

戦略的創造研究推進事業 CREST
研究領域「高度メディア社会の生活情報技術」
研究課題「連想に基づく情報空間との対話技術」

研究終了報告書

研究期間：平成 13 年 12 月～平成 19 年 3 月

研究代表者：高野 明彦
(情報・システム研究機構 国立情報学研究所
連想情報学研究開発センター長・教授)

1 研究実施の概要

1. 1 プロジェクトの概要

あらゆる文書情報の電子化が進み、それらの有効活用なしに、充実した知的生活は考えられなくなっている。しかし、現在の電子化された情報空間は、高信頼な知識獲得手段としてはもちろん、創造性を刺激するメディアとしても極めて不十分であり、遠く書籍に及ばない。「情報爆発時代の情報技術」などという問題の捉え方に端的に示されているように、現在日本で盛んに研究されている情報技術は、扱う情報の量的な増加にいかに対処するかという観点ばかりが重視され、人が電子情報を活用する場面での質的な問題は見過ごされがちである。しかし、当然ながらこの質的な問題こそが情報技術の社会的な価値を決める。

人が自分の考えを深めたり、新しい着想を得たりするために電子情報を活用するとき、そこで扱う情報の総量や網羅性はそれほど重要ではない。むしろ、望んでいるような思考の飛躍や気づきが自分の頭の中で実際に起きるかどうかが問題なのである。これは、数時間を費やして Google がオートマジカルにサジェストする「重要」なサイトを片っ端から眺めて歩くウェブサーフィンよりも、大した内容でもない1冊の新書を批判的に読む読書の方が自分の考えや発想を鍛えるのに役立つことから明らかである。

読書の有効性は、情報を理解しやすい構成・順序で摂取することにより、その情報についての系統的な文脈が頭の中に形成されるためだと考えられる。少なくとも良書はそうのように書かれている。その文脈の移り変わりに沿って、著者の示すアイディアと読者である自分の頭に浮かぶ問いかけをぶつけ合わせることが、新しい自分だけのオリジナルな発想の手がかりとなる。ここに読書の醍醐味がある。

本研究は、電子情報空間においても、このような読書に匹敵する効果的な情報体験の場を実現し、思考や議論を深めるための情報技術を確立することを目的とした。取り扱う情報の量ではなく、その情報利用環境で得られる経験の質を追及した。

我々のアプローチは「連想の情報学」という標語で要約できる。ここでは2つの「連想」を考えている。1つは利用者である人間の頭の中で起きている「連想」である。人は何らかの問いかけを外から受けると、無意識のうちにそれに対応する記憶が連想的に想起され、いくつかの文脈がウォームアップされる。1つ目はこの連想である。2つ目は、電子情報空間における「連想」を指す。統計的処理の安定する数万件以上の文書集合について、文書間や単語間の類似性を確率的計量で測ることにより、与えられた文書と類似する文書の収集や、文書集合の内容を要約する単語群の自動抽出が可能となる。これが2つ目の連想である。「連想の情報学」では、これらの2つの「連想」がどのように創造的に相互作用可能かを研究している。

5年間の研究では、この新分野に計算機科学の手法によりアプローチして、膨大な情報に基づく連想過程の計算的な基礎づけや、情報空間の内容的概観をフィードバックして人の連想を効果的に刺激する対話手法を開発した。この情報技術を実用上も意味のある文書情報に適用して、連想的対話環境を特徴とする **12種類の情報サービス** を実際に構築した。そのうちの **8サイトは一般公開** され、ユニークな情報サービスとして多くの

ユーザを集める人気サイトになっている。特に、**Webcat Plus**、**文化遺産オンライン**、**新書マップ**、**Book Town じんぼう**、**闊病記ライブラリー**の5サイトは、それぞれの分野の情報サービスとして国内 No. 1 の評価を得ている。

さらに、2006 年 7 月には、これらの特徴ある情報源を自在に組み合わせて仮想的に 1 つの情報源として連想探索できる技術を開発して、「**想—IMAGINE Book Search**」を開発した。この「想・IMAGINE」は、電子情報空間での新しい「読書環境」と呼ぶに相応しいシステムである。本研究プロジェクトで追及してきた実用的な情報技術としての「連想の情報学」研究の最大の成果となった。

1. 2 研究の構想

「連想の情報学」というアプローチの技術的な基礎は、我々が 1996 年以来研究開発してきた **Dua/NAVI** および **GETA** という 2 つのシステムに求められる。これらは本研究提案時にすでに完成しており、本研究では、これらの研究的蓄積をベースに、**GETA** を基本コンポーネントとして活用することで初めて実現可能となる高信頼な連想的情報アクセスシステムの構築を目指した。さらにそれらを実用上も意味のある本物のコンテンツに適用して、社会的に意義のある情報サービスの立ち上げに繋げることを最優先した。これにより提案技術の実用性を実証するとともに、**上質の情報サービスを一種の公共財として社会へ提供すること**を最大の目標とした。

この研究方針は、論文等の学術的成果を偏重する普通の基礎研究プロジェクトにならないようにとの自省を込めて設定した。論文では研究成果の実用性を謳っているが、実用性のまったくないおもちゃのようなデータへの適用実験しかクリアできない「基礎技術」や、研究プロジェクト終了と同時にデータ更新が止まってしまう「実証サイト」の構築には意味がないと考えた。本研究の目的は、我々の構築する連想情報学が従来の基礎研究とは一線を画する本物の実用的情報技術であることを実証することとした。

1. 2. 1 [参考] **Dua/NAVI** と連想計算エンジン **GETA**

(1) 連想的文書検索システム **Dua/NAVI**

連想計算に基づく情報アクセスの有効性を示すため、1996 年より我々は連想的文書検索システム **Dua/NAVI** (デュアル連想検索) を開発してきた。その特徴は、検索時にキーワードではなく文書(群)を入力して類似の文書を検索する文書連想検索と、検索結果として得られる文書群の内容的把握を助ける特徴語グラフ(自動要約機能)の有機的連携である。新聞、百科事典、学術論文誌など実用規模 DB への適用を通じて、原始的な形とはいえ、**Dua/NAVI** は創造的相互作用の場を提供することに成功し、「連想の情報学」の有効性を示した。注目すべき学術情報検索システムとして Google などとともに英国 Nature 誌でも紹介された (Vol. 405, pp. 112-115, 11May2000)。Stanford 大学図書館で注目されるなど世界レベルのオリジナリティだったと思うが、専用アプリが必要で標準ブラウザ上では動作しないなどの限界があり、実用システムとしては普及しなかった。この反省が本研究での実証システム構築に生きている。

(2) 汎用連想計算エンジン GETA

DualNAVI の研究を通じて、連想計算（文書群同士、単語群同士、文書群-単語群間の類似性関連性計算）こそが最も基本的で重要な計算機構であると確信した我々は、大規模な文書集合について各種の連想計算を高精度かつ高速に実行できる汎用連想計算エンジンを開発した（IPA独創的先進的情報技術プロジェクト、平成11～13年度）。

GETA (**G**eneric **E**ngine for **T**ransposable **A**ssociation) は、1千万件規模の文書コーパスを対象に、文書連想検索や文書群要約など創造的相互作用の実現に必要な連想計算を高精度かつ高速に処理するソフトウェアである。その性能の秘密は、インデクス情報の圧縮法と分散処理のためのインデクス分割方式にある。

GETA は「連想の情報学」における各種実験システムの構築ベースとして、文書連想検索、文書クラスタリング、語彙連鎖抽出などの研究や商用システムの実装に広く利用されている。**GETA** の最新版はオープンソース形態で2002年7月から無償公開している (<http://geta.ex.nii.ac.jp/>)。GETA はその後も進歩を続けており、現在の最新版は64bitアーキテクチャに対応し、簡便に連想計算のWebサービスを立ち上げる仕組みも付加されている。**GETA** は同種のソフトウェアの中で恐らく世界一の性能であると考えている。

1. 3 研究アプローチ

「連想の情報学」の実用性を実証するため、**GETA** の連想計算を各種の情報源に適用して、連想的対話環境を特徴とする次の12個の情報サービスを構築した。（*印は一般公開済みのサービス、並びは構築時期の順）

- *Webcat Plus（国立情報学研）：大学図書館1000館の図書目録、985万冊
- ・XML電子辞典システム（NPO連想出版）：岩波辞典7種の統合検索
- *文化遺産オンライン（文化庁・総務省）：博物館・美術館100館、文化財5000点
- *新書マップ（NPO連想出版）：テーマ別選書ガイド、1030テーマ、9000冊
- ・pictopic 毎日 PhotoBank（NPO連想出版）：テーマ別報道写真、1000テーマ
- *Book Town じんぼう（NPO連想出版）：神保町ポータル、165店、日々在庫29万冊
- *千夜千冊マップ（NPO連想出版）：松岡正剛氏の書評、1144夜
- ・Stanford マップ（NPO連想出版）：Stanford大学のDB4種
- ・文化遺産遊歩館（文化庁）：テーマ別文化財ガイド、100テーマ、700点
- *明治古典会遊歩館（NPO連想出版）：テーマ別古書オークション出品、100テーマ
- *闘病記ライブラリー（NPO連想出版）：病気別闘病記ガイド、57分類、700冊
- *想-IMAGINE Book Search（NPO連想出版）：Wikipedia等の異種情報源7種

1. 3. 1 連想する情報サービス

上記で構築したいわゆる「連想する情報サービス」は次のような特徴を持っている。

(1) 専門知識を必要としない情報サービス

連想する情報サービスの第1の特徴は、特別な専門知識がなくても利用できることで

ある。キーワードはもちろん、普通の文章を丸ごと質問として指定して連想検索できる。検索結果を要約する関連語（トピックワード）が表示されるので、そこから次の検索語を選ぶことができる。また、検索結果から複数の正解を選んで再検索することにより、所望の情報を検索できる（関連性フィードバック）。これらの特徴により、情報空間に奥行きや安心感を与える対話技法といえる。

(2) 関連情報サービスとの動的連携機能（図 1.3.1）

異なる情報源がそれぞれ連想する情報サービスとして立ち上がっていれば、それらの間の相互連携は極めて簡単かつ精度がよい。それぞれの情報サービスでは、自然文による連想検索がサポートされているので、1つの情報源についての検索結果を関連語群で要約し、それを受け渡すことにより、ユーザの意図を汲み取りながら異種情報源へ渡り歩くことが可能になる。さらには、複数の情報源に同じ質問文を投げて、返される検索結果の関連語だけを見て、どの情報源が所望の情報を保有していそうか判定することも可能になる。

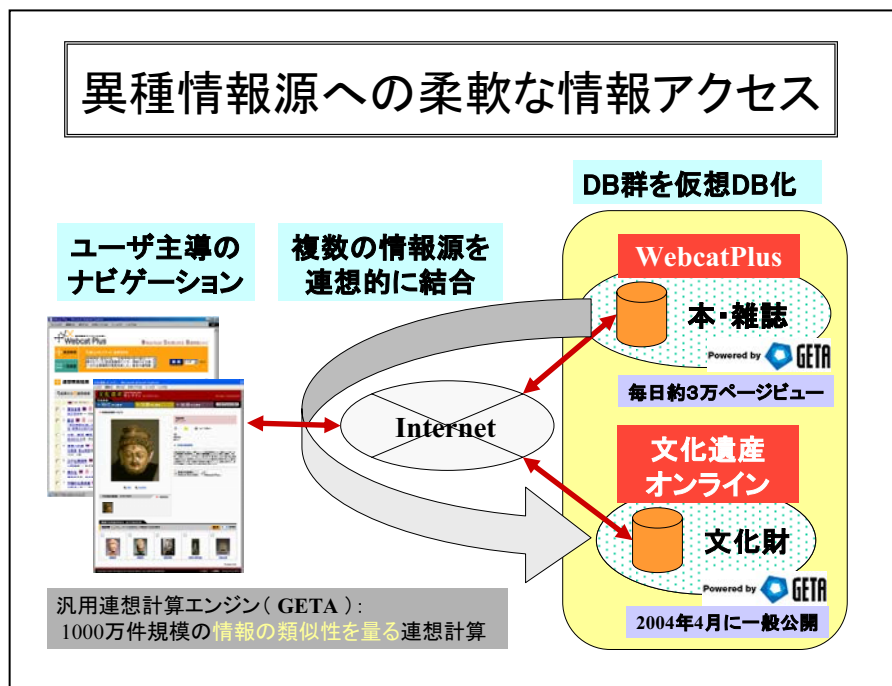


図 1.3.1 異種情報源への柔軟な情報アクセス

1. 3. 2 「連想の場」としての図書館リファレンス

上記で構築した「連想する情報サービス」を活用した新しい図書館のリファレンスコーナーのサービス構築を行っている。2007 年 5 月にリニューアル開館予定の東京都千代田区立千代田図書館のリファレンスコーナーに、我々の構築した「想・IMAGINE」や新書マップを活用した情報探索システムが設置される予定である。新書マップ収録の新書・選書に RFID タグを貼り、それをリファレンスツールとして、リーダーにかざすだけで各種の情報サービスが利用できる環境を構築している（図 1.3.2）。これにより図

書館リファレンスは次のように生まれ変わると考えられる。

(1) 情報に文脈（コンテキスト）を与える場

本の情報を本以外の情報源に関連づけることができる。例えば、利用者が選択した本を手がかりに、関連する新聞記事、報道写真、美術館・博物館収蔵品等々を連想的に収集できる。逆に、本以外で得た情報を本の情報に関連づけることもできる。これにより、報道情報の裏を取ったり、自分では理解できない情報から、より入門的な文章で書かれた書籍を見つけて理解を深めることも可能になる。これにより、ユーザは自分の興味をどこまでも掘り下げることができ、また、いつでも理解可能な情報に辿りつけるので、情報空間に奥行きと安心感が与えられる。

(2) 情報に自分だけの文脈を発見する場

人の記憶は言葉のネットワーク（文脈）として表現されていると考えられる。その意味で、人が思い浮かべる言葉の文脈は常にその人の記憶から紡ぎだされた個人的なものである。文脈を受け取り文脈を返す連想とは、まるで指紋のようにその人に固有のものといえる。多様な情報源を使ってユーザの連想を引き出す新しい図書館リファレンス環境は、情報に自分だけの文脈を発見する場ともいえる。



図 1.3.2 新書で引く新書マップインタフェース

1. 3. 3 情報に文脈を与える「想・IMAGINE」

「想・IMAGINE」は、連想計算の Web サービスが利用可能な情報源を複数集めて、そこから関連情報を連想検索で一気に収集できる。さらに、収集された情報の中から気に入ったものにチェックを入れて、それらを基点に関連情報を再収集できる。見かけは

A9 の単なる横断検索に似ているが、収集された情報を使って、次の情報探索の基点を自在に作り上げられる点が本質的に新しい。Google や Yahoo の次にくる情報源発見・利用技術であると考えられる（図 1.3.3）。

「想」IMAGINE

想いを連ね、想いを映し、その響きに耳を澄まそう



図書館の蔵書DB、古書店の在庫DB、テーマ別新書ガイド、百科事典、博物館収蔵品DB、新聞記事DB、報道写真アーカイブなど、さまざまなジャンルの情報源から、関連する情報を連想計算で収集して一覧表示します。その中からさらに深く知りたい情報を指定して、それを基点に連想を深めることが可能です。従来の検索では不可能だった、多様な情報によって自分の想いを確かめながら深めていける情報探索が経験できます。人間の連想を刺激して発想力を高める、まったく新しい情報利用環境をお楽しみください。

図 1.3.3 情報に文脈を与える「想・IMAGINE」

2 研究構想及び実施体制

2. 1 研究構想

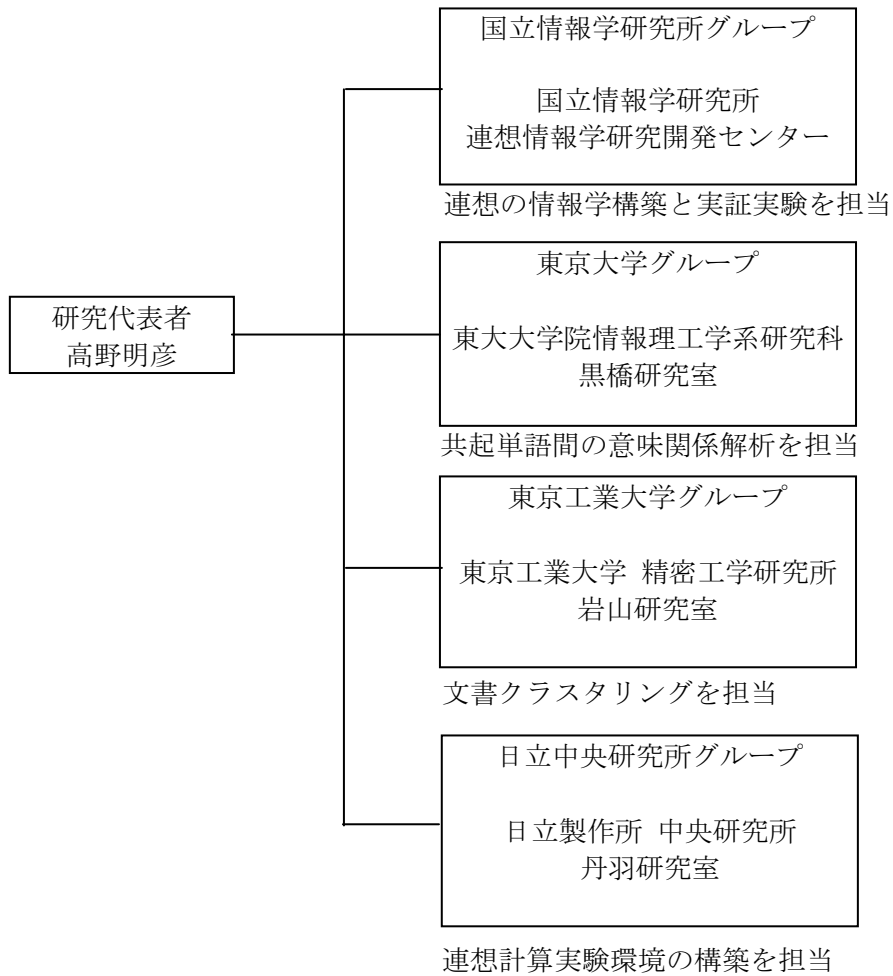
2. 1. 1 公共財としての情報サービスの構築

上で述べたように、本研究の主要なアウトプットとして、世の中で広く使われる実用的な情報サービスを構築・提供することを目指した。現在、電子情報空間の利用手段は、Google や Yahoo などの商用システムに大きく依存していて、商業主義とは独立に信頼性の担保されている公共財のような情報サービスで実用的なものは非常に少ない。SNS や Blog の普及により個人の日記のような情報が爆発的に増え、平均的な電子情報の品質は極端に劣化している。これに対する一つの対抗手段として、本研究の目指した信頼に足る情報源の公共財として情報サービス構築の意義は大きいと考える。

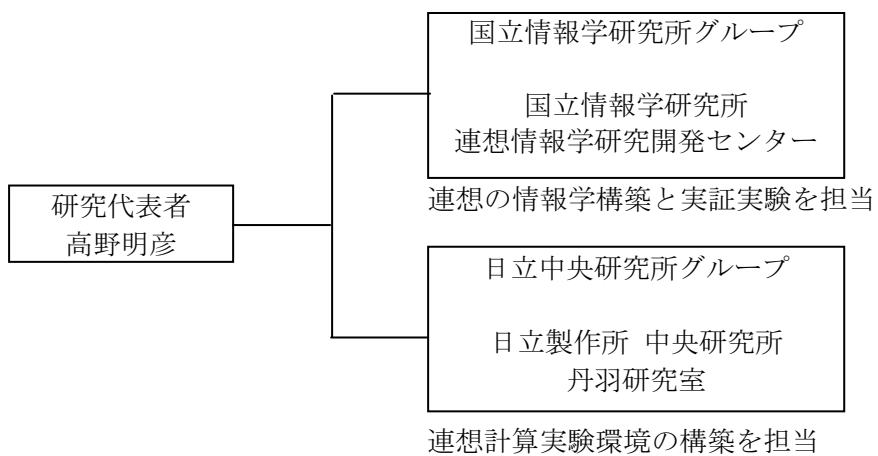
2. 2 実施体制

プロジェクト開始後 1 年半は、4 チームに分かれて要素技術の検討を行ったが、後半は国立情報学研究所に研究資源を集約して、情報サービス構築に専念する体制とした。この集中により、多くの有用な情報サービスを構築できたと考えている。

[H13.12～H15.3] :



[H15.4～H19.3] :



3 研究実施内容及び成果

上記の実施体制の項目に書いたように、主な研究は国立情報学研究所のチームで推進したため、この項はそのチームについてのみ報告する。

3. 1 連想の情報学構築とその実証実験

(国立情報学研究所 連想情報学研究開発センター)

「連想の情報学」の実用性を実証するため、GETA の連想計算を各種の情報源に適用して、連想的対話環境を特徴とする次の12個の情報サービスを構築した。以下、各システムの概要を紹介する。

3. 1. 1 NII 図書情報検索サービス Webcat Plus (<http://webcatplus.nii.ac.jp/>) 公開中

国立情報学研究所が運営する図書情報検索サービス (図 3.1.1)。全国の大学図書館 1000 館の所蔵図書情報に加え、1986 年以降に発行された日本語図書 96 万冊の目次・概要を含む BOOK データベース、英語圏で現在入手可能な英語図書 568 万冊に関する情報 Nielsen Book Data などが検索できる。自由な自然文からの「連想検索」と書誌情報を指定しての「一致検索」をサポートしている。日本語図書・英語図書のどちらについても、これだけのタイトル数が目次や概要の情報で検索できるサイトは他に見当たらない。データベースの検索機能を利用しているように見える「一致検索」についても、性能面から GETA を利用して実装している。文字列 N-gram を単語として GETA のインデックスを作成している。国立情報学研究所が提供する学術情報サービス GeNii の中核となっている。



図 3.1.1 Webcat Plus

3. 1. 2 XML 電子辞典システム

岩波書店の7つの辞典（情報科学辞典、6つの小事典）をまるで一つの辞典のように引いて活用できるシステム。それぞれの項目の本文中、辞典の見出し語については自動的にリンクが張られる。連想検索による関連項目検索、KWIC 検索なども完備している。項目本文を使って、他のデータベースを自動的に連想検索する機能もある。図 3.1.2 の例では、岩波書店発行の書籍から関連書を検索している。

XML で準備された辞書であれば簡単に追加可能であり、連想機能を十分に活用した辞書利用環境として一定の実用性を備えている。しかし、人名解説などの項目は、生没年、生誕地、国籍、業績などの記述が定型になっているために、人物の項目の関連項目としては、人物ばかりが挙げられてしまうなどの問題点は残った。辞典全体の項目数が合計7千と少ないことも原因のひとつであった。



図 3.1.2 XML 電子辞典システム

3. 1. 3 文化遺産オンライン (http://bunka.nii.ac.jp/) 公開中

国立情報学研究所が構築し、文化庁・総務省が運営する日本の文化遺産情報のポータルサイト。2004年4月27日に一般公開した。現在、全国100館以上の博物館・美術館から提供された文化財5085点を連想検索できる。

文化財についての知識がなくても調べられるように、時代から探す、分野から探す、地域から探すという3つの軸で検索できる。検索結果は5×4=20枚の写真で一覧できる。写真をクリックすると個々の文化財についての詳しい情報が読め、さらに大きい写真を眺められる。個々の文化財の説明ページの下には、もっとも関連性が高い文化財5つが並ぶ。気に入った写真にチェックを入れて連想すれば、それらを中心に関連文化財が集められ、さながらマイ企画展が開かれる。

説明文の下のWebcat Plusボタンを押すことにより、関連書をWebcat Plusで連想検索できる。これは連想する情報サービスの動的連携(図1.3.1)の実例になっている。



図 3.1.3 文化遺産オンライン

3. 1. 4 新書マップ (http://shinshomap.info/) 公開中

2004 年 6 月、連想検索を応用した新しい情報サービス「新書マップ」を立ち上げた。現在流通している約 7000 冊の新書・選書が約 1000 のテーマに分類され、テーマごとに本のリストと読書案内を読むことができる。トップページで自由な長文やキーワードを与えて検索すると、それに関連する 10 テーマが選ばれて、星座図のように配置された「マップ」が現れる。円内のテーマノードをクリックすると本のリストが背表紙写真とともに表示される。円の周囲に配置された 12 個の関連語ノードをクリックすると、その語で再検索され、また新しいマップが返される。テーマのページは、拾われている本繋がりに関連テーマへリンクされている。本のページでは、表紙写真とともに目次や概要を読むことができる。

基本は人手で丹精して作られた静的な 1000 個のテーマだが、それを連想計算でユーザの興味に応じて動的に集めてマップを作ることにより、無限の深みを演出している。Flash を用いた動きのあるマップの見せ方と、書棚のアフォーダンスを想起させる背表紙写真の組み合わせにより、リアルな書棚では絶対できない「書棚ブラウズ」を可能にしている。書棚で見るテーマ一覧では、ジャンルの近いテーマの書棚写真を展開して眺めることができる。「世界で一番長い書棚」と呼んでいる。

情報の品質を保証しながら、連想機構を用いてユーザ要求への適応性を高める手法には普遍性があり、これこそが次世代ディレクトリサービス技術であると考えられる。

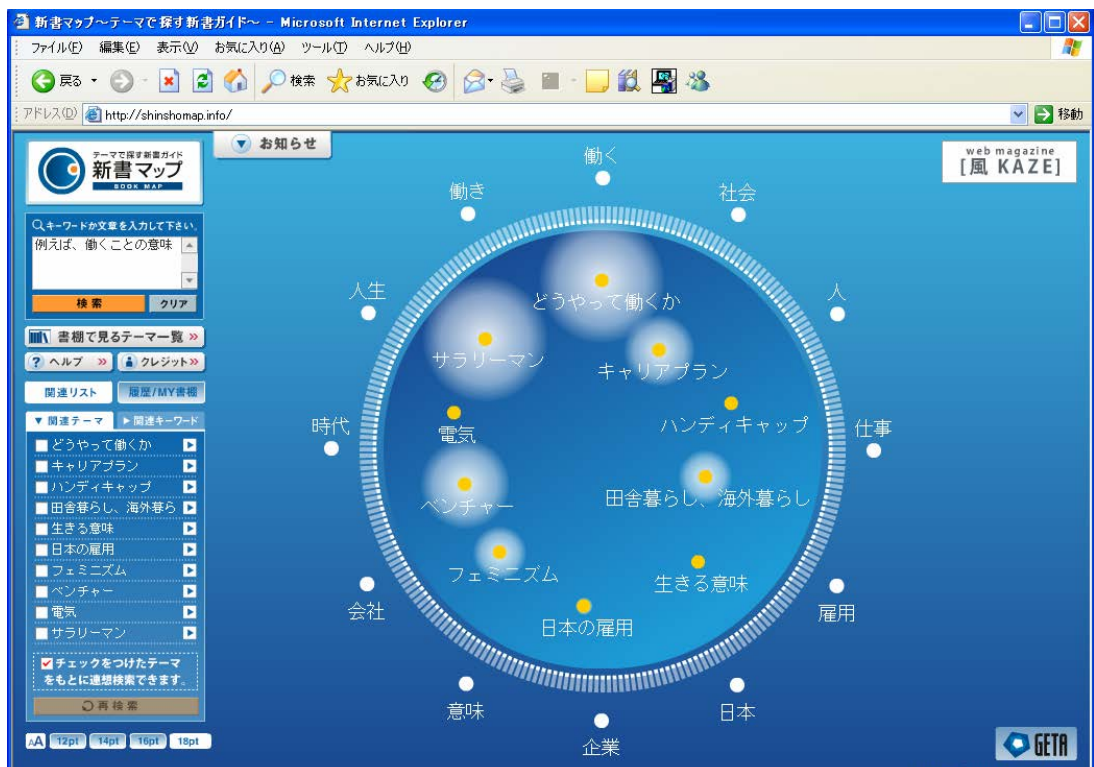


図 3. 1. 4 新書マップ(1)

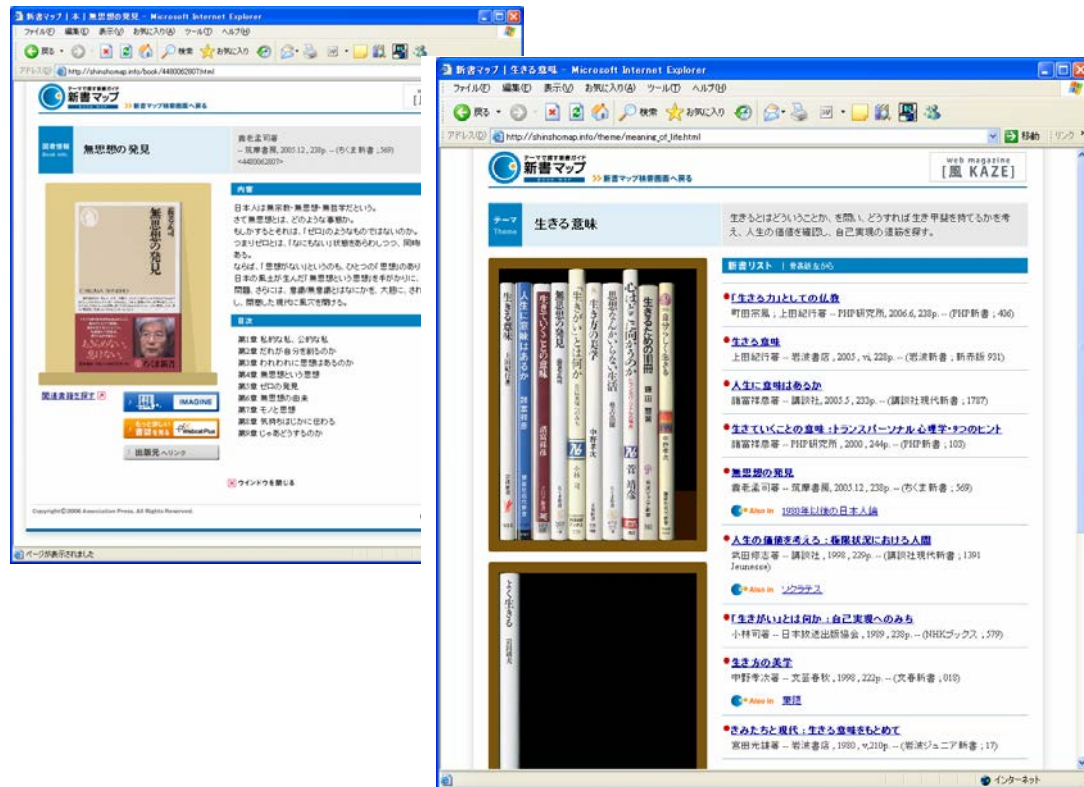


図 3.1.5 新書マップ(2)

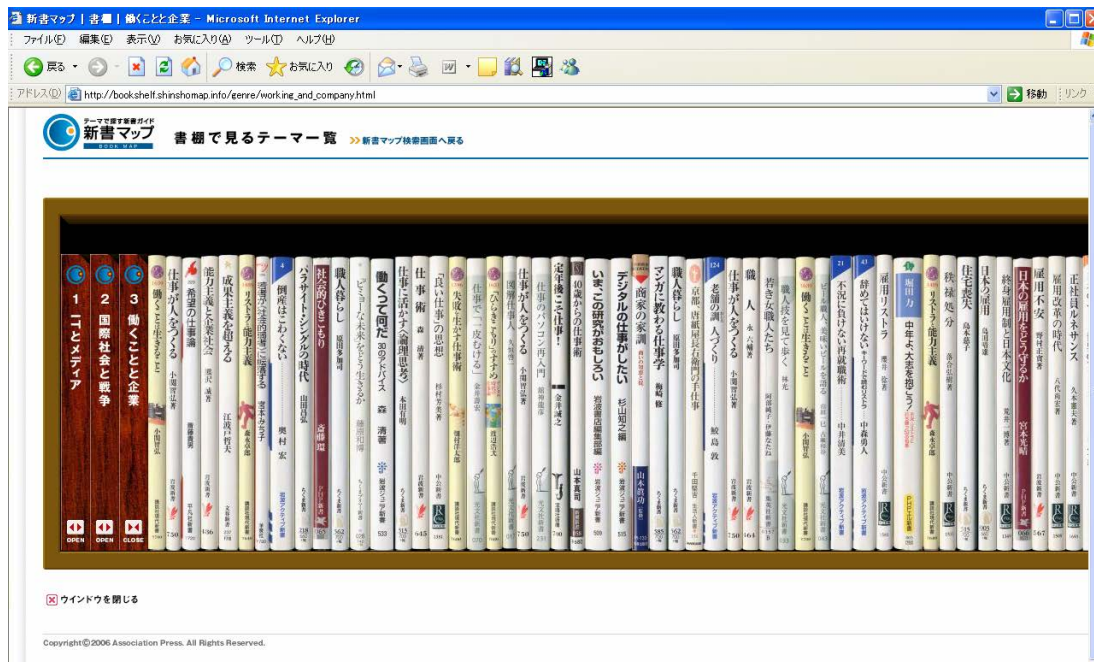


図 3.1.6 新書マップ(3)

3. 1. 5 pictopic 毎日 PhotoBank

毎日 Photobank は、毎日新聞社が保有している報道写真 25 万枚のデータベースで、特に明治～昭和初期について定評がある。ここからテーマ 1000 個をベテラン記者の手で作ってもらった。各テーマには 10 枚ずつの写真が選ばれている。これを新書マップと同様に連想検索可能にしたのが、pictopic 毎日 PhotoBank である。写真の一覧性やスライドショー機能を取り入れたインターフェースになっている。



図 3. 1. 7 pictopic 毎日 PhotoBank (1)



図 3. 1. 8 pictopic 毎日 PhotoBank (2)

3. 1. 6 Book Town じんぼう (http://jimbou.info/) 公開中

古書店 170 店を擁する世界一の本の街、神田神保町の古書店連盟の全面的な協力を得て、神田神保町のオフィシャルポータルサイト「Book Town じんぼう」を立ち上げた。個々の古書店の紹介ページ、古書店の在庫情報を一括して連想検索可能な古書データベース、得意分野からお目当ての古書店を探して地図に表示するジンボウナビなどの機能を備えている。神田神保町にある膨大な知的資源への案内役として、多くのユーザに支持され、人気サービスとして定着しつつある。



図 3.1.9 JIMBOU(1)



図 3.1.10 JIMBOU(2)

3. 1. 7 千夜千冊マップ (http://senya.pictopic.info/) 公開中

松岡正剛氏の書評サイト「千夜千冊」に掲載中の 1144 夜分の書評を、新書マップと同じ Flash インタフェースで連想検索可能にした。思いがけない本の間の関係性に気づかされる。



図 3.1.11 千夜千冊マップ

3. 1. 8 Stanford マップ、公開準備中

Stanford 大学図書館との共同研究の一環として、**GETA** の連想計算サーバを Stanford 図書館内に立ち上げ、試行データベース4つに適用した。連想の基本機能を確認するために、新書マップと同様のインタフェースを提供している。このインタフェースでは、連想検索する対象のデータベースを切り替えられるようにしているため、異なるデータベース間の連想的な動的連携についても確認できる。

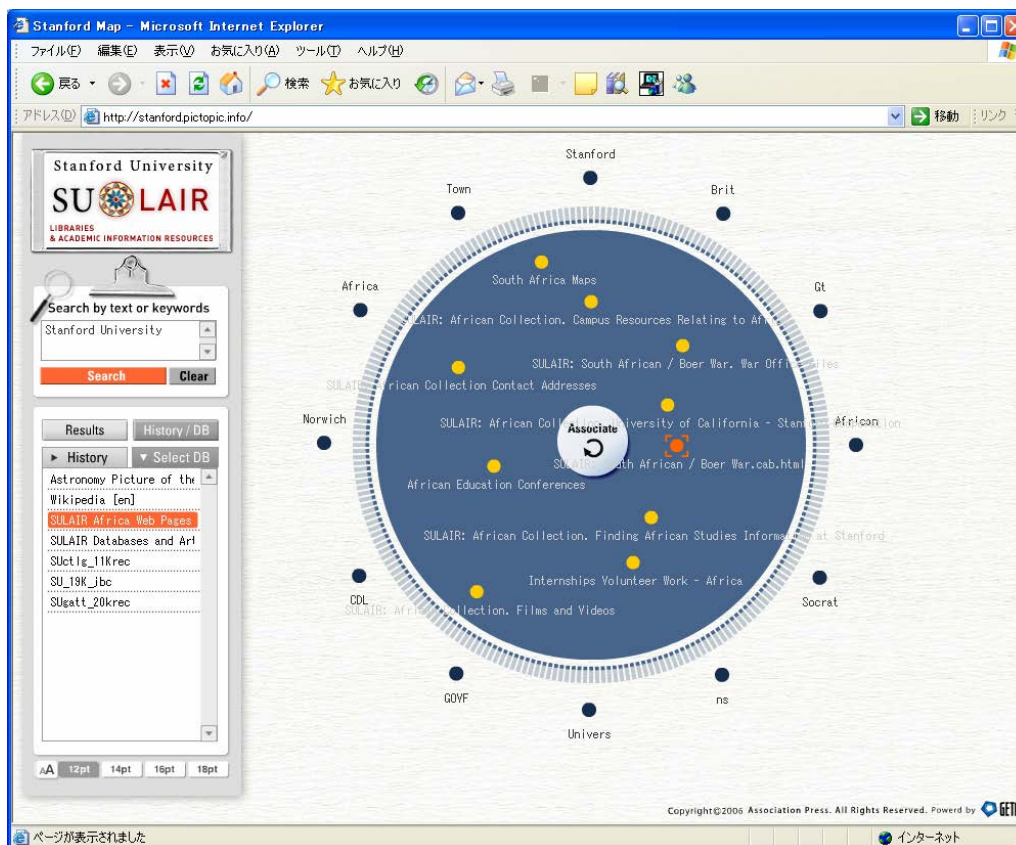


図 3.1.12 Stanford マップ

3. 1. 9 文化遺産遊歩館、公開準備中

文化庁の依頼により、文化遺産オンラインに収録されている文化財をテーマに分類した形で閲覧できるリーフレット型のナビシステム「遊歩館」を作成した。現在100テーマに700点ほどの文化財が人手で分類されている。文化財の詳細情報については、対応する文化遺産オンラインのページへ飛ばす形で提供している。この遊歩館と文化遺産オンラインの関係は、ちょうど新書マップと Webcat Plus の関係に似ていて、人間が作った雑音のないきちんとした情報の文脈を、連想検索で機械的に広げられる世界へリンクしていく構造になっている。

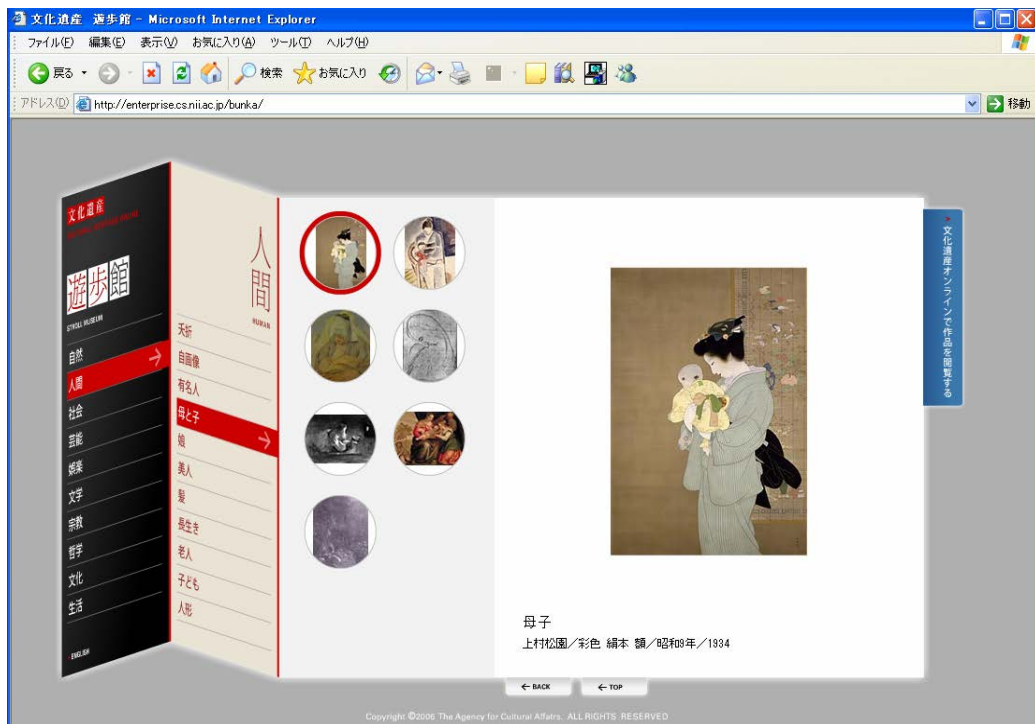


図 3. 1. 13 文化遺産遊歩館



図 3. 1. 14 文化遺産オンライン

3. 1. 10 明治古典会遊歩館、(<http://meiko.jimbou.info/yuuhokan/2006/>) 公開中

遊歩館のインタフェースを生かして、毎年七夕に神保町で開かれる明治古典会出
展品データを使って、古書オークションの世界を紹介するサイトを作った。人手で
100 テーマ 1000 作品を選び、リーフレット型のナビシステム「遊歩館」で閲覧で
きるようにしている。

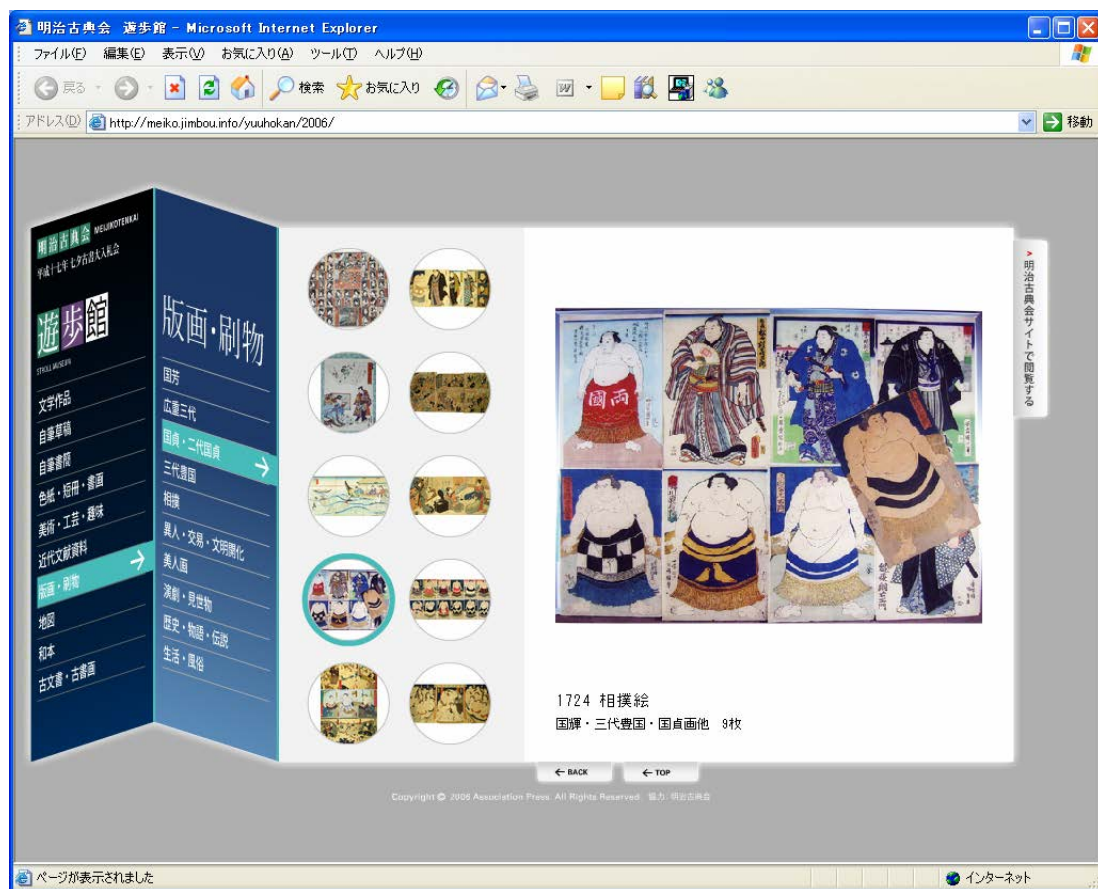


図 3. 1. 15 明治古典会遊歩館

3. 1. 1 1 闘病記ライブラリー (<http://toubyoki.info/>) 公開中

医学系図書館司書や看護学研究者などの有志からなる健康情報棚プロジェクトでは、公共図書館に良質の闘病記を分類して収める「闘病記文庫」設置推進運動を進めている。インフォームドコンセントが広がる中、患者や家族は心の拠り所として闘病記を求めるが、良質の闘病記を見つけるのは難しい。タイトルから何の病気か分からないものも多く、従来の図書分類では、バラバラに配架されてしまう。新書マップの出発点が「各社の棚に分かれて内容で探すのが難しい」新書や選書を何とかうまく活用できないかということを出し、上記プロジェクトに協力して、Web上に仮想闘病記文庫を作ることになった。我々が文部科学省から受託している研究プロジェクトからも書籍購入費等の援助が得られることになり、本研究でも仮想的な読書インタフェースを作るという部分について貢献することとした。

本の概要の読み取り、病気への分類等の専門的作業は、健康情報棚プロジェクトのメンバーにお願いして、我々はWeb上にどれだけ使い心地のよい仮想書棚と読書環境を作れるかに専念した。なお、今回は目次だけでなく本文の前書きの部分もスキャンして、出版社や著者の許諾が得られた本については、それらをWeb上で読めるように公開している。

2006年6月12日に一般公開したが、NHKニュースや大手新聞で大きく取り上げていただき、社会的な関心の強さを日々感じている。その後、リアルな闘病記文庫を設置済みの鳥取県立図書館や聖路加看護大学健康情報コーナー（るかなび）などでは、併せて闘病記ライブラリー専用端末を置いて、リアルとバーチャルの両方から調べられる環境を提供している。



図 3. 1. 16 聖路加看護大学にて日野原理事長へデモ

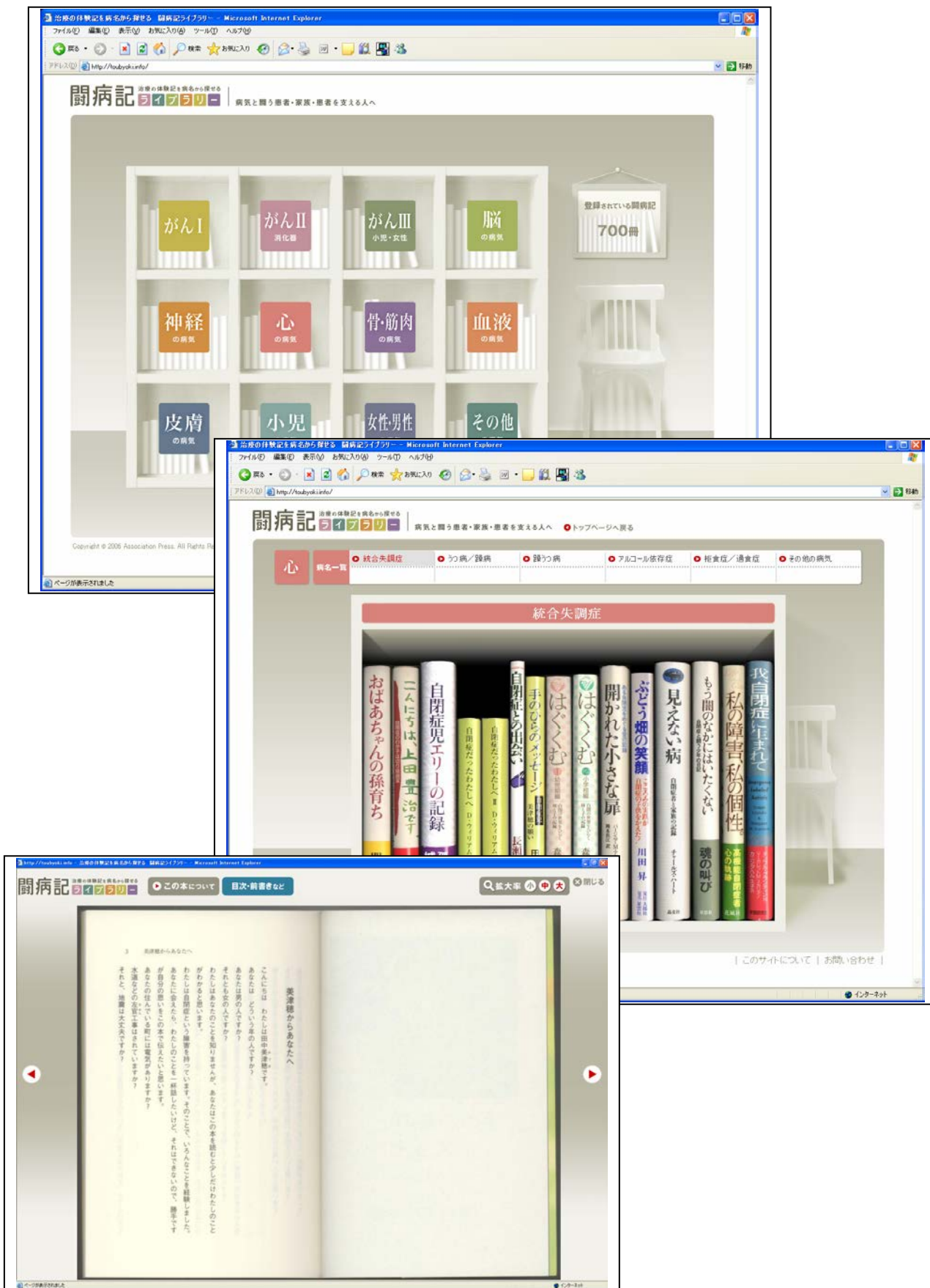


図 3. 1. 17 闘病記ライブラリー

3. 1. 12 想—IMAGINE Book Search (http://imagine.bookmap.info/) 公開中

これまで紹介した情報サービスは、ほとんどすべて連想計算の Web サービスを利用して構築されている。「想・IMAGINE」は、これら Web サービスが利用可能な情報源を複数集めて、そこから関連情報を連想検索で一気に収集できる。さらに、収集された情報の中から気に入ったものにチェックを入れて、それらを基点に関連情報を再収集できる。AJAX を使って情報源の増減をスムーズにできる点や見かけは A9 の単なる横断検索に似ているが、収集された情報を使って、次の情報探索の基点を自在に作り上げられる点が本質的に新しい。

この「想・IMAGINE」技術を使って、2006 年 7 月に、想—IMAGINE Book Search というサービスを公開した。現在は、Webcat Plus、新書マップ、Book Town じんぼう、文化遺産オンライン、ウィキペディア、松岡正剛の千夜千冊など 7 つの情報源が使える。今後、情報源を増やしていく計画である。



図 3.1.18 「想—IMAGINE Book Search」

3. 2 研究成果の今後期待される効果

新書マップは、企業の関心も高く、いずれは独立採算でサービスを維持していけると期待できる。出版社との信頼関係も得られたので、今後は出版社から直接書籍についての信頼できる情報を自動収集する仕組みなどを検討していきたい。

文化遺産オンラインを通じて得られた博物館や美術館の方々との信頼関係は貴重なもので、すでにいくつかの博物館・美術館とは直接共同研究を開始する計画もある。国のすべての文化財情報を文化遺産オンラインに集めるという文化庁の構想には無理があるとしても、分散型での同種のサービスには可能性が見えてきている。

JIMBOU は、神田神保町のポータルとして完全に定着した。NPO 神田学会が進めている町おこしの運動などとも連動しながら、JIMBOU 専用の PC を町に配備するなど、よりコミュニティと直結した情報サービスの形を模索する。

以上、いずれも発展が期待される情報サービスであるが、我々は「想・IMAGINE」が、それらのどれよりも大きなインパクトを与えていると考えている。索引情報すら一ヶ所に集中させることなく、それらを連想的に融合できる「想・IMAGINE」は、Google や Yahoo の次にくる情報源発見・利用技術であると考えられる。Stanford 大学図書館は「想・IMAGINE」の先駆性にいち早く注目し、さっそくこの技術を使って 8 百万冊の蔵書カタログ検索を行うことを計画している。また、次のステップとしては、Google Print プロジェクトで全文スキャン+文字認識している蔵書データについて、想・IMAGINE 技術を適用することを検討している。

国内でも、人文系の史料編纂所や博物館、国会図書館等で作成・維持されているデータベースについて、連想計算サーバを提供して、組織やコミュニティの枠を越えて「想・IMAGINE」技術で、コンテンツが出会い、ユーザの想いと出会う場所を提供していく計画である。

フィールドとしては、新千代田図書館を始めとする公共図書館のリファレンスコーナーが有望である。図書館であれば、有料コンテンツへのアクセスをサイトライセンスという形で問題なくクリアできるので、「想・IMAGINE」の軸として有料・無料の情報源が隣り合いながら、お互いに競い合うこともできる。専門辞典を数多く制作してきた大手出版社も「想・IMAGINE」技術に興味を示しており、実験用にこれまで門外不出だった辞典の XML データの提供も受けている。本当に価値のある知識がこうして「想・IMAGINE」技術によって、再び多くの人たちの日常的な知的活動に生かされる日が来ることを信じている。

4 研究参加者

4. 1 国立情報学研究所グループ（連想の情報学構築の研究）

氏 名	所 属	役 職	研究項目	参加時期
高野 明彦	国立情報学研究所	教授	連想の情報学	H13年12月～ H19年 3月
西岡 真吾	国立情報学研究所	特任教授	連想計算エンジン	H17年 4月～ H19年 3月
丸川 雄三	国立情報学研究所	特任助手	文化財情報の連想計算	H15年 4月～ H19年 3月
湯原 葉子	国立情報学研究所	CREST 研究員	連想する情報サービス	H14年 5月～ H19年 3月
小池 勇治	国立情報学研究所	プロジェクト 研究員	連想計算の Web サービス	H17年 9月～ H19年 3月
萱島 礼香	国立情報学研究所	CREST 研究員	連想する情報サービス	H17年 2月～ H18年 6月
荒井 紀子	国立情報学研究所	CREST 研究員	連想する情報サービス	H17年 6月～ H19年 3月
絹川 博之	東京電機大学	教授	XML・Web サービス	H13年12月～ H19年 3月
小池 勇治	東京電機大学	修士課程	XML・Web サービス	H15年 4月～ H17年 3月
藤波 淑子	国立情報学研究所	CREST 研究員	連想する情報サービス	H16年 7月～ H17年 6月
林 祐子	国立情報学研究所	CREST 研究員	連想する情報サービス	H16年 8月～ H17年 7月
佐藤 一郎	国立情報学研究所	助教授	エージェント・システム	H13年12月～ H17年 3月
村田 剛志	国立情報学研究所	助教授	情報空間の可視化	H13年12月～ H17年 3月
古山 宣洋	国立情報学研究所	助教授	情報空間のアフォーダンス	H13年12月～ H17年 3月
神門 典子	国立情報学研究所	教授	ジャンル分析	H13年12月～ H17年 3月
江口 浩二	国立情報学研究所	助手	Web 文書検索	H13年12月～ H17年 3月
黒橋 禎夫	東京大学	助教授	共起単語間の意味関係解析	H15年 4月～ H17年 3月
黒橋 禎夫	京都大学	教授	共起単語間の意味関係解析	H18年 4月～ H19年 3月
岩山 真	東京工業大学	客員助教授	文書クラスタリング	H15年 4月～ H17年 3月

4. 2 東京大学グループ（共起単語間の意味関係解析の研究）

氏名	所属	役職	研究項目	参加時期
黒橋 禎夫	情報理工学研究科	助教授	共起の意味関係解析	H13年12月～ H15年 3月

4. 3 東京工業大学グループ（文書クラスタリングの研究）

氏名	所属	役職	研究項目	参加時期
岩山 真	精密工学研究所	客員助教授	動的クラスタリング	H13年12月～ H15年 3月
丸川 雄三	精密工学研究所	助手	コーパスデータ整備	H13年12月～ H15年 3月

4. 4 日立中央研究所グループ（共通実験環境整備の研究）

氏名	所属	役職	研究項目	参加時期
丹羽 芳樹	中央研究所	主任研究員	共通実験環境整備	H13年12月～ H19年 3月
久光 徹	中央研究所	主任研究員	共通実験環境整備	H14年 5月～ H19年 3月
西岡 真吾	中央研究所	主任研究員	共通実験環境整備	H14年 5月～ H15年 8月

5. 招聘した研究者等

氏 名（所属、役職）	招聘の目的	滞在先	滞在期間
Michael A. Keller （Stanford 大学図書館長、 HighWire Press 発行人）	共同研究打合せ、 「本の未来・未来の本に ついてのシンポジウム」 招待講演	国立情報学研究所	H17 年 3 月～ H17 年 3 月

6 成果発表等

6. 1 原著論文発表 (国内誌7件、国際誌0件)

1. 高野明彦 「古書街で情報サービスの未来を想う」, 情報管理, Vol. 49 (2006), No. 7.
2. 高野明彦 「インターネットと情報の信頼性」, 情報管理, Vol. 49 (2006), No. 3 p.142-144.
3. 高野明彦、西岡真吾、丹羽芳樹、連想に基づく情報アクセス技術—汎用連想計算エンジン GETA を用いて—、情報の科学と技術, Vol.54, No. 12, pp. 634—639, Dec. 2004.
4. 高野明彦、連想に基づく情報空間との対話技術、心理学ワールド 27号, Oct. 2004.
5. 小池勇治、矢島匡人、高野明彦、絹川博之、Web サービスを利用した総合電子文書検索システム、FIT2004、Sept. 2004.
6. 丹羽芳樹、高野明彦、あいまい検索の技術と応用、雑誌『日本語学』 2004 年 2 月号、Vol.23 No. 2.
7. 小池勇治、高野明彦、絹川博之、複数辞典の鳥瞰が可能な XML 電子辞典システム、FIT2003、September 2003.

6. 2 その他の著作物

1. 高野明彦 「本のデジタル化は我々に何をもたらすか」, 文部科学時報, 2006.2.
2. 高野明彦 「インターネットで詩心を育てる」, 「情報と文化」2005 年冬号, 新潟日報社, 2005.12.
3. 高野明彦 「『連想』で本の未来は開けるか」, 「出版情報」2005 年 9 月号, 講談社, 2005.9.
4. 高野明彦 「古書の街、神保町で、本の未来を考える」, 「情報と文化」2005 年夏号, 新潟日報社, 2005.6.
5. 高野明彦 「バイオ・情報の最前線」, 丸善ライブラリー情報研シリーズ 5 (分担執筆), 2005.5.
6. 新書マッププレス (高野明彦) 編著、「新書マップ」、日経 B P、Nov. 2004.
7. 高野明彦、「文化遺産オンライン」試験公開版システムの概要、文化庁月報, 2004 年 7 月号.

6. 3 学会発表 (国際学会発表及び主要な国内学会発表)

6. 3. 1 招待講演 (国内会議8件、国際会議4件)

1. Akihiko Takano. “GETA, Associative Search Engine for Future Digital Library (Invited talk)”, The China Education and Research Conference, 2006.5.17-18, Beijing.
2. kihiko Takano. “Information Access by Association”, Keynote speech, The 21st Digital Asset Management Seminar, 2006.1.6, Hong Kong.
3. Akihiko Takano, “Mozume: Associative Information System for Cultural Heritage (Invited talk)”, International Conference on Digital Archive Technologies (ICDAT2004), Taipei, March 2004.
4. Akihiko Takano, “Association Computation for Information Access (Invited talk)”, Discovery Science, 6th International Conference, DS 2003, LNCS Vol. 2843, pp.33-44, Springer-Verlag, 2003.
5. 高野明彦、情報を発想力に変える連想エンジン (招待講演), 第5回 情報科学技術フォーラム FIT2006、福岡大、2006.9.6.
6. 高野明彦、”Information Access based on Association” (招待講演), 駒場、2006.7.26. 情報計算化学生物学会 2006 年大会.
7. 高野明彦、「情報を発想力に変える連想エンジン」 (基調講演), 2006.6.8, 国立情報学研究所オープンハウス 2006.

8. 高野明彦 「連想計算による情報アクセス」 (基調講演), 2005.11.10, 平成 17 年度全国公共図書館総合・経営部門研究集会.
9. 高野明彦、連想に基づく情報空間との対話技術 (招待講演), 言語処理学会第 9 回年次大会併設ワークショップ[メディア/モダリティ統合における言語処理]、横浜国大、2003.3.21.
10. 高野明彦、連想に基づく情報空間との対話技術 (招待講演), 2003 年情報学シンポジウム、日本学術会議講堂、2003.1.16.
11. 高野明彦、連想に基づく情報空間との対話技術 —連想の情報学を目指して (招待講演), 第 1 回 情報科学技術フォーラム FIT2002、東工大、2002.9.25.
12. 高野明彦、連想に基づく情報空間との対話技術—連想の情報学を目指して (招待講演), 人工知能学会第 16 回全国大会, 国立情報学研、2002.5.31.

6. 3. 2 口頭発表 (国内会議 17 件、国際会議 6 件)

1. Akihiko Takano. “Association Engine GETA for New Digital Library”, SUN Microsystem COE Meeting on Digital Library, 2006.3.9, New York City.
2. Akihiko Takano. “Associative Theme Map for New Directory Service”, The 2nd Franco-Japanese Workshop on Information Search (ISIP2005), 2005.5.11, Lyon.
3. 高野明彦 「連想の場としての本の世界」, 第 4 回 NII 国際シンポジウム「本の未来・未来の本についてのシンポジウム」, 2005.3.25.
4. Akihiko Takano, “Information Access based on Association”, The Third Workshop on Programmable Structured Documents, Jan. 2005.
5. Yuzo Marukawa, Akihiko Takano, Akio Takamizawa, “Mozume: Associative Information System for Cultural Heritage”, Nara Symposium for Digital Silk Roads (The second NII International Symposium), Nara, December 2003.
6. Akihiko Takano, Shingo Nishioka, Osamu Imaichi, Yoshiki Niwa, “Association Computation for Information Access”, APLAS2002, 上海交通大学, 2002.11.30.
7. 丸川雄三、「文化遺産オンラインにおける作品情報収集システムの構築と東京国立博物館との連携実験について」、東京国立博物館公開研究会「博物館情報学の構築」、2006.7.19.
8. 高野明彦 「文化遺産オンラインの目指す情報発信」, 2006.6.3, アート・ドキュメンテーション学会公開シンポジウム「ミュージアム・ドキュメンテーションの新時代」.
9. 高野明彦 「連想する情報サービス」, 2006.5.11, 毎日ユニバーサロンアクセシビリティセミナー「HCI としてのサーチエンジン」.
10. 高野明彦 「DB の新しい価値を引き出す情報サービスについて」, 2006.3.27, CCJP シンポジウム『誰でも簡単に使える著作権表示』の未来について考える.
11. 西岡真吾 「XML Information Set の C 言語向け実装」, 第 8 回プログラミングおよびプログラミング言語ワークショップ論文集 pp. 34-42 (2006.3).
12. 高野明彦 「連想する情報サービス」, 2006.2.2, PAGE 2006.
13. 高野明彦 「連想する情報サービス」, パネルディスカッション, 2006.1.26, 国立国会図書館公開シンポジウム「デジタル時代における図書館の変革」.
14. 高野明彦 「図書館システムでの RFID タグ付き新書マップ棚の活用」, 2005.11.8, SUN Japan ERC.
15. 高野明彦 「連想計算による情報アクセス」, 2005.10.26, OSPG 第 3 回基盤技術研究会セミナー.
16. 高野明彦 「連想に基づく情報アクセス」, 2005.10.25, DATABASE Tokyo 2005.
17. 高野明彦 「GETA システムの紹介」, 2005.9.30, NII サイエнтиフィック・オープンソース・ソフトウェアデー.

18. 高野明彦 「図書街プロジェクトと千夜千冊マップ」, NICT 図書街シンポジウム@東京国際フォーラム, 2005.9.8.
19. 西岡真吾 「トピックビューワーの実装」, 夏のプログラミング・シンポジウム「プログラミング道～その原点.その道を極める～」報告集 pp. 65-70 (2005.9).
20. 高野明彦 「本探しのための連想技術について」, 日書連 情報化推進・デジタルコンテンツ研修会, 2005.7.7.
21. 高野明彦 「連想計算による情報アクセス」, 電機連合図書館研究会, 2005.5.27.
22. 矢島匡人、小池勇治、高野明彦、絹川博之、XML 電子新書システム、FIT2004、Sept. 2004.
23. 西岡真吾、The Design of the Generic Engine for Transposable Association (汎用連想計算エンジン GETA の実装)、PPL2004, March 2004.

6. 4 特許出願（国内 0 件、海外 0 件、その他 0 件）

なし

6. 5 受賞等

6. 5. 1 新聞報道・テレビ報道

1. 闘病記雷雨ラリー完成——病種別にネットで検索（読売新聞：2006 年 8 月 16 日夕刊 2 面）
2. 情報クリック：一気に本を検索（東京新聞：2006 年 8 月 7 日夕刊 5 面）
3. 闘病記 700 冊、ネットで検索——無料サービス始まる（中国新聞：2006 年 6 月 21 日朝刊 [くらし] 13 面）
4. YAHOO!検索 時事ワードランキング（2006.6.12～6.18）（読売新聞：2006 年 6 月 21 日夕刊 2 版 9 面）
5. Patients' insights offered on Net (International Herald Tribune/Asahi：2006 年 6 月 13 日 28 面)
6. 闘病記のデータベースを公開（NHK ニュース：2006 年 6 月 18 日, 18:03～18:05, 他 3 回）
7. 闘病記 700 冊、簡単に検索——ネットで疾患別紹介（日本経済新聞：2006 年 6 月 12 日夕刊 3 版 20 面）
8. 闘病記 700 冊を一覧——病気ごと解説も（朝日新聞：2006 年 6 月 12 日朝刊 14 版 3 面）
9. 2006.4.22 7:34-7:44, NHK 総合 おはよう日本（首都圏版）, “神田神保町——「本の街」活性化の取り組み”.
10. 2006.2.26, NHK BS2 週刊ブックレビュー696 号, 特集「書店の新しい波」にて Book Town じんぼうの紹介・インタビュー.
11. 朝日新聞：2005 年 11 月 18 日夕刊 2 版 6 面, 縦横コラム「古書」——「寄り道」の楽しさサイトでも.
12. 2005.11.6, The Japan Times, “Find the right book without even going to Jinbocho”.
13. 東京新聞：2005 年 10 月 28 日夕刊 [東京] 27 面, 古書店サイト好評——神田神保町在庫 10 万冊の情報掲載.
14. 日本経済新聞：2005 年 10 月 27 日朝刊 [東京・首都圏経済] 35 面, 街全体が一つの書店——神田神保町の 190 店が連携.
15. 毎日新聞：2005 年 10 月 23 日朝刊 11 版 11 面, 「本の街」のサイト.
16. 2005.10.16 18:07, NHK 総合ニュース（全国版）, “神田の古書 検索データベース公開”.
17. 読売新聞：2005 年 9 月 5 日夕刊 2 版 4 面, 愛書家 夢のネット検索——目利きが選び連想で綴る、NPO 発足.
18. 2005.4.6, 日刊工業新聞 朝刊 1 面, “知の社会連携---専門知識を共有財に”, 進化する日本力（第 3 部）.

19. 2005.3.2, 毎日新聞 夕刊 5 面, “人模様 IT 使い若者を古書店に”.
20. 2005.2.28, 日刊工業新聞 朝刊 33 面, “国立情報学研究所 無償ソフトで社会貢献 古書店グループと連携 情報共有システムも公開”.
21. 2005.1.22, 岐阜新聞 夕刊 2 面, “新書 ネットで簡単検索 7400 冊テーマ別に 目次や概要も閲覧可能”.
22. 2005.1.21, 中国新聞 夕刊 2 面, “新書探し ウェブ本棚---7000 冊テーマ別で簡単に、高まる人気 見て楽しい工夫も”.
23. 2005.1.21, 東奥日報 夕刊 3 面, “「新書マップ」好評---テーマで選んで簡単検索 7000 冊登録、読書ガイドも”.
24. 2005.1.21, 岩手日報 夕刊 8 面, “天球型検索サービス「新書マップ」が人気---画面で「本棚サーフィン」テーマ別 7000 冊簡単に”.
25. 2005.1.20, 山形新聞 夕刊 2 面, “テーマでらくらく検索 ウェブ上 7000 冊「新書マップ」人気呼ぶ”.
26. 2005.1.20, 信濃毎日新聞 夕刊 1 面, “テーマ 1000 項目 新書検索---情報学研開発「蔵書」7000 冊横断的に”.
27. 2004.11.26, 読売新聞 夕刊 4 面, “書籍情報ネット化「新書マップ」情報社会、読み解くカギに”.
28. 2004.11.17, 日本経済新聞 夕刊 16 面, “プリズム現代 第 20 集 本の周りで 9 本探しを楽しんで”.
29. 2004.11.7, 中国新聞 朝刊 7 面, “論説室から 新書の今昔”.
30. 2004.7.23, 科学新聞 週間 2 面, “知的資産の活用支援 ソフト技術基盤構築へ”.
31. 2004.7.2, 毎日新聞 夕刊 8 面, “新書マップ めくもりのある読書案内...新書 7000 点、テーマ別に並べる仮想本棚”.
32. 2004.6.30, NHK ニュースおはよう日本 (8:15) , “新書データベース 30 日稼働”.
33. 2004.6.30, NHK ニュースおはよう日本 (6:13) , “新書データベース 30 日稼働”.
34. 2004.6.26, 朝日新聞 夕刊 1 面, “新書・選書 ネットで無料検索... 7 千冊、千のテーマに分類”.

6. 5. 2 その他

1. 闘病記専門 700 冊のブックガイドサイト (ITmedia News : 2006 年 6 月 12 日)
2. 2005.12.2, ITmedia ニュース, 日本人のアタマを救え——書籍検索でネットに“知の信頼性”を.

6. 6 その他特記事項

なし

7 研究期間中の主な活動

7. 1 ワークショップ・シンポジウム等

年月日	名称	場所	参加人数	概要
平成 17 年 3 月 25 日	「本の未来・未来の本 についてのシンポジ ウム」	学術総合センター 一橋記念講堂	約 300 名	書籍の電子化や電子的な 読書環境に深く関わるプ ロジェクト推進者の参加 を仰ぎ、技術的なトレンド や国際的な連携実現のた めにクリアすべき課題な どについて議論した。「本 の未来」と「未来の本」に ついての刺激的なフォー ラムとなった。NII 第 4 回 国際シンポジウムとして 開催した。 http://www.nii.ac.jp/intsympo/04/04_j.html

8 結び

5 年前、連想計算という基本アイデアとそれを実装した GETA の準備が整ったタイミングでこの研究ファンドをいただけたことは、本当に幸運であった。新しい研究環境において、研究チーム作りから行うことができ、当初掲げた「連想の情報学」というキャッチフレーズに相応しい実証的プロジェクトを数多く進めることができた。提案時には Google を「連想の情報学」で引いてもヒットがなかったが、今では最初の数ページを我々のグループの研究的実践が埋め尽くすところまでできた。4 月には、国立情報学研究所内に「連想情報学研究開発センター」という組織まで立ち上げることができた。

研究の推進に当たっては、ほとんど学術論文を書こうとせずに、ひたすら実用的なシステムの構築を続ける我々のスタイルが特異で、いろいろご心配をおかけしたと感じている。個人的には、当然、学術論文の価値を否定するものではないが、本研究のフェーズを考えると、この 5 年間は二兎を追わずにひたすら連想情報学の実用性を研ぎ澄ますことに専念して正しかったと思う。最後の年に「想・IMAGINE」という、大きなブレークスルーを得られたことで、このアプローチの正しさが証明されたと信じている。数年後には、必ずや「想・IMAGINE」でコンテンツ流通のパラダイムを変えてみせるという意気込みで、引き続き精進を続けるつもりである。