社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－FI－81（1）
2005／11／2

情報資源間の関連を考慮したリンキングシステムの構築

萬代 浩子　松村 敦　宇陀 則彦

筑波大学大学院　図書館情報メディア研究科

近年、学術情報とその情報提供機関が増加し、情報が様々なところに分散した状態で存在している。そのため、利用者は必要な情報資源を探すところからはじめ、不慣れなシステム上で求める情報が得られるまで検索を繰り返さなければならない。そこで本研究では、このような情報探索の負担を減らすため、インターネット上の多様な情報資源をあらかじめリンクしておき、単にボタンをクリックするだけで関連情報を辿れるリンキングシステムを構築した。本システムでは、利用者の探索行動が途切れないよう、検索結果からキーワードを自動的に切り出し、動的に関連情報をリンクするようにした。また、ガイドマップと呼ばれるインタフェースを用意することで、多様な情報資源を常に選択できるようにした。

Construction of a Resource Linking System based on Relationships between Information Resources

Hiroko BANDAI　Atsushi MATSUMURA　Norihiko UDA

Graduate School of Library, Information and Media Studies
University of Tsukuba

Recently, huge amounts of scholar information are distributed over the internet. In order to obtain the necessary information, researchers must repeat the retrieval using various inexperienced systems. In this paper, we propose a novel linking system, which connects various information resources on the internet in advance, and provides users the list of links to the related information resources. Our system extracts keywords automatically from the retrieval results, and generates links among information resources dynamically using these keywords. The system also provides navigation interface called “guide map”. Using our linking system, users can select appropriate resources at any time.

1.　はじめに

近年、学術情報**とその情報を提供する機関が増加し、情報が様々なところに分散して存在するようになった。そのため、利用者は目的の情報がどこにあるのか分からず、入手するのに多大な時間と手間がかかってしまう。例えば、利用者は学術情報を探す際、論文データベースを検索したり、OPAC を調べ図書館で論文のコピーを取ったり、求める情報に関連した Web サイトなどのインターネット情報資源や書籍を検索する必要がある。

その上、目的の情報を見つけた後も、それに関連する情報が必要となることが多い。そのような場合には、利用者は同じ情報資源内でキーワードを変えて再検索するか、他の情報資源にアクセスしなおさなければならない。それどころか、利用者が関連した情報を含んでいそうな適切な情報資源を知らない場合、情報資源を探すところからやり直す必要がある。このように、従来のシステムの枠組みでは情報探索の流れが断続的で、利用者に余計なストレスを与えている。

現在の WWW 上では、データベース内検索のような特定の範囲内であろうが、不特定範囲であろうが、全て検索技術を用いて解決しようとする傾向があるが、近年、検索とは異なる手法として注目を集めているのが、リンク技術である。これは利用者が不特定の範囲から情報を探す場合、検索技術より有効であると言われて

**本研究では、「学術情報」という語を「研究で必要とする情報」という意味で用いる。

いる[1][2]。利用者の要求とそれを満たす情報をリンクしておけば、利用者の状況に適したサービスを提供することができるからである。

そこで、本研究では、多種の情報資源をリンクし、多様な関連情報へのアクセスを常に提供するリンキングシステムを構築した。利用者は本システムによってリンクされた情報資源を検索して結果を連鎖的に得ることができる。また、利用者が検索をしている間、関連情報は常に提示される。これにより、利用者は関連情報を得るために情報資源間を移動したりキーワードを再入力したりすることなく、多様な情報を得ることができる。

2. システム概要

本システムの全体図を図 1 に示す。本システムは利用者が情報資源自体を探索したり、検索し直さずに済むよう、インターネット上にある様々な情報資源をあらかじめリンクしておくことに特徴がある。本システムでリンクした情報資源は、現在のところ文献情報データベース、書籍データベース、Web サイトデータベース、ウェブログ記事データベースの 4 つである。

文献情報データベースには、筑波大学知的コミュニティ基盤研究センターの知的コミュニティ情報システム[3]で使われている図書館情報学関連の書誌情報データベースを用いた。このデータベースには、図書館情報学関連の論文やインターネット情報資源のメタデータが含まれている。本システムでは、このメタデータの一部であるタイトル、著者名、識別子、概要のフィールドのデータを使用し、システム内のデータベースに格納した。

書籍データベースには、Amazon.co.jp[4]を使用した。 Amazon.co.jp は外部システムからのアクセスを受け付け、結果を返す Web サービスを行っている。学術情報を収集する際、書籍の情報収集は一般的に行われているためリンクした。Amazon にリクエストを送信すると、大･小の表紙画像の URL や中古品販売での価格など、様々な情報が含まれた結果が返ってくる。本システムでは、そのうちのタイトル、著者、ISBN、価格、小サイズの表紙画像 URL のデータを使用した。

Web サイトデータベースには Google[5]を用いた。Google も Amazon.co.jp と同じように Web サービスを提供している。インターネット上には研究論文やデータなどの学術情報が公開されていることも多く、学術情報収集の際にインターネット情報資源を探すことは必要である。Google もリクエストを送信すると、様々なデータが返ってくるが、本システムでは Web サイトのタイトル、URL、Web サイトの概要データを使用した。

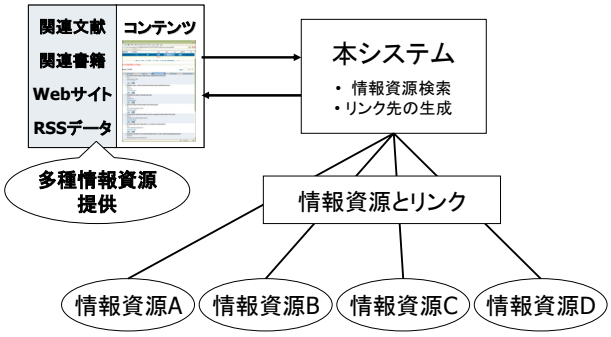

図 1. システム全体図

また、最近はウェブログなどによる個人の情報発信が盛んである。研究者がその専門分野の最新の動向やプロジェクトなどを記事にしているウェブログが多くなっている。研究活動にウェブログを活用するための製品も発売されている[6]。本システムでは、ウェブログのデータを収集しデータベースに格納することによって、ウェブログの記事情報を情報資源の 1 つとした。ウェブログは普段利用者が読んでいるものを複数登録できる。これにより、利用者は検索対象とする記事データを容易にカスタマイズできるようになり、自分の興味にあった記事データ情報資源を作ることができる。本システムでは、記事のタイトル、URL、詳細、作成日付、作成者等のフィールドのデータを使用し、システム内のデータベースに格納した。

3. システムの実際

3.1. インタフェース

本システムのインタフェースは、ブラウザの画面右側の結果表示部分と左側の関連情報リンク一覧から構成される（図 2）。一般的に、要求を満たす情報を 1 つ見つけた後で、関連情報がほしいという場合には、キーワードを入力し検索しなおすか、情報資源を移動して再検索をする必要があるが、これでは関連情報を求める際、検索の流れが途切れてしまう。

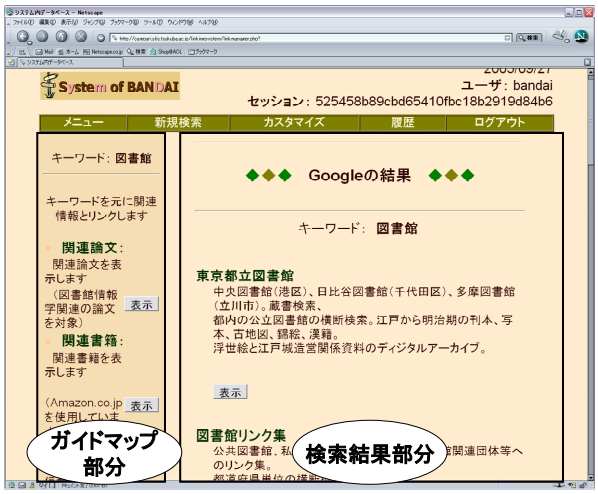

図 2. システム画面

そこで、本システムでは右側に検索結果を表示し、左側に関連情報資源へのリンクボタンを表示することで、利用者の検索の流れを途切れさせないようにした。この左側の表示部分は、利用者を様々な情報資源内の関連情報へ案内するためのものなので、“ガイドマップ”と呼ぶ。

ガイドマップの下部には情報資源内にある関連情報へリンクされたボタンを表示し、最上部には、関連情報へリンクする際に用いるキーワードを表示した。このようなインタフェースにすることで、利用者は連続的に関連情報にアクセスできる。

3.2. システム実行例

ここでは、図 3 に示した実行例に従って、本システムの具体的な操作性について説明する。

(1) ログイン画面

本システムにアクセスすると、システムへのログイン画面が表示される。利用者はユーザ名とパスワードを入力して、システムにログインする。

(2) 新規検索画面

ログインすると最初に表れるのは新規検索画面である。ここでは、検索語入力フォームにキーワードを入れ、情報資源を 1 つ選択し、検索を行う。図の例では、キーワードに「電子図書館」と入力し、Web サイト検索を行う場合を示した。

(3) 結果一覧表示画面

新規検索の結果は画面右側の結果一覧画面に表示される。それと同時に、画面左側のガイドマップには、関連情報にリンクされた他の情報資源の一覧が表示される。

図の例では画面右側に Web サイト検索結果一覧が表示されるとともに、ガイドマップには(2)で使用した Web サイト検索以外の、文献データベース、書籍データベース、ウェブログ記事データベース内にある関連情報へのリンクが表示される。どのようなキーワードでリンクされたか利用者に分かるよう、最初に入力した「電子図書館」というキーワードをガイドマップ上部に表示した。続いて、ガイドマップにあるボタンをクリックすると、選択した情報資源内の関連情報一覧が表示される。

(4) 一次情報表示画面

(3)で得た結果一覧から求める情報を 1 つ選択すると、画面右側にその一次情報が表示される。この図は、青空文庫のサイトを選択したところである。左側のガイドマップには、右側で選択したサイトに関連する情報がリンクされている。ガイドマップの上部には、一次情報のタイトル部分を形態素解析で切り出した “青空”と“文庫”が表示されている。本システムは切り出したキーワードを用いて、リンクしてある情報資源内の関連情報にアクセスするボタンを提示する。

新規検索を行う場合は、画面上部に表示されているナビゲーションバーの「新規検索」をクリックすることでいつでも(2)の新規検索画面に戻ることができる。また、関連情報を探すには、ガイドマップを利用して(3)と(4)の画面を行き来し続けたり、(3)の一覧画面から、他の情報資源の関連情報一覧を表示させたりできる。

以上が本システムの基本的な使い方であるが、このように関連情報に連続的にアクセスすると、途中どのような情報資源にアクセスしたか分からなくなる[7]。そこで、アクセスした情報の履歴をログアウトするまで残すようにし、ナビゲーションバーの「履歴」を選択することでいつでもアクセスした情報資源の履歴を見ることができるようにした。

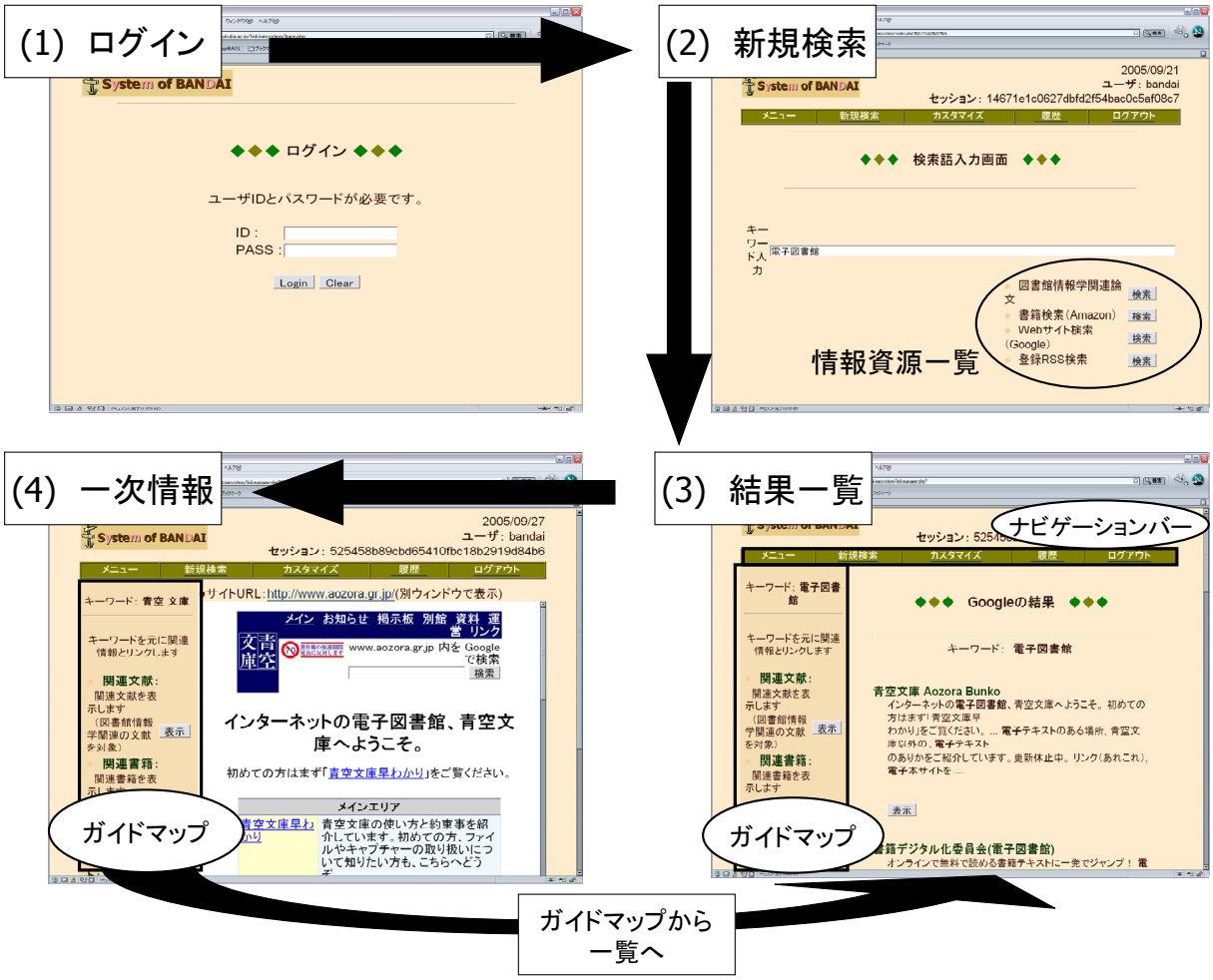

図 3.　システムの流れ

4.　システム構成

本システムの構成を図 4 に示す。検索や情報資源の選択など、関連情報を求める要求が発生すると、その要求は HTML の POST メソッドでキーワードと共に Link Resolver に送信される。Link Resolver では、送られてきたキーワードの処理や文字コードの変換を行い、情報資源を検索する関数 Link Manager を呼び出す。Link Manager は利用者の要求に応じて適切な Linker を選択し、キーワードを渡す。Linker は実際に関連情報を検索するとともに、その情報に関連する情報をリンクするためのキーワード抽出を行う。

Linker が取得した結果は、結果表示テンプレートとガイドマップテンプレートを通してブラウザに表示される。テンプレートを介すことで表示を統一し、どの結果にも常にガイドマップを容易に表示させることが可能となった。ガイドマップには、表示した結果に関連した他の情報資源の情報へのリンクが張られており、Linker で処理した関連情報抽出のためのキーワードが埋め込まれている。

本システムの開発には以下のものを使用した。

- OS：　Windows XP
- Web サーバ：　Apache 1.3
- DBMS：　MySQL 3.23.51
- 開発言語：　PHP 4.3.2
- 形態素解析器：　WinCha 2.3.3

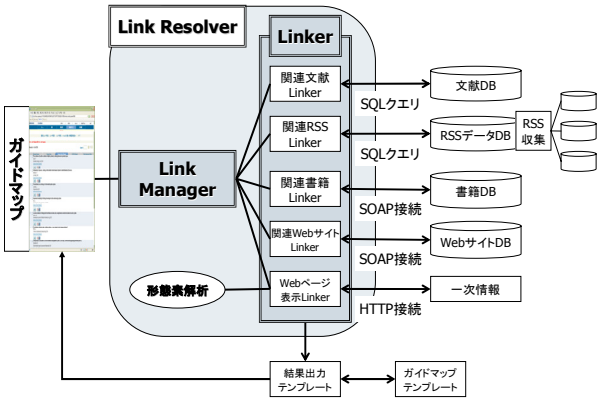

図 4. システム構成

5. 考察

5.1. 検索過程の省略について

既存の検索システムを用いて情報を探索する場合、利用者はキーワードの入力と検索結果の評価を繰り返し行う。すなわち、検索過程が探索過程の大部分を占めるといってよい。それに対して、本システムは検索過程を省き、リンクを辿るためのボタンをクリックするだけで情報にアクセスできる。このため、利用者の本システムの操作感は既存のものとは相当異なることが分かった。本節では、一般的なシステムを使う際の利用者行動と、本システムを使用した場合の行動がどう異なるかについて考察する。

既存のシステムでは、キーワードを考えて再検索を行ったり、関連情報を探すということが利用者側の判断に委ねられている。一方、本システムでは関連情報があらかじめリンクされているので、利用者はキーワードを入力したり、情報資源を移動したりする必要がない。これは一見便利なように見えるが、システムの操作中に、ガイドマップ上部に表示された関連情報をリンクするためのキーワードを見ずにボタンをクリックしてしまうと、利用者は元となる情報と表示された結果一覧の関連が分からないという問題が起こることが分かった。たとえば、3.2.で示した実行例で、青空文庫のサイトから関連情報を得るとする。ここで“青空”と“文庫”をキーワードとすると、利用者は青空文庫に関連した情報が得られるであろうと想像する。しかし、キーワードを意識せずに関連情報一覧を得た場合、表示されたものがどのように関連しているのか理解できない。そのため、利用者は関連情報を得られたと感じず、いたずらにクリックを繰り返すことになる。今後は表示を大きくするなどしてキーワードをアピールし、利用者にどのような関連情報がリンクされてくるのかを示す必要があるだろう。

5.2. 探索終了の判断について

一般的な検索システムは、「検索 → 検索結果一覧 → 詳細情報(一次情報表示)」という流れになっているものが多い。1 つの検索結果が提示されると、そこでシステムはさらにリンク先を提示することはなく、処理が終了する。

それに対して、本システムは一次情報を表示させた後、それに関連するリンクを提示するため「検索 → 検索結果一覧 → 詳細情報 → 関連情報一覧 → 詳細情報 → ･･･」という流れになる。つまり、本システムには探索終了を示す明示的なサインがない。このことは多くの関連情報にアクセスできるという利点がある一方で、利用者は探索を止める機会を見失ってしまうという問題があることが分かった。そこで今後は、履歴を見やすく提示するなどして、利用者が自然に探索を終了できるインタフェースを検討する必要がある。

5.3. キーワード抽出方法について

関連情報をリンクする仕組みとして、現在は元となる情報のタイトル部分の名詞をキーワードとして抽出している。しかし、タイトル部分の名詞にはキーワードとしてふさわしくないものが含まれていることがある。たとえば、関連元とする情報のタイトル部分が「はてなダイアリー－図書館情報学とは」だった場合、形態素解析を行うと“ダイアリー”、“図書館”、“情報”、“学”がキーワードとなる。この場合、“ダイアリー”は関連語としては不適切であり、関連情報をリンクできない。

このように、企業名やサービス名などが関連情報の内容と関係のない名詞が含まれていたり、タイトル内の名詞が多すぎて関連情報が絞り込まれすぎる場合、正しく関連情報へリンクできない。今後キーワードを切り出す方法と、対象とする部分がタイトル部分で良いのかを検討する必要がある。

5.4. ガイドマップについて

現在のシステムがリンクしている情報資源は 4 個だが、今後これらが増えた場合、ガイド

マップに表示する情報資源の種類をその検索結果に応じて適切に変化させる必要がある。

例えば、医学分野のデータベースで胃癌に関する情報を探している時に、経済学のデータベース内に関連情報を求めることはないであろう。システムが扱う情報資源が増えてきた場合、関連情報が存在しない情報資源にリンクさせる可能性が高くなる。このため、ガイドマップに表示する関連情報を選択するための基準を検討する必要がある。

5.5 検索とリンキングの連携

インターネット上の代表的なサーチエンジン Google は、フルテキスト検索の手法に加えて、リンクの参照関係も加味した検索システムである。すなわち、Google は検索にリンクを利用したシステムである。それに対して、本システムはリンクすることに検索を利用したシステムで、Google とは逆である。そういう意味で、本研究は検索システムとも既存のリンキングシステムとも全く異なるタイプのシステムを開発したと言える。このシステムが情報探索にどのくらいの効果があるのかは検討が不十分であるが、今後の改善次第では応用範囲の広いシステムとして展開する可能性があると思われる。

6. おわりに

本研究では、利用者の求める多種の情報を常にリンクすることによって、関連情報を提供するシステムの構築を行った。本システムにより、多様な情報資源を結びつけ、利用者に関連情報を常に提供できるようになった。また、ガイドマップインタフェースを用いることで、利用者に多種の情報資源内の情報を常に提示することを可能とした。今後の課題は、現在のシステムの持つ様々な問題を改善することである。検索とリンクを組み合わせて利用者の状況に応じた適切なサービスを提供するために、機能やインタフェースを改良していくことである。

参考文献

[1] Marcia J. Bates. Speculations on Browsing, Directed Searching, and linking in Relation to the Bradford Distribution. Emerging Frameworks and Methods: COLIS4, 2002, p.137-150.

[2] Sanjay J.. Searching vs. Linking on the Web: A Summary of the Research. (online), available from <http://usability.gov/searchlinkfinal1.pdf>, (accessed 2005-09-30).

[3] 筑波大学知的コミュニティ基盤研究センター. “筑波大学：知的コミュニティ基盤研究センター Web ページ”. (オンライン), 入手先 <http://www.kc.tsukuba.ac.jp/rkcs/>, (参照 2005-09-30).

[4] Amazon.co.jp. (オンライン), 入手先 <http://www.amazon.co.jp/>, (参照 2005-09-30).

[5] Google. available from <http://www.google.com/>, (accessed 2005-09-30)

[6] i-HIVE Communication. “学術研究系ブログ アカデミックブログ”. (オンライン), 入手先 <http://www.comlog.jp/academic/>, (参照 2005-09-30).

[7] Marcia J. Bates. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review, vol.13, 1989, p.407-424.