

学術情報ポータル「Academic Navi U-Tokyo」の紹介

デジタル・ライブラリ係

はじめに

平成17年1月、情報基盤センター図書館電子化研究部門および図書館電子化部門では、学術情報ポータル「Academic Navi U-Tokyo」、通称「AcaNavi（アカナビと読む）」の提供を開始しました。今回は、このAcaNaviについてご紹介します。

1 . AcaNavi とは

AcaNaviとは、東京大学内のWebサイトで公開されている学術情報へのナビゲートを行う、学術情報ポータルです。学内には、数多くのWebサイトが公開され、研究成果をはじめ、研究者情報、学習情報、大学そのものに関する情報など、さまざまな学術情報が発信されています。情報基盤センター図書館電子化部門では、学内外を問わず、こうした学術情報のうち、研究成果情報を中心に「インターネット学術情報インデックス (IRI)」に収録し¹⁾、効率的かつ効果的な検索サービスを提供してきました。しかし、学内で生産され、発信される学術情報は日々増加・更新されており、これらの情報を一括して検索できるシステムが必要ではないかとも考えてきました。そこで、図書館電子化研究部門および図書館電子化部門では、学内で公開されている学術情報を定期的に収集し、データベース化して公開するシステム「Academic Navi U-Tokyo」を開発しました。2005年6月末現在、586,638件のデータが登録されています。



図1 Academic Navi U-Tokyo のトップページ

(URL : <http://acanavi.dl.itc.u-tokyo.ac.jp/>)

2 . AcaNavi の特徴

AcaNaviの最大の特徴は、収集した学術情報を自動分類システムにより自動的に分類し、公開

できる機能を有していることです。具体的には、収集した学術情報、すなわちWebページからテキストを抽出し、これをAcaNaviで用意した分類知識データベースと比較することにより、そのWebページを該当するカテゴリに分類するというものです。現在用意されているカテゴリは、「東京大学研究者紹介」でも用いられている「研究分野」です²⁾。分類知識データベースには、それぞれの研究分野カテゴリに特徴的な参照データが収録されており、これを元に学術情報は各カテゴリに分類される仕組みになっています。これら参照データは、必要に応じて追加・更新できるので、新しいカテゴリを追加したり、分類精度を向上させるといったことも可能です。

なお、学内の学術情報は、毎月1回(現在は5日)クローラーによって自動的に収集されます。すべての学術情報を収集・分類し、検索用インデックスを作成して、データベースに反映されるまでにおよそ1週間を要しています。

3 . AcaNavi を検索する

AcaNavi には、収録されたデータすべてを対象に検索するキーワード検索と研究分野に沿って検索するカテゴリ検索の2つが用意されています。



図2 Academic Navi U-Tokyo の検索結果画面

キーワード検索では、好きな語を入力して検索できます。また、複数の検索語を論理演算子を使って入力することにより、検索結果を絞り込んだり、逆に検索結果の集合を拡大したりできます。具体的には、複数の検索語を半角スペースで区切って入力すると「すべての語を含む検索 (AND 検索)」に、半角パイプ(|)で区切って入力すると「いずれかの語を含む検索 (OR 検索)」に、半角エクスクラメーション(!)で区切って入力すると「特定の語を含まない検索 (NOT 検索)」になります。

キーワード検索によって得られた検索結果には、「おすすめカテゴリ」も表示されます。これは、入力されたキーワードから、関連のあるカテゴリを類推して表示する機能です。ここから該当のカテゴリに分類された学術情報にアクセスすることもできます。

もう1つの検索機能であるカテゴリ検索は、AcaNaviの特徴を生かした検索方法です。AcaNaviの研究分野カテゴリは階層構造をなしており、最大3階層となっています。目的の、あるいは興味・関心のあるカテゴリのリンクをたどっていくことにより、検索結果を一覧することができます。各Webページは、1つあるいは複数のカテゴリに分類されるので、さまざまな研究分野からアクセスすることが可能となっています。

4 . AcaNavi の今後

サービスを開始したばかりのAcaNaviですが、今後、解決すべき課題として2つを考えています。まず1つ目は、AcaNaviの最大の特徴である自動分類の精度の向上です。自動分類する際に用いられる分類知識データベースには、それぞれのカテゴリに分類するための参照データが含まれています。この参照用のデータは、あらかじめ用意されたものではなく、こちらでいくつものサンプルデータを収集し、これをもとにAcaNaviに「学習」させることによって獲得させることができます。したがって、自動分類の精度を向上させるため、カテゴリごとに最適な分類ができるようなサンプルデータを用意し、「学習」させることが1つ目の課題です。

もう1つの課題は、検索インタフェースの改良です。現在利用できる検索は、キーワード検索とカテゴリ検索の2つです。カテゴリ検索では、提示されたカテゴリのリンクをたどることによって、目的の学術情報にアクセスすることができますが、キーワード検索の場合、検索するための語を自ら入力しなければなりません。しかしながら、より適切な学術情報を検索するために、どのような語を入力すればよいかを考えるのはなかなか難しいものです。そこで、AcaNaviに収録されているデータを参考にしながら、カテゴリごとに特徴的な語、あるいは関連性の高い語をあらかじめ提示し、そこから選択することによって検索できる仕組みを検討しています。その際、図書館電子化研究部門で開発された「言選Web」の活用も視野に入れていきます³⁾。

おわりに

図書館電子化研究部門および図書館電子化部門では、学内外を問わず、より効率的で効果的な学術情報の収集、組織化および発信に取り組んでいます。今回サービスを開始した学術情報ポータル「Academic Navi U-Tokyo」もその1つです。このポータルサイトを多くの人にご活用いただき、ご意見・ご要望等をお寄せいただけたら幸いです。

注

- 1) インターネット学術情報インデックス. http://resource.lib.u-tokyo.ac.jp/iri/url_search.cgi
- 2) 東京大学研究者紹介. http://www.adm.u-tokyo.ac.jp/IRS/Navi/BaseFrame_Cover_j.html
- 3) “ 専門用語 (キーワード) 自動抽出システム ” のページ. <http://gensen.dl.itc.u-tokyo.ac.jp/>