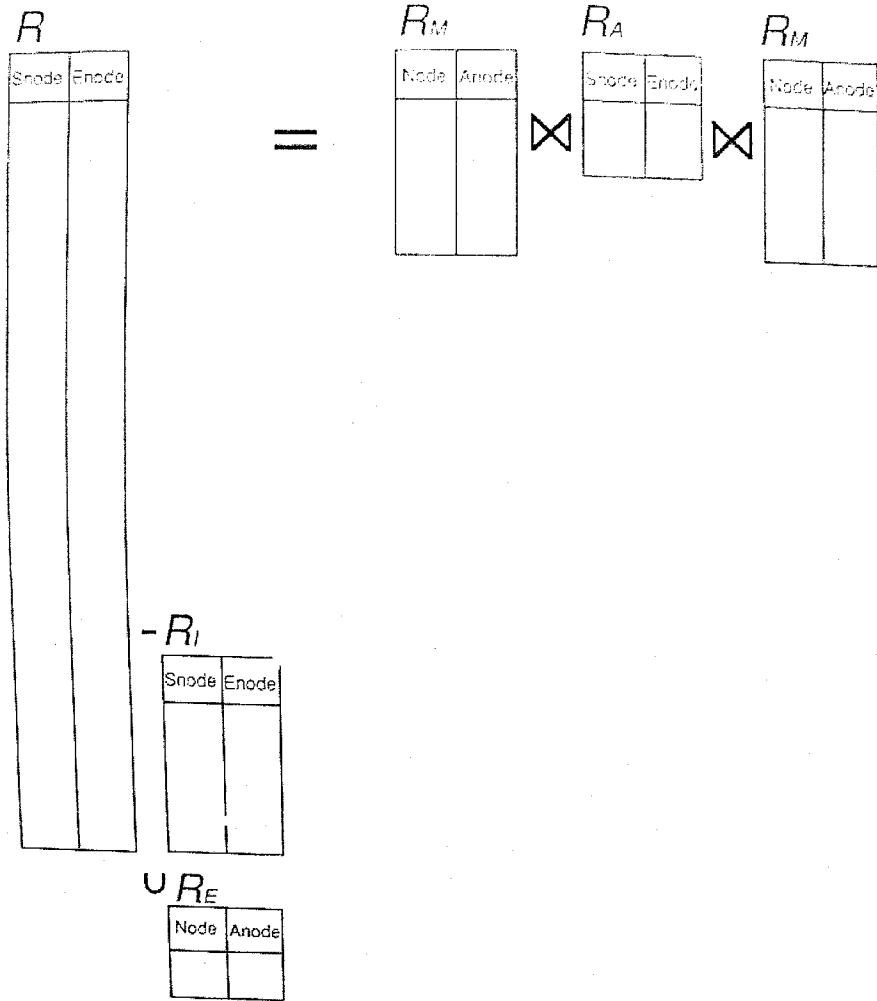


図2-7 式の図的表現



#### 2. 5. 4 ACE クラスタリングの計算量

ここでは、ACE クラスタリングの問題が、NP-困難であることの略証を示す。それゆえ、問題解決のため、何らかの探索的アルゴリズムが不可欠となる。

定理 1 : ACE クラスタリングの最適解を求める問題は、NP-困難である。

証明 : 問題  $P_i$  が、NP-困難であることを証明するためには、既知の NP-完全問題  $P'_i$  から  $P_i$  への多項式オーダーの変換式が存在することを示せば十分である [GaJo 79]。ここでは、まず、既知の NP-完全問題として、グラフ分解問題 [GaJS76] をとりあげる。そして、次に、グラフ分解問題から、ACE クラスタリング問題への多項式オーダーの変換式を示す。

グラフ分解問題の設定は、以下の通りである。

入 力 : 偶数個のノード  $n$  をもつ、重み付き有向グラフ

出 力 : 等分割されたノードを有する 2 つのグラフで、両者をまたぐカット数が最小となるもの

次に、ACE クラスタリングがグラフ分解問題を特殊なケースとして含んでいることを示す。この種の証明法は、restriction と呼ばれているものであり、既に多項式オーダーの変換式であることが証明されている。ここで、グラフ分解問題を考えると、その出力結果は、以下の 2 種類に類別される。

- (1) 両クラスタグラフ自身に、自己ループとしての集約リンクが存在し、かつ、両クラスタグラフ間に集約グラフが存在しない場合
- (2) その他の場合

ケース (1) の場合、グラフ分解問題は、ACE クラスタリングの場合と同等のコスト関数を有することとなり、ACE クラスタリングの特殊ケースとなる。すなわち、本コスト関数を用いることが、負の例外領域 (大部分が 1 の領域) のリンク数を最大化することと、正の例外領域 (大部分が 0 の領域) のリンク数を最小化することに相当するためである。

ケース (2) の場合、2 種類の理由がある。一方は、クラスタグラフ内に集約リンクが存在しないことによる場合、もう一方は、両クラスタグラフ間に集約リ

## 第2章 ハイパーメディアの抽象化と物理設計

リンクが存在する場合である。また、両方の理由が重畳されているケースもある。いずれの場合にせよ、個々のサブグラフに対して、グラフ分割問題を再帰的に、すべてのグラフがケース(1)の場合になるか、もしくは、クラスタサイズが1になるまで、適用することができる。この一連の操作は、高々 $\log n$ のオーダーである。

明らかに上記のステップは、クラスター数  $p$  が固定されている場合において、ACE クラスタリングの一最適解を規定することに相当する。したがって、グラフ分解問題は、ACE クラスタリングの一特殊ケースに相当する。それゆえ、ACE クラスタリングは、NP-困難である。

### 2. 5. 5 ハイパーメディア物理設計の分析モデル

本節では、ACE クラスタリングの分析モデルにつき検討を行なう。これにより、最適解における種々のパラメータの推定を行なうことができる。問題を簡単化するため、以下の2つの仮定をおく。

- (1) 分割された各クラスター内のノード数は等しいとする。
- (2) 負の例外領域における負の例外リンクの出現確率と、正の例外領域における正の例外リンクの出現確率は等しいとする。

本モデルを定式化するため、以下のパラメータを用いる。

$n$  : 入力グラフのノード数

$p$  : 入力グラフのリンク数の比率

(すなわち、 $pn^2$  が入力グラフのリンク数を示す。)

$q$  : 分割領域中の例外リンクの平均出現確率

(仮定(2)より、負の例外領域と、正の例外領域における例外リンクの平均出現確率は等しいとする。)

$s$  : 集約グラフ  $G_A$  のノード(クラスター)数、 $E_A$

(仮定(1)より、各クラスタサイズは同一とする。)

$a$  : 集約グラフのリンク数

$\alpha$  :  $s$  と  $n$  との比率、 $\alpha = s/n$

## 第2章 ハイパーメディアの抽象化と物理設計

本クラスタリングの目的が、例外リンクの出現割合を下げることにあるので、以下のような、条件が設定できる。

i)  $0 \leq q < p \leq 1$

ii) また、入力リンク数の総和は、集約リンクを構成する個々のリンク数の総和と正の例外リンク数の総和とを加えたものであるので、以下の条件が満足される。

iii)  $pn^2 = a(1-q)\left(\frac{n}{s}\right)^2 + (s^2 - a)q\left(\frac{n}{s}\right)^2$

それゆえ、

$$a = \{(p-q)/(1-2q)\} s^2$$

集約リンクの平均出現確率  $p'$  は、以下の式で与えられる。

$$p' = (p-q)/(1-2q)$$

これより、 $p$  と  $q$  とをパラメータとして、集約リンク数を推定することができる。集約リンク数  $p'$  に対する条件は、以下のように表わされる。

i)  $p < 0.5$

$$0 \leq q < p \rightarrow p \geq p' > 0$$

ii)  $p = 0.5$

$$p' = 0.5 \text{ (for all } q)$$

iii)  $p > 0.5$

$$0 \leq q < 1-p \rightarrow p \leq p' < 1$$

たとえば、 $p = 0.2$  で、 $q$  が 0.1 に減少されたとき、集約リンクの平均出現確率  $p'$  は、12.5% である。

次に、ACE クラスタリングを用いることが、物理 DB 設計に対して有効になるための条件を考える。これは、出力レーションのタプル数が入力レーションのタプル数より小さくなる場合である。それゆえ、以下の条件を満たさなければならない。

$$|R_M| + |R_C| + |R_I| + |R_E| < |R|$$

本節で定義したパラメータを用いて書き直すと、以下の式が得られる。

## 第2章 ハイパーメディアの抽象化と物理設計

$$n + a + qn^2 < pn^2$$

また、 $\alpha$ を用いることにより、以下の式を得る。

$$\alpha < \sqrt{[(1-2q)(1-1/\{n(p-q)\})]}$$

それゆえ、効率的なクラスタサイズと適用可能解は、以下の条件に応じて推定することができる。

$$s < n \sqrt{[(1-2q)(1-1/\{n(p-q)\})]}$$

### 2. 5. 6 主記憶DBにおける最適格納コスト

主記憶 DB の物理設計では、格納効率が重要である。ここでは、データベースサイズ（出力リレーシジョンのタプルの総数）の最適解を推定する。

データベースサイズ  $S$  は、以下の式で計算される。

$$\begin{aligned} S &= |R_M| + |R_A| + (|R_I| + |R_E|) \\ &= n + a + qn^2 \end{aligned}$$

最適のデータベースサイズを推定するためには、まず例外リンクの平均出現確率  $q$  と、集約グラフの平均ノード数  $s$  との関係を見る必要がある。ここで、 $s$  が 1 の時、明らかに  $q$  は  $p$  に一致する。また、 $s$  が  $n$  の時、出力グラフ中には、例外が存在しないので、 $q=0$  となる。それゆえ、 $q$  と  $s$  には、負の相関がある。 $q$  は、 $s^2$  のオーダー、あるいは、 $\alpha^2$  であるので、リンクの出現割合が一様の場合で、 $n$  が比較的大の場合、以下の関係を仮定することができる。

$$q = p(1-\alpha^2)$$

$\alpha$ を用いて書き直すと、以下の式のように書き表せる。

$$\begin{aligned} S &= n + [(p\alpha^2)/\{1-2p(1-\alpha^2)\}] n^2 \alpha^{2+} pn^2(1-\alpha^2) \\ &= n + pn^2 [ \{(1-2p)\alpha^4 - (1-4p)\alpha^2 + 1-2p\} / \{2p\alpha^2 - 2p + 1\} ] \end{aligned}$$

一方、 $\partial S / \partial \alpha = 0$  という条件により、 $\alpha$ の最適解は、以下の通りとなる。

## 第2章 ハイパーメディアの抽象化と物理設計

$$\alpha_{opt}^2 = \{\sqrt{(1-2p)} - (1-2p)\}/2p$$

すなわち、 $\alpha_{opt}$  は、入力グラフのリンク出現確率  $p$  のみの関数であり、入力グラフのノード数  $n$  には依存しない。 $p$  の値が、1 に比して十分小の時 ( $p \ll 1$ ) は、 $\sqrt{(1-2p)} \approx (1-p)$ 、及び  $\alpha_{opt}^2$  は約 0.5 となる。また、 $\alpha_{opt}^2$  が 0 に近づくと、 $p$  は、0.5 に漸近する。

入力グラフが、比較的大規模とみなせる場合 ( $n \gg pn^2$  かつ  $\sqrt{(1-2p)} \approx (1-p)$ )、以下の最適条件が導かれる。

$$|R_{A|opt} + |R_{I|opt} + |R_{E|opt} \approx (3/4)pn^2$$

入力データベースサイズは、 $pn^2$  であるが、ACEクラスタリングを用いることにより、約 25%軽減される。図2-8 は、 $\alpha^2$  に応じて、入力、及び出力のデータベースサイズがどのように変化するかを比較したものである。

もちろん、最適解の結果は、 $q$  と  $\alpha$  との関係に依存する。もし、入力グラフのリンク出現に相関が高ければ、一層コンパクトなグラフ表現を得ることができる。また逆に、入力グラフのリンク出現がランダムであれば、圧縮効率は低い。実際のデータに適用した結果は、2. 5. 9にて説明を行なう。

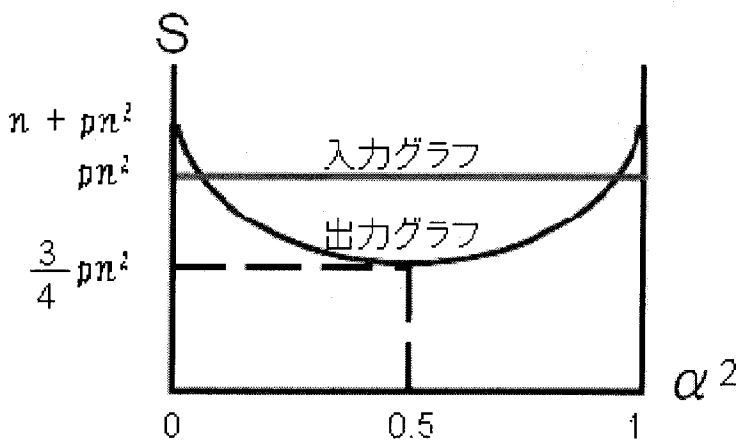


図2-8 データベースサイズSとストレージ関数 $\alpha^2$

2. 5. 7 効率的な検索のための二次記憶管理

次に、二次記憶上の物理設計を考える。この場合は、二次記憶からの検索コストが重要であり、その大部分はページフェッチングに要する時間である。以下では、入出力コストを軽減するための効率的な二次記憶管理構成について検討する。図2-9に示すように、集約グラフを表す隣接行列を、2次元配列として構成できる。各配列要素には、集約リンクビット、当該集約リンクに対応する例外リンクを格納したブロック数、及び先頭ブロックIDを記述する。例外リンクを格納したブロックは、図2-9のようにシーケンシャルに格納できるので、1回の入出力操作で、全てのブロックをフェッチすることが可能である。また、各配列要素は、4バイトで十分である。

もし、配列型集約インデックスと、メンバーシップリレーション $R_M$ とが、主記憶に格納可能ならば、平均入出力コストは、低減される。もし、主記憶に格納できなければ、ロードするために1回の入出力操作が必要である。さらに、例外リンクが存在する場合に限って、例外リンク情報ロードのための入出力操作が追加される。2. 5. 9では、この追加の操作がどの程度具体的に必要なのかについて、実験データをもとに検討する。

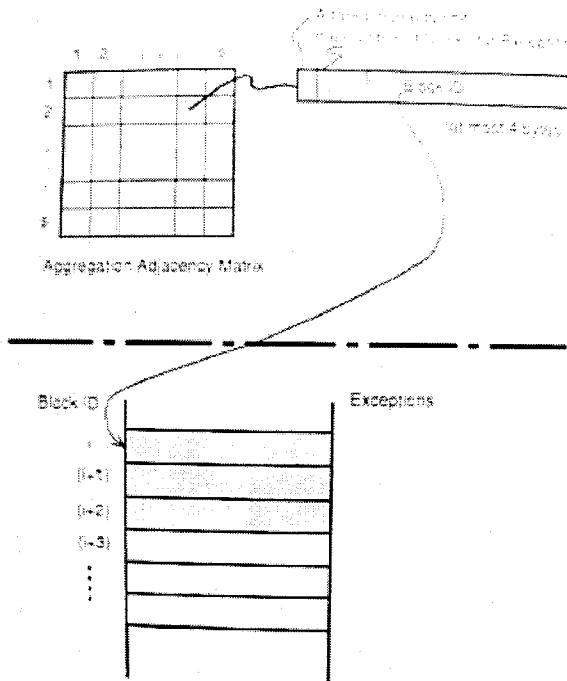


図2-9 ACEクラスタリングの2次記憶

### 2. 5. 8 ACE クラスタリングを実現する一探索的アルゴリズム

本節では、ACE クラスタリングの最適解を見いだすために、カーニハン＝リンアルゴリズム[KeLi70]に基づく探索的手法につき説明を行なう。カーニハン＝リンアルゴリズムは、有名なグラフ分割アルゴリズムである。偶数個のノードを有する重み付き有向グラフを入力とし、カット数が最小となるような等ノード数のグラフに2分するグラフ分割手法である。

ここでは、カーニハン＝リンアルゴリズムを ACE クラスタリング問題に適用するために、次の拡張を行なった。まず、クラスター数を2個以上でもよいようにすることであり、しかも一般には、各クラスターサイズが同一でなくてもよいことである。具体的には、各ステップにおいてダミークラスターを考慮して、分割、併合等の操作がシミュレートできるようにしたことである。このアルゴリズムを以下に示す。

### An Extended Kernighan-Lin Algorithm for ACE

(Step1) Initial clustering step

Make all nodes belong to one cluster  $C_1$  (the number of clusters:  $k = 1$ )

(Step2) Repeating step

While no updating occurs do

Make  $C_{k+1}$  as a dummy cluster

for  $i:=1$  to  $n$  do (the number of nodes:  $n$ )

Choose some unselected node and call it  $v_i$

Let  $j_c$  be the cluster of  $v_i$  (i.e.,  $v_i$  in  $C_{j_c}$ )

for  $j:=1$  to  $k+1$  and  $j \neq j_c$  do

Calculate the cost when  $v_i$  moves into  $C_j$

Select the pair of  $(v_i, C_j)$  if the movement makes the  
best benefit (i.e., largest decrease in cost)

end

Add  $(v_i, C_j^{(i)})$  to the list of movement with the best  
benefit for this group of  $n*k$  alternative cases

end

Find  $l$  ( $0 \leq l \leq n$ ), s.t.  $\text{sum Cost}_l \rightarrow \max$

Perform the translations  $(v_1, C_j^{(1)}), (v_2, C_j^{(2)}), \dots, (v_l, C_j^{(l)}), ;$

that is move  $v_i$  into cluster  $C_j^{(i)}$

Rearrange clusters ( $k$  may change to  $(k-1)$  or  $(k+1)$ )

end



### 2. 5. 9 実験結果

本節では、2種類のタイプのデータに本手法を用いた結果を説明する。1つは、実際のハイパーテキストデータであり、ノード間に相関があるものである。もう1つは、ランダムにリンクが生成されたグラフで、グラフ理論アルゴリズムのテストベツトになっているものである。

#### (1) Hypertext on Hypertext

“Hypertext on Hypertext” [Hype89] は、ハイパーテキスト構造をもつ典型的なシステムである。8件のハイパーテキスト関連論文[CACM88]を、ハイパーテキスト構造化し、HyperCard [Good87] 上で利用できるようにしたものである。基本構造は、313のカード情報（ノード）とそれらのリンク情報、並びにインデクシング情報で構成されている。本手法を“Hypertext on Hypertext”に適用する際に、以下のルールを用いた。

- (A) 同じキーワードを持つカードは、相互にリンクがはられているものとみなす。
- (B) “hypertext”と“hypermedia”というキーワードは、全体の1/3以上に出現しているので、省略する。

オリジナルなシステムでは、集合操作的なコマンドを用意していないので、関連するカード間に過剰なリンクがはられる可能性がある。この可能性を防ぐために、インデックスカードが用いられている。このため、悪影響として、関連するカード間でも、直接ナビゲーションを行なうことができない場合がある。そこで、この問題を避けるため、ルール(A)を採用した。

また、情報検索分野でいう検索精度をあげる（適合率と再現率をあげる）ため、ルール(B)を採用した。これは、全ての論文が、ハイパーテキスト/ハイパーメディアに関連があるため、このようなリンクがあまり意味を持たないことが明らかであるためである。

表2-1にシミュレーションで用いた入力データと、解析結果を示す。結果は、主記憶上のデータベースサイズの軽減にも、また二次記憶上のページフェッチングにも有効であった。カード間の相関が非常に強いため、入力グラフのリンク数の軽減は、77.2%となった。二次記憶格納構造に対しては、ブロックサイズを

## 第2章 ハイパーメディアの抽象化と物理設計

4K(512byte)とすると、配列型集約インデックスは、高々6ブロック(26 times 26 times 4 / 512)ですむことがわかる。

また、配列型集約インデックスのうち、2/3の要素が例外リンクのないものである。それゆえ、配列型集約インデックスが主記憶に格納されれば、わずか1回の入出力操作で必要な情報を検索することができる。しかも、その操作は、例外リンクが存在している場合のみであり、平均1/3程度に軽減される。

一方、オリジナルのグラフでは、24ブロック(313 times 313 / (512 times 8))が占有される。従って、以下の2種類の実現が考えられる。すべてのブロックを主記憶に常駐するか、もしくは、毎回の操作に対して、1回の入出力操作を行なうかである。いずれの場合も、本手法適用後の構造の方が妥当である。

図2-10に“Hypertext on Hypertext”から得られた、大局的な関連構造を示す。5つのカード以上からなるクラスターを考慮すると、8個の主要な集約ノードが得られる。しかも、その各々に具体的な名前付けが可能である。それゆえ、ACEクラスターリングは、効率的な物理DB設計に適用できるだけでなく、効果的なオーバビューダイアグラムを提供できる可能性がある。

また、8個の主要な集約ノードのうち、3個が、単一概念間の組み合わせ概念(KMS&Note Cards, Node/Links&Note Cards, Note Cards&Query/IR)である。従来のグラフ分割アルゴリズムでは、抽出されにくかった複合的集約ノード間の関係が抽出されている。さらには、“Hypertext on Hypertext”全体構造の関係が、Low Relevant Cardsも含めて、R. KumarのTrawlingで得られたWebの構造とも類似性のある全体構造となっている。

### (2) 標準ランダムグラフ

もう1つのシミュレーション例は、Johnson [JASM89]の標準ランダムグラフである。標準ランダムグラフは、 $n$ と $\sim p$ という2つのパラメータにより規定される。パラメータ $n$ は、対象グラフのノード数である。また、パラメータ $\sim p$  ( $0 < \sim p < 1$ )は、任意の2つのノードがリンクを構成する確率である。

表2-1に、入出力結果を併せて示す。このようなランダムグラフの場合でさえも有効であることがわかる。入力データベースサイズの軽減は、12.3%となった。また、このランダムグラフの入力、並びにシミュレーション結果の構造を、

## 第2章 ハイパーメディアの抽象化と物理設計

図2-11、図2-12、図2-13に示す。これにより、ACEクラスタリングによる関連の集約化の様子がある程度把握できる。

### 2.5.10 分析モデルとの比較

2.5.5において、最適解におけるデータベースサイズの削減は約25%にのぼることを述べた。ここでの仮定は、入力グラフのリンクの存在が単一分布に従うこと、及び制約式の関係に従うことであった。実験結果と、分析モデルとの差異は、主として2つの理由が考えられる。1つは、シミュレーションに用いたグラフのリンク分布が、計算式から実際には、ずれていることである。式  $q = F(\alpha^k)$  を考えた時、"Hypertext on Hypertext"のグラフは、 $k < 2$  であり、ランダムグラフの場合は、 $k > 2$  であったと想定される。もう1つの理由は、探索的アルゴリズムを用いざるを得ないため、シミュレーション結果は、局所最適解であるためである。

また、興味深い知見としては、集約リンクを構成している個々の入力リンク数と、正の例外リンク数との比率である。分析モデルでは、この比率は1であり、バランスしている。一方、実験結果では、相関の高い場合は1より大きく、ランダムの場合は、1より小さい。このように、ACEクラスタリングを用いた縮約の程度は、負の例外リンクの領域（集約リンク）をいかに抽出できるかに依存する。

	Hypertext on Hypertext	Random Graph
The number of nodes : $n$	313	124
The number of edges : $e$	8461	2642
The fraction of input links : $p$	8.6%	15.5% ( $\bar{p} = 16\%$ )
Aggregation nodes :  VA	26	58
Aggregation links :  EA	138	254
Exclusive links :  Ee	94	110
Inclusive links :  Ei	1440	1742
Average fraction of exceptions : $q$	1.5%	12.0%
The fraction of non-exception areas	30.8%	40.4%
DB size $S = n +  EA  +  Ee  +  Ei $	1925	2230
The reduction gain : $1 - (S/e)$	77.2%	12.3%

表2-1 実験結果

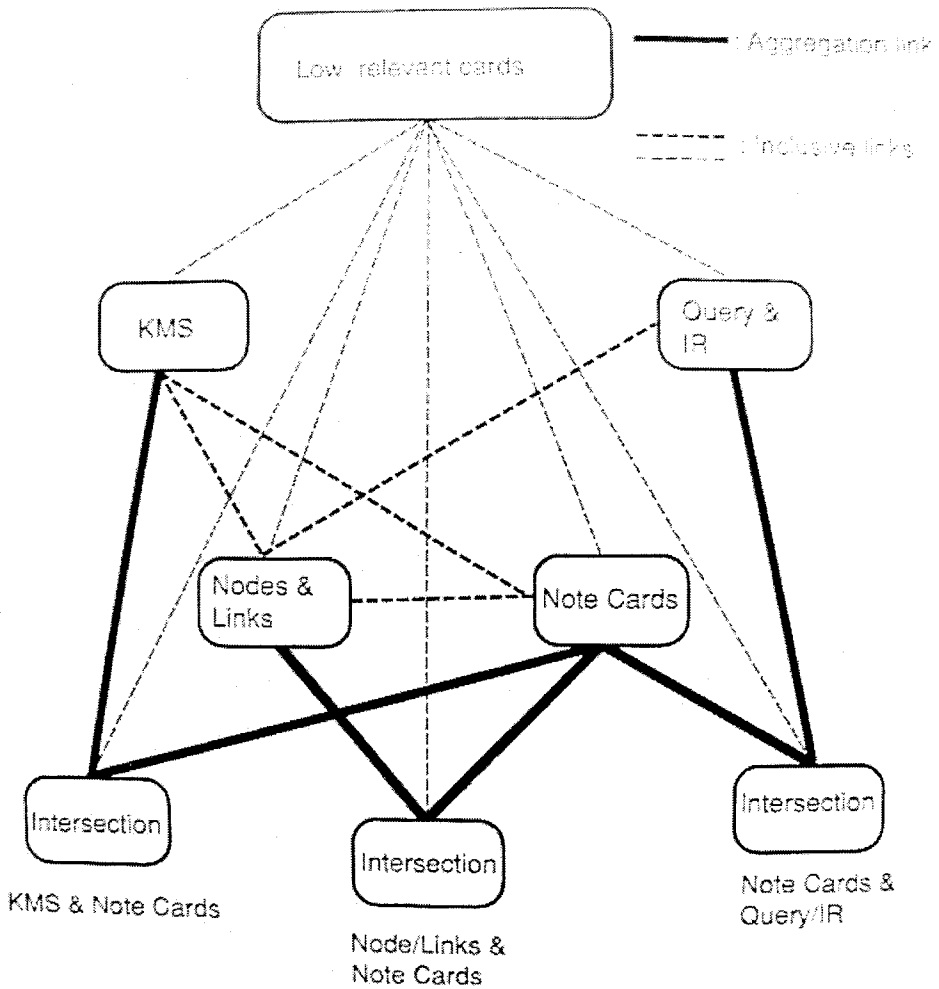


図2-10 “Hypertext on Hypertext”の大局構造  
(クラスサイズが5以上)

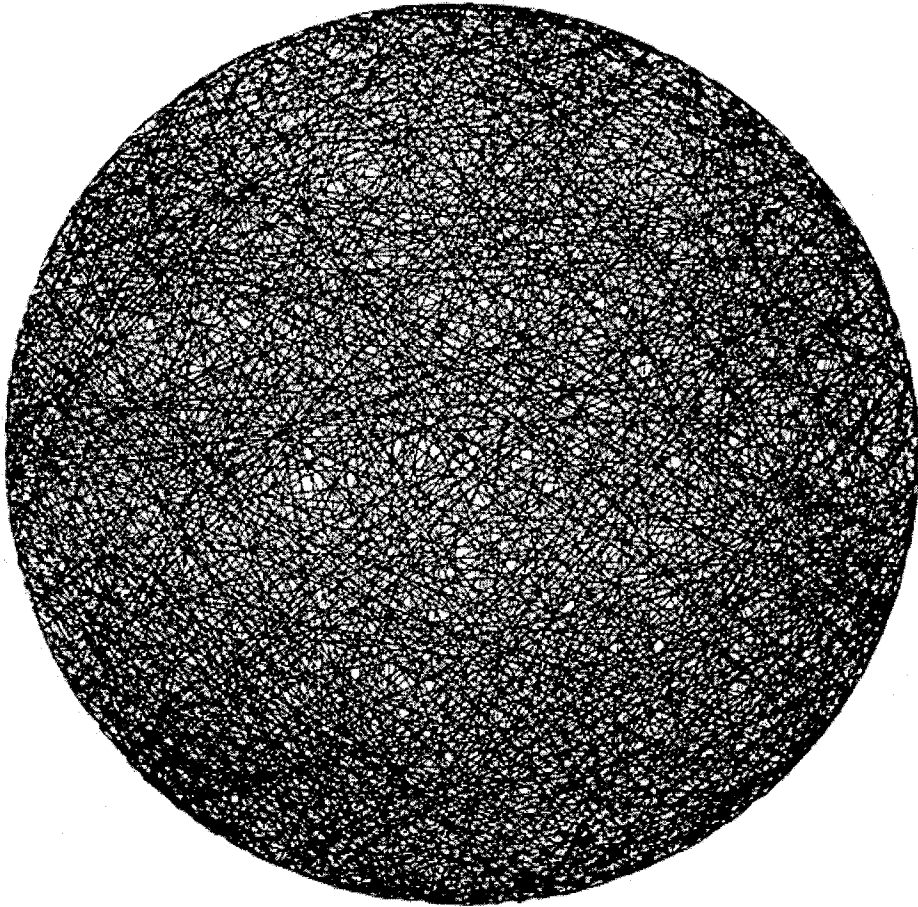


図2-11 ランダムグラフの入力データ

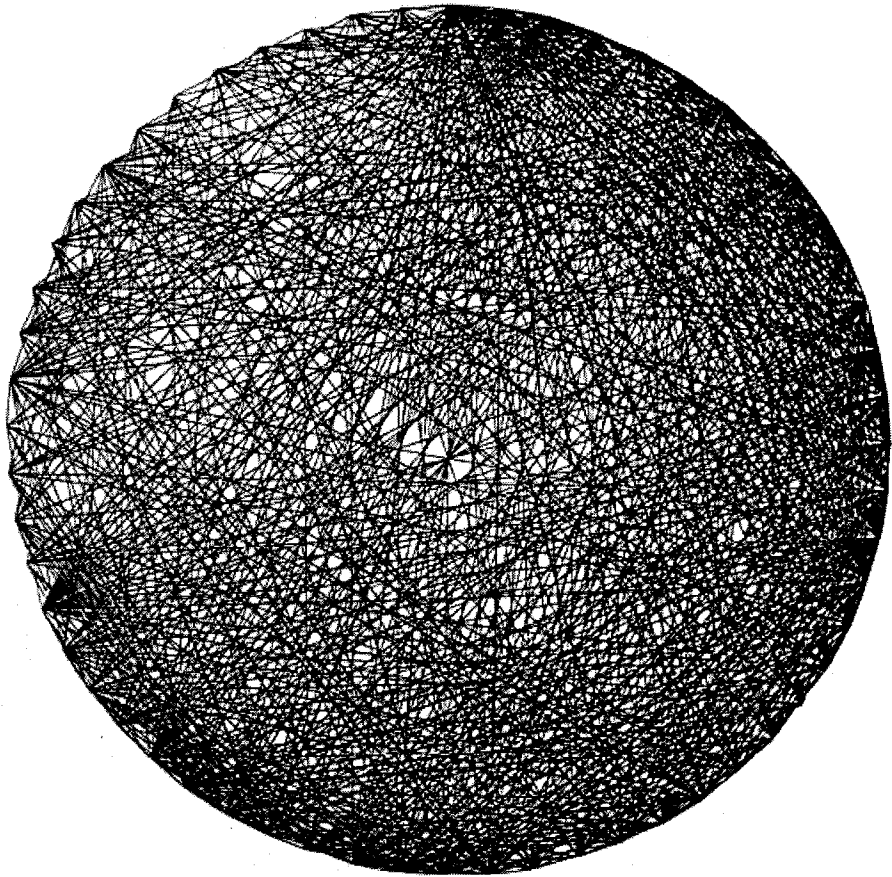


図2-12 ランダムグラフの結果例  
(Aggregation Graph + Inclusive Links)

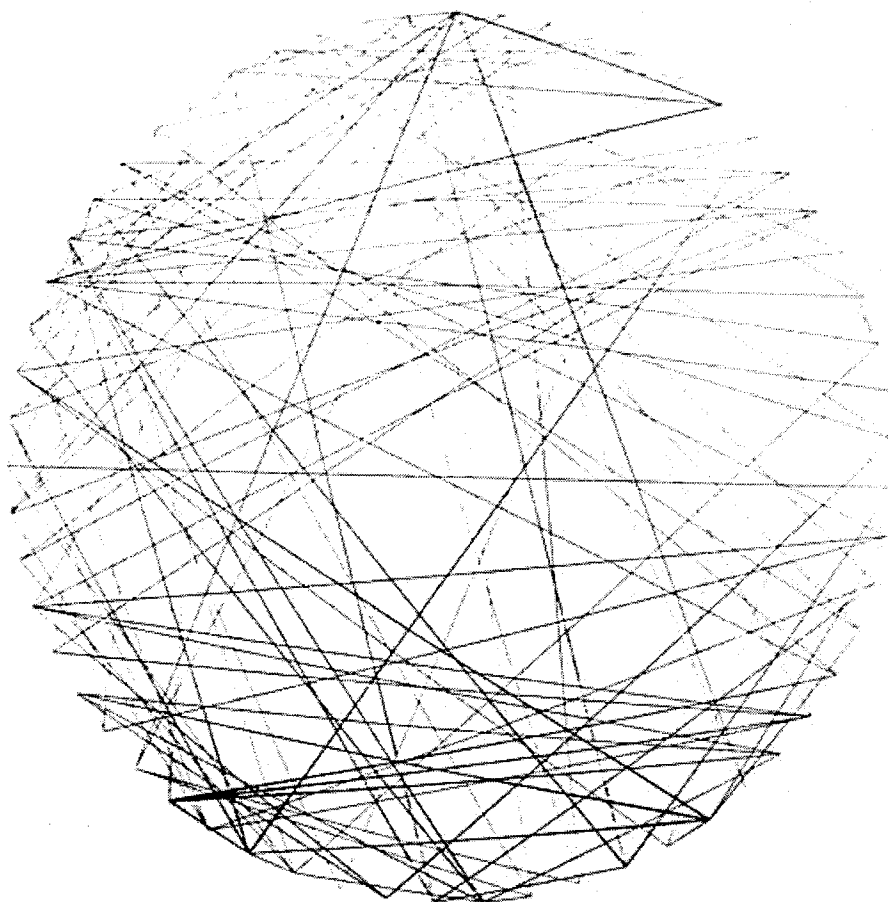


図2-13 ランダムグラフの結果例  
(Aggregation Graph)

### 2. 5. 11 今後の検討課題

本章では、ACE クラスタリングの改善と、より一般的なハイパーテキスト／ハイパーメディア構造への適用についてまとめる。

#### (1) 重畳的／再帰的 ACE クラスタリング

入力ノードと出力クラスターとのメンバーシップリレーションは、多対1に制限して議論を進めてきた。この制限は、単純な出力構造にするため、メンバーシップリレーションのサイズを入力ノード数  $n$  と同じにするためである。このような単純な構造は、二次記憶上の効率的な検索と主記憶上のコンパクトな表現に有効であった。しかし、ある種の特殊な場合では、単純な多対1のクラスタリングと例外の考慮というアプローチよりも有効な表現方法が存在する。1つの拡張は、重畳的にクラスタリングを行なうことである。すなわち、メンバーシップリレーションの多対1関連の制約をとり、多対多関連とすることである。このような重畳型クラスタリングは、ある入力ノード集合が複数のクラスターと相関が高い際に有効である。

もう1つの拡張は、再帰的にクラスタリングを適用することである。再帰的にクラスタリングを適用することにより、集約グラフと例外グラフとの系列が抽出される。このようなグラフの分解は、多変量解析における主成分分析法とアナロジーがあるものである。主成分分析と同様、重要な成分から順に一種の直交分解が行えるとみなすことができる。特に、同じ構造の集約グラフで再帰的に集約グラフの抽出が可能であれば、有益な物理構造が抽出できると想定される。より詳細な分析評価（特に、入力のグラフ構造のタイプとの関連等）が、今後の課題である。

#### (2) 探索的アルゴリズム

2. 5. 4で述べたように、ACE クラスタリングは、グラフ分割問題と類似性があり、NP-困難な問題である。このため、5章で説明した探索的アルゴリズムは、実現の一方法にすぎない。他の探索的アルゴリズム等に関する説明は、[JAMS89]に詳しい。Simulated Annealing は、代替的なアルゴリズムである。これは、最適解の精度は良くなる可能性があるが、時間がかかるアルゴリズムである。また、別のアルゴリズムは、Greedy アルゴリズムに基づくものである。このアルゴリズムは、一次元探索手法なので、局所最適解に陥りやすいが、比較的早いアルゴリズムである。それゆえ、対話的操作が必要なオーバビューダイアグラムへ



の適用が考えられる。ACE クラスタリングを適用したより効果的なオーバビューダイアグラムの実現は今後の課題である。

### (3) 一般的なハイパーテキスト構造

任意のハイパーテキスト構造へ適用するためには、参照リンクと同様、構成リンクについても考慮しなければならない。ここでは、統合化のため2種類の方法につき検討する。1つは、構成リンクを縮退させて、グループノードとして取り扱う場合である。実際のシステムでは、構成リンクが比較的少ない場合も多いので、このような取り扱いでも有効である。

また、もう1つの方法は、全ての関連を階層構造的に取り扱うアプローチである。これは、逆に参照リンクを階層構造的に再構成して、形式的には構成リンク同様に取り扱うアプローチである。いったん、階層構造に変換されると、種々の効率的な方法が提案されているので、統一的に取り扱える。ACE クラスタリングは、参照リンクをノードの属性情報（クラスタ名）に変換させる方法とも考えられるので、このような、参照構造の階層構造化にも有用である。

我々は、以下の2つの理由により、後者の階層化アプローチがより有用であると考え。1つは、階層化させることにより、より効率的なアルゴリズムが適用できるだけでなく人間の認知の観点からも有効であると考えためである。また、もう1つは、より高次のレベルで、ハイパーテキスト構造と、データベースモデル諸機能との統合がはかれると考えるためである。このような統合が実現できれば、情報の利用効率が向上する。

## 2.6 本章のまとめ

本章では、従来のハイパーメディア構造から高レベルの関連を抽出するクラスタリング手法、ACE クラスタリング (Aggregation Clustering with Exceptions)、並びにその実現アルゴリズムを提案した。また、本手法を物理データベース構造の設計に適用した際の分析モデルの規定、及び、実データを用いた実験評価を行なった。

本手法を適用した実験結果では、データベースサイズの格納効率の改善、並びに二次記憶におけるページフェッチングの改善がはかられた。またこの実験結果により、本手法は、ハイパーメディア構造に対する物理格納構造の設計に寄与するだけでなく、ヒューマンインタフェースとしてのオーバビューダイアグラムの実現も適用できると想定される。本手法を一般化することにより、ハイパーテキスト構造とデータベースとを統合するための鍵となるスキーマ抽出を行ない、ひいては、より高次の意味的構造を規定していくことができる。

ハイパーメディアの抽象化技法は、後年の Web コミュニティ発見のアルゴリズムとも関連が深い。Web コミュニティは、大規模ハイパーメディアとしての Web 情報に対して、構造・ビヘイビア分析を行うことによる情報の偏在性、相関性を引きだすことに成功している。今後ますます、情報フィルタリングや、情報信頼性向上の側面での利用が期待される。

# 第3章 抽象化によるハイパーメディアの 論理設計

## 3. 1 ハイパーメディア論理設計の必要性

2章では、ハイパーメディア抽象化の概念を、物理構造設計と捉えて説明したが、抽象化の意味するところは、このようなレベルに留まるものではない。本章では、ハイパーメディア抽象化の概念を論理設計にも適用させて、大局的な関係情報を論理レベルで表現することにより、大規模なハイパーテキスト／ハイパーメディアの構築と情報共有化に寄与する。加えて、各分野における固有問題とその解決方策につき、プロトタイプシステム開発による機能検証を行い、かつ、情報モデル特性からの考察を行う。

論理設計モデルの構築・運用は、ハイパーメディアの設計はもとより、データベースをはじめ、一般的なソフトウェア開発において、必須のプロセスである。ハイパーメディア設計に際しては、グラフ構造に基づく関係表現の設計自由度と、不必要な関係構造を増やさないという意味での制約とのバランスをとり、リンクの爆発を招かない高品質のハイパーメディアシステムを実現する必要がある。関連する情報をリンクで結んで情報提供するハイパーメディアは、一般の利用者にも受け入れやすい考え方であるが、一方では、非常に不均質なハイパーメディアが結果的に生成されたり、また、リンク先がない、あるいは間違っけてリンクされたりしているという低品質なハイパーメディアになるケースも多い。このような品質劣化の問題を未然に防ぎ、かつ、ハイパーメディアの簡便な編集、操作を生かす設計手法の提供が重要となる。

本章では、3つの観点から、抽象化によるハイパーメディアの設計表現手法について提案する。第一は、データベースの論理設計手法であるE-Rモデルをハイパーメディアの特性を考慮して援用する方式である。リンク規定に対して、品質が均質化され、また、データベース問い合わせ手法を用いたリンク操作の多様化が図られる。第二は、例外を含んだ関係の抽象化設計手法の提案である。この提案は、E-

### 第3章 抽象化によるハイパーメディアの論理設計

Rモデルの考え方をさらに発展し、関係の情報表現にスケーラビリティを与える方策である。第三は、リンクの自動生成の一方式として、言語規約によるデータベースからのハイパーテキスト／ハイパーメディア生成手法の提案である。この手法は、論理設計として構築した概念空間の情報を、利用者のもつハイパーメディアの表現空間に射影、マッピングする方式である。

このようなハイパーメディア論理設計の今日的意義としては、品質管理のできるWeb情報の作成、編集、メンテナンスに寄与していることである。今日では、人手のみに頼らず、体系的にリンク構造を構築し、かつ、リンク構造の分析、評価、メンテナンスを行うWeb Content Managementシステムが実現されている。本章で提案しているハイパーメディアのリンク自動生成の考え方は、人手とコンピュータとによる処理を協調させて効率化をはかる半自動的スケーラブルなアプローチの先駆けとなるものである。

#### 3. 2 関連研究

ハイパーメディアの論理設計手法については、古くから種々の言及、提案がなされてきている。1945年のVannevar Bush [Bush45]以来、Douglas Engelbart, Ted Nelson, Andy van Dam, Nicole Yankeovich, Frank Halaszなどの研究者がハイパーテキスト、ハイパーメディアの設計指針について足跡を残している。特に、Halaszの“Seven Issues”: Revisited [Hala91]に、詳細がバランスよくまとめられている。

ハイパーメディアの論理設計についてのアプローチとしては、“Literalists”と“Virtualists”との2派に大別される。前者は、ハイパーメディアの構造（リンク構造）は、明示的に生成され、明示的な構造に基づいてナビゲーションを行うという考え方に基づく設計を行う。後者は、ハイパーメディアの構造（リンク構造）は、コンテンツ、データ、属性で表現されるノードの相対的關係の中で暗示的に生成され、このような暗示的、動的生成された構造に基づいてナビゲーションを行うという考え方に基づく設計を行う。両者は、応用分野にも依存するが、各々一長一短がある。前者は、一般に、直感的でわかりやすい反面、利用操作が限定される。後者は、その逆の特性である。

本章で提案する抽象化によるハイパーメディアの論理設計は、後者の流派に属する。ハイパーメディアをアウターDBと捉え、データベースの論理設計方式の良さを踏襲しつつ、“Virtualists”としてのハイパーメディア実現に寄与する方式である。また、中長期のゴールとして、スケーラビリティのあるハイパーメディアのリンク自動生成をめざす先駆的研究として位置づけられる。設計プロセス効率化のアプローチとして、BieberらのRelation-Navigation分析、Rossiらのオブジェクト指向によるハイパーメディア設計 [AsSi99] などがある。また、田中 [TaQi92]、Bodner [BoCT97]らの問い合わせによるリンク生成などがある。本論文で提案した手法は、他の手法と比較し、美術、芸術、観光案内など、多次元の観点からナビゲーションを行う応用システムに適した手法である。本手法の考え方は、ハイパーメディアの設計・ナビゲーションのみならず、多次元情報分析ツールとしてのOLAP (On-Line Analytical Processing)操作の先駆的位置づけの研究となっている。

また、ハイパーメディアの品質を高める方策としては、自動生成をめざしたアプローチとは別に、標準化による品質向上方策もあげられる。WuilらのOpen Hypertextの活動、WebコンソーシアムにおけるSMILなどの活動があげられる [AsSi99, etc.]。このようなハイパーメディアの品質向上は、XMLとして、メタデー

### 第3章 抽象化によるハイパーメディアの論理設計

タを含むコンテンツの総合利用を促進すると共に、Web Content Management システム、Web サービスの基本機能として、現在のシステム構築に活かされている。

### 3. 3 ハイパーメディア論理設計の課題と要件

#### 3. 3. 1 課題

ハイパーメディア論理設計の課題を以下のように整理する。

##### (1) 柔軟な設計指針の欠如

プレゼンテーション型のようなハイパーメディアを構築する場合、まず素材としてのメディア収集・電子化を行う。しかるのちに、個々のメディアの関連づけを行う。個人の情報管理ツールとしてハイパーメディアを利用する場合も、おおむねこのような製作過程である。一方、情報検索・分析を主眼としたナビゲーション型の大規模ハイパーメディアでは、データベース概念モデリングのようなスキーマの設計を行い、その後、個別メディアデータのアサインを行う。このように、種々の設計要請があるにもかかわらず、従来の構築システムでは、柔軟な設計がなされていない。

##### (2) 同期と関係づけの問題

マルチメディアは、静止画・動画、音声などの個別メディアを時間的、空間的に連動して複合化させ、複合メディアを用いたことによる相乗効果を目的とする。したがって、実時間メディアにおける同期処理や、空間メディア間による割り付け処理が本質となる。一方、ハイパーメディアでは、マルチメディア情報の意味的關係を陽に指定すること（ハイパーテキスト構造化）に重点がある。しかし、同期と関係づけの問題は、本来切り離せない表現であり、より単純なモデル表現が望まれる。

##### (3) 利用・編集面の不備

電子化情報のメリットは、格納された情報を効率的・効果的に利用できることである。この点で、従来のナビゲーションツールは、いまだ不十分である。ナビゲーションツールの不備は、ひいては情報空間における迷子の問題(Disorientation Problem)を引き起こす。一方、編集時における関連づけやインデクシングのコストが多いため、編集。電子化の自動化方策が必要である。

#### 3. 3. 2 ハイパーメディア論理設計の要件

以上の問題をふまえ、スケーラビリティのあるハイパーメディアシステムの論理設計としての要件を整理する。

##### (1) Top-down と Bottom-up 設計の提供

従来のデータベース設計管理は、Top-down 設計であり、全体のスキーマ設計を行ったのち、個々のインスタンス情報を作成する。一方、従来のハイパーメディア設計では、個々のマルチメディアオブジェクト（URL ページ等）を作成した後、全体の関連づけを行うという Bottom-up 設計である。一般に、大規模定型的な対象であれば、Top-down 設計が有効であり、中小規模で非定型的な対象であれば Bottom-up 設計が有効である。したがって、初期の段階で対象が少なければ、Bottom-up 設計を行い、対象量の増加に応じて、Top-down 設計への移行が行えることが必要である。あるいは、Top-down 設計を行った対象でも、更新後に、規模が小さくなれば、Bottom-up 設計的な利用が行えるという柔軟な設計指針が必要である。

##### (2) 属性表現とグラフ構造表現との変換

問題点(2)で述べた、同期と関係づけの問題は、見方を変えれば、E-R-A モデルのような属性表現と、ノードとアークとで構成されるグラフ構造表現との相互変換とみなすことができる。すなわち、オブジェクトとオブジェクト間との間接的／直接的関係を、両方もしくは、いずれか一方の構造表現で実現できることが、問題点(2)の解決策の1つとなる。ハイパーグラフを基本としたモデル[WaSh90, HaKa90]や、時間軸オブジェクトを考慮した MOL モデル[TaMH91, TaMH92]などでも試みられている。ハイパーメディア構造表現の一般的な取り扱いが課題である。

##### (3) 多様なナビゲーションの提供

Halasz の”Seven Issues Revisited”[Hala91]でも述べられているように、種々のリンク表現に基づくナビゲーション（リンクを利用した情報検索機能）と提供することが必要である。リンク表現は、ノード間リンクを陽に指定する外延的リンク(Extensional Link)だけでなく、同じ属性をもつノードを間接的に結合する内包的リンク(Intentional Link)をも対称とする[DeRo89]。また、書誌的・概念的関係でなく、メディア情報（たとえば、画像の形状、色等）に立脚したナビゲーションを提供する必要がある[HiHa92]。



さらに、対象が大規模になった場合でも同様にハイパーメディア処理的取り扱いができるようなナビゲーションを提供する必要がある。このような大規模ナビゲーションには、グラフ構造のクラスタリングが重要な技術となる[BoSh91, HaKW91-1]。

#### (4) インデクシングの自動化

2次情報の自動作成や、プレーンテキストからのハイパーテキスト変換も、編集効率化を図る上で重要な要因である。対象がマルチメディアとなるハイパーメディアの場合には、書誌的項目やテキスト情報のみでなく、静止画、動画、音等のメディア情報を利用したインデクシングも考慮する必要がある[HiHa92]。さらに、個別メディアエディタとのカップリング（たとえば、埋め込み型システムによる構成）も、インタラクティブな関係づけを行う際に重要となる。

## 3. 4 拡張 E-R モデルによるハイパーメディアデータモデル

### 3. 4. 1 ハイパーメディアプラットフォーム Himotoki

本節では、前節でまとめた要件をもとに、実用規模を想定したナビゲーション型のハイパーメディアプラットフォーム Himotoki について記述する[HHTH96-1]。

#### (1) 拡張 E-R モデルを基本とした構造化

Top-down と Bottom-up 設計の提供、並びに、多様なナビゲーションの提供を行うために拡張 E-R モデルを基本構造とする[BaCN92]。拡張 E-R モデルとは、Entity (実体)、Relationship (関連)、Attribute (属性) からなる E-R モデルに加えて、is-a, is-part-of の関係も考慮するモデルである。is-a 関係では、上位クラスの属性を下位に継承し、is-part-of 関係では、継承関係のない階層関係か、もしくは、下位クラスの属性情報を上位に伝播する。

拡張 E-R モデルを採用した理由は、以下の通りである。第一に、汎用性・拡張可能性である。データモデル的には、従来から利用されているものであるが、今後の新機能／ツール群を取り込むことを考慮して、モデルのオープン性・普遍性を重視する。第二に、規模に依存しないスケーラブルなナビゲーションを実現するためである。すなわち、ノードの集合としての実体、リンクの集合としての関連とが対応するためである。これはまた、ハイパーメディアのクラスタリング[BoSh91, HaKW91-1]とも親和性が良い。また、単にデータ蓄積・管理にデータベース技術を利用するだけでなく、ナビゲーション機能にデータベース操作機能を援用するためである。

#### (2) 概念系と表現系との分離

ビジュアルインタフェースレベルの明確な区別と同様、内部のモジュール構成においても概念系と表現系処理との分離を行う。これは、データベース技術とハイパーメディア双方の良さを生かすためである。ここで、概念系処理とは、ハイパーメディアの論理・意味的関係に関する処理である。具体的に、書誌・キーワードなどで表現されるオブジェクト間の直接／間接的関係を記述する。ハイパーメディア的には、ドキュメント間の関係を記述することに相当する。

一方、表現系処理では、ハイパーメディアの表現・レイアウト構造、イベント処理、画像処理等のメディア処理を行う。この処理は、従来の中小規模のハイパーメディアで実現されてきた操作インタフェースが対応する。

### 第3章 抽象化によるハイパーメディアの論理設計

概念系と表現系との分離の利点は、概念系の情報管理が抽象化されることによるモデルの単純化である。また、異種メディアの導入容易性や、処理ハードウェアを分離して全体の効率を上げる等の拡張も行いやすい。

#### (3) メディアインデクシング

ハイパーメディアの効率的編集には、二次情報としてのインデクシング技術が重要である。インデクシングには、キーワードのようなテキストメディアのみならず、静止画・動画当の非構造メディア情報への対処も可能である。非構造メディアに対する二次情報利用の際は、インデクシングだけでなく、マッチング処理も重要となる。ピクチャインデックスを用いた画像検索[HiKa92, HiHa92]は、静止画メディアに対するインデクシング手法の1つである。利用者は、画像の形状、色、領域等の情報を手がかりに、対象情報の検索、ナビゲーションが行える。

#### (4) ハイパーテキストクラスタリング

百科事典や、マニュアルのように対象が大規模の場合や、Bottom-up 設計を支援するためには、クラスタリング技術が不可欠である。クラスタリングによる情報構造化は、2次情報、あるいは、高次情報の自動抽出と位置づけられる[SaHC92]。このような構造化手法を用いると、検索の高速化、オーバビューダイアグラムの実現といった大規模ハイパーテキストに有効な種々の方策が実現できる。

### 第3章 抽象化によるハイパーメディアの論理設計

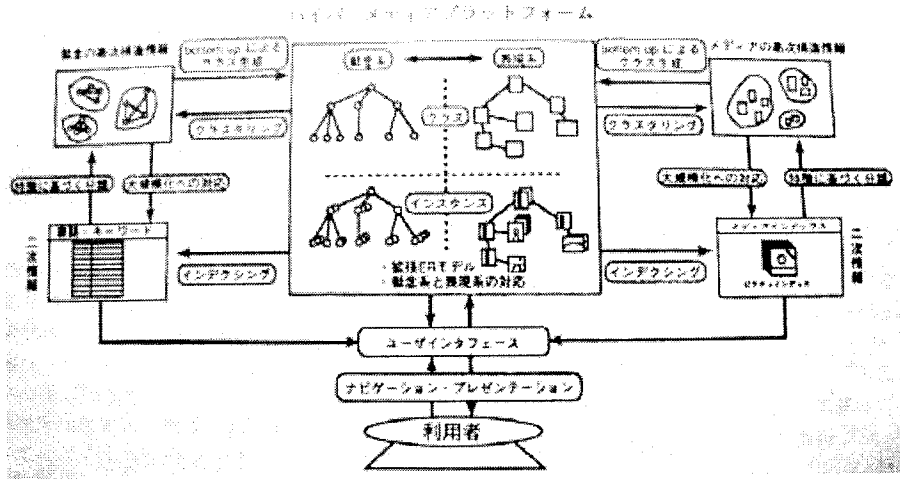


図3-1 ハイパーメディアプラットフォーム Himotoki

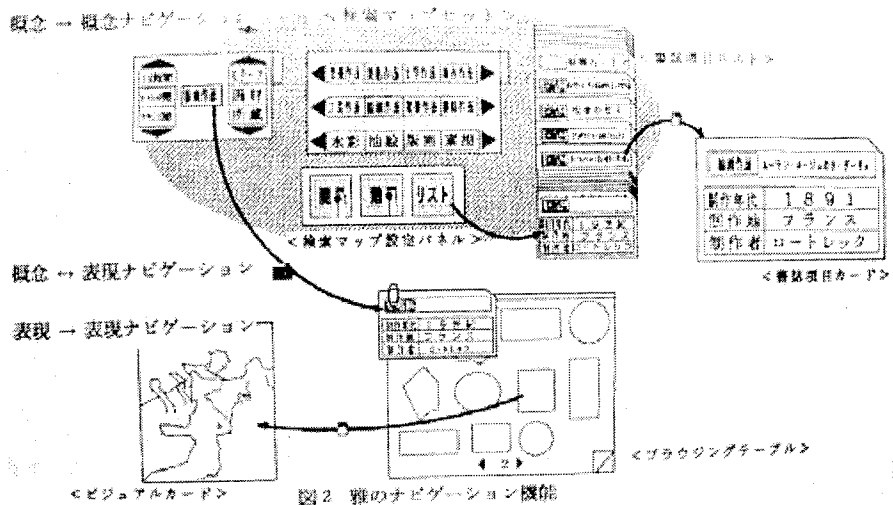


図3-2 Himotokiのナビゲーション機能

#### 3. 4. 2 システム評価と課題

本節では、ハイパーメディアプラットフォーム Himotoki によるシステム構築を行い、システム評価と課題について説明する。

##### (1) ナビゲーション機能の実装

Himotoki では、図3-1、図3-2のように概念系のツール・データと、表現系のツール・データとがあるため、下記の4つのナビゲーション機能の実装を行った。

**概念→概念ナビゲーション：** E-R モデル上の関連・実体図、is-a 階層図、is-part-of 階層図等の検索マップセット（スキーマブラウザ）を基本に、概念間の関連オブジェクトを検索・提示する。書誌項目リスト（インスタンスリスト）を併用して、スキーマ間、スキーマ・インスタンス間、インスタンス間のナビゲーションを実現する。

**概念→表現ナビゲーション：** 概念インスタンスに対応する表現インスタンス（ビジュアルカード）を提示する。概念インスタンスには、種々のメディア表現があり、逆に、表現インスタンスにも複数の概念が関連し得るので、両者の関係は、多対多の関係となる。複数の表現インスタンスが対応すれば、一覧性を生かしたブラウジングテーブルにより、複数候補の一覧提示を行う。

**表現→概念ナビゲーション：** 表現インスタンスに対応する概念インスタンス（書誌項目カード）を提示する。複数の概念インスタンスが対応すれば、同様に、一覧提示を行う。

**表現→表現ナビゲーション：** 表現インスタンス間のナビゲーションである。従来の静的リンクに相当する直接遷移だけでなく、対応する概念インスタンスの情報を利用した集合操作的ナビゲーションも提供する。また、ピクチャインデックスを用いた類似画像検索も、本ナビゲーションとして実装する。

ナビゲーションには、常に現時点对象とするクラススキーマ、及び、インスタンスが存在する。これをフォーカスと呼ぶ。フォーカスに関連するデータは、フォーカスの移動に伴い、更新される。ある時点でのフォーカス内容をつねに表示しておきたい場合には、スナップショット化する。

#### (2) オーサリング機能

オーサリング機能としては、各クラススキーマレベルの編集・更新、並びに、各インスタンスレベルの編集・更新の各機能を提供する。簡便かつ操作性のよいインタラクティブ処理、並びに、大規模データの効率的な一括編集としてのバッチ処理双方を用意する。

#### (3) ユーザインタフェース

ユーザインタフェースの基本は、マウス等を利用した簡便な操作インタフェースの提供にある。ユーザドリブンなイベント処理として、マウスボタンクリック、ドラッグイン（対象をある領域に引きずり入れる）、ドラッグアウト（対象をある領域から引きずり出す）の3つ、もしくは、これらの組み合わせですべての操作が行えるように実装する。また、概念クラス/インスタンス、表現クラス/インスタンス等の対象は、全てウィンドウとして管理され、各ウィンドウにイベントが組み合わされて処理が決まる構成となっている。

#### (4) パリ観光案内システムの開発と評価

図3-3は、ナビゲーション・オーサリング評価用として構築したパリ観光案内ハイパーメディアの拡張 E-R モデルである。通常のリンクに沿ったハイパーテキストナビゲーションに加えて、5W1H 的な問い合わせをベースにしたナビゲーションが行える。

例では、「場所」、「人物」、「建造物」、「事件」、「美術作品」、「年代」の6つのベースエンティティを基軸としている。約 600 程度の概念インスタンスと対応する表現インスタンス、並びに、約 2,000 程度のリンクからなる。「人物」、「建造物」、「事件」には、is-a 関係で、よりきめ細かい実体を定義している。これは、実体間の関連をより細かく規定し、精度のよいナビゲーションを実現するためである。たとえば、ある建造物に対して関係が深かった人物という関係がある。これに対し、人物が建築家であれば、設計するよいうより詳細な関係を規定することができる。また、建造物が墓地であれば、埋葬するという関係も規定することができる。

パリ観光案内システムの開発を通じての評価点は、このようなきめ細かい指定を行うことにより、再現率・適合率の高いナビゲーションが実現できることである。図3-4は、静的リンクを利用したナビゲーション例である。フランスの表現オブジェクト（地図）をベースに、シャンゼリゼ、凱旋門とナビゲーションを行っている

### 第3章 抽象化によるハイパーメディアの論理設計

る。また、シャンゼリゼの関連実体テーブルを表示して、シャンゼリゼ近傍にある凱旋門以外の建物も一覧することができる。たとえば、「チュイルリー公園」、「コンコルド広場」などである。

図3-5は、関連・実体図、is-a階層図を用いたスキーマナビゲーション例である。画家にまつわる関連実体や属性を見たり、画家以外のタイプの人物にまつわる情報をナビゲーション行ったりすることができる。図3-5では、スキーマブラウザの1つとして関連・実体図を用い、ゴッホ作品をナビゲーションし、「オーベールの教会」を提示した例である。関連・実体図では、関連する実体情報だけでなく、フォーカスの属性（この場合は、「流派」）も同様に表示している。

一方、このような観光案内ハイパーメディアシステム開発を通じての課題は、変化の激しい構造変更に追従しにくいことである。観光案内の応用システムは、一般的には、構造変化の少ない応用である。また、仮にあったとしても、追記型の対応で処理できるケースが多い。このような状況では、課題が顕在化しにくかったが、対象データの更新が多い場合までを含む統合的な対応については、今後の課題である。

第3章 抽象化によるハイパーメディアの論理設計

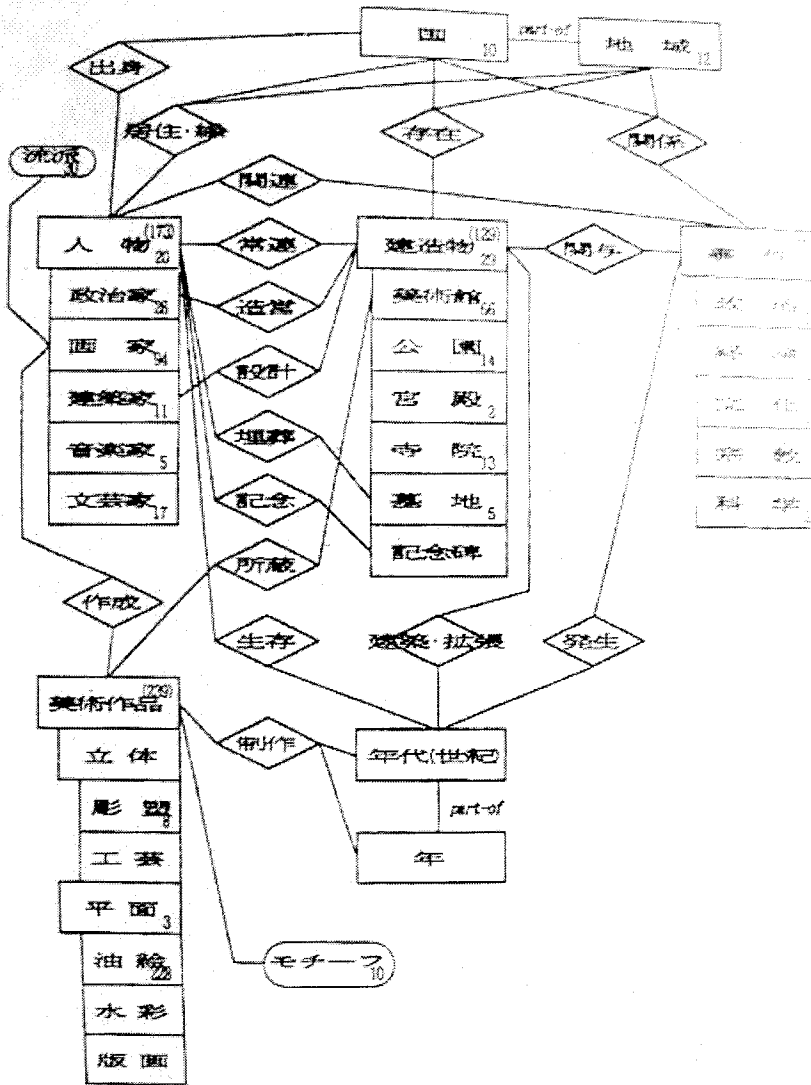


図3-3 パリ観光案内ハイパーメディアの拡張 E-Rモデル



### 第3章 抽象化によるハイパーメディアの論理設計

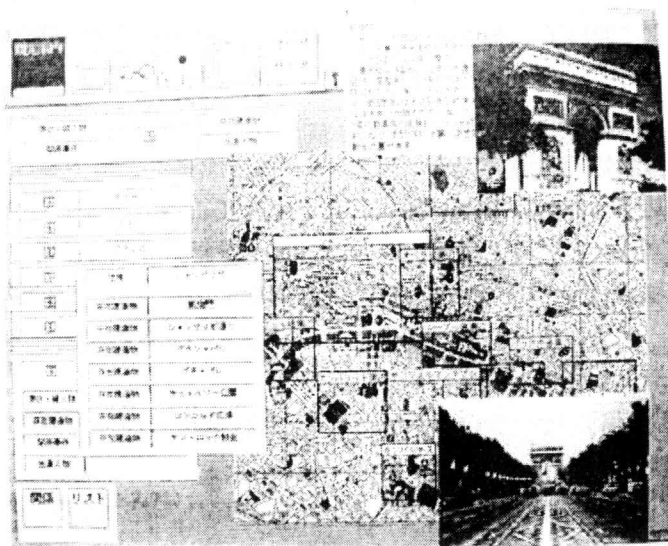


図3-4 静的リンクを用いたナビゲーション例

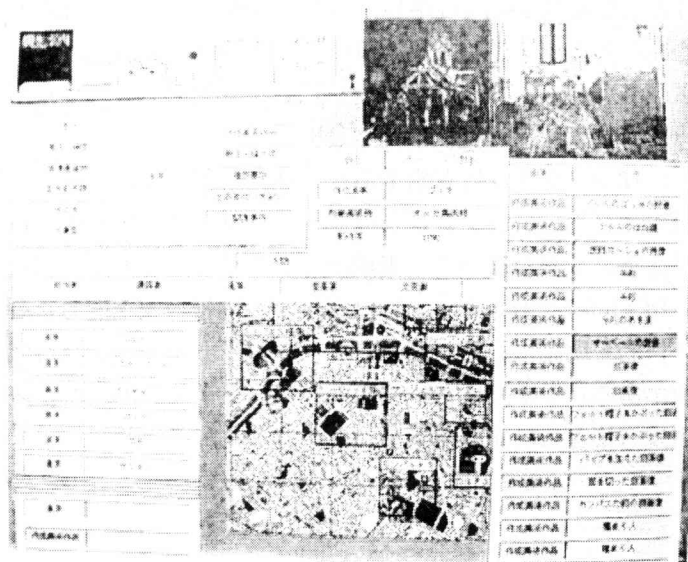


図3-5 スキーマブラウザを用いたナビゲーション例

### 第3章 抽象化によるハイパーメディアの論理設計

上述したように、トップダウン的な論理設計方法論については、運用上の課題も顕在化されつつある。拡張 E-R モデルを基本とするハイパーメディア論理設計を用いたシステム開発において、トップダウン的な論理設計方法論の今後の課題を以下にまとめる。

#### (1) メディア分散処理環境の実現

本システムの基本的な考え方は、各メディアの特性に応じた機能処理分散をはかることにある。概念系と表現系情報における処理・操作の分離はその一例である。この結果、マルチメディアデータの取り扱いが抽象化され、モデルの単純化がはかられた。今後は、この考え方を発展拡張させ、メディア分散処理環境の実現をはかり、性能の向上や、システム構築の便宜性をはかる。

#### (2) 大規模化処理

数万ノード以上の大規模ハイパーメディアへの適用のためには、ハイパーテキストクラスタリングの技法を援用し、適正規模にグルーピングする処理や帰納的にスキーマを規定していく処理が必要となる。

また、キーワード、画像インデクシングの自動化技術を援用して、リンクの自動生成、自動メンテナンス実現する必要がある（Web マイニング技術）。

また今回、スキーマブラウザとして、E-R モデル上の関連・実体図、is-a 階層図、is-part-of 階層図表示を実現した。しかし、より大規模なデータの認識・把握のためには、これらを統合したオーバビューブラウザが必須となる。

#### (3) ナビゲーションガイドの提供

本システムでのリンクタイプは豊富であり、種々の関連づけ表現能力は高い。したがって、一般のナビゲーション、プレゼンテーション型応用ハイパーメディアへの適用が可能である。しかしながら、表現能力が高くなった半面、利用者に対するアクションの負担も同時に増大する。したがって、情報の表現能力を高くするだけでなく、意図・目的を明確にしたナビゲーションガイドのような支援機能も合わせて提供する必要がある。シナリオベースのハイパーメディア[OgHK90, OgHa91]との統合も、その実現方策として重要である。

### 3. 5 関係抽象化によるハイパーメディア論理設計

前節では、拡張 E-R モデルを用いたハイパーメディアの体系的設計について説明した。本節では、この考え方をさらに発展し、関係の情報表現に対するスケーラブルな設計指針— 関係抽象化 —について言及する。関係抽象化の考えかた自体は、2章で説明した Hypertext Aggregation の概念と同じである。しかし、2章ではその考え方をハイパーテキスト、ハイパーメディアの物理設計の効率化に適用したものであるのに対し、本章では、関係抽象化の考え方をハイパーテキスト・ハイパーメディアの論理設計に提供し、設計時・編集時の効率化に寄与する。

#### 3. 5. 1 データベース設計における抽象化の考え方（従来のデータモデル設計）

データベース抽象化(Database Abstraction)の考え方は、実世界の意味情報をコンピュータ上の処理可能な形式に変換する際に、基本的なものである。同様の考え方は、人工知能やソフトウェア工学における設計技法にも存在する。ここでは、従来の設計手法の代表例としてデータベース抽象化の考え方を概観する。

##### (1) Aggregation

Aggregation は、実体（オブジェクト）に関連する属性を、より上位の属性として抽象化する操作である。たとえば、関係モデル(Relational Model)では、表の列にある複数の属性を統合してまとめる操作の総称をさす。言いかえれば、属性のタイプがより複雑な構造まで含めて対処でき、利用者は、その中の任意のレベルの粒度で抽象化された属性を利用できる。具体的には、属性は、“is-part-of”関係を用いた抽象化階層として構成される。複合されたオブジェクトに対しては、Composite Object として表現・操作される。また、構造モデル(Structural Model)では、Ownership Relation が Aggregation とみなすことができる。

通常、抽象化階層では、下位のレベルに位置する実体の属性がその上位のレベルに位置する実体の属性から継承されるという継承関係は存在しない。しかし、下位のレベルの属性からボトムアップ的に上位のレベルに当該属性が利用されるという関係(property inheritance)が存在するケースもある。

Aggregation の関係は、複数の異なった利用者が、各々の興味のレベルに応じて情報活用できるという点で有用である。また、情報利用時だけでなく、情報編集時の

効率的設計にも寄与する。しかしながら、この抽象化の考え方は、主として、E-Rモデルにおける実体(Entity)に適用することを想定している。

#### (2) Generalization

Generalization は、複数のオブジェクトが、基本オブジェクト (Generic Object) として、共通の属性を持つ場合の抽象化である。たとえば、関係モデルでは、表の行の複数の要素 (tuple) の共通の属性をまとめて統合するものである。また、オブジェクト指向モデルでは、この抽象化操作は、Superclass - Subclass 間の関係に相当する。複数の行の情報は、"is-a"関係を用いた Generalization 階層構造として表現される。Generalization の逆操作は、Specialization である。また、構造モデルでは、Subset Relation が、Generalization とみなすことができる。

Aggregation と比較した際の Generalization の特徴は、Generalization 階層構造においては、トップダウン的な属性の継承機構が存在することである。すなわち、上位クラスの属性は、共通属性として、下位クラスの属性として継承利用される。もし、属性名が両者で同じであれば、下位クラスの属性値が用いられる。

Generalization の利点は、高次構造の意味表現を効率的かつ簡潔に表現できることである。情報の利用時、編集時双方でメリットがある。しかし、Generalization も Aggregation 同様、E-R モデルにおける Entity に適用することを想定したものであり、直接関係に適用することを想定していない。

#### 3. 5. 2 Database Abstraction の例

映画データベースを例に、Database Abstraction(Aggregation, Generalization)の適用例を説明する。図3-6は、Relational Model における抽象化の一例である。また、他の抽象化タイプとしては、Version と Time Abstraction とがあげられる。特に、CASE、CAD 応用では有用である。本論文では、ハイパーメディアを対象としているので、対象の範囲外とする。

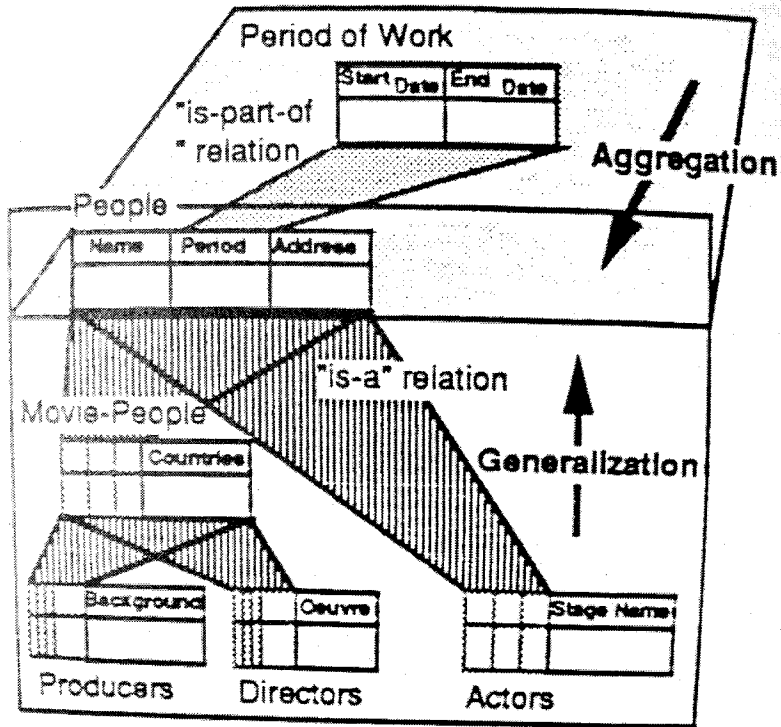


図3-6 Aggregation と Generalization



図3-7 映画と俳優・女優との関係 (多対多)

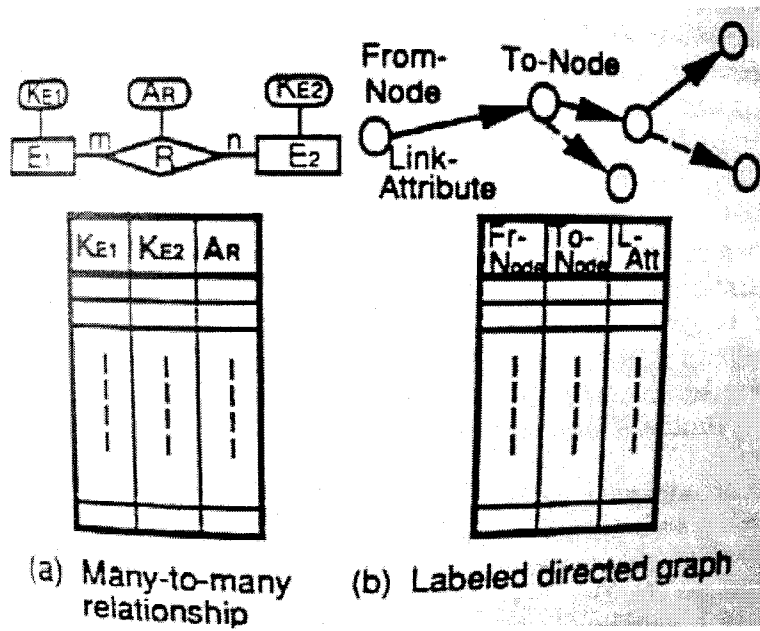


図3-8 関係の一般的属性

#### 3. 5. 3 関係の抽象化設計 (Relationship Abstraction)

本節では、E-R モデルにおける抽象化の考え方を関係(Relationship)にまで拡張した設計指針について説明する[HaKG91-2]。このような抽象化の考え方が規定できれば、ハイパーメディアの関連 (Association) への適用も可能である。これは、E-R モデルにおける自己完結関係表現(Self Relationships)の場合に相当する。

関係表現の抽象化の目的は、過度な関係表現を防ぐことである。過度な数の関係表現が存在すると、ナビゲーションに際して、見通しが悪くなり、情報空間の迷子問題を一層助長する結果となる。また、他の効用としては、関係自体に着目した抽象化を行うことにより、関係表現における高次の意味構造を抽出することができる。このような高次意味構造が抽出できれば、構造を考慮した種々の問い合わせ操作が可能となり、結果として高品質な情報検索・ナビゲーションを提供することができる。

まず、一般的な関係の特性について考える。関係の特性は、実体の特性と比較すると、以下のような差異がある。

- (1) 関係モデルにおける関係表の表現においては、キー（一意にタプルを規定）は、2つの外部キーの組み合わせである。

関係の表  $R_R$  は、基本的に下記の3つタイプの属性からなる。

- － 対応する2つの実体表の各々の外部キー (Foreign Key)
- － 関係の属性（もしくは、関係の抽象化属性）

すなわち、2つの外部キーが存在することにより、関係表は、対応する実体表における更新に影響を受ける。また、この情報表現は、一般には、ラベルつき有向グラフの情報表現に相当する。

- (2) 相対的に関係表における属性の数は少ない。

実体表と比較すると、一般に、関係表における属性の数は少ない。したがって、Functional Dependency のような意味的制約関係を関係表に直接適用するのは、困難である。

- (3) 関係表にあるタプル（行）の数は、相対的に大きい。

関係表にあるタプルの数は、対応する実体表のタプルの積に比例する。

従来の Database Abstraction で関係に対する取り組みがあまりなされてこなかった背景としては、関係に対する抽象化そのものが、大規模ハイパーメディアや Web の到来以前は、あまり有用性が見出せなかったからである。関係に付随する属性があまりなく、抽象化そのものの概念を適用できる場面が限られていたためである。また、上述したように、関係表の特性が、実体表の特性とは異なり、種々の制約がある。このため、概念そのものを単純には一般化できなかつたことにも起因する。例外のケースとしては、関係表における行の分離のみに主眼がおかれてきた。

#### 3. 5. 4 関係の抽象化に対する要件

関係表現の設計とメンテナンスを効率的に行うためには、下記に述べる3つの要件を満たす必要がある。まず、本質的要件としては、記述表現の単純化である。記述表現が逆に複雑かつ量的に多くなってしまつては本末転倒である。これは、単なる情報の分離以上の意味をもつ。第二に、どんな操作に対しても、分解構造にたいする操作が不変なことである。逆に、最構成された情報は、オリジナル情報に比べて、冗長性も無矛盾性もないことが要件である。第三に、抽象化された各構造に意味的名前づけが行えることである。この要件は、関係抽象化の操作が、実世界の意味ある処理と対応づいていることに相当する。

##### (1) Augmentation

関係属性やリンクタイプには、明示できる意味(Visible Meaning)と暗示される意味(Invisible Meaning)とが存在する。一例として、俳優・女優とその配役のデータベースの設計について考える。両者の関係は、一般には多対多の関係であり、関係名は、演技する(“casts in”)である。しかし、背後に種々の暗示された意味がある。ある俳優や女優が配役された背景には、種々の理由があり、たとえば、俳優・女優自身の所属組織、監督の選好等々がある。これらの属性は、多くの場合、実体間の関係に付随する。

ここで提案する Augmentation は、図3-9に示すように対応する実体間の属性を介して関係を対応づける関係抽象化操作である。これは、関係を分解し、属性間のアグリゲーション操作を抽出することである。すなわち、関連(Relationship)に対応する実体の属性に置き換えて、同等の関係情報を実行させる。このような代替表現とその操作により、一層コンパクトで見通しのよい関係表現を実現することができ



る。これは、実世界の関係表現は、多くの場合、かなり相関の強い関係で規定されていることに起因する。

このような実体に対する属性とその属性値の規定に際しては、2章で説明したACE クラスタリングもその一実現手法である。関係表が、関係のクラスタリング操作により、実体の新規属性として表現される。実体の新規属性に対する属性値のパターンは、一般には、オリジナルの関係表現に比べて、かなり少ない。一旦このような属性と属性値とが抽出・付与されると、もとの関係は、抽出属性間でのJoin演算で再構成できる。このような属性が規定されて、意味のある対応づけがなされれば、オリジナルな関係自体になんらかの理由づけを見い出すことが可能となる。以下、Augmentationとしての可能なパターンを整理して列挙する。

#### (A) Translated Attribute Types

ある可能な変換としては、関係情報が、実体の既存の属性 (Visible Attributes) に変換できることである。すなわち、対象関係と、対応実体のある既存の属性とに強い相関がある場合である。もしこのような既存属性が存在しない場合には、関係が表現されるための別の意味があると仮定し、新規属性(Invisible Attributes)を規定する。このような属性は、実体の要素のクラスタリングにより規定可能である。

#### (B) Mapping Patterns

関係と実体の属性との間には、3種類のマッピングパターンが存在する。第一は、Overlap Mapping と呼び、関係の要素と対応の実体属性の値とに多対多の関係が存在する場合である。第2は、Non-overlapping Mapping with Exception である。これは、ある種の例外を考慮すれば、関係要素と対応の実体属性値とに多対1の関係で実現されるものである。第3は、Recursive Non-overlapping Mapping と呼び、関係要素と対応の実体属性値との間に複数の多対多の関係が存在する場合である。

#### (C) The Relationship Among Attribute Values

同一実体の複数の属性が関与して、関係を表現する場合がある。このような場合は、関係抽象化を行うこと自体が複雑化することになるため、除外する

### (2) Globalization

第二の関係抽象化操作は、関係表の適用範囲の程度によって規定される抽象化操作である。すなわち、ある種の関係は、全体構造として捉えることができ、別の関係は、局所的構造として捉えようとするものである。また、局所的構造は、全体構

### 第3章 抽象化によるハイパーメディアの論理設計

造で全てを表せないときの補完的、補償的構造としてみなすこともできる。このような認識をすれば、関係の全体一局所階層構造を抽出することが可能となる。これは、前節で示した実体に対する Generalization-Specialization 階層構造と対比できる。

Globalization は、図3-10に示すように、全体一局所の関係構造を抽出する操作である。これは、関係表のタプルに対するグルーピング操作であり、実体表に対するタプル水平操作(Generalization)に対応する。また、上位クラスの関係の属性が継承されるので、下位クラスの関係表に具体的なオーバーライドする記載がない限り、冗長な記載をおこなわなくて良い。

関係表に対する Globalization と実体表に対する Generalization との違いは、下記の2点である。まず、Globalization における属性は、不変であり、属性値のみが更新される。これは、全体一局所階層関係の構造が、例外を表す構造をもとにした階層関係になっているためである。また、Globalization のインスタンスは、Generalization とは異なり、どのレベルにも存在しうる。(Generalization の場合は、最下層のレベルのみに現れる。)

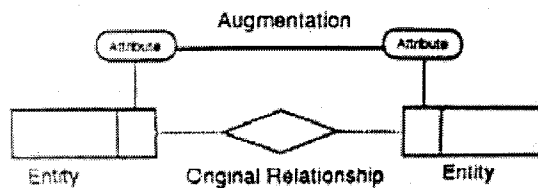


図3-9 Augmentation

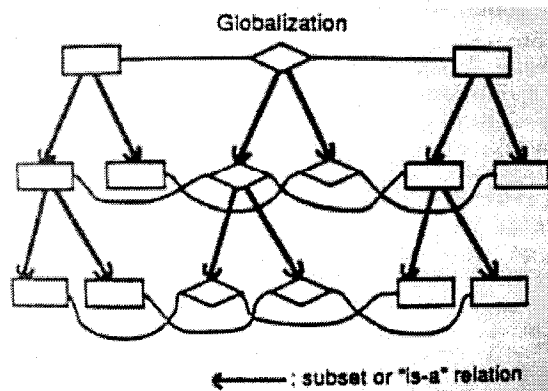


図3-10 Globalization

### 3. 5. 5 ハイパーメディア論理設計手順

本節では、関係抽象化の概念を援用したハイパーメディア構造設計について言及する。

#### (1) マクロフレームワーク設計と素材抽出

グローバルノード設計によるマクロ構造の設計がまず重要である。このフェーズでは、ある種のインプリシットな意味的制約を規定することに相当する。また、全体フレームワークの中で、どのノード、リンクやパスが重要で焦点をあてるべきかの判断を行う。このようなプロセスは、作成者・編集者の意図規定や利用者へのガイド支援として表現するものである。このような全体概要レベルでのシナリオ検討フェーズがなければ、全体として有機的な質の高い情報共有を行うことが困難となる。

#### (2) ノード設計

次に個々のノード設計を行う。このフェーズでは、データベース抽象化の考え方が有用である。また、“is-a”関係による階層化が有用である。ハイパーメディア設計では、対象情報を表現するノードとそのアクセス手がかりとの対応づけが必要である。

#### (3) リンク設計（ボトムアップ設計）

次に、ノード間のリンクを規定する。リンクタイプとしては、静的リンク、動的リンク、バーチャルリンク等がある。大部分のリンク設計はボトムアップ型である。リンクは、まず、局所的に整合性がとれる形態で設計規定される。その結果として、全体構造としてのハイパーメディアが構築される。このリンクづけ方策は、ハイパーメディア設計・更新の本質的なプロセスである。

#### (4) 全体制約との整合性検査

一般によい設計プロセスを実現することは、どのフェーズからでもフィードバック評価が行えることである。特に、ハイパーメディア設計の場合、マクロな全体フレームワークと個々のノード・リンク設計との比較、評価が重要である。このようなプロセスが、ハイパーメディアの利用を促進させる。しかしながら、特別のツールがない状況では、このような全体構造の制約と個別構造との比較、評価は、作成者の設計スキルに大きく依存する。過度のリンク構造になることを防ぐためには、ルールを用いたり、遷移律を用いたりして冗長なリンクを削除すること

も可能である。しかし、リンク設計におけるよい設計基準を示す効率的な方法は無い。

#### (5) 再構成

最後のプロセスは、得られた整合性検査の結果をもとにした再構成のフェーズである。このような再構成プロセスにより、いままでは、健在化していなかった重要な関係等を明示し、作成者、編集者の意図をより明確に規定して情報提供することが可能となる。このようなハイパーメディア設計は、構造がより大規模になり、また、利用者のアクセス頻度が一層多くなるにつれ、より重要になるものである。

関係抽象化手法は、上記5つのステップにおいては、特に、(4)の評価のフェーズで有用である。関係の抽象化技法により、マクロフレームワークと個々のリンク設計との対比を効率的に行うことが可能となる。この場合は、どの箇所が異なつて、どのような改造を行えばよいのかの指針を与えてくれる。また、設計者が、マクロ設計時に特段の制約を与えていない場合にも有効である。

3. 5. 6 関係抽象化による論理設計の評価と課題

上述の記載は抽象的レベルであったので、具体的に映画データベースに適用して、関係抽象化の具体的な意味を説明する。図3-11はもとの映画データベースに対し、関係抽象化の考え方を導入して設計したスキーマ例である。すなわち、もとの映画データベースのスキーマは、各実体表（映画と俳優・女優）に付随する新規属性とそのサブクラスとして分解表現できる。

スタンフォード大学で管理運用しているデータベースに適用し、評価を行った。これは、6,250 の映画、5,300 の俳優・女優、22,000 の配役、1,450 の監督からなる（評価時）。映画のカテゴリと俳優・女優のカテゴリ、両カテゴリの組み合わせ、並びに、例外的組み合わせの表現により、意味的対応付けの可能でかつ関係構造が単純化された関係構造を見通すことが可能となった。たとえば、アクション映画群とアクション映画に出演する俳優グループとの関係などである。

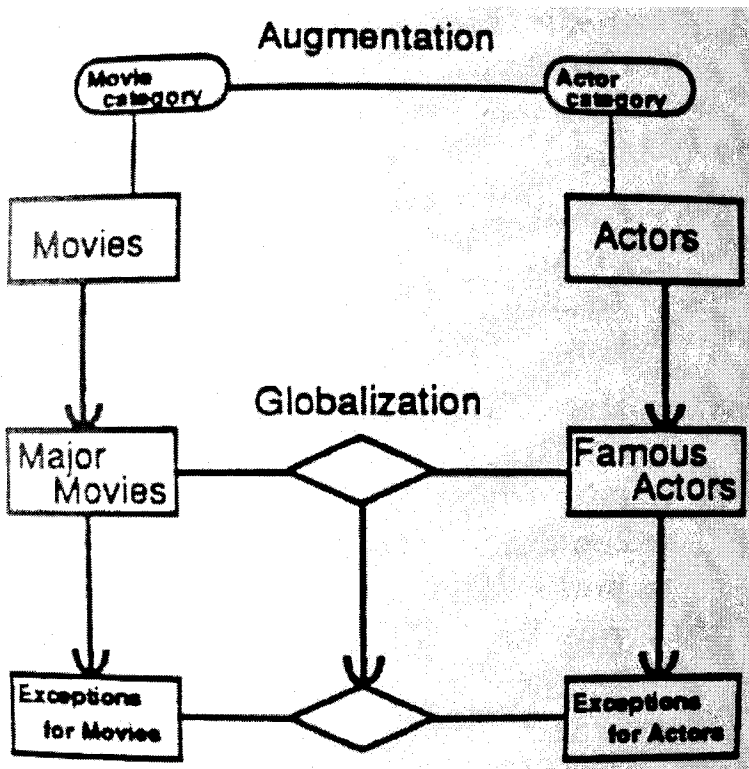


図3-11 関係抽象化の例

#### 3. 6 言語規約によるデータベースからのハイパーテキスト生成

前節では、拡張 E-R モデルによるスキーマ設計を明確にしたハイパーメディアの設計について説明した。このような設計ができれば、以降のナビゲーションやシステムの拡張時に効力を発揮する。本節では、このような考え方をさらに発展させて、スキーマ設計された情報をもとに、ハイパーテキストのリンク情報を効率的、体系的に生成(Hypertext Projection)することについて説明する[HaBo94]。

一般には、ハイパーメディアのオーサリングとブラウジング双方のプロセスを効率化することはトレードオフの関係にあり、相応のコストが必要である。しかし、大規模なハイパーメディアの構築・運用に関しては、オーサリングとブラウジング双方のコストを効率化することが求められる。近年、企業内での情報共有、意思決定支援手段として必須となっている企業内ポータルシステム、知識管理システムでは、特に顕著である。

本節では、データベース管理技術とハイパーメディア情報検索技術とを統合させ、データベース管理の情報空間から、ハイパーメディア情報を自動生成するアプローチについてデータモデル、操作言語面から説明する。このようなアプローチの特徴は、

- (a) ハイパーメディアに対する良構造設計が導入できる
- (b) ハイパーメディアのノードとリンクとが自動生成できる
- (c) アプリケーションの編集、更新が効率的、体系的に行える
- (d) 情報検索(問い合わせ)メカニズムが改善される
- (e) 複数のアプリケーションに対するデータ共有が促進される
- (f) ハイパーメディアデータの冗長性、矛盾性、セキュリティ、信頼性等の諸機能が改善される

##### 3. 6. 1 ハイパーメディア DB の設計指針

企業の社内外向けに情報発信される企業ポータルや、各種 EC サイトでは、大規模かつ更新の多い Web 情報をいかに高品質に保ち、情報提供できるかが必須の要件である。このようなミッションクリティカルな用途では、Web コンテンツ管理をより一層、体系的に行う必要がある。このような大規模ハイパーメディアの効率的編集・利用には、アドホック的ハイパーメディア設計から、より構造的設計への

### 第3章 抽象化によるハイパーメディアの論理設計

転換が不可欠である。このような構造的設計の有益な実現形態として、データベースにおけるスキーマとスキーマ言語の援用があげられる。ハイパーメディア編集への応用としては、まず、データベーススキーマ設計に基づくデータ・コンテンツ管理、次にテンプレート型設計言語によるハイパーメディアアプリケーションの生成（射影）という2段階のプロセスにより適用を考える。

データベースモデル活用のメリットは、データ管理に体系的な手法が活用できること、コンテンツ・データ自身の管理とその表現形態（ハイパーメディアは1つの表現形態、メタファーと考えられる）を分離して設計・利用できることである。さらには、

- 整合性のある Web 情報ページレイアウト設計の実現：同じタイプに属するデータに対しては、同じレイアウトスキームを適用できる（たとえば、EC サイトの商品情報提示等）
- 自動リンク付与：データベースにおける属性情報をもとにしたオブジェクト（タプル）間の関係は、間接・暗示的な関係(implicit relationship)である。これに、制約・ルールを考慮したアプローチにより、直接・明示的な関係(explicit relationship)を生成できる。自動リンク生成・付与は、人手によるアドホックなリンク付与に比べて、リンクの張り間違えや、ジャンプ先のないリンク(dangling link)になることを防ぐことができる。
- 編集・更新の簡便化：表示ページもテンプレートにより自動生成されるので、編集・更新の際には、一括、体系的に実行することができる。また、リンクもリンク定義を用いて自動生成されるので、追加、削除が矛盾なく実行できる。
- 柔軟な情報表現形態への対応：ハイパーメディアのリンク構造は、固定的、普遍的なものではなく、誰が、どのような目的で提供・利用するかにより柔軟に変化するべきものである。たとえば、作成者・編集者の提供する意図をもったリンク構造と、利用者が欲する別の観点・目的からのリンク構造とは、必ずしも一致しない。このようなハイパーメディアにおけるビューの考え方の導入は、データベース、ハイパーメディア双方のデータモデルを統合することのメリットである。

### 第3章 抽象化によるハイパーメディアの論理設計

Literalists と Virtualists との融和：ハイパーメディア学術領域では、Literalists 学派（リンク構造は、直接的な関係で規定され、ナビゲーションは、編集時に規定されたリンクをトラバースすることにより実現）と Virtualists 学派（リンク構造は、コンテンツ間の間接的な関係で規定、もしくは計算され、ナビゲーションは、利用時にオンデマンドに生成されたリンクをもとに実現）とがある。両者は、各々、利点と問題点があるが、本提案手法では、両者の利点を生かした統合、融和をはかることができる。たとえば編集者は、作成時に、「すべての 17 世紀の画家を誕生日順にソートして、一括してリンク付与」を行ったり、「政治家とその政治家が関与した事件・イベントとに一括してリンク付与」をおこなったりすることができる。

#### 3. 6. 2 Hypertext Projection

Hypertext Projection という考えかたを導入する（図3-12参照）。これは、間接的な関係を表す relation から、直接的な関係を表す link を抽出するものである。Hypertext Projection は、以下の3つのステップからなる。

##### (1) Relation の規定

Relational model による関係表現を行う。Relational model を用いる理由は、間接的な関係をシンプルかつ完備に表現できるためである。ここでは、relational table における tuple を1つの基本ノードとみなし、tuple に付随する attribute(属性)を用いて間接的な関係を表現する。

##### (2) Relation からの Node の規定

relational table や tuple から対応する Node を規定する。Node は、言いかえれば、relational table や tuple の仮想化 (virtualization) である。すなわち、tuple については、具体的実体として表現するものに対応し、relational table 全体については、tuple の全体ビューとして表現する。

##### (3) 制約を規定することによる Link の規定

最後のステップでは、生成された Node 間の直接的関係 (Link) を生成する。生成に際しては、始点 Node 集合の規定、終点 Node 集合の規定、両集合間の制約 (クエリによる規定) を行う。



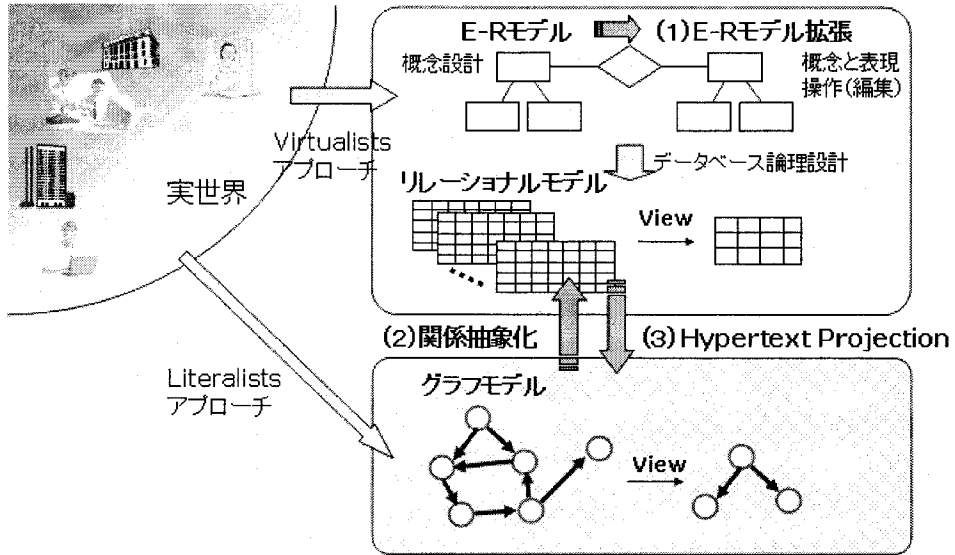


図3-12 Hypertext Projection の概念図

### 3. 6. 3 Hypertext Projection 変換言語 (HPTL: Hypertext Projection Translating Language)

本節では、Hypertext Projection 機能を実現するための変換言語 (HPTL) について説明する。これは、relational model で表現される情報を hypertext model で表現される情報に射影する操作である。変換言語の規定においては、SQL を基本とした言語体系により規定する。例として美術史 DB をとりあげ、画家、年代、作品等の関係を示すハイパーテキスト構築を目的に説明する。

#### (1) Node 生成の一般的構文

まず、Node 生成の一般的構文は以下の通りである。

```
CREATE NODE [<Relation>]
[SELF:
  [NAME = <string>];
  [TEMPLATE = <template-name>];
  [ASSOCIATE <attribute-commalist>
    <field-commalist>]];

[CHILD:
  [NAME = {<string> | attribute} ];
  [TEMPLATE = <template-name>];
  [ASSOCIATE <attribute-commalist>
    <field-commalist>]];
```

ここで、[]内のアーギュメントはオプションで、<>内の情報が適宜埋められる。また、|で分離された{}内のアーギュメントは、その中の1つが選択されるものとする。"Relation" とは、relational model の Relation であり、"string"とは、任意の文字列を表す。"template-name"とは、template の名前であり、"attribute", "field"は、各々 relation で規定された属性と、template で規定されたフィールドをさす。また、"commalist"とは、コンマで区切られた複数の要素のリストを意味する。"attribute-commalist"と"field-commalist"との要素の数は同じである。

このような構文を用いると、以下の2種類のタイプのノードが生成される。

Composite node: 所与の relation から直接規定される複合ノード

### 第3章 抽象化によるハイパーメディアの論理設計

(上記 SELF で記述された箇所)

Basic node: 当該 relation に属する tuple から規定される (複数の) ノード  
(上記 CHILD で記述された箇所)

もし、SELF.NAME が省略されていれば、この名前は、"relation"と同じになる。また、TEMPLATE が省略されていれば、当該ノードはノードの存在のみを意味し、非表示でブラウズできなくなっている。さらに、ASSOCIATE が省略されていれば、"attribute"と"fields"との間接的関係を示している。

以下に例をあげて説明する。Painter の relation で、name, birth, death, photo, biography が属性であるとする。この relation から Composite node としての "Painter"とそれに属する、各 tuple を代表する Basic node と以下の言語表現で生成する。もし、最初の Painter relation の tuple 数が 10 であれば、合計 11 の node(Composite node を含む)が生成される。

#### 例 : Relation "Painter"からのノード生成

```
// Create node from relation Painter
```

```
CREATE NODE Painter
```

```
SELF: // Composite node
```

```
TEMPLATE = "index.temp"; // will be an index
```

```
ASSOCIATE = (name, birth), (name, date);
```

```
CHILD: // Nodes from tuples
```

```
NAME = name;
```

```
TEMPLATE = "painter.temp"
```

```
ASSOCIATE = // Relation → Template
```

```
(name, birth, death, photo, biography);
```

```
(name, born, died, picture, description)
```

また、特定の集合のみを表現したい際 (たとえば、フランスの画家に限定) は、データベースにおける View の操作を用いて規定する。

```
// Creates a view FPainters for the database. For convenience uses the same names as the painter
```

### 第3章 抽象化によるハイパーメディアの論理設計

```
// template attributes
```

```
CREATE VIEW FPainters (name, born, picture, description)
As SELECT Painter.name, Painter.birth, Painter.death, Painter.photo, Painter.biography;
    FROM Painter;
    WHERE Painter.country = "France"
```

```
// Create nodes from the view
```

```
CREATE NODE FPainters
    SELF:
        TEMPLATE = "browser.temp"; // Graphical browser
    CHILD:
        NAME = name;
        TEMPLATE = "painter.temp"
```

このような表記法で、表現すべきノード集合が生成される。図3-13に示すように template の一例である "painter.temp" からノードが生成される。Web ページでは、各 tuple に相当する情報が、ページ内のスロットとして表現される。また、図3-13に示すように、"browser.temp" の template から、具体的な graphical browser を生成する。しかし、このフェーズでは、ノードのリストアップまでであり、どのようにナビゲートするリンクを作るかは、次のステップである。

#### (2) Link 生成の一般的構文

Link 生成の一般的構文を以下に示す。

```
CREATE LINK [link-name]
    SOURCE:
        NAME = <node-name>;
        IN {SELF | CHILD>;
        [ ANCHOR <field> ];

    DEST:
        NAME = <node-name>;
        IN {SELF | CHILD>;
        [ ANCHOR <field> ];
```

```
DIRECTION {FORWARD|BACKWARD|BIDIRECTIONAL};  
[ WHERE <constraint-list>;
```

ここで、"link-name"はリンクタイプを記述する。SOURCE と DEST は各々リンクの始点と終点のノード名である。もし、IN 節において、CHILD が規定されれば、ノードの下位ノードにリンクが付与される。そうでなければ、当該ノードにリンクが付与される。ここで、終点ノード (Destination) は、ANCHOR を有する。ANCHOR とは、テンプレートの中のどのフィールドがリンク先なのかを指定するものである。これは、リンクの方向が、BACKWARD もしくは、BIDIRECTIONAL の時に必要である。WHERE 節は、リンク指定の制約を規定する。また、WHERE 節には、ノードの全てのノードの属性 (SOURCE.name) を用いることもできる。

#### 3. 6. 4 Hypertext Projection の評価と展望

##### (1) 画家ハイパーメディアへの適用と評価

一例として、画家のハイパーメディアへの適用を考える。たとえば、全ての「Impressionists (印象派)」から「Post-impressionists (後期印象派)」への関係をリンクとして表現する。Link の属性を BIDIRECTIONAL とし、「影響を与えた」、「影響を受けた」の意味でのトラバースを可能とする。この場合の記述を下記に示す。

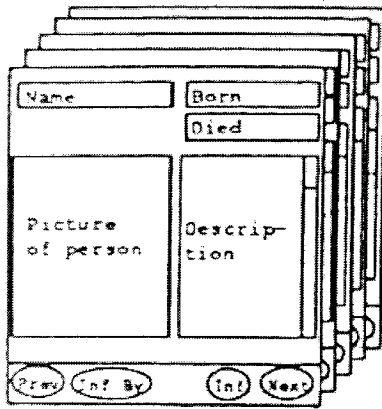
```
CREATE LINK Influenced  
SOURCE:  
  NAME = FPainters;  
  IN CHILD;  
  ANCHOR "Inf"  
DEST:  
  NAME = FPainters;  
  IN CHILD;  
  ANCHOR = "Inf By"  
DIRECTION BIDIRECTIONAL;  
WHERE SOURCE.school = "Impressionism",  
      DEST.school = "Postimpressionism"
```

### 第3章 抽象化によるハイパーメディアの論理設計

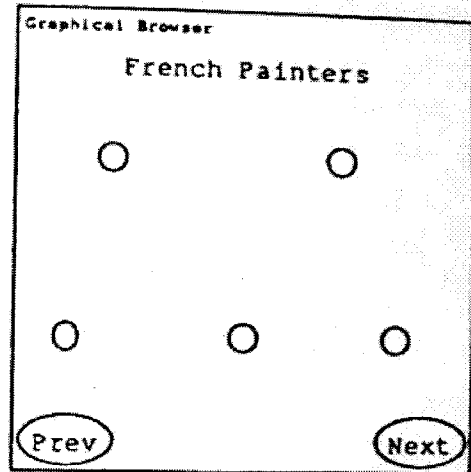
ここで、リンクは、1対1である必要はない。多くの場合、1対多の関係となる。"Inf"のボタンをクリックした際、どのノードにまずジャンプするのかなどの問題は、ユーザインタフェース上の問題である。一例として、可能なすべてのジャンプ先をリストとして表示することも可能である。

このように、ノードとリンクの定義により、FPainter ノードを使ってブラウジングを行うことができる。また、この際、全てのノードがリンク可能である必要はない。また、"next"ボタンを指定することにより、全ての対象となる画家を、たとえば、誕生順にブラウズすることができる。(図3-13における点線のリンクに相当。)このようなソーティングは、データベースのビューにより指定することができる。すなわち、"next"ボタンを ANCHOR として定義し、DIRECTION の値を FORWARD とし、制約として、WHERE 節に、SOURCE.next = DEST (ここで、next は、ノードの間接的な属性として指定) とすれば、規定できる。さらには、first (生誕順で最初の画家)、last (生誕順で最後の画家)、number (対象となる画家の人数、DEST のノードの数) が指定できる。

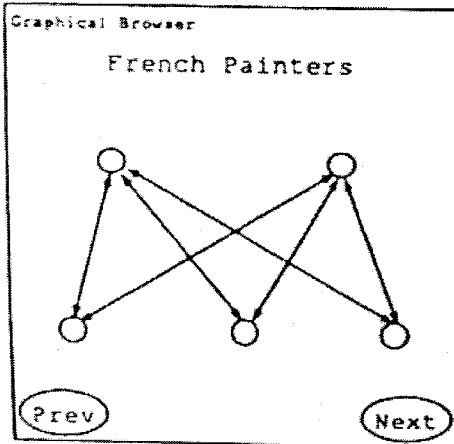
以上、述べたように、あらかじめテンプレートとして言語規約に基づいてデータベースからハイパーメディアへの自動生成が行えることは、当面の実現可能な研究アプローチである。かつ、現状の技術開発で精度よく実現できる策といえる。



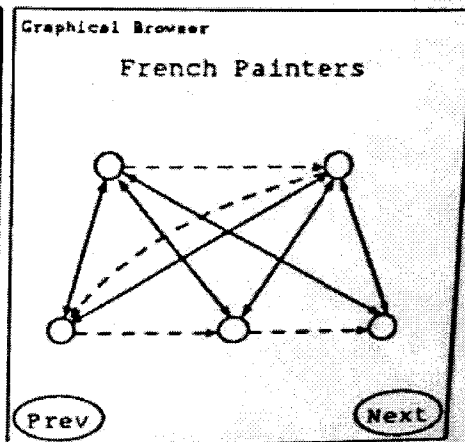
(a) French painters' nodes are created.



(b) No links between nodes yet.



(c) "Influence" relation added. "Select all" constraint used.



(d) "Next" added. "Select one" constraint used.

図3-13 データベースモデルからハイパーメディアの生成

#### (2) Hypertext Projection の展望

本節では、ハイパーメディアの自動作成を支援する方式として、全体モデルプロセスと言語表現によるハイパーメディア生成を説明した。まず、通常のリレーショナルデータベースモデル (relational model, E-R model, etc.) によるデータモデル化を行い、次に、データベースによる概念空間での情報表現から、ハイパーメディアによる表現空間での情報表現への射影 (Hypertext Projection) の操作により、ハイパーメディアを生成するものである。これは、ハイパーテキストクラスタリング (ハイパーメディア表現情報から、属性を追加することによるデータベース的概念情報への変換) の逆の操作に相当する。また、データベースのビュー操作、ハイパーメディア内のアグリゲーション操作と組み合わせて、自然な形態でのデータベース操作とハイパーメディア操作とを融合することが可能となる。

また、ハイパーメディアへの射影操作に対しては、SQL 的な言語表現により、データベースからの概念レベルでの情報を、ハイパーメディアとしての表現レベルでの情報を生成する。ハイパーメディアへの変換言語 (HPTL: Hypertext Projection Translating Language) の特徴は、単なる変換だけでなく実行時におけるナビゲーションを支援する機能を提供することである。したがって、編集時のハイパーメディア自動作成効率化に寄与するだけでなく、ナビゲーション・利用時の機能拡張を提供することができる。このようなフォーマル言語表現によるアプローチは、次世代大規模ハイパーメディアシステムの構築に有用である。



### 3. 7 本章のまとめ

本章では、ハイパーメディア抽象化の概念を論理設計プロセスに適用させて、大規模なハイパーテキスト／ハイパーメディアの構築と情報共有化を行う手法と提案した。具体的には、下記の3つの観点から、抽象化によるハイパーメディアの設計表現手法について提案した。

#### (1) E-Rモデル援用したハイパーメディア論理設計とその利用

拡張 E-R モデルを基本とした構造化・ナビゲーション方式により、リンク品質が均質化され、リンク操作の多様化が実現された。また、処理・モデルの単純化を目的とした概念系と表現系処理との分離により、設計容易化が実現された。大規模データハンドリングのための高次構造情報を考慮し、観光案内ハイパーメディアプロトタイプを構築し、基本機能の有用性を確認した。

#### (2) 例外を含んだ関係の抽象化設計手法

E-R モデルの考え方をさらに発展し、関係の情報表現にスケーラビリティを与える方策を提案した。Augmentation と Globalization というハイパーメディア固有の操作を規定することにより、過度な関係表現を防ぎ、見通しのよいハイパーメディアシステム構築に寄与する。特に、関係表現において強い相関関係が見出される場合に有効な設計手法の考え方である。映画データベースからのハイパーメディア化に適用し、有用性を確認した。

#### (3) 言語規約によるデータベースからのハイパーメディア生成

対象コンテンツからのハイパーメディアシステムの自動構築をめざし、論理設計として構築した概念空間の情報を、どのようにして利用者のもつハイパーメディアのメタファーとしての表現空間へのマッピング法を提案した。具体的には、Hypertext Projection という操作を提案し、言語規約による実現手段を示した。

このようなハイパーメディアの論理設計の今日的意義としては、品質管理のできる Web 情報の作成、編集、メンテナンスに寄与していることである。今日では、人手のみに頼らず、体系的にリンク構造を構築し、かつ、リンク構造の分析、評価、メンテナンスを行う Web Content Management システムが実現されている。本章で提案しているハイパーメディアのリンク自動生成の考え方は、人手とコンピュータとによる処理を協調させて効率化をはかる半自動的スケーラブル化アプローチとして、生かされてきている

### 第3章 抽象化によるハイパーメディアの論理設計

# 第4章 大規模ハイパーメディア構造可視化とブラウジング

## 4. 1 ハイパーメディアの構造理解と可視化技法の必要性

本章では、全体構造把握のための視覚的表現(Visualization)への適用を説明する。ハイパーメディア構造の部分構造を、図式的に表現するグラフィカルブラウザは、有用である。ちょうど、旅行者が地図をたよりに旅行するのと似ている。オーバビューダイアグラムは、概略構造を表示するグラフィカルブラウザに相当するが、情報空間が大規模でハイパーテキスト構造の各ノードとリンクを同時に表示できないような場合に不可欠である[Niel90]。また、全体と部分との関係が同時に把握できるのも、情報空間の空間的認知に有用である。

構造視覚化の意義としては、直接表示による全体構造理解という目的が一義的であるが、加えて、目標とする情報コンテンツへ到達しやすくするガイドラインや、文脈(コンテクストコンテキストで、情報解釈の多義性・あいまい性を減少させる意義も重要となる。本章では、ハイパーメディアの大規模性への対処と、文脈を考慮した多義性・あいまい性への対処という要件を満たす2つの可視化方式について説明する。

本章では、まず、全体と部分との関係を同時に表現でき、かつ大規模なハイパーメディアにもスケーラブルに対応できる方式(Focus+Context Views)について説明する。これは、魚眼図的な表現手法のアナロジーとして、グラフ構造表示にその考えかたを援用したものである。ハイパーメディアにおける重要ノード(Landmark Node)を如何に規定するかが鍵となる。Focus+Context Viewにおける文脈とは、ある時点での全体構造のスナップショットを提示して、現在閲覧している情報の位置づけを明確化することである。これにより、情報空間の迷子問題や、情報利用時の多義性・あいまい性を軽減する。

また、多次元分類に基づくブラウジングにおける文脈情報を考慮した方式(トラバース)を提案、プロトタイプによる検証を行う。ここで、トラバースによる文

## 第4章 大規模ハイパーメディア構造可視化とブラウジング

脈とは、異なった観点を活用したナビゲーションというプロセスを重視し、多面的な理解を行わせることにより、情報理解の多義性・あいまい性の軽減をはかるものである。情報空間における羅針盤の役割のような機能である。これらの情報可視化技法を、応用により使い分け、ハイパーメディア、Web の情報理解を深めることをめざす。

## 4. 2 関連研究

Focus+Context View は、局所的な詳細情報と全体コンテキスト双方のバランスをとって複数の抽象化レベルを表示するオーバビューダイアグラムである。関連表示形態としては、Guided tour [Trig88] や、魚眼図表示(Fish Eye View)がある[Frun86]。魚眼図表示は、物理的構造の限られた表示オーバビュー表示を行うのに対し、Focus+Context View は、Landmark Node による論理的構造に着目した概要表示手法である。また、このような論理構造の概要表示手法は、データ構成法に着目した表示手法として、いくつか提案されている。たとえば、Perspective Walls, Document Lens, Table Lens, Hyperbolic Browser などである[AsSi99]。これらは、木構造等のデータ構造に着目した表示手法であり、本章で提案する方式は、Web 情報を含む大規模ハイパーメディアに対する概要表示手法という点で異なる。

また、本研究の表示技術は、複雑な大規模ハイパーメディアのネットワーク構造を如何に単純化し、意味ある構造を抽出するかを目的とする。ハイパーテキストクラスタリング技術はその実現要素技術である。文献[MuHa97]に記載されているアプローチは、ハイパーメディアネットワーク構造におけるトポロジーとテキスト情報双方を活用して、意味のある構造を Web から抽出している。本研究の特徴は、全情報空間を均質に解析するのではなく、特定の Landmark Node を抽出し、有益なハイパーメディア構造を抽出することにある。また、従来手法は、スケーラビリティの問題が解決されていないものも多く、提案手法は、このような問題点の解決を支援するものである。

一方、ハイパーメディアにおけるコンテキスト情報を活用した検索・ナビゲーション支援方式としては、是津らの Aspect 発見 [ZeKT04]、中島らのブックマークへのコンテキスト情報の応用 [NOST00]などの研究があげられる。本論文が、多次元分類構造を基本とした「観点」を導入しているのに対し、是津らは、主にキーワードの解析によるコンテキスト情報の分析利用であるが、ハイパーメディア型情報への適用として関連が深い。

### 4. 3 Focus+Context Views

Web における主要問題は、情報空間(Hyperspace)における迷子問題である。これは、利用者が情報空間をナビゲーションする際になにも制約がないことに起因し、いま自分がどこにいるのか、またどこにいけば希望の情報に到達できるのかがわからなくなる。いつの解決法は、情報空間の全体概要表示と、その中での現在点と到達点の明示を行うことである。同様に、利用者が URL を指定して、ある情報から別の情報へジャンプする場合にも、ノードの近傍の表示は有効である。

本節では、Web 文書に対するオーバビューダイアグラムの一例として、Focus+Context View を提案し、概説する [MuHa97]。これは、特定ノードへの焦点をあてた表示手法である。Focus View は、対象ノードと直接リンクされた近傍ノードとの表示を行うものである。リンクは、参照、非参照双方のリンクが存在する。また、Context View は、全体の概要コンテキストを把握するために活用する。これは、大規模構造になると全体表示が複雑かつ困難になるため、途中の重要なノード(Landmark Node)とそれらのパスのみに着目した表示手法である。Focus+Context View とは、両者の機能を兼ね備えた表示手法である。これは、ちょうど道に迷った人間が、まず近傍の状況を把握し、それから地理的に重要な建造物を確かめようとする作業に似ている。

他のオーバビューダイアグラム手法は、当該の情報空間を表示し、当該位置の把握と進むべき方向性の把握を支援する。しかし、Web 情報のようなレベルの大規模情報になると、全体情報を同じ情報密度のまま表示するのは、計算量的にも、認知的にも得策ではない。また、2次元や3次元表示を活用しても、その表示複雑度は解消されるものではない。また、グラフ構造の配置問題もノード数とリンク数とが大きくなるため、複雑になる。

ここに、Focus+Context View の意義がある。考慮すべき点は、Landmark Node の規定である。Landmark Node の特性としては、ある程度の数で、利用者が情報空間全体を把握できるような理解ができる必要がある。地理的なナビゲーションメタファーに準拠し、Landmark を順に理解しながら町を探索するようなアナロジーと同様である。また、通常 Landmark node の総数は少ないので、オーバビューダイアグラムが複雑になりすぎることはない。

### 4. 3. 1 Landmark Node の発見・規定

Focus+Context View を作成する際に、Landmark Node の規定が重要な役割を果たす。ここでは、Web 情報の利用者は、ランダムに Web をアクセスするのではなく、意味的にまとまった Web 局所情報(Web Locality)を基本に情報アクセスするものとする。このような Web 局所情報は、一般的には、物理的に共通な場所（同じ Web サイト、Web サーバ）に格納されている場合が多い。

Web 局所情報に対する Landmark Node の抽出に際しては、

- ハイパーリンク情報（ページ間リンク構造）
- アクセスパターン情報（一定期間どの程度の情報アクセスがあったかの程度）
- 階層構造情報（Web ページの階層構成情報）

の3つの情報を用いる。以下詳細に記述する。

#### (1) グラフ構造の結合程度

文献[MuHa97]にあるように、Landmark Node は、一般には、Landmark Node でないノードよりも結合度が多い傾向にある。これは、航空会社のハブと同様である。経験的には、ハイパーテキスト中に現れるキーワードと主題の関連程度と、リンク長2までのノード集合（Second-order Connectedness）とは高い相関がある。この知見を援用し、リンク長2以内の非参照ノードの数(Back Second-order Connectedness)も、良い Landmark Node を抽出する際の指標となる。また、リンク長1の参照ノード数、非参照ノード数も Landmark Node 発見に活用する。

#### (2) アクセス頻度

グラフ構造の結合程度だけでは、静的なグラフ構造の解析のみに留まっており、Landmark Node 抽出に際しては、その断片的な情報しか提供していない。たとえば、電子マニュアルの Web サイトでは、目次ページのリンク結合度が高い。しかしながら、意味のある Landmark の採択としては、必ずしもよい選択とはいえない。またリンク構造としては、ハブ型になっていなくても、ブラウジングの指標として有益なページも存在する。これらの問題に対処するために、利用者のアクセス頻度情報も活用する。明らかであるが、アクセス頻度の高いノードが Landmark Node としてふさわしいものとする。

### (3) Web サイト階層情報

第3の情報は、Web サイトにおける階層構造情報の援用である。Web サイトにおけるページ情報は、階層構造型のファイルシステム構造により構成されており、一般的には、上位の階層に位置するページほど、Landmark Node として採用するに適切な情報となっている。たとえば、典型的な大学のホームページでは、トップページに、教育、学科、教員等の概要情報、もしくは、これらへのリンク情報が記載されている。階層の下位に属するページほど、特定の個別情報の記載となっている。それゆえ、上位のページほど、より一般的な情報が掲載され、また、Web 局所情報を理解するには適切である。それゆえ、Landmark Node としては適切である。すなわち、階層の上位（深さ）に位置するページ（ノード）ほど重要であると認識する。ここで階層構造における深さ情報は、URL における階層構造記述により規定できる（区切り記号/で判断）。

上記の3つの視点による Landmark Node 抽出のアルゴリズムについて整理する。

SOC(Second-Order Connectedness): リンク長2の参照ノード数

BSOC(Back Second-Order Connectedness): リンク長2の非参照ノード数

O: リンク長1の参照ノード数

I: リンク長1の非参照ノード数

A: アクセス頻度

D: 階層深さ

### アルゴリズム

$$\text{struct\_importance} = (I + O) * w_1 + (\text{SOC} + \text{BSOC}) * w_2$$

ただし、 $w_1 + w_2 = 1$

(デフォルト値は、影響力を考慮して、 $w_1$ を $w_2$ より大きめに設定。)

$$\begin{aligned} \text{重要度} = & \text{struct\_importance} / (\text{Max struct\_importance}) * \text{conn\_w} \\ & + A / (\text{Max A}) * \text{access\_w} \\ & + \text{depth\_w} / D \end{aligned}$$

ただし、 $\text{conn\_w} + \text{access\_w} + \text{depth\_w} = 1$

各要素の値を正規化して重みづけする。

もし、重要度が事前に規定された閾値 (cutoff value) よりも大きければ、対応するノードを Landmark Node として規定。(初期値としては、0.1に設定。)



パラメータ設定に関する考察：Landmark Nodeに関連するパラメータを変更すれば、Focus+Context Viewに現れるノードも変化する。たとえば、Landmark Node 閾値を上げると、Landmark Node の総数は減少する。たとえば、Georgia 工科大学の College of Computing のサイトでは、閾値が 0.1 の時は、42 の Landmark Node が現れるが、閾値を 0.5 に増大させると、2 個に減少する。閾値と Landmark Node との関係を図 4-1 に示す。

しかしながら、地理的に適切な Landmark Node を抽出するためには、各 Web 局所情報に対して、個別に閾値をチューニングする必要がある。特に、広域の Web 情報空間の概要表示に際しては注意が必要である。たとえば、局所的に、アクセス頻度が非常に高く、また、結合度が非常に高い局所情報が存在する場合には、その Web 局所情報空間に対して、より高い閾値に規定する必要がある。したがって、Web 局所情報空間毎に階層的な適用を行う。

また、閾値自体を使わず Landmark Node を抽出するには下記の 2 つの代替案がある。1 つは、総数  $n$  を規定し、重要度順に  $n$  個の Landmark Node を抽出する方法である。また、別の代替案は、比率  $n\%$  を規定し、重要度上位の  $n\%$  を Landmark Node とする方法である。どの方法にしても、各 Web 局所情報空間の特性を考慮して規定する必要がある。

また、リンク結合度情報、アクセス頻度情報、階層情報間の重み付けの規定も重要である。基本的なアプローチとしては、ヒューリスティック的に規定せねばならないが、現実のデータ構成や、システム運用構成に忠実であるべきである。たとえば、階層の深さ情報は、Web 情報の作成者側でのみ規定される情報であり、リンク結合度情報のように、作成者、利用者双方の情報を援用したものではない。この点で、重み付けの値は、相対的に低く設定すべきである。また、アクセス頻度情報は、Web 局所情報の特性を規定する上で重要な情報であり、かつ、グローバルな利用者が関与して決定される情報である。このような特性を勘案し、デフォルトの重み付けファクターとして、アクセス頻度情報の重み付け 0.5、リンク結合度情報の重み付け 0.4、階層深さ情報の重み付け 0.1 としてシステム構築を行う。

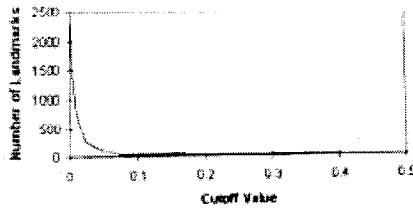


図4-1 Landmark Node に対する Cut-off 値の変化

#### 4. 3. 2 Focus+Context View の実現機能

次に、Focus+Context View の実現機能について説明する。下記のような個別 view を設定できる。

**In-view:** In-view は、各 Landmark Node から対象ノードにいたるまでのパスの表示である。この表示は、局所の詳細情報(Focus)とグローバルコンテキスト情報からなる。グローバルコンテキスト情報は、全体表示を過度に複雑化させないために、Landmark Node と対象ノード間の関係（リンク情報）のみを表示させる。具体的には、まず、各 Landmark Node から対象ノードへの最短パスを見つける。もし最短パスが存在しない場合には、当該 Landmark Node を In-view に表示させない。

また、もしある Landmark Node から対象ノードへのパスが存在しても、大規模上空間の把握に役立たないケースは除外するほうが望ましい。たとえば、このようなパスの途中に結合度の高いノード（hub/authority）がある場合である。このような場合の多くは、高結合度ノードと Landmark Node 間、ならびに、高結合度ノードと対象ノード間には、意味的関連があるが、Landmark Node と対象ノード間には、直接的な意味的関連がない。したがって、Context View から除外しても大きな影響はなく、構造を単純化できる。一方、このような高結合度ノードから対象ノードへの最短パスは、必ず In-view として表示させる。また、一般に、In-view における Landmark Node が多くなりすぎる場合は、対象表示リンクにおいて、始点のノードの重要度が終点のノードの重要度に対し、同程度以上になるケースのみを表示させる。

**Out-view:** Out-view は、対象ノードから各 Landmark Node にいたるまでのパスの表示である。Out-view も In-view と同様、局所の詳細情報（Focus）とグローバルコンテキスト情報からなる。グローバルコンテキスト情報の規定に際しては、In-view で

## 第4章 大規模ハイパーメディア構造可視化とブラウジング

議論したケースと同様、パスが存在し、途中に高結合度ノードがなく、また、パス中のリンクにおいて、終点のノードの重要度が始点のノードの重要度と同程度以上になるようなケースのみを Out-view 内に表示させる。

**Focus+Context View** パス発見の計算量：グラフ構造（トポロジー）が抽出されれば、最短パスの規定は、 $n$  をノード数、 $l$  をリンク数とすると、depth-first search 等のアルゴリズムを援用して、計算量のオーダーは  $O(n+l)$  となる。もし、 $k$  個の Landmark Node が規定されていれば、全体の計算量のオーダーは、 $O(n+l)k$  である。 $k$  が相対的に小さければ、全体の計算量も少ない。

図4-2に、Focus+Context View の例を示す。中央の丸が現在フォーカスしているノードであり、隣接のノードが小さい四角で表現されている。また、Landmark Node が周辺の大きい四角で表現されている。このような骨格を、フォーカスの移動と共に追従、表現させることにより、情報空間の適切な把握、理解の支援をさせる。また、図4-3は、Focus+Context View の機能を、Web イメージ検索システムに適用させた例である。全体と部分との同時表示機能を実現すると共に、問い合わせ結果の視覚的表示、グルーピングを行わせる意味としても活用している。例は、バスケットボールに関する URL をグルーピングし、視覚表示させた例である。

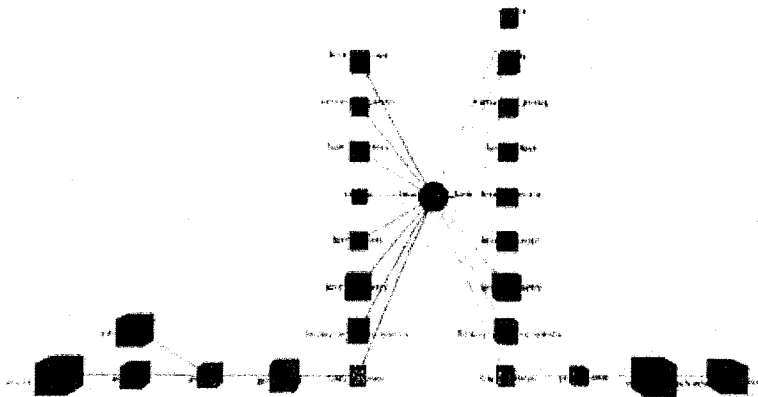


図4-2 Focus+Context View の例

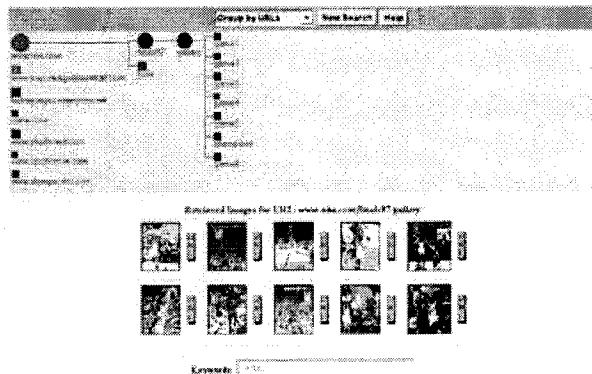


図4-3 Focus+Context View のイメージ検索システムへの適用

### 4.3.3 評価と考察

狭義の情報検索という行為は、利用者が検索手がかりについてよく理解している場合には、有効である。しかし、利用者がどのような情報を検索したいのかという目的が明確でなかったり、検索手がかりについての想起性、選択一意性がはっきりしなかったりする場合には、ブラウジングの方が有効となる。Web 局所情報の理解については、後者のケースがあてはまる。ブラウジングに際しては、対象情報空間の視覚表示は有効なフィードバックツールである。しかしながら、構造が大規模で複雑な場合、全てのノードやリンクを表示することは本質を理解するという意味ではむしろ妨げとなる。ここに Focus+Context View を用いる意義がある。

Web 情報空間に対するオーバビューダイアグラムの利用は、種々の点でメリットがある。まず、直感的に全体像を把握しやすい。また、利用者は、各注目ノードのアクセス頻度の状況も概観できる。さらに、当初期待をしていなかった情報にも接することができる。たとえば、評価対象とした、ジョージア工科大学の College of Computing のサイトのケースでは、コンピュータサイエンスとは直接関連のない、詩のページのアクセス頻度がかなり高いことがわかった。これは、利用者は、最初から詩のページの存在を知ってアクセスするだけでなく、ブラウジングの中から文脈情報を用い、興味のあるページにアクセスできるという一例である。

また、利用者は、Focus+Context View をブラウジングツールとして、情報空間を概観するだけでなく、特定のノードをクリックすることにより、該当ページを直接参照することもできる。

### 4. 4 スキーマブラウザ

本節では、3章で説明したスキーマの考え方を援用し、大規模ハイパーメディアの構造表現とそのナビゲーションへの適用について説明する [HHTK96-1]。

#### 4. 4. 1 スキーマブラウザ実現の背景

##### (1) ブラウジングメタファーの連続性

読み進む際の操作面、及びインタフェース上の制約により、利用者の思考を妨げないようにする必要がある。特定に項目だけでなく、全体の情報の把握を行いたい場合や、要点の拾い読みをする場合には、読み進み方を連続的に行わせることが重要である。ここで、読み進み方の連続性とは、フィルタリングとブラウジングとの区別を利用者に意識させず、スムーズに希望の関連情報をを提示させていける点、及び遷移前後の検索、提示情報になんからの関係があることである。

##### (2) 動的なカスタマイゼーション

また、閲覧時の個人属性、用途、目的等に応じて情報のカスタマイズ（選択、整理）ができることも必要である。本のアナロジーでは、ある本を読んでいる際に、必要に応じて、辞書、関連文献、別の観点から整理した書類などを捜してよむことに相当する。ハイパーメディアでは、このような「本」を必要に応じて動的に構成できる。これは、あらかじめフィルタリングされた情報空間内のブラウジングから、関連フィルタリング空間をブラウジングする拡張となる。

##### (3) 「観点」の導入による情報提示

必要な情報をもれのないように読みたい場合や、関係のない情報を読まないようにしたい場合には、ただ漠然と読み進んでも効果が薄い。1つの有効な手段は、「観点」を用いて、これをガイドに読み進んでいく方法が考えられる。たとえば、時間順や、地域順に整理して読んだり、ある主題との関連という観点から情報を捉えなおしたりする場合である（図4-4参照）。

##### (4) 詳細の程度に応じた情報提示

観点の導入による読み進み方が同レベル情報間の遷移とするならば、情報内容をマクロ/ミクロにみる読み方も相補的に必要である。たとえば、概要情報を読みな

がら、興味に応じて詳細情報を読む場合である。このような概念の遠近法的利用も、必要な情報を適切に知るという点で重要である。

### (5) 表示メディアの物理的制約からの開放

紙のメディアでは、情報の表示単位がメディアの物理単位（ページ）と同一になる。これに対して、電子メディアでは、スクロール、拡大／縮小、要約／詳細化の機能を用いることにより、情報の表示単位が必ずしも媒体の物理的特徴に制約されることはない。従って、情報有効利用のために、意味内容に即した論理構成から情報を扱う単位を規定し、表示構造と独立にすべきである。

### (6) 履歴の保存

著者が情報をどのように読めばよいかをガイドしたり、別の読者や過去の自分がどのように読んでいたかを知ったりするためには、読書履歴の保存は有効である。この場合、何を見たかに加えて、どのようにして情報を見たかが重要となる。たとえば、読み進む際の観点の指示や、情報メディア表現の変更等が後者に相当する。このような情報の有効かつように加えて、現在、読者が情報空間中のどの部分を読んでいるかをしるためにも、履歴の保存は必要である。

## 4. 4. 2 集合操作リンクに基づくスキーマブラウザ

前述した考え方に基づく情報検索方式として、集合操作リンクに基づくスキーマブラウザを提案し、説明する。これは、観点、及び複合実体表現を取り入れたハイパーメディアの動的生成リンクとその利用である。

(1)、(2) からフィルタリングも一種のブラウジングとみなし、ブラウジング主体の読み進み方のパターンを提供する。さらに、(3)、(4) を考慮して、観点、及び複合実体の表現が可能なデータモデルを用いる。また、このようなデータモデル・操作を用いることにより、(5)、(6) を満たすことが可能となる。

4. 4. 3 集合操作リンクを実現するデータモデル

(1) 観点の導入

観点を導入するため、時間、内容など複数の実体属性の情報空間を規定する。ここで、観点とは、時間や主題等に対応する見方であり、情報空間を規定する1つの軸に相当する。各実体は、複数の観点で構成される多次元情報空間中の1つの点、もしくは、部分領域として表現される。

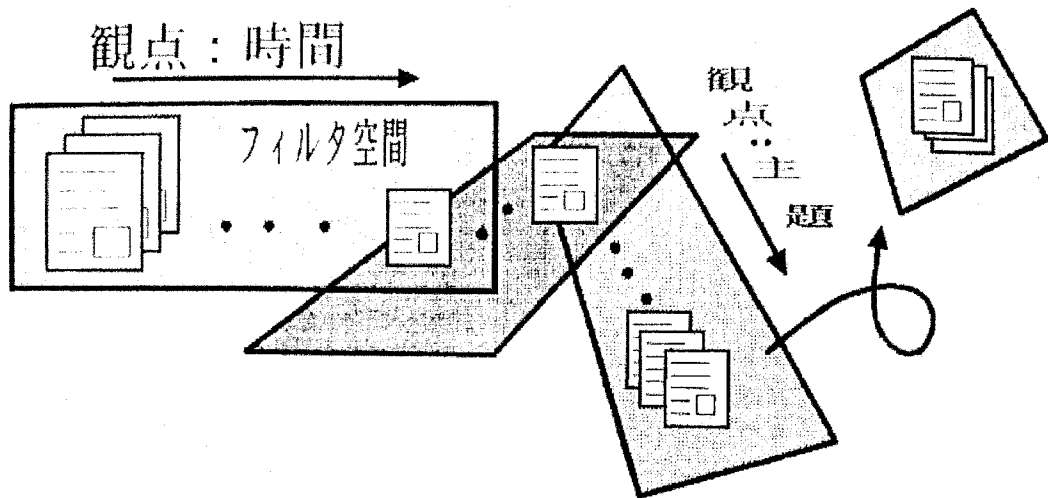


図4-4 観点を導入したナビゲーション

(2) 実体属性項目間の関係表現

前述の観点の複数個用意し、各々に対して、階層型もしくはネットワーク型の2種類のタイプにより、実体属性項目間の関係表現を行う。

**階層型属性表現：** 分類のように属性項目間が階層構造となっているデータ表現である。通常、属性項目（分類項目）値は、個数が少なく、あらかじめ設定されている。多次元情報空間の基本フレームとなる属性表現である。

**ネットワーク型属性表現：** キーワード/シソーラスのように、属性項目間に特定の関係づけがなされていないか、あるいは、ネットワーク構造となっているデータ表現である。たとえば、文献検索における、著者名やタイトル名などの書誌項目がこれに相当する。この場合、著者名やタイトルといった「観点」は指定されているが、属性項目（書誌項目）値は、情報編集時に定める。一般に、階層型属性項目に比べて、属性項目値の数が多い。

図4-5に、美術百科辞典の場合の観点、および、属性の関係表現の例を示す。

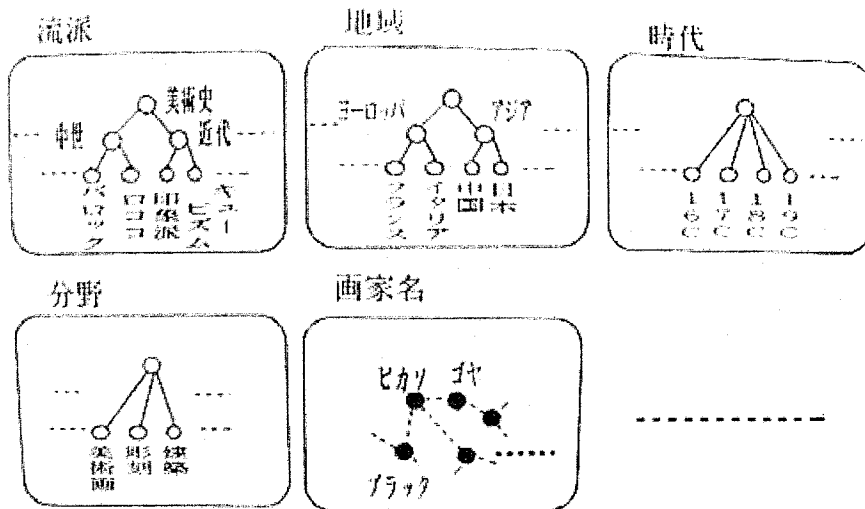


図4-5 観点及び属性の関係表現の例



### (3) 複合実体の表現

情報のアクセス単位を柔軟にするために、情報内容の抽象度をレベルに応じた実体表現を行う。たとえば、要約と要約された詳細項目等の情報表現である。ここでは、is-part-of の関係を用いて関連づけ、複合実体を構成する。

#### 4. 4. 4 トラバース (当該の読み進み方) のパターン

上述の表現構造に沿った場合の、トラバース (当該の読み進み方) のパターンについて説明する。読み進みかたの連続性を考慮して、表示情報の遷移に際しては、なんらかの関係があるもとする。また、読者の動的な目標に対処するために、下記の読み進み方のパターンを用意する。遷移パターンの類型化は、フィルタ空間 (検索空間)、及びその中に位置する現在の情報属性との変化に着目して行う。

##### (1) フィルタ空間不変、かつ表示情報変化

利用者の対象とする検索空間は変わらず、その中で、該当する情報を表示し、読み進んでいく場合である。逐次ブラウジングのように、紙の本におけるぱらぱらめくりと同様の機能や、見出しやキーワードに沿って読み進むカテゴリーブラウジングがある。

**逐次ブラウジング：** フィルタ空間内にある情報集合を逐次的に表示する。あるいは、トラバース軸に沿って整理されている情報集合を、提示情報単位に表示する機能である。紙の本におけるページ単位の情報提示 (ページめくり) 機能と同様である。

**カテゴリーブラウジング：** フィルタ空間内にある情報集合を、トラバース軸の見出し項目単位 (分類項目やキーワードなど) に提示する機能である (図4-6参照)。たとえば、時間軸 (18世紀、19世紀、 ) や、地域軸 (日本、アメリカ、フランス、 ) の見出し項目に沿って、拾い読みを行う場合である。

##### (2) 表示情報不変、かつフィルタ空間変化

(1) の場合とは逆に、現在表示している情報は変わらないが、利用者の想定するフィルタ空間が変化する場合である。読書中に関連情報が知りたくなる状況とは、利用者の欲求する検索情報空間が変化することに相当する。電子の本では、このような関連フィルタ空間のあいだを遷移する場合も、広義のブラウジングとみなすこ

## 第4章 大規模ハイパーメディア構造可視化とブラウジング

とができる。フィルタ空間の変化を行うトラバースとしては、遷移前後のフィルタ空間の関連性を考慮して、下記の4つの機能を用意する。

**制約トラバース：** 観点の変更を伴うトラバースで、変更前の見出し項目を制約として、変更後のトラバース軸の方向に、フィルタ空間を広げる機能である。たとえば、「イタリア」美術百科情報を時間軸に沿って読み進み、画家セガンチーニが表示されたとする。このとき、同時代の他地域の画家としては、誰が活躍していたか（たとえば、画家のゴッホなど）を探す状況が制約トラバースの利用に相当する（図4-7参照）。制約トラバース後のフィルタ空間は、次式で表される。

$$F = (\text{トラバース前のフィルタ条件} - \text{変更後のトラバース軸に関するフィルタ条件}) \cap \text{変更前の見出し項目}$$

**ズームイン：** フィルタ空間を縮小し、該当情報を絞り込んで、読み進む場合に相当する。具体的には、階層型のトラバース軸において、表示見出し項目もしくは、その下位項目をフィルタ条件に組み込むことに相当する。たとえば、情報を国別に読み進んで、「イタリア」に関する情報が提示されている際に、ローマやフィレンチェなどイタリアの地域別に絞り込んで読みを深める場合にあたる。（図4-8参照）

**ズームアウト：** ズームインの逆操作である。フィルタ空間を拡大し、該当情報を広げて、マクロに読み進む場合に相当する。ズームアウトとは、見出し項目の上位項目をフィルタ操作に組み込む操作である。

**表示情報属性のフィルタリング化：** 表示情報属性値をフィルタリング条件に置き換え、フィルタ空間を表示情報属性の規定する空間と一致させる機能である。形式的なトラバース操作であるが、関連情報を探し出す際の基本操作である。

### (3) フィルタ空間変化、かつ表示情報変化

第3のトラバースは、フィルタ空間、表示情報空間双方が、遷移により変化する場合である。紙の百科辞典の利用法を勘案すると、代表的な遷移機能は、キーワードを介して、関連情報、詳細情報と参照するクロスリファレンスである。

## 第4章 大規模ハイパーメディア構造可視化とブラウジング

クロスリファレンス： キーワードを介して、関連情報を参照する。この時のフィルタ条件は、両者に関連するキーワードとなり、表示情報は、参照された関連キーワードとなる。（図4-9参照）

また、形式的であるが、フィルタ空間、表示情報共に変化しないトラバースも考えられる。たとえば、トラバース軸の変更のみの操作や、トラバース軸のスケーリングの変更（粗くする、詳細化する）操作などである。

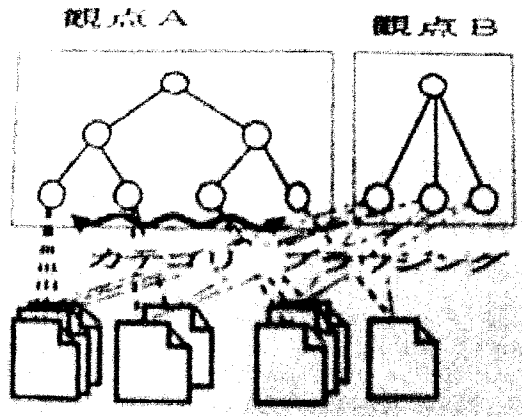


図4-6 カテゴリーブラウジング

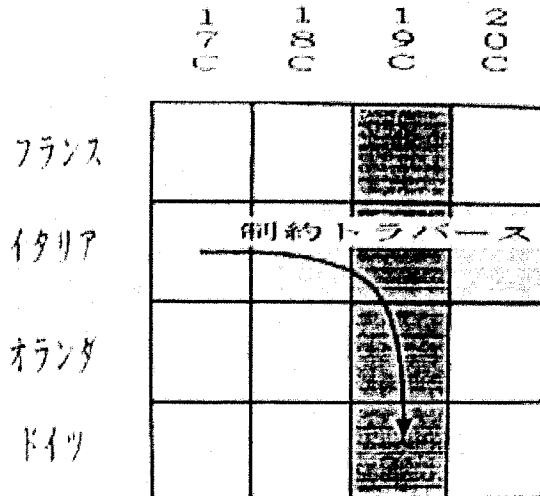


図4-7 制約トラバースの例

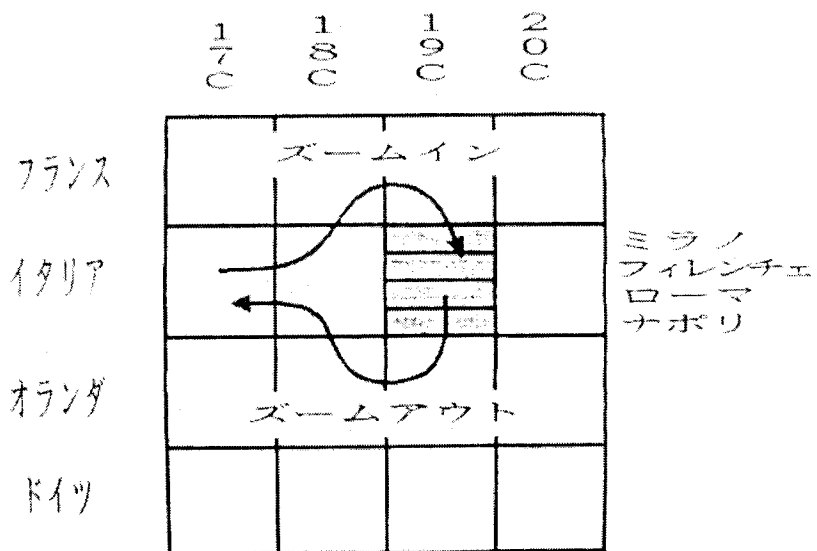


図4-8 ズームイン/ズームアウトの例

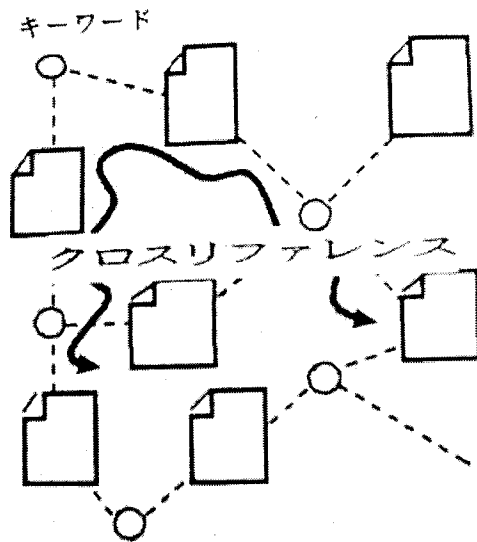


図4-9 クロスリファレンス

### 4. 4. 5 評価と考察

本稿では、スキーマブラウザを電子美術館システムに適用した際の評価と考察について述べる。まず、トラバースの効用は、下記のような操作を読み手が簡単にできることである。

- a. 「イタリア」の美術百科事典を構成（フィルタ空間＝「イタリア」）し、ダ・ビンチのページをめくる。この時、時代をトラバース軸とする。（図4-10参照）
- b. カテゴリブラウジングにより、時代軸に沿って読み進む。たとえば、「19世紀」の見出し項目により、セガンチーニのページが現れる。
- c. ここで、利用者が同時代の他地域の画家について知りたくなった場合を想定する。ここで、視点を地域に変えて、制約トラバースを行う。たとえば、「オランダ」では、ゴッホをはじめとした同時代の画家が現れる。この時のフィルタ空間は、「19世紀」と変更されている。
- d. また、文中の語句を直接指示することにより、クロスリファレンスが行える。（通常のハイパーメディアによるナビゲーション）

従来の情報検索手法と比べて、円滑な対話操作により必要な情報の検索だけでなく、探索的な情報ナビゲーションにも有用であることが認識された。

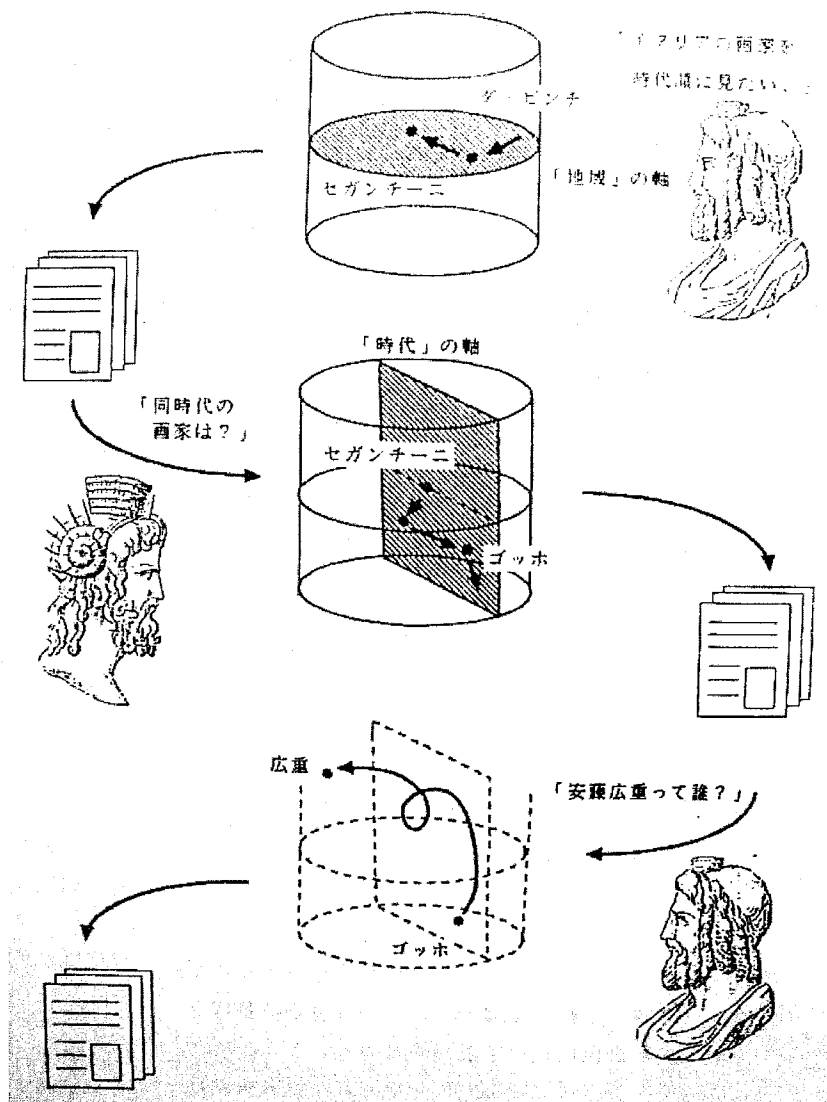


図4-10 トラバースの例

### 考察：ブラウジングによる印象の多様性

ハイパーメディアを用いて人間の発想、理解を支援する第一の効果としては、同じ表現、内容から、状況に応じていくつもの解釈が引き起こされることである。たとえば、美術ハイパーメディアにおいて、「モネ」の作品を見る際に、印象派の作家の系列の中から「モネ」を選び作品を見ていく場合と、同じ主題のもとに描かれた「モネ」の作品を比較しながら見ていく場合とでは、同じ絵画でも受ける印象が異なる。これは、明らかに、当該の絵画のみから印象を受けているだけではなく、一連の提示（ブラウジング）系列とそれに含まれる当該の絵画との関係にも依存するためである。すなわち、提示系列における当該絵画の位置づけや、特徴的差異が強調されるため、提示系列が異なれば、同じ画像も異なった印象を受けるものと想定される。テーマを想定した展示会や、雑誌の特集号はこのような効果を念頭において実現されているが、作成・運営労力はかなり大きい。ハイパーメディアの場合、比較的簡単に手持ちの素材を用いて実現可能である。

このような効果は、静止画におけるコラージュ（画像を組み合わせることにより、全体であるまとまったイメージを打ち出すが、必ずしも全体のイメージが個々のイメージの加算とは限らない）、動画におけるモンタージュ（前のシーンの印象に依存して、次のシーンの印象が変化する）と同様の効果である。

以上のように、スキーマブラウザは、多面的な印象を与え、探索的な情報ナビゲーションに適した検索・提示方式であるといえる。

### 4.5 本章のまとめ

本章では、ハイパーメディアの大規模性への対処と、文脈を考慮した多義性・あいまい性への対処という要件を満たす2つのハイパーメディア構造可視化方式について提案、プロトタイプについて言及した。

まず、全体と部分との関係を同時に表現でき、かつ大規模なハイパーメディアにもスケーラブルに対応できる方式(Focus+Context Views)について提案、説明を行った。特に、Context View において、明快な構造を抽出、明示するための Landmark Node の発見、抽出が重要である。Landmark Node の抽出に際しては、グラフ構造情報、アクセス頻度情報、階層構造の深さ情報を用いて行い、実際の全体概要構造の表示に際しては、最短経路パスの存在を活用して、計算量が少なく、かつ構造理解の見通しのよい方式を提案した。これは、魚眼図的表現手法のアナロジーとして、グラフ構造表示にその考えかたを援用したものである。また、ある時点での全体構造のスナップショットを提示して、現在閲覧している情報の位置づけを明確化することにより、情報空間の迷子問題や、情報利用時の多義性・あいまい性を軽減させる。

また、スキーマを規定したブラウジング方式として、多次元分類に基づく文脈情報を考慮したトラバース操作を提案、プロトタイプによる検証を行った。トラバース操作は、異なった観点を活用したナビゲーションというプロセスを重視し、多面的な理解を行わせることにより、情報理解の多義性・あいまい性の軽減をはかるブラウジングである。

これらの方式は大規模 Web 情報に対するオーバビュー表示機能だけでなく、Web 検索エンジンに対するブラウジング支援ツール、情報検索ツールとしての利用が行える。これらの情報可視化技法を、応用により使い分け、ハイパーメディア、Web の情報理解を深めることをめざす。



## 第5章 Augmented Hypermedia

### － 編集・利用効率向上をめざすシステム構成 －

#### 5. 1 Augmented Hypermedia 構成の背景

ハイパーメディアのメタファーは、Web 情報利用の定着により、今日では電子メディアコンテンツのアクセスに対する共通のヒューマンインタフェースとして広く認識されてきている。リンクを介したナビゲーションは、先験的知識をもたずに利用者が情報をアクセスする直接的な行為である。また、著者・編集者は、簡単なノード・リンクモデルや半構造化モデルにより、効率的にコンテンツを作成、編集することができる。

Web の出現・定着以前と以降とにおいての顕著な差異の1つは、編集者と利用者との役割が変化してきたことである。Web の出現以前では、個人向けシステムを除いては、ハイパーメディアシステムの編集者グループとその利用者グループとは異なっていた。システム構築・運用に際しては、編集者側のリソースがボトルネックとなっていた。その結果、商用システムの継続的運用には制約ができ、結果として、電子百科事典や CD-ROM 出版等の同一コンテンツが大多数の利用者に共有利用されるか、あるいは、編集コストがあまり負担にならないという中小規模の応用に限られていた。

Web の出現・定着以降、Web というハイパーメディアに対しては、各利用者が専門の編集者と同等の編集者になりえることが可能になった。この利用者の役割変化が、編集者のリソースボトルネックを飛躍的に改善した。しかし、編集効率だけでなく、対象情報の利用効率も含めた双方を改善することは、一般にトレードオフの関係にあるため、いまだ途上の段階である。したがって、Web を含むハイパーメディア情報の中から、編集効率を保ったまま、如何に情報利用効率を増大させるかが、さらなる情報活用の重要な指針となる。

本章では、編集・利用プロセスに着目した Augmented Hypermedia というシステム構成について説明する。また、Augmented Hypermedia の考え方に基づくハイパーメ

## 第5章 Augmented Hypermedia — 編集・利用効率向上をめざすシステム構成 —

ディプロトタイプ開発を通じ、編集・利用双方の効率化を向上する方策について説明する。

## 5. 2 関連研究

利用者のプロフィールや利用ビヘイビアを考慮した付加価値のある適応型 Web システムの研究としては、Brusilovsky らの提案した Adaptive Hypermedia [Brus96, etc.] があげられる。Adaptive Hypermedia の要件としては、利用者モデルがあり、この利用者モデルをハイパーメディアへ適用する機能を有するハイパーメディアシステムである。このようなシステム構築のための設計手法や利用者モデル構築技術等の研究が進展している。たとえば、適応システム構築のための意味ある個人属性の抽出手法、利用者のふるまいに応じた評価手法等である。応用としては、教育用途のハイパーメディア、オンライン情報提供システム、オンラインヘルプシステム、検索型ハイパーメディア、ハイパーメディアのパーソナライゼーションなどがあげられる。

本論文におけるアプローチでは、利用ビヘイビアに基づく利用者モデルと、リンク解析・内容解析に基づくデータモデルとの統合による差別化に重点をおき、関係情報の共有、精緻化、抽象化機能を提供する。利用者モデルとデータモデルという異なった観点のモデル・属性情報を一元的に直接評価することは困難である。しかし、ハイパーメディアの編集コストと利用便益とを最適化するという目標のもとに、両者の統合を進める。

また、システム構築面での 3 層構成的なモデルとしては、Wiederhold らの提唱している Mediator [Wied92] や ASP モデルなどが関連する。従来のシステムでは、クライアント-サーバ型の 2 層構成的なモデルが主流であったが、運用・サービスの効率化のため、利用者と 1 次情報としてのコンテンツ・機能を提供するサーバとの間に、主に 2 次情報を統合処理するレイヤーの必要性が提唱されている。運用・サービスの効率化は、サーバ側、クライアント側、各々の処理能力、データ容量や、ネットワークの遅延、配信速度等に依存する。広域分散システムとしての Web システムの構成としては、適切な構成である。

筆者らの開発した Web 情報検索システム AMORE では、このようなアーキテクチャの特性を生かし、比較的更新頻度の少ない応用に対して、種々のハイパーメディア機能の統合を進めることが可能となっている。しかし、Web のような広域分散システムにおいて、高アクセス頻度、高頻度データ更新環境下でのスケーラブルかつ応答性能のよいシステム構築運営は、関連研究活動も含め、今後の課題である。

### 5. 3 Augmented Hypermedia

図5-1は、従来ハイパーメディアと比較した場合の Augmented Hypermedia のプロセスを表したものである。従来のハイパーメディアでは、ハイパーメディアを作成・編集する人々と、得られたハイパーメディアを利用する人々との集団とは異なり、各々別々の役割を担っていた。主な利用は、関連情報間にリンクづけして参照・ナビゲーションする電子出版形態であり、CD-ROM のようなパッケージを情報流通媒体として用いていた。このような環境では、利用(Utilization)の規模は、一部の例外を除いてはそれほど大きくならない。従って、効率的な編集・利用を実現するためには、編集(Authoring)コストを抑える必要がある。しかし、編集コストを単に抑えるだけでは、あまり魅力的なハイパーメディアが作成できず、結果として、利用が広がらないという負のスパイラルに陥りがちとなる。

Augmented Hypermedia とは、情報利用からえられた情報を新たな編集情報として活用し、ハイパーメディアの編集・利用効率を高める機構である [HaHi98]。利用者の行動様式、コンテンツ解析、オントロジー情報等を活用して、関係情報の共有、精緻化、抽象化プロセスを導入し、編集・利用効率を向上させる。このようなプロセスを実現することにより、編集・利用プロセスに対する正のスパイラルを作り出し、付加価値を生み出す高品質ハイパーメディアの発展をめざす。以下、関係情報の共有、精緻化、抽象化の各々のフェーズにおける詳細説明を行う(図5-2参照)。

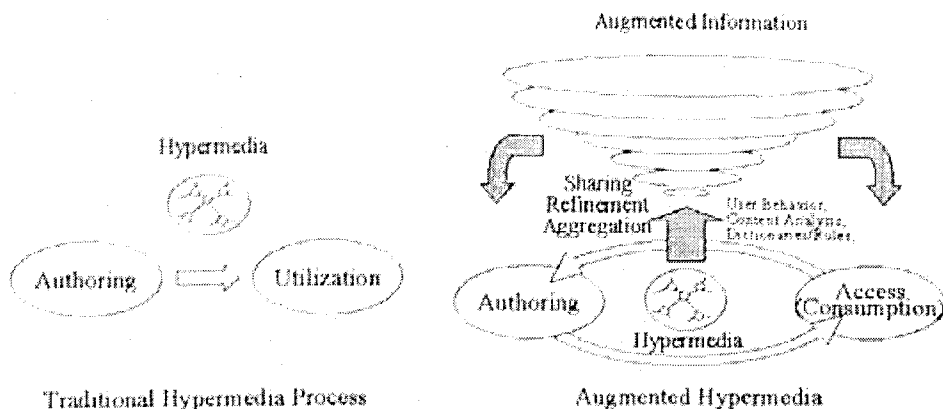


図5-1 Augmented Hypermedia のコンセプト

### 5. 3. 1 関係情報の共有

ハイパーメディアにおける知識共有は、ある利用者の意味ある関係構造を、他の利用者也利用できることにある。たとえば、Web ブックマークの共有、ハイパーメディアの利用履歴（ヒストリ）、シナリオ（パス）等はその一例である。専門性ある利用者の情報アクセス、利用履歴・選好情報は、他の利用者の情報活用をより活性化させ、グループとしての情報利用アクティビティを増大させる。

編集者の作業タスク（オーサリングコスト）を増大させずに、付加価値を生む構造を抽出・共有するために、まず、利用者の行動ビヘイビアを分析するアプローチをとる。具体的には、ある利用者のナビゲーションのパターンを抽出・分析する方法と、ナビゲーションした結果のコンテンツリストを抽出・分析する方法とがあげられる。前者の方法、すなわち、利用者のナビゲーションパターンを用いる方法は、近年のデータベース問い合わせ技術を用いることにより進展している。このようなナビゲーションパターンは、利用者のナビゲーションプロセスを通じて抽出される。ナビゲーションパターンを効率的に抽出・表現するために、宣言言語的なアプローチをハイパーメディアにも適用されている。たとえば、WebSQL, WebOQL, WebLog, W3QS などが Web データの問い合わせ言語として提案されている。また、半構造化データに対するデータベースアクセス言語も提案されている。たとえば、LORE, Strudel, UnQL などである。しかし、これらのプロジェクトは、知識共有を目的とするナビゲーションな操作への適用を想定したものではない。

筆者らのグループで開発した Powerbookmarks は、WQL という宣言的問い合わせ言語によりブックマークの共有利用の実現とその活用促進をはかるシステムである [LCAH99]。Powerbookmarks のインタラクティブでナビゲーションなインタフェースを用いることにより、利用者は、ブックマークの共有をはかることができる。また、ブックマーク URL の定期的検査、検索、再構成を支援する。他のブックマークシステムとは異なり、アクセス制御のある情報共有環境を実現でき、また、ナビゲーション結果の保持、編集を行うことができる。

### 5. 3. 2 関係情報の精緻化

関係情報の精緻化は、コンテンツ分析や Web マイニング技術を用いてより正確な関係構造を抽出・表現することにある。たとえば、筆者らが開発したビデオハイパーメディアにおける Moving Hot-Spot（動的オブジェクトに対応するボタン）を非

構造メディアデータから同定するのは、その一例である。意味のある関係情報を規定するために、メディア情報に対する文書構造解析技術、オブジェクト認識技術やオブジェクト追跡技術を活用する。たとえば、MVD(Multivalent Document)モデルや、COIR(Content-Oriented Image Retrieval)イメージインデクシング技術などである。このような要素技術をハイパーメディア2次情報として利用し、関係情報の精緻化を行う[H MOL97]。

### 5. 3. 3 関係情報の抽象化

関係情報の抽象化 (Aggregation) は、特に大規模なハイパーメディアに対してもスケーラブルな処理を適用する意味で重要である。共通の汎化構造を同定することにより、利用者の情報利用効率の向上をはかる。基本的考え方は、第2章で述べたとおりであるが、以下では、特にイメージ情報に処理の考え方を適用した場合について説明する。

AMORE(Advanced Multimedia-Oriented Retrieval Engine) [MuHH99]は、筆者らのグループで開発した Web 情報に対するイメージ検索システムである。AMORE では、Primary Object という考え方を導入し、イメージに対するクラスタリング、ナビゲーション技術の提供を行っている。これは、イメージメディアから抽出される複数の基本要素 (水平線、垂直線、矩形、円、楕円形等) をもとに、対象イメージ集合をクラスタリングし、整理しなおそうとする試みである。このような関係は、ハイパーメディアに対する、新たな関係情報タイプの付与とその抽象化に相当する。

また、大規模で複雑なハイパーメディア情報をそのまま表示するには、構造複雑度が高すぎるので、関係情報の抽象化は有用である。第4章で説明した Focus+Context View などがその具体的例である。

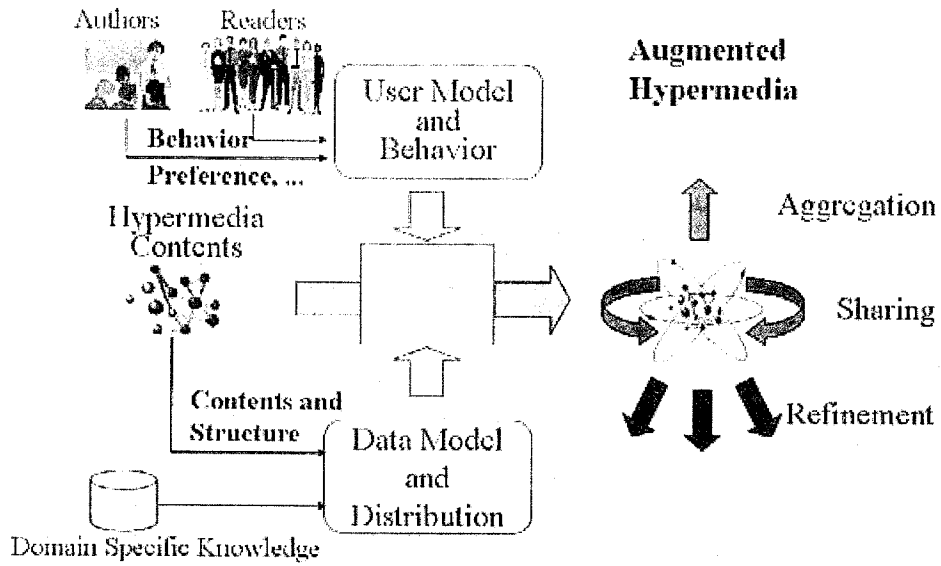


図5-2 Augmented Hypermedia 基本機能構成

## 5.4 システム構成

本節では、Augmented Hypermedia の必要機能である、関係の共有化、精緻化、抽象化を効率よく実現するためのシステム構成につき言及する。これらの機能をスケラブルに実現することがポイントである。

### 5.4.1 3層構成のシステムアーキテクチャ

図5-3に、Augmented Hypermedia を効率的に実現するシステム構成を示す。通常のクライアント-サーバ形式ではなく、クライアントとサーバとの間に Mediation Server をおく3層構成のアーキテクチャが自然であると考えられる。この理由は、2層のクライアント-サーバ形式では、サーバ側で、関係の共有化、精緻化、抽象化といった機能を実現するのが非常に複雑になるためである。（また、当然、大局的処理を行うので、クライアント側での処理は、制約がある。）また、サービス機能の追加、変更においても、従来の2層のクライアント-サーバ形式では、あまり効率的ではない。

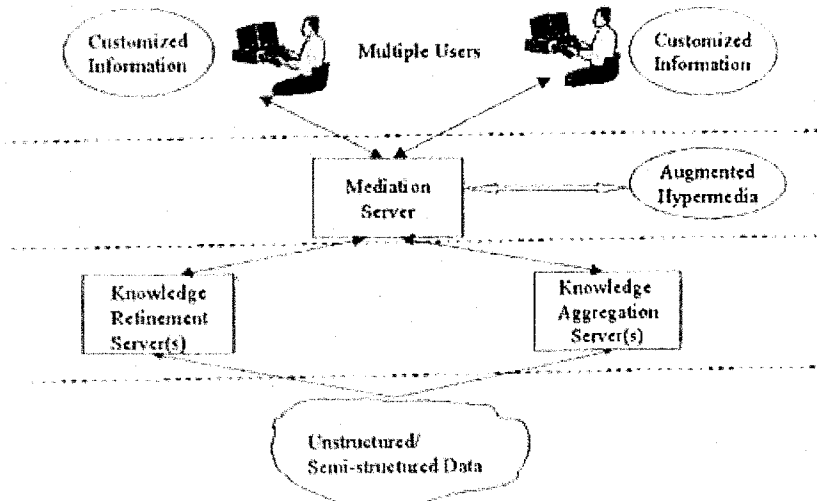


図5-3 Augmented Hypermedia システム構成



ここでは、関係の共有化、精緻化、抽象化という各操作を分散化し、これらのサーバ間の処理結果を統合する構成を採用する。この意味で3層構成のシステムアーキテクチャとなる。まず、非構造、もしくは、半構造のハイパーメディアコンテンツは、知識抽象化サーバと知識精緻化サーバにより分析される。Mediation Sever が、プライマリーサーバ間のコーディネータとして機能する。利用者は、個別サーバにアクセスする必要はなく、このコーディネータのサーバとコミュニケーションをとって情報アクセスを行う。ここで統合されたハイパーメディアコンテンツが Augmented Hypermedia である。各クライアントサイトは、利用者側の選好をもとにカスタマイズされた情報を提示する。クライアント側は、他の利用者に対する共有に対して、指定することができる。この情報は、Mediation Server 側で処理、利用され、さらに今後の情報再利用に活用される。

図5-4は、Augmented Hypermedia の3層構成モデルの詳細である。Individual User Layer, Mediation Layer, Individual Data Analysis Layer からなる。Individual User Layer は、利用者のクライアント端末側で対応する動的ビヘイビア分析、表示等の機能を提供する。Mediation Layer は、情報共有、情報統合のためのグローバル処理をつかさどるレイヤーである。ハイパーメディア分類、検索、提示機能の処理を行う。Individual Data Analysis Layer は、Mediation Layer への情報入力のため

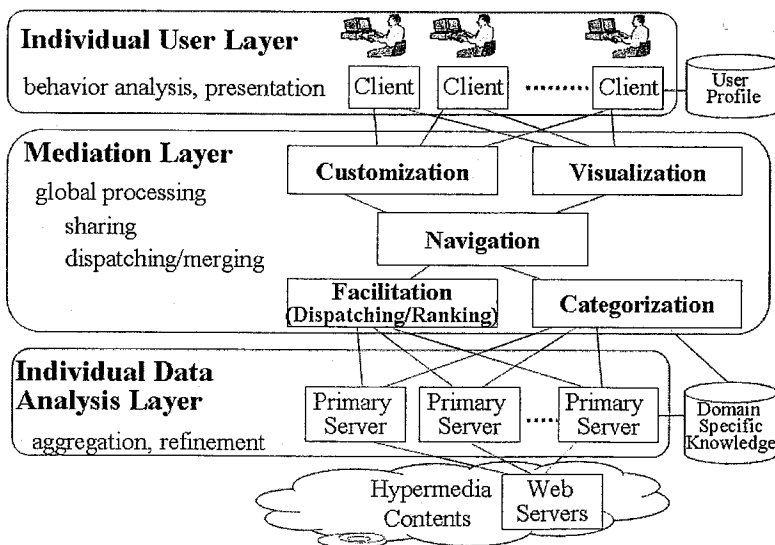


図5-4 Augmented Hypermedia 3層構成の詳細

のデータ分析統合、精緻化処理を行うレイヤーである。オリジナルデータとしての1次情報は、インターネットを介して存在する Web サーバに蓄積されている情報である。

#### 5. 4. 2 AMORE によるシステム構成例

3層構成のシステム構成法は、筆者らが開発した AMORE (マルチメディア Web 情報検索システム) にも援用した。図5-5に示す。収集された Web ページは、イメージと関連する周辺テキスト情報とに分離して、各々解析される。この後、対応するイメージに対して、周辺テキスト解析結果を付与し、情報ナビゲーション時に活用する。AMORE のサーバは、Mediation Server としてテキスト処理とメディア処理とのコーディネーション、統合処理を担う。

このような構成により、利用者は、対象とするイメージ情報のナビゲーションに際し、視覚手がかり情報からでも、テキストの意味的情報からでも、また、両者の組み合わせ情報からでも、アクセスすることができる。加えて、問い合わせの精緻化処理、オーバビュー表示処理等のサーバを順次追加することが、効率的に行える。このような3層のシステム構成の利点は、このような新規機能のサービスを付加する際に、各サービスがライトウエイトな機能に規定できるので、統合が効率的に行えることにある。

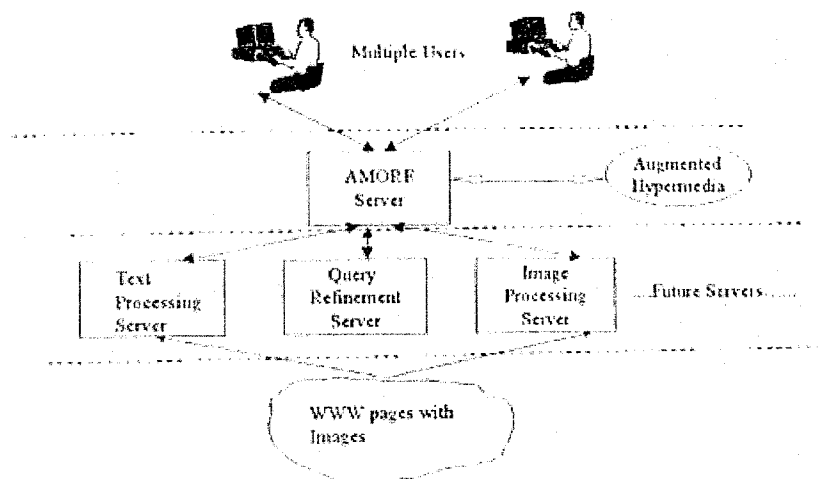


図5-5 AMORE システムアーキテクチャ

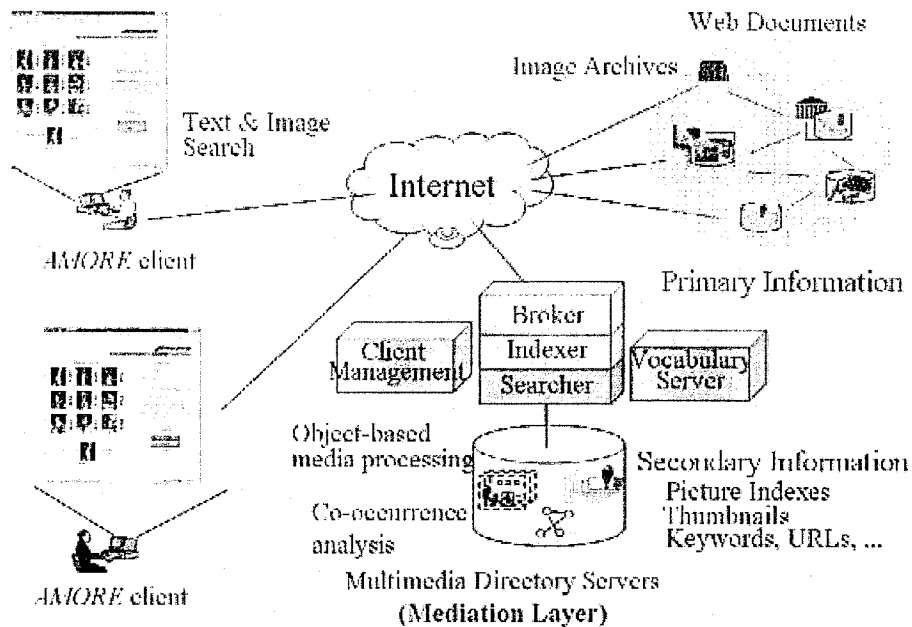


図5-6 AMOREシステム構成

図5-6は、米国美術館で、運用実験が行われた具体的システム構成である。Mediation LayerとしてMultimedia Directory Serverを開発した。ここで分散されて蓄積されているWebベースの芸術・美術作品情報に対し、2次情報（キーワード、シソーラス、サムネール、イメージインデックス等）を一括集中管理し、自動分類、検索機能を提供している。

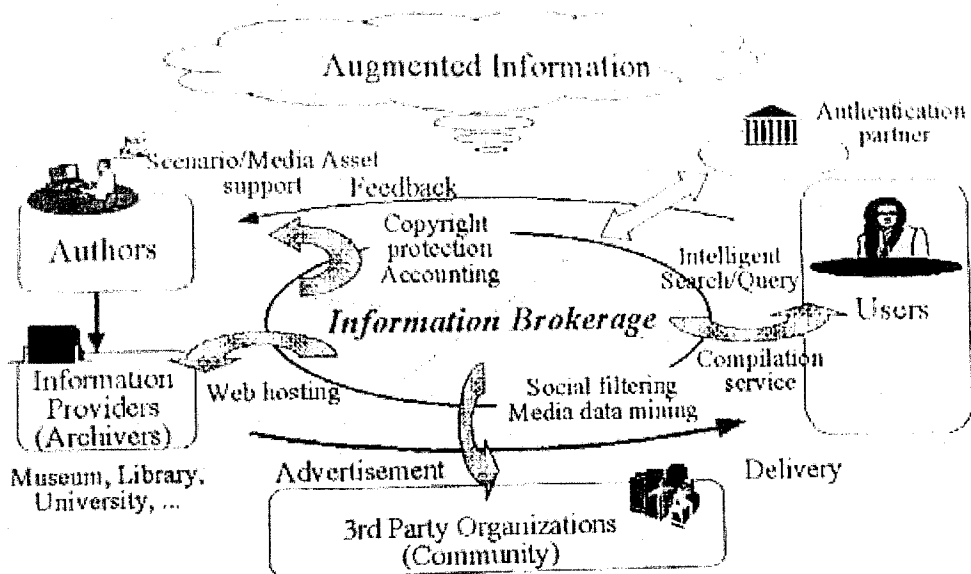


図5-7 Augmented Hypermedia 応用

図5-7は、Augmented Hypermedia の考え方を一般化し、Information Brokerage のシステムへの応用をめざしたものである。情報提供者と情報利用者との間の信頼性、問い合わせサービス、著作権管理等の付加機能を有し、安心して情報活用が行えるハブ環境を提供する。

## 5. 5 Augmented Hypermedia の利用に関する知見

ハイパーメディア応用システム開発事例をふまえ、Augmented Hypermedia を活用した人間の知的活動に関しての知見をまとめる。

### (1) 新しい関係づけの発見

ハイパーメディアのリンクづけ方策にも依存するが、なかには、利用者自身がしなかった、あるいは、作成者自身も想定しなかった情報間の関係づけが見出される。これは、利用者にとって新しい関係づけの発見であり、単なる情報検索から、新たな発想を刺激する機能となりうるものである。たとえば、集合操作リンクを用いると、利用者は、非常に多数の属性の組み合わせの中からトラバース操作により、情報を提示させていく。

### (2) メディア表現によるコミュニケーションの役割分担

初期のハイパーメディアシステムは、個々のマルチメディアオブジェクト（ノード）をリンクで結び提示するという単純な構造である。この場合、内容と表現メディアは一体化されており、表現メディア固有の特徴を生かしきれていない。これらを陽にとりいれたハイパーメディア化は、人とコンピュータ、コンピュータを介した人と人とのコミュニケーションの情報流通を向上させる。

### (3) 情報表現／操作の連続性

ハイパーメディアの良さは内部表現と外部表現／操作との親和性が高いところにある。このような親和性を保つことは、利用者の思考遮断を防いだり、おぼろげな検索手がかりも活用できたりする可能性があるなどの理由で有益である。

### (4) 人間の記憶と検索手がかりとの関係

人間の短期記憶・長期記憶の特性を考慮したハイパーメディアはまだ実現されていない。しかし、Bush 以来のハイパーメディアの目標は、記憶の外在化を行って情報の有効活用をはかることにあり、人間の記憶との対応関係は避けて通れない課題である。このような構成指針はまだないが、人間のチャック（人間が認識／記憶できるオブジェクトの個数は約7個）を考慮した提示方法が提案されている段階である。今後は、人間の記憶と親和性のよい情報手がかり／表現法が検討される必要がある。

## 5.6 本章のまとめ

本章では、ハイパーメディアの編集・利用効率向上をめざすシステム機能、構成として Augmented Hypermedia の考え方を提案し、効用について議論を行った。従来のハイパーメディアでは、ハイパーメディアを作成・編集する人々と、得られたハイパーメディアを利用する人々との集団とは異なり、各々別々の役割を担っていた。しかし、Web のような情報編集・提供者と、情報利用者との垣根が低くなっている環境では、情報利用からえられた情報を新たな編集情報として活用するなどして、ハイパーメディアの編集・利用効率を高める機構が必要となる。Augmented Hypermedia では、利用者の行動様式、コンテンツ解析、オントロジー情報等を活用して、関係情報の共有、精緻化、抽象化プロセスを導入し、編集・利用効率を向上させる。このようなプロセスを実現することにより、編集・利用プロセスに対する正のスパイラルを作り出し、付加価値を生み出す高品質ハイパーメディアの発展をめざす。

本章では、関係情報の共有、精緻化、抽象化という側面から機能を検証し、3層構成としての Augmented Hypermedia のシステムアーキテクチャを規定した。3層構成のシステム構成法は、筆者らが開発した Web ベースのイメージ検索システムにも適用し、米国美術館での運用実験でその有用性を確認した。今後の展望としては、Augmented Hypermedia の考え方を一般化し、Information Brokerage のシステムへの応用をめざすことがあげられる。情報提供者と情報利用者との間の信頼性、問い合わせサービス、著作権管理等の付加機能を有し、安心して情報活用が行えるハブ環境を提供するものである。

## 第6章 まとめと今後の展望

### 6. 1 高品質ハイパーメディア研究のまとめ

本論文では、Web 情報活用のしかたの優劣による情報利用格差が増大してきている状況をふまえ、大規模な情報洪水に埋もれて、最新の役に立つ必要な情報をなかなか探しだせないという問題（情報空間の迷子問題）、リンクの爆発を防ぐためにどう編集すればよいかという問題（ハイパーメディアの設計問題）、ハイパーメディアを介した書き手と読み手との認識共有に関する問題（情報のトラスト問題）との解決に焦点をあてた。技術的な側面から課題解決に寄与する部分に限定し、ハイパーメディアの関係（リンク）に関する高品質化についての研究開発とその成果・意義についてまとめた。Web のような大規模ハイパーメディアを対象としたコンテンツの品質を高め、情報リテラシーの習熟度の違いのある種々の利用者を想定して、情報活用効率をあげる仕組みが必要とされる。これらの問題は、大規模情報を扱うデータベース・データモデルからのアプローチと、人間を系に含む情報処理サイクルモデルからのアプローチとの統合による解決が重要となる。

#### 6. 1. 1 高品質ハイパーメディアのモデル、機能、プロセス

高品質ハイパーメディア実現に際しては、ハイパーメディアの編集・利用効率をあげるモデル、機能、プロセスに焦点をあて、特に、下記の4点から詳細化した。

##### (1) 大規模ハイパーメディア関係構造の分析と把握

まず、現状のハイパーメディア関係構造を分析し、見通しのよい関係構造の抽出をはかった。高品質化のためのハイパーメディア関係構造の分析と把握に対する基本的指針として、関係構造に基づく抽象化の提案と導入を行った。これは、ハイパーメディアのリンク構造のなかから、完全2部グラフを発見し、マクロなリンク構造の抽出を行うものである。ただし、完全2部グラフのみに着目するとまとまった大局的な構造抽出が困難なので、例外を考慮したマクロなリンク構造の抽出を行う。例外のコストを考慮することにより、一元的な評価基準で、比較的まとまった大きさの大局構造を抽出することが可能となる。このような関連情報の抽象化操

作により、コンテンツの共有理解を促進させたり、オーバビューダイアグラムの実現手段としたりして、情報空間の迷子問題解決に寄与する。

本論文で詳述した関係構造の抽象化の考え方は、現在の Web コミュニティ（共通性ある Web サイト群）の発見に発展、展開される。これは、Fun と呼ばれる Web サイトの集合から Center と呼ばれる Web サイトの集合への完全 2 部グラフの抽出を基本とした、同じ興味をもつコミュニティの発見を行う方式である。より大きなクラスター間の大局的な関係の流れを把握しようとする、本論文で提案しているような欠損リンクを含む例外を考慮する必要が顕在化する。

### (2) ハイパーメディア論理設計に対する指針と構築

次に、これから高品質な関係構造を有するハイパーメディアを設計、構築するにはどのような指針がよいかについて説明した。データベース的なアプローチの良さを取り入れたハイパーメディア設計、関係構造に着目した概念設計、ハイパーメディア関係構造の自動生成についてプロトタイプ開発を例に説明した。具体的には、データベース設計時に有用な E-R スキーマを援用した関係構造の抽象化設計、ハイパーメディア抽象化の概念を具体的プロセスに拡大適用させて、関係構造に着目した抽象化概念設計、データベースからハイパーメディアの関係構造を自動生成する Hypertext Projection について説明した。体系的なアプローチにより、人為的に陥りやすいリンク情報の品質劣化（規定し忘れ、張り間違い、変更し忘れ、消去し忘れ等）を防ぐ。このような Web、ハイパーメディアコンテンツの信頼性向上策は、現在では、Web Content Management という製品領域で実現、運用されてきている。

### (3) 大規模ハイパーメディアに対する情報提示と検索・ナビゲーション

さらに、高品質ハイパーメディアを用いた情報提示と検索・ナビゲーション機能への適用について、ツールの提案とプロトタイプ開発を行った。Focus+Context View は、ハイパーメディア全体構造把握のための視覚表現であり、対象情報の構造的、意味的近傍にある情報を具体表示するだけでなく、情報空間におけるブラウジング指標として意味ある情報（Landmark Node）とその関係も表示させる。Landmark Node 抽出に際しては、リンク構造、アクセス頻度、Web サイトにおける階層の深さ情報等を組み合わせて重要度判断を行うものである。また、集合操作リンクと呼ぶ一種の動的リンクモデルに基づくハイパーメディアナビゲーションについても説明した。ナビゲーションの際に、「観点」というガイドを導入することにより、編集者の意図と利用者の検索行為との差異を吸収し、付加価値のある検索ナビゲーションを実現している。



### (4) ハイパーメディアを自律的に発展、運用させるための仕組み

最後に、得られた高品質ハイパーメディアを、自律的に発展、運用させていくための機能、システム構成について議論した。Augmented Hypermedia とは、情報利用からえられた情報を新たな編集情報として活用し、明示的にハイパーメディアの利用効率を高めるものである。利用者の行動様式、コンテンツ解析、オントロジー情報を活用して、関係情報の共有、精緻化、集約化等をはかる。このようなプロセスを実現することにより、新しい付加価値のあるハイパーメディア構造の発展拡張をめざす。また、Augmented Hypermedia のシステム構成としては、アドホック的、個別的な関連づけの対応だけではなく、体系的・スケーラブルな処理が行えることを考慮し、提案の3層のアーキテクチャについて、具体的プロトタイプを交えて説明した。

### 6. 1. 2 情報空間の迷子問題に対する対処と知見

本論文では、人間を系にいたしたハイパーメディア情報モデルを明示・共有する意識が欠如していたとの仮説をもとに、高品質ハイパーメディアの実現アプローチを明確化し、情報空間の迷子問題に対する対処を試みた。具体的には、1次情報・2次情報の明示・視覚化（大規模情報空間の客観的かつ直感的な把握）と、利用者自身の検索目的の明確化（情報空間の視覚化、フィードバック化による利用者自身の検索目的の明確化・修正）を考慮したアプローチである。このような考え方を念頭におき、本論文では、関係構造の抽象化（第2章）と、その考え方の設計、利用（検索・提示）フェーズへの適用（第3章、第4章）により、情報空間の迷子問題に対する方策を検討した。以下、情報空間の迷子問題に対する対処と知見に関してまとめる。

本論文では、具体的には、トップダウン並びにボトムアップによる情報空間全体把握のための構造化技法、情報空間の構造視覚化表現、迷子問題に対処するための種々の検索、ナビゲーション技法について言及した。一般に編集コストと利用の便益とはトレードオフの関係にあるので、効率のよいバランスのとれた対応方策を実現することが肝要である。以下、各項目に対する知見をまとめる。

#### (1) 大規模ハイパーメディアの見通しのよい全体構造抽出

まず、大規模ハイパーメディアがどのような構造になっているかを分析することが必要である。本論文では、仮想的な意味のあるノード、並びにリンクの集合を

見出すという抽象化操作(アグリゲーション)を第2章で説明した。この考え方は、現在の Web Community の考え方に位置づけられる。一般に、ボトムアップ的なクラスタリング操作に基づくアグリゲーションでは、客観的な指標を用いた階層構造化により、スケーラブルな全体構造抽出が行える。半面、1つのクラスターが、大規模な強連結成分となり、他のクラスターが、小さなサイズのクラスターとして抽出されがちである。現在の Web 情報全体の構造解析や、第2章で解析した Hypertext on Hypertext でも、同様の現象が現れている。

このような、詳細分析が行いにくいサブ構造といくつかの微小構造に分解されるのは、例外処理を考慮していないためである。本論文で提案したハイパーメディア抽象化方式は、ある程度の例外の存在を認め、サブ構造がバランス化されたノード、リンククラスターを抽出する。また、効率的な探索アルゴリズムにより、スケーラブルな実現手法を提案した。また、トップダウン的な方法として、上記問題を回避できる。3章で説明した拡張 E-R モデルによる情報構造化はその一例である。

また、大規模ハイパーメディアの構造分析としては、アグリゲーションだけでなく、特定の重要ノードのみを抽出し、骨格のフレームワーク、プロセスを抽出することも有用である。第4章で説明した、Focus+Context View で用いる Landmark Node は、この考えに基づく全体構造抽出方式である。この方式では、抽象化を行わず、オリジナルのノードもしくはリンクの中から重要なものを選択し、他の情報を隠蔽することにより、結果的にシンプルでわかりやすい全体構造抽出を行うものである。

ここで、抽象(アグリゲーション)操作がよいか選択(セレクション)操作がよいかは、対象情報の意味的単位(Information Unit)による。一般に、まとまった意味的単位が1つの Web ページに対応している場合は、選択操作により Landmark Node のような代表ノードを抽出することが有効である。また逆に、複数の Web ページである意味的なメッセージを表現している場合は、抽象化操作による構造抽出がよりよい傾向を引き出す。たとえば、情報フィルタリングにおける特定サイト集合の抽出とその影響、波及効果判定などの応用である。

### (2) 大規模情報空間の視覚化表現

大規模情報空間の視覚化は、情報空間の迷子問題を解決する一番直接的な方法である。これは、日常生活で迷子にならないために、地図を活用することに似ているが、物理空間の視覚化とは異なり、表現の自由度が高い。理解、想起しやすい視

覚表現を規定することに加えて、表現上の制約が少ないことを利用した情報空間固有の視覚化表現を実現することが重要である。

本論文では、このような条件を満たす大規模情報空間の視覚化表現手法について説明した。たとえば、第4章の Focus+Context View では、物理空間における魚眼図表示のアナロジー（近傍は詳細表示、それ以外は程度に応じた概略表示）により、表示画面、表示紙面上の制約をクリアした表現方法である。また、データの更新や、利用者の視点に応じて、Landmark Node の構成やそれらの関係表現を変更できることも、情報空間の迷子問題解決に有効である。また、第2章のハイパーメディア抽象化操作によるオーバビューダイアグラムや、第3章のスキーマブラウザによるオーバビューダイアグラムも、大規模情報空間の概要表示手法として有用である。

これらの手法は、スナップショットとしての大規模情報空間表示を的確に表現する。反面、更新頻度の高い情報空間表示に対しては計算量が多いため、精度を保ったままでのリアルタイム追従性の実現は今後の課題である。

### (3) 検索・ナビゲーションの多様化

コンテンツ1次情報・2次情報の構造分析とその視覚化表現は、大規模情報空間の客観的かつ直感的な把握が行える手法である。次のステップとしては、得られた情報視覚表現を用い、利用者自身の検索目的を明確化させること、検索目的・行為の修正を行わせること、また、これら操作を支援する検索・ナビゲーション機能を多様化させることが必要である。

本論文では、集合操作リンクを用いた情報空間のトラバースの操作がこのような検索・ナビゲーションの一例である。利用者は、インタラクティブな検索・ナビゲーション操作の中から、対象情報を閲覧すると同時に、自身の検索目的、問い合わせを精緻化させる。また、Augmented Hypermedia 構成の一例として述べた AMORE Web イメージ検索システムの例では、リンクタイプとして、キーワード、分類項目のような2次情報だけでなく、色、形状といったイメージの形態情報や、周辺説明情報と組み合わせた2次情報も利用し、検索・ナビゲーションの多様化を行っている。AMORE システムの利用評価では、視覚的、形態的な手がかりによるナビゲーションが、情報空間の迷子問題を改善するトリガーになり、インタラクティブのプロセスを経て、意味的に関連のある情報コンテンツを引き出すという有用性も、米国美術館での実運用を通じて報告されている。以上述べたように、情報空間の迷子問題の解決に向けては、全体状況の把握とともに、利用者の

検索モデルに対するフィードバックと、柔軟な問い合わせ変更・更新処理の実現が有効である。

### 6. 1. 3 ハイパーメディアの設計問題に対する対処と知見

大規模ハイパーメディアの設計に際しては、ノードの設計だけでなく、多数のリンクを効率よく制御・利用する設計指針も重要である。本論文では、リンク数を軽減することによる物理構造設計と、トップダウン的なハイパーメディア論理構造設計とをを行い、リンクの爆発による関係構造の制御困難性を回避し、ハイパーメディアの品質向上を進めた。

#### (1) ハイパーメディアの物理構造設計への寄与

2章で述べたハイパーメディアのクラスタリング手法 (ACE クラスタリング) を適用することにより、格納すべきリンク数を軽減することができる。分析モデルによれば、約 25%の格納サイズを軽減することができる。クラスタリングされた集約ノード間のリンクが多く、相関があるほど、格納サイズの圧縮率が高くなる。従来のグラフ分割方式では、入力隣接行列 (ノードの接続関係を表現) にたいし、ブロック対角成分のみに、接続関係 (リンク) を纏めるようなクラスタリング方式である。これに対し、本論文での提案手法は、非ブロック対角成分も同様に考慮して、接続関係 (リンク) を纏め、集約リンクの有無をはっきりさせる方式である。従って、最適化のための探索範囲が広くなり、結果として、より効率的な物理構造設計に寄与できる。

#### (2) ハイパーメディアの論理構造設計への寄与

3章で述べたトップダウン的な論理構造設計を用いることにより、不必要なリンクの付与を回避し、制御・利用しやすいハイパーメディアを構成することができる。トップダウン設計にはいくつかの方法があるが、本論文の寄与は、リンクスキーマを、ノードのスキーマから派生して規定させるのではなく、E-R モデルの拡張や、リンク関係自身のスキーマを定義して、論理設計を行うことにある。このようなトップダウン設計を事前に行なうことにより、リンクの爆発を未然に防ぐことができる。

また、言語規約に基づくデータベース (関係モデル) からのハイパーテキスト構造生成は、編集者が個別のリンクを規定しなくてよいため、編集負荷の軽減に貢献する。これは、事前に言語規約によって、あらかじめ変換テンプレートを作る方式である。現在、深刻になっている深層 Web に対する編集効率化の先駆的活

動と位置づけられる。今後は、機械学習等の技術等を活用し、自律的に生成すべきリンクを認識し、リンクの自動生成が行える方式への発展が期待される。

### 6. 1. 4 情報のトラスト問題に対する対処と知見

情報流通コンテンツの高信頼性方策は、一般には大きな課題である。本論文では、高品質ハイパーメディアを情報流通媒体とみなし、如何にコンテンツの信頼性の程度を計量化し、どう信頼性を維持し、また信頼性が損なわれたときにどう改善していくかという観点で、情報のトラスト問題に対する対処と知見とをまとめる。

#### (1) コンテンツ信頼性の指標化（2次情報からの1次情報の重要度規定）

コンテンツの信頼性は、利用者がコンテンツ自身から判断するというよりは、コンテンツの作成者（組織）、評価者（組織）の信頼性から判断される性質のものである。一般には、2次情報と総称されるコンテンツにまつわるコンテキスト情報、環境情報の信頼性、確実性、重要度等を総合して、対象コンテンツの信頼性が判断される。たとえば、発信源の不確かな口コミ情報よりは、大手の新聞社、テレビからの発信情報の方が、コンテンツ信頼性が高いといえる。また、複数ソースからの情報を、重要度を加味して統合化できれば、さらに信頼性が増す。

高品質ハイパーメディアの研究開発では、ノードの重要度を評価する尺度として、種々の2次情報（コンテキスト情報）を活用した方式を提案、実現した。たとえば、Focus+Context View における Landmark Node の規定では、リンク構造、アクセス頻度、Web サイトにおける階層の深さ情報を組み合わせて重要度を判定した。これらの情報は、客観的な恣意性のない尺度であり、統計的に統合結果を活用することができる。また、集合操作リンクを用いたトラバースでは、観点（多次元分類の分類軸）というコンテキスト情報を用いて、利用者がインタラクティブにナビゲーション対象のコンテンツの重要度を判断しているとみなすことができる。このように、ハイパーメディアにおけるコンテキスト情報を用いた1次情報の重要度規定は、コンテンツ信頼性の指標化の1つと位置づけられる。

現在 Weblog と総称される Web コンテンツに対する評価、コメントづけ機構が脚光を浴びているが、このような Weblog 属性情報や、評判情報を利用した Web コンテンツの信頼性判断は、当該研究の延長上にあり、さらなる発展が期待できる領域である。

### (2) 高品質ハイパーメディア設計・機能による信頼性維持

次に、いったん得られたコンテンツに対する信頼性をどう維持していくかという課題解決が重要である。通常は、コンテンツの信頼性については、利用者は特段の意識もなく情報を利用していることが多い。しかし、いったん何らかの原因で、対象コンテンツに対する信頼性が得られなくなった場合、利用者のコンテンツ信頼性に対する事象が顕在化する。たとえば、ハイパーメディア情報を利用する場合には、人為的に陥りやすいリンクの規定し忘れ、張り間違い、変更し忘れ、消去し忘れ等のリンク構造品質の劣化や、また、情報空間の迷子問題も間接的に、コンテンツ信頼性を落とす要因である。

このようなコンテンツの信頼性を落とさずに、維持・構築していくためには、ハイパーメディア設計時に、このような不確実要因をできる限り除去しておくことが望ましい。高品質ハイパーメディアの設計技法では、3章で述べたトップダウン的ハイパーメディア設計、論理構造の関係抽象化、Hypertext Projection などの設計技法を活用し、上述のような信頼性をそぐ要因を、設計時に除去、軽減している。このような考え方は、現在の Web Content Management の機能として、実現されている。

### (3) 情報フィードバック・自律化による信頼性再構築

さらに、コンテンツ信頼性の知見としては、仮に信頼性が損なわれたときにどう改善していくかという観点での向上施策も重要である。情報信頼性構築だけでなく、情報セキュリティの向上施策についても同様であるが、システム全体で完全な運用を想定することは現実的ではない。システムの脆弱性は、相対的には常に存在し、仮に信頼性が低下しても、対応できる自律的、フィードバックのあるメカニズムが必要である。たとえば、情報作成・編集者と情報利用者との間の情報目的意識や意図の違いは、日常的に存在するが、そのまま放置しておけば、情報解釈の違い、コミュニケーションの不備が、結果的にコンテンツ信頼性の低下につながる。

本論文では、Augmented Hypermedia においても言及したように、継続的な編集者と利用者との間の情報フィードバックや、システムの自律的情報蓄積により、ハイパーメディア情報を高品質に保つモデル、ツール、機能の実現に努めてきた。このような体系的なアプローチが、コンテンツの信頼性再構築に寄与できるものと考えられる。人間同士のコミュニケーションにおいては、不確実な情報に基づく信頼性低下を防ぐため、もとの情報発信者と確認を行ったり、関連情報からの分析をしたりして信頼性向上に努めている。このような日常生活の中での信頼性構築・維持のアナロジーを、Web やハイパーメディア情報処理に援用することも有用である。

第6章 まとめと今後の展望

以上、上述した環境課題、具体的解決施策、高品質ハイパーメディアの機能についての対応付けを図6-1にまとめる。

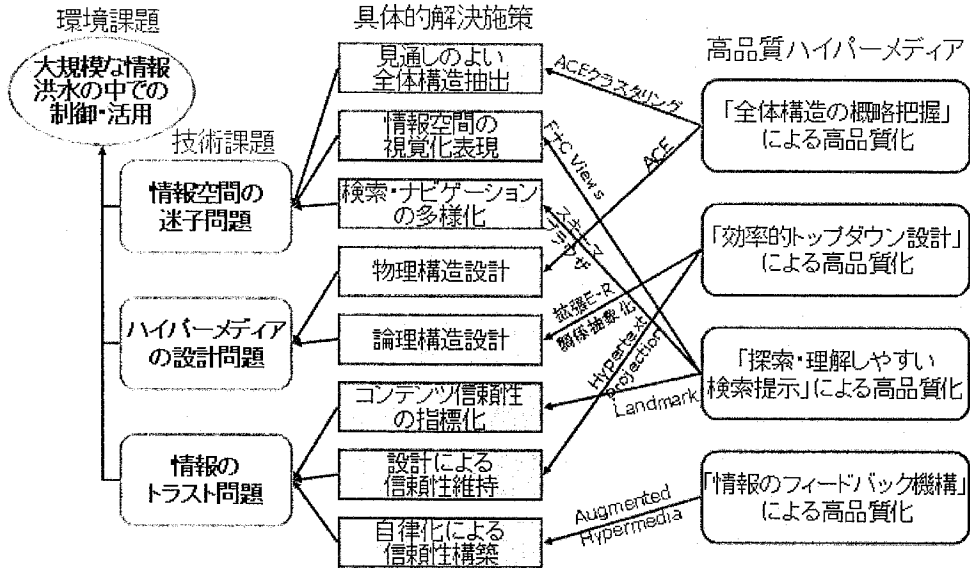


図6-1 課題に対する具体的解決施策のまとめ

## 6. 2 今後の研究展望

本節では、高品質ハイパーメディア研究活動に基づく今後の研究展望につき言及する。

### (1) グラフ理論と機械学習理論との統合発展（基礎アルゴリズム）と応用評価

近年、SVM(Support Vector Machine)や Graphical Model 等をはじめ、グラフ理論と機械学習理論との統合化が進展している。このよう基礎アルゴリズム進展の中で、自律型ハイパーメディアに対する理論的基盤構築や、動的変動に追随する Web マイニング、Web コミュニティ研究が今後の課題である。特に、グラフ理論に加えて機械学習理論を用いることにより、オンライン学習による適応型関係抽象化技法への応用が期待される。

### (2) ハイパーメディアデータモデルの発展に同期した関係品質向上

Web システムのアーキテクチャとしては、利用者を想定した HTML ベースのコンテンツアクセスモデルから、機械間での XML ベースのアクセスモデル (Web Service) へと進展している。また、対象となるコンテンツも、未分化な Web モデルから、被評価対象の Web と評価情報としての Weblog(Blog)とに分化したモデルが成熟化してきている。Weblog は、トラックバック情報といった双方向性を内在するモデルでもあり、一般に、このようなハイパーメディアデータモデルの発展に同期した関係品質向上方策は今後の課題である。従来のノードとリンクというシンプルなハイパーメディアデータモデルから、より情報空間の迷子問題や情報のトラスト問題を解決しやすいデータモデルへと移行できると考えられ、従来モデルとのコンパティビリティのある処理技法も今後の課題である。

### (3) ハイパーメディア・Web 情報活用に対する社会科学的側面からの研究

中長期的には、ハイパーメディアの情報科学的側面からの分析だけでなく、社会科学的側面からの分析、貢献が一層重要になると考えられる。すなわち、人間を系に含むハイパーメディアを考慮した場合、人間の行動様式 (ビヘイビア) 分析の高度化が必須である。すなわち、情報利用側の人間としてだけではなく、系の対象となる情報単位として人間の特性、アクションを含むビヘイビアを分析し、ハイパーメディア情報として統一的に扱うことが重要となる。このような人間のビヘイビア分析は、ビジネスインテリジェンスや情報監視系の応用では既に適用が開始されているが、ハイパーメディアの高度化、高品質化にも適用が望まれる。



## 第6章 まとめと今後の展望

さらには、近年のソーシャルネットワークに代表される信頼性を考慮した人的ネットワーク構築支援も重要な研究方向性である。コンテンツのトラストから、人的ネットワークを含むネットワーク化されたコンテンツに対するトラスト評価が一層重要になる。このような流れは、本論文で説明した高品質ハイパーメディアの発展的延長であり、今後の有望な研究課題である。社会、コミュニティにおいて、安心して快適な情報共有や情報活用を行える基盤構築が望まれる。

### 6. 3 結言

社会システムとは、一般に、「もの」と、「もの」と「もの」との「関係」（ないしは、「ふるまい」）によりシステム定義される。非常に抽象的であるが、単純かつ普遍的な考え方である。システム工学や計量経済モデルなど、種々の異分野でも同様の考え方に立脚してモデル構築、システム構築がなされてきている。Web情報やハイパーメディアも、このような理解においては、社会システムの一形態ととらえることができる。社会システムに対しては、階層化の考えなどを導入して、大規模かつスケラブルな適用、分析を可能としてきた。

筆者は、ハイパーメディアの高度化という研究に従事して、特に、社会システムにおけるループや、フィードバックのプロセスの重要性について認識を新たにした。グラフ理論的には、強連結成分 (SCC: Strongly Connected Component) に対する処理に対応するが、本質的に解析、把握が困難な部分である。他方、このような本質的に把握、困難な部分の中にこそ、本質的な情報が内在しており、このような地挑戦的課題への対処が、筆者の研究開発の原点である[HaKa84]。今後は、情報处理的側面と社会科学的側面双方に立脚した社会システムの高度化をめざし、研究開発の進展に寄与していく所存である。

## 謝 辞

本論文の作成にあたり、御指導、御鞭撻を賜りました京都大学大学院情報学研究科 田中克己教授、石田 亨教授、京都大学学術情報メディアセンター 喜多 一教授に謹んで感謝の意を表します。

また、適切な御教示を賜りました京都大学大学院情報学研究科 故上林彌彦教授に心より謝意を申し上げます。

本研究の遂行にあたり、様々な御支援、御教示を賜りました NEC 渡辺孝次郎氏、片山 博氏、平田恭二氏、Sougata Mukherjea 氏、Wen-Syan Li 氏、Rodrigo Botafogo 氏、笠原 裕氏、高野 元氏、川崎成人氏、小川隆一氏、原田浩明氏、田中栄市郎氏、館村純一氏、Stanford 大学 Gio Wiederhold 教授、Jeffrey Ullman 教授、Arthur Keller 教授、UC Santa Barbara 校 Divyakant Agrawal 教授、Arizona 州立大学 K. Selcuk Candan 助教授、並びに NEC 中央研究所、NEC Laboratories America, Inc. の関係各位に謹んで感謝の意を表します。

また、本研究を遂行するにあたりご配慮いただきました NEC 中央研究所 國尾武光執行役員、海老野征雄支配人、並木淳治支配人、日本電気特許技術センター 渡辺久恒社長、早稲田大学 後藤 敏教授、法政大学 山本昌弘教授、NEC 情報システムズ 真名垣昌夫執行役員常務、千葉大学 阪田 史郎教授に謹んで感謝の意を表します。

また、当初の研究開発より適切な助言を賜りました東京大学 喜連川 優教授、大阪大学 西尾章治郎教授、御茶ノ水女子大学 増永良文教授、Stanford 大学 Hector Garcia-Molina 教授、UC Berkeley 校 Lawrence Rowe 教授、Robert Wilensky 教授、UC Santa Barbara 校 Edward Chang 教授、Vienna 大学 Dimitris Karagiannis 教授に謹んで感謝の意を表します。

本研究の原点となりましたグラフ理論に基づく研究開発とその社会システムへの応用に際し、ご指導、ご鞭撻を賜りました東京大学 茅 陽一名誉教授に心より感謝の意を表します。



## 主要文献

- [MuHH99] Mukherjea, s., Hirata, K., and Hara, Y. "AMORE: A World-Wide Web Image Retrieval Engine," World Wide Web, Vol.2, No.3, 1999, pp.115-132.
- [MuHa99] Mukherjea, S., and Hara, Y. "Visualizing World-Wide Web Search Engine Results," Information Visualization, 1999, pp. 400-407.
- [HaHi98] Hara, Y., and Hirata, K. "Augmented Hypermedia: Issues on System Integration and Usability Study," FODO, 1998.
- [MuHa97] Mukherjea, S., and Hara, Y. "Focus + Context Views of World-Wide Web Nodes," ACM HYPERTEXT, 1997, pp.187-196.
- [MuHH97] Mukherjea, S., Hirata, K., and Hara, Y. "Towards a Multimedia World-Wide Web Information Retrieval Engine," WWW6, 1997, pp.1181-1191.
- [HHTK96-1] Hara, Y., Hirata, K., Takano, H., and Kawasaki, S. "Hypermedia Database "Himotoki" and Its Applications," IEEE ICDE, 1996, pp.372-379.
- [HaBo94] Hara, Y., and Botafogo, R. "Hypermedia Databases: A Specification and Formal Language," DEXA, 1994, pp.521-530.
- [HaKG91-1] Hara, Y., Keller, A., and Wiederhold, G. "Implementing Hypertext Database Relationships through Aggregations and Exceptions," ACM HYPERTEXT, 1991, pp.75-90.
- [HaKG91-2] Hara, Y., Keller, A., and Wiederhold, G. "Relationship Abstractions for an Effective Hypertext Design: Augmentation and Globalization," DEXA, 1991, pp.270-274.
- [HaKa90] Hara, Y., and Kasahara, Y. "A Set-to-Set Linking Strategy for Hypertext Systems," ACM COIS, 1990.

## その他関連文献

- [HaKa84] 原, 茅, 「強連結成分を有するシステムのグラフ自動作図化手法」, 計測自動制御学会論文集, Vol.20, No.6, 1984, pp.506-513.
- [LCHH01] Li, W., Candan, K., Hirata, K., and Hara, Y. "Supporting Efficient Multimedia Database Exploration," VLDB Journal, Vol.9, No.4, 2001, pp.312-326.
- [HMLH00] Hirata, K., Mukherjea, S., Li, W., and Hara, Y. "Integration of Image Matching and Classification for Multimedia Navigation," Multimedia Tools and Applications, Vol.11, No.3, 2000, pp.295-309.

- [LVAH99] Li, W., Vu, Q., Agrawal, D., Hara, Y., and Takano, H. "Powerbookmarks: A System for Personalizable Web Information Organization, Sharing, and Management," WWW8, 1999, pp.1375-1389.
- [LCHH98] Li, W., Candan, K., Hirata, H., and Hara, Y. "Hierarchical Image Modeling for Object-based Media Retrieval," DKE, Vol.27, No.2, 1998, pp.139-176.
- [HMLO98] Hirata, K., Mukherjea, S., Li, W., Okamura, Y., and Hara, Y. "Facilitating Object-based Media Retrieval," TAPOS, Vol.4, No.4, 1998, pp.261-283.
- [HMOL97] Hirata, K., Mukherjea, S., Okumura, Y., Li, W., and Hara, Y. "Object-based Navigation: An Intuitive Navigation Style for Content-oriented Integration Environment," ACM HYPERTEXT, 1997, pp.75-86.
- [HTOH96] Harada, K., Tanaka, E., Ogawa, R., and Hara, Y. "Anecdote: A Multimedia Storyboarding System with Seamless Authoring Support," ACM MULTIMEDIA, 1996, pp.341-351.
- [HHTK96-2] Hirata, K., Hara, Y. H. Takano, and S. Kawasaki, "Content-oriented Integration in Hypermedia System," ACM HYPERTEXT, 1996, pp.11-21.
- [HSH93] Hirata, K., Hara, Y., Shibata, N., and Hirabayashi, F. "Media-based Navigation for Hypermedia Systems," ACM HYPERTEXT, 1993, pp.159-173.

## 参考文献

- [AgBJ89] Agrawal, R., Borgida, A., and Jagadish, H. V. "Efficient Management of Transitive Relationships in Large Data and Knowledge Bases," ACM SIGMOD'89, 1989, pp. 253-262.
- [AsSi99] Ashman, H., and Simpson, R. "Computing Survey's Electronic Symposium on Hypertext and Hypermedia," ACM Computing Surveys, Vol. 31, No. 4, December, 1999.
- [Berg01] Bergman, M. "The Deep Web: Surfacing Hidden Value," 2001, <http://www.press.umich.edu/jep/07-01/bergman.html>
- [BoCT97] Bodner, R., Chignell, M., and Tam, J. "Website Authoring using Dynamic Hypertext," Webnet, 1997, pp. 59-64.
- [Borg85] Borgida, A. "Language Features for Flexible Handling of Exceptions in Information Systems," ACM Trans. on Database Systems, Vol. 10, No. 4, 1985, pp. 565-603.
- [BoSh91] Botafogo, R., and Shneiderman, B. "Identifying Aggregates in Hypertext Structures," ACM HYPERTEXT, 1991, pp. 63-74.

- [BoRS92] Botafogo, R., Rivlin, E., and Shneiderman, B. "Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics," ACM Trans. on Information Systems, Vol. 10, No. 2, 1992, pp. 142-180.
- [Bota93] Botafogo, R. "Cluster Analysis for Hypertext Systems," ACM SIGIR'93, 1993, pp. 116-125.
- [Brus96] Brusilovsky, P. "Methods and Techniques of Adaptive Hypermedia," User Modeling and User Adapted Interaction, 1996, pp. 87-129.
- [BrPa98] Brin, S., and Page, L. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," Computer Networks, Vol.30, 1998, pp.107-117.
- [Bush45] Bush, V. "As We May Think," Atlantic Monthly 176, 1945, pp. 101-110.
- [CACM88] "Special Issue on Hypertext," Communications of the ACM, Vol. 31, No. 7, 1988.
- [ClGa90] Clifton, C., and Garcia-Molina, H. "Indexing in a Hypertext Database," Proc. of the 16th VLDB Conference, 1990, pp. 36-49.
- [CoBe88] Conklin, J., and Begeman, M. L. "gIBIS: A Hypertext Tool for Exploratory Policy Discussion," ACM Trans on. Office Information Systems, Vol. 6, No. 4, 1988, pp. 303-331.
- [Conk87] Conklin, J. "Hypertext: An Introduction and Survey," IEEE Computer, Vol. 20, No. 9, 1987, pp. 17-41.
- [CrNo89] Cruz, I. F., and Norvell, T. S. "Aggregative Closure: An Extension of Transitive Closure," Proc. of the 5th Int. Conference on Data Engineering, 1989, pp. 189-204.
- [Fein88] Feiner, S. "Seeing the Forest for the Trees: Hierarchical Display of Hypertext Structure," Proc. of ACM Conference of Office Information Systems," 1986, pp. 205-212.
- [FILG00] Flake, G., Lawrence, S., and Giles, L. "Efficient Identification of Web Communities," KDD, 2000, pp. 150-160.
- [FrCo89] Frisse, M. E., and Cousins, S. B. "Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project," Proc. of the Hypertext'89, 1989, pp. 199-212.
- [Furn86] Furnas, G. W. "Generalized Fisheye views," Proc. of ACM CHI'86, 1986, pp. 16-23.
- [GaJS76] Garey, M. R., Johnson, D. S., and Stockmeyer, L. "Some Simplified NP-Complete Graph Problems," Theory of Computer Science, Vol. 1, 1976, pp. 237-267.

- [GaJo79] Garey, M. R., and Johnson, D. S. "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman and Company, 1979.
- [Good87] Goodman, D. "The Complete HyperCard Handbook," Bantam, 1987.
- [HaKa88] Hara, Y., and Kaneko, A. "A New Multimedia Electronic Book and Its Functional Capabilities," User-oriented, Content-based, Text and Image Handling (RIAO), 1988, pp. 114-123.
- [Hala91] Halasz, F. "Seven Issues: Revisited," ACM Hypertext'91 Keynote Talk, 1991.
- [HoMI03] How Much Information? 2003,  
<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>
- [Hype89] "Hypertext on Hypertext," ACM Press Database and Electronic Products Series, 1989
- [JAMS89] Johnson, D. S., Aragon, C. R., et al. "Optimization By Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning," Operations Research, Vol. 37, No. 6, 1989, pp. 865-892.
- [KeLi70] Kernighan, B. W., and Lin, S. "An efficient heuristic procedure for partitioning graphs," Bell Systems J., Vol. 49, No. 2, 1970, pp. 291-307.
- [KRRT99] Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A., "Trawling the Web for Emerging Cyber-Communities," Computer Networks, Vol. 31, 1999, pp. 1481-1493.
- [LaGi99] Lawrence, S., and Giles, C. L. "Accessibility of Information on the Web," Nature, Vol. 400, 1999, p. 107.
- [MoBB90] Moline, J., Benigni, D., et al. (eds.) "Proceedings of the Hypertext Standardization Workshop," National Institute of Standards and Technology, 1990.
- [NOST02] Nakajima, S., Oyama, S., Sumiya, K., and Tanaka, K. "Context-Dependent Web Bookmarks and their Usage as Queries," WISE, 2002, pp. 333-344.
- [Niel90] Nielsen, J. "Hypertext and Hypermedia," Academic Press, 1990.
- [PaSt82] Papadimitriou, C. H., and Steiglitz, K. "Combinational Optimization: Algorithms and Complexity," Prentice-Hall, 1982.
- [RoMN81] Robertson, C. K., McCracken, D., et al. "The ZOG Approach to Man-Machine Communication," Int. J. of Man-Machine Studies, Vol.14, 1981, pp. 461-488.
- [SmSm77] Smith, J. M., and Smith, D. C. P. "Database abstractions: aggregation and generalization," ACM Trans. on Database Systems, Vol. 2, No. 2, 1977, pp. 105-133 .



- [FaQi92] Tanaka, K., and Qian, Q. "Two-Level Schemata and Generalized Links for Hypertext Database Models," *Future Databases*, 1992, pp. 231-239.
- [Tomp89] Tompa, F.W. "A Data Model for Flexible Hypertext Database Systems," *ACM Trans. on Information Systems*, Vol. 7, No. 1, 1989, pp. 85-100.
- [Trig88] Trigg, R. H. "Guided Tours and Tabletops: Tools for Communicating in a Hypertext Environment," *ACM Trans. on Office Information Systems*, Vol. 6, No. 4, 1988, pp. 398-414.
- [Ullm84] Ullman, J. D. "Computational Aspects of VLSI," *Computer Science Press*, 1984.
- [UtYa89] Utting, K., and Yankelovich, N. "Context and Orientation in Hypermedia Networks," *ACM Trans. on Information Systems*, Vol. 7, No. 1, 1989, pp. 58-84.
- [Wied87] Wiederhold, G. "File Organization for Database Design," *McGraw-Hill*, 1987.
- [Wied92] Wiederhold, G. "Mediators in the Architecture of Future Information Systems," *IEEE Computer*, Vol. 25, No. 3, 1992, pp. 38-49.
- [WoKi87] Woelk, D., and Kim, W. "Multimedia Information Management in an Object-Oriented Database System," *Proc. of the 13th VLDB Conference*, 1987, pp. 319-329.
- [YHMD88] Yankelovich, N., Haan, B. J., et al. "Intermedia: The concept and the construction of a seamless information environment," *IEEE Computer*, Vol. 20, No. 1, 1988, pp. 81-96.
- [ZeKT04] Zettsu, K., Kidawara, Y., and Tanaka, K. "Aspect Discovery: Web Contents Characterization by their Referential Contexts," *APWeb*, 2004, pp. 738-743.