

A Study on Ubiquitous Information System for Utilization of Highly Distributed Information Resources

高度分散情報資源活用のための ユービキタス情報システムに関する研究

プロジェクトリーダー 安 達 淳
国立情報学研究所 教授



図1. 研究打合わせ風景



図2. OCRによる文書画像処理

1. 研究の目的

インターネットの普及に伴い、多様な情報がネットワーク上に分散して蓄積されるようになってきている。しかし、ネットワーク上の情報源は、その表現メディア、内容、記述の粒度、質、システム構成などさまざまな点で多様性を持っている。そのため、利用側からみると混沌とした情報源であり、利用に適した情報源とは言い難いのが現状である。マルチメディアを蓄積する技術については、近年、電子図書館システムの開発も進められているが、(1) データの所在が不明である、(2) データを蓄積するコストが大きい、(3) データを組織化するコストが大きい、(4) 蓄積されたデータを活用する手法が確立していない、など研究開発の必要な課題が多い。本研究は、ネットワーク上に分散する多様なメディアの情報資源の活用を、情報の解析、獲得、蓄積・交換、提供という4つの切り口で研究することによって、その所在を意識せずに自由に利用できるような次世代の情報ネットワークシステムの実現を目的としている。特に、多くの拠点に分散設置されたサーバ群が情報を提供するようなアーキテクチャを採用する方針を採り、利用者からは情報の所在を意識することなく、あたかも情報が広域ネットワーク全体に遍在する(ubiquitous)ようなイメージでとらえられるシステムを目指している。

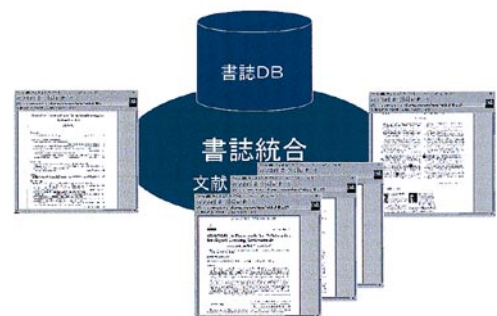


図3. 書誌統合のイメージ

2. 研究の内容

(1) 情報の分析

ユービキタス情報システムにおいて情報を活用する場合、利用者が必要に応じて情報を効率よく収集することが重要になる。テキスト情報は、計算機を用いて情報を収集する場合に中心的な役割を果たす。本研究では、既存のテキストを電子化する手法とともにそれを解析することによって、言語の特性、特にその統計的な性質を解明し、情報検索に適応する研究を進めている。学術情報に焦点を当て、情報検索に重要な役割を果たす専門用語の出現に関する統計的モデルの構築、文書の内容を端的に表すタイトルの言語構成モデル、学術的な文献における文書の機能構造のモデルについて研究を進めるとともに、これらのモデルの正しさや各種情報検索手法の有効性を検証するためのコーパスの構築を行っている。

(2) 情報の獲得

ユービキタス情報システムにおいては、情報は組織ごとに自律分散的に蓄積される。そのため、情報の表現レベルは多種多様となり、最も低レベルの画像における画素や文字の集まりから、SGML文書のような構造化された文書までさまざまなレベルが存在する。また各種粒度のマルチメディア情報の統合が課題になっている。これらの情報源に対して統一的なアクセスを実現するためには、画素や文字のようなデータの単位から、画像中に現れるオブジェクトやテキスト中の単語、パラグラフ、機能要素などの情報の単位を抽出することが必須となる。本研究では、データから情報単位を自動的に抽出する手法に関する研究を行っている。現在、ニュース映像を対象として動画像からその基本的なオブジェクトの一つである登場人物の情報を自動的に抽出する研究および文書画像から文書の基本構成要素である書誌情報を自動的に抽出する研究を進めている。

(3) 情報の蓄積・交換

ネットワーク環境での活用を前提とすると、オープンな情報資源の中から必要な情報を蓄積しているサイトを探索し、そのサイトが保有する大量の情報の中から必要な情報を高速に選択し、そのデータを利用者のもとに効率よく届ける技術が必要になる。本研究では、ユービキタス情報システムのプラットフォームとなる、データベースシステムおよび分散情報処理システムの研究開発を行っている。機能構造および文書構造を操作可能な全文検索エンジンの研究開発および高精度検索を可能とする構造化インデキシングについての研究を進めている。また、自律的に分散した電子図書館システムを構成するための情報オブジェクトの同定、識別、アクセス等の管理機能や通信方式に関する研究を実施している。

(4) 情報の提供

情報の活用において、マルチメディア情報に整合性を持つヒューマンインタフェースの高度化が円滑な情報活用には不可欠である。本研究では、利用者の目的に応じた情報の組織化とわかりやすい提示方法など、情報の可視化を中心に研究を行っている。情報の自動的な組織化の一環として、文書における出現頻度を特徴量として用いた専門用語のクラスターリング、文書-用語空間の可視化に基づいた情報空間の発見的探索、多言語情報に対する統一的な情報検索インタフェースを実現するための日英索引語の同義語の自動抽出についての研究を進めている。

3. 研究の体制

期 間：1996年10月～2001年3月

構 成：プロジェクトリーダー：安達 淳（国立情報学研究所・教授） コア

メンバー：3名、研究協力者：14名（内ポスドク2名） サブプロジェクト構

成：総括3名、分析4名、獲得3名、蓄積・交換2名、提供4名、評価2名

実施場所：国立情報学研究所 千代田区一ツ橋2-1-2

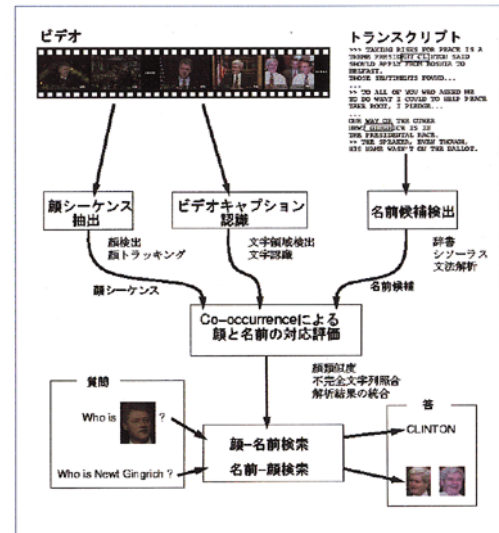


図4. ニュース映像からの人物の同定・検索

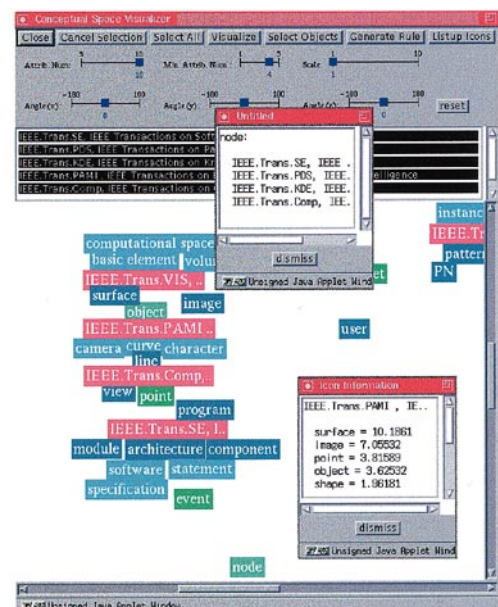


図5. 情報の可視化