目指せ！ネットエスパー

インターネット新検索術

インターネットの中に潜む膨大な情報たち。そこから欲しいものを見つけ出して活用するには、もはや１検索サイトだけでは役に立たない。ネットの海を自在に泳ぎ、必要な情報をすぐ取り出して活用する、そんな「ネットエスパー」に変身すればインターネットの利用価値は無限大になる。この連載で「ネットエスパー」に変身するスタートを切ろう！

二木麻里（アリアドネ運営）　Jump ariadne.ne.jp
Illust : Ebisu Yoshikazu

第11回 古典文学の資料を探す

秋の夜長も本格的になってきた。「あ、いつか読んだあの本が読みたい」と夜中に思ったら、自分の部屋からインターネットで探してみるといいかもしれない。近所の書店の棚になくても、オンラインでなら読める文学作品はたくさんある。ことに著作権の切れた古典作は、しばしばボランティアの入力により無料アーカイブに掲載されている。また、紙の本では検索しにくい作中の一節を、インターネットなら楽に探し出せるという利点もある。「あの作品のあの個所はどこだっけ？」というときは、オンラインのデジタルデータで探してみるのがおすすめだ。

1 源氏物語の女性談義はどの章に？

「……だめだ」。編集部の隣の席で、ナカムラ先輩がぶ厚い本を放り出した。

「どうしたんですか」と訊ねるエリカさん。

「探せないんだ、僕もむかし高校の授業でやったきりだし。あのね、源氏物語に女性談義のくだりがあったろう」

いわれてみるとおぼろげな記憶が。「……理想の女性はどういうものか、男性同士でえんえん語りあうのでしたっけ」

「そうそう。筆者が引用してきたんだけど、本人も資料をひっくり返しているうちに出典部分がわからなくなっちゃったって言うんだ。“今は、ただ、品にもよらじ。容貌をばさらにも言はじ”。最後は相手の生まれも顔も、もうどうでもいいっていう……。文章はすごくおもしろいんだけど典拠を確認しないと使えない。どの巻かわからない？」

「ぜんぜんだめです。よかったらこれから図書館に行って……」

「その暇があればいいんだけど、今日印刷所に原稿を渡すんだよ。あと2時間くらいしかない」

うわっ、あんな大長編小説の、どこになにが出てくるかなんてどう探せばいいの？　手がかりは“今は、ただ、品にもよらじ。容貌をばさらにも言はじ”だけ。あとは女性談義。探しものの出発点はやはり検索サイトかな。でも時間がないんだっけと思い返して、メタ検索サイトを使うことにする。複数の検索サイトを一度に探してくる機能だ。英語圏ではずいぶん発達したが、日本語のメタ検索サイトはまだ多くない。最近スタートしたWAKANO日本版 Jump を試してみた。

WAKANO JAPAN
Jump www.wakano.co.jp

メタ検索サイト：複数のサーチエンジンを使用して検索してくれるサイトのこと。代表的なものにMetacrawler Jump www.metacrawler.comやSavvySearch Jump www.search.comなどがある。

インターネットマガジン／株式会社インプレスR&D
©1994-2007 Impress R&D

2 メタ検索サイトで本文を探す

とにかく本文のテキストが必要なのは確かだ。WAKANOの単純検索で日本語ウェブ文書を指定し、まずは「源氏物語」と入力してみる。すぐにステータスリポートの小さいウィンドウがポップアップし、Yahoo! Japan、LYCOSJapan、OCNnavi、goo、infoseek japanなど7本の検索サイトで一斉に探しはじめたことがわかる。

まもなく戻ってきた結果を見ると、ページ上部に「ベスト参考サイト」として、いくつかのウェブサイトが掲示されている。その下には別枠で「実時間自動分類」として、7つの検索サイトからのヒット結果がリストされてくる。この「実時間自動分類」は同一ヒットの重複も多いが、「ベスト参考サイト」は10本のみ。興味深いことに、英語圏で発信されている『日本語テキストイニシアティブ』Jump が最上位だ。源氏物語の資料としてここがベストかどうかはむずかしいところだが、サイト自体はバージニア大学とピッツバーグ大学による有名なジョイントプロジェクトで、英語圏では数少ない本格的な日本文学研究用アーカイブである。この「ベスト参考サイト」はWAKANO内部のディレクトリー分類によるデータ掲示だろう。内部データベースが「日本ドメイン」ではなく「日本語」で区分されているらしい。

なお、「実時間自動分類」に出てくるサイトには、ヒットした検索サイトの名前が（LYCOS Japan）、（infoseek Japan）のようにカッコで示されている。

こうして、全体に外部の検索サイトによる検索結果と、WAKANO自体のローカルなディレクトリー構造部分とを連結した結果を表示していることがわかる。これはこのサイトの特徴で、分類も表示方法も独自のスタイルを持っている。検索速度はまだ遅いが、おもしろい。

さてトップページに戻って、検索窓の下にあるディレクトリー分類を調べてみると、さらにこのサイトでは分野別ディレクトリーがそれぞれ独立した「メタ検索エンジン」として機能していることがわかる。複数の専門サイトを同時に探してくるのだ。どのサイトを検索するかという構成も指定できる。

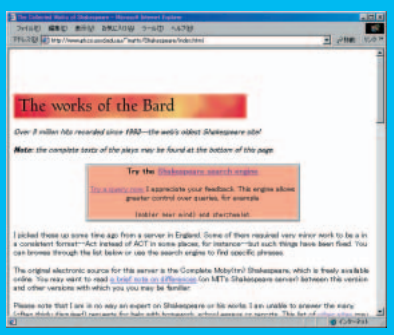
世界的にも、メタ検索エンジンはエンジンの総本数やエンジンの構成、分野などをユーザーがカスタマイズできるありかたが普通になっている。たとえば慣れないと扱いにくい巨大ヒット数の検索サイトはとりのぞいて使うこともできる仕組みだ。

「実時間自動分類」結果では、どこの検索エンジンからヒットしたかがわかるようになっている。

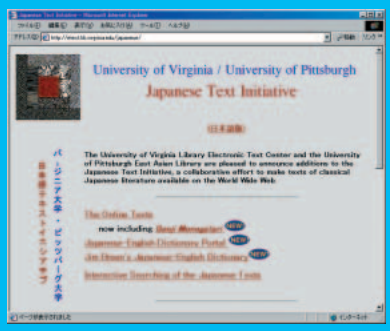
最強！シェイクスピア全作品の検索サイト

あらゆる古典作品のなかで、もっとも引用されることの多い作品群は？ そう、おそらくシェイクスピアだ。あの引用回数に勝てるのは聖書くらいではないだろうか。ええと、あのセリフはどこだっけ、と探し回った経験のある方も多いだろう。今回ご紹介するサイトThe Works of the Bard（バード：シェイクスピア全作品）Jump はシェイクスピアのアーカイブとしては、屈指である。シェイクスピアの全作品をテキストデータで擁しているサイトはほかにもあるが、ここは全作を通じて横断検索ができるthe Shakespeare search engine（シェイクスピア・サーチエンジン）で名高い。単語を入力すれば、初期の喜劇からあの長大な歴史劇まで全部通して探してくれる。

筆者は以前Green Sleeves（グリーンスリーブス）という一語を探して疲れはてたことがある。Alas my love, you do me wrong…で始まるあの有名な歌は、ビクトリア朝時代に一世を風靡した流行歌。シェイクスピアにも言及がある、とそこまでは読んだものの書籍で探しても見つけられなかった。シェイクスピア・サーチエンジンに「袖」（sleeves）とだけ入力して、出てきた時のうれしさは忘れられない。原文の大意は「言葉と本心とは大違い、聖歌とグリーンスリーブスくらい違うんだから！」とあった。『ウィンザーの陽気な女房たち』の第二幕第一場という箇所だ。



The Works of the Bard
Jump www.gh.cs.usyd.edu.au/~matty/Shakespeare/



日本語テキストイニシアティブ
Jump etext.lib.virginia.edu/japanese/

アーカイブ：関連する情報やファイルをまとめて保管しておくところ。また保管されているもの自体を指すこともある。

インターネットマガジン／株式会社インプレスR&D
©1994-2007 Impress R&D

3 資料同士の関係をきちんと把握しよう

さて、それでは推薦された『日本語テキストイニシアティブ』にいってみよう。たしかに「源氏物語」とある。しかし、テキスト本文などは日本のリソース『源氏物語の世界――定家本「源氏物語」本文の表現と解釈』 Jump からのリンクである。すなわち内容的には『源氏物語の世界』の制作者、高千穂商科大学の渋谷栄一氏による研究成果なのだ。このサイトは原文と現代語訳、さらにローマ字版が添えられた屈指の構成で、『日本語テキストイニシアティブ』は事実上このサイトのポインターという位置関係だったことがわかる。幸い出典が明快に示されていて助かるが、ユーザー側もインターネットではURLをよく見て、オリジナルの「ソース」がどれで、眼前のページとどういう関係になっているのかをきちんと確認する必要がある。

ことにテキスト類のアーカイブサイトは外部リンクによって作品集を構成しているものも多いため、注意が必要だ。きちんとしたアーカイブサイトやリソースリストサイトであればリンク元は明確に記されているが、検索サイトのヒット結果だけを見ると「真の制作者」は時にわかりにくい。強いていうならこのWAKANOでも、せっかく「源氏物語」としてベストサイトを挙げるなら、積極的な理由がないかぎり、リンク先でなくオリジナルを優先して欲しかった。

『源氏物語の世界――定家本「源氏物語」本文の表現と解釈』
Jump www.sainet.or.jp/~eshibuya/

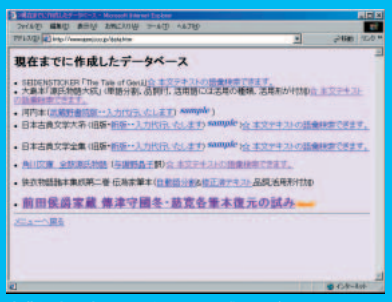
4 源氏物語の「全文検索」を使ってみる

さて、作品中の個所を知るには、全文からテキスト検索をかけるのが早い。そうでないと全五十四帖を一巻ずつ探してこなければならなくなる。全文検索が行える研究サイトを見つけるには、たとえばいちばん最初の検索結果をもとに絞り込みをするのが早い。探しに出た先で必要なものが見つからないときは「検索の幹に戻る」のが基本原則だ。

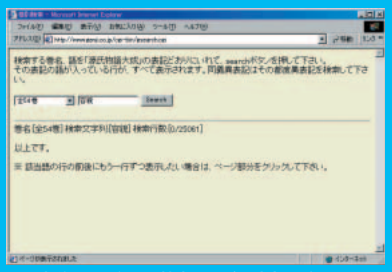
WAKANOに戻って「源氏物語 テキスト検索」あるいは「源氏物語 コーパス」とキーワードを足してみる。「源氏物語 コーパス」で『古典総合研究所』など複数のデータベースが出てきた。この『古典総合研究所』 Jump01 を見ると、全文を通して検索ができるとある。さっそく「今は、ただ、品にもよらじ。容貌をばさらにも言はじ」から言葉を選んで入れてみる。今、品、容貌……。だが出てこない。該当文はゼロ、あるいはあきらかに違う文脈だ。動詞の活用も試す。よらじ、言はじ。

だめだ。ほんとうにあるのかしら？ 筆者の思い違いでは？

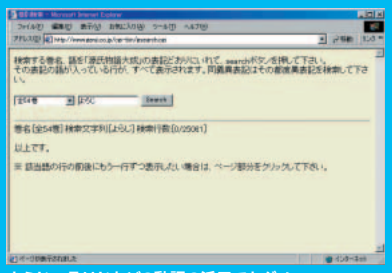
見ると、このサイトには与謝野晶子による現代文の源氏物語があり、やはり全



古典総合研究所　現在までに作成したデータベース。
Jump01 www.genji.co.jp/data.htm



品、容貌などの名詞で検索したが、該当はゼロ。



よらじ、言はじなどの動詞の活用でもダメ。

コーパス：語形変化などを知るための研究用テキストデータベースの形式。

インプレス・ウェブマガジン／株式会社インプレスR&D
©1994-2007 Impress R&D

文検索できるとある。やってみよう。時間がせまってくる中、並行資料として与謝野源氏をためしてみる。「今」「品」「容貌」といった原文の単語を晶子が踏襲してくれていれば、可能性はある。

4帯 0039-18 「ですからもう階級も何も言いません。容貌もどうでもいいとします…」 Jump 02

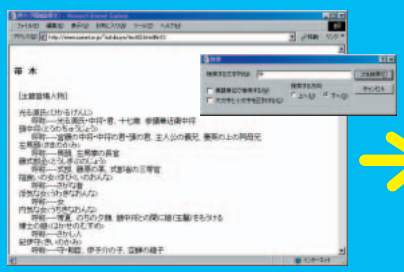
あった。「容貌」で出てきたのだ。内容的に見て、おそらくここだろう。これが原文と合えば、全巻を順に探さなくてすむ。

さきほどの『源氏物語の世界――定家本「源氏物語」本文の表現と解釈』に戻って「帚木」（ははきぎ）の巻 Jump 03 を見てみる。信頼できるリソースが見つかったら不必要にほかのリソースを使い散らさず、できるだけ1個所を深く活用してみよう。とくにこのサイトのように複数のデータベースで並行検索が行えるなら、それに越したことはない。主資料で結果が出なくても、あわてずに副資料に当たろう。

さて「帚木」をひらき、今度はブラウザーの機能を利用してそのページに検索をかける。コントロール＋Fキーを押すと検索窓が出てくるので「今」とキーワードを入れる。微妙な表記の違いがあるとヒットしないので、ここでは必要最短のキーワードにしておく。

すると、ページ中程の一行が表示された。第一章　雨夜の品定めの物語（中略）女性論、左馬頭の結論「今は、ただ、品にもよらじ。容貌をばさらにも言はじ」 Jump 04 。間違いなくここだ。めでたしめでたし。

与謝野源氏物語を「容貌」で検索した結果。
Jump 02 www.genji.co.jp/cgi-bin/ysearch.cgi



『源氏物語の世界――定家本「源氏物語」本文の表現と解釈』の中の「帚木」のページに行き、「今」で検索してみた。
Jump 03 www.sainet.or.jp/~eshibuya/text02.html

目的の文脈に無事に到達！
Jump 04 www.sainet.or.jp/~eshibuya/text02.html#in13

5 今月のポータルキット

文学作品の語句や出典個所を調べたいことは、専門家でなくてもよくある。ただ検索サイトで直接、作品本文の語句を入力して該当個所が出てくることはあまり期待できない。概してアーカイブサイトは階層構造が深く、ルートページから何層も降りたアンカーディレクトリーに本文が納められている場合が多いためだ。文書が圧縮されていることもある。また今回の源氏物語のように大長編であれば、1つの作品が多数のファイルに分割されて格納されている。全文検索機能が貴重なのはこうした事情もあるのだが、全文データベースが入手できなければ、複数のファイルを順に検索していく。このとき、分割されているファイルごとに別々のブラウザーを広げてあらかじめ全章を開いておき、順に参照していって検索を行うとよい。検索字句がキープされているので、意外なほどすばやく行える。また作品を探す手順の中で、別バージョンがないかどうか、サブリソースをチェックしておくことも忘れずに。今回のように現代訳があれば参照もできる。

1	手がかりを確認
2	メタ検索サイトで本文を探す
3	資料同士の関係をきちんと把握しよう
4	源氏物語の「全文検索」を使ってみる

二木麻里（ふたき まり）
上智大学外国語学部卒。翻訳家。社会・人文科学系の国内外資料を案内した総合サイトARIADNEを運営。著書に『思考のためのインターネット』（筑摩書房ちくま新書）など。

アンカーディレクトリー：ルートディレクトリーに収められたルートページなどからハイパーリンクでジャンプが定義されたディレクトリーを指す。

インターネットマガジン／株式会社インプレスR&D
©1994-2007 Impress R&D

［インターネットマガジン　バックナンバーアーカイブ］ ご利用上の注意

このPDFファイルは、株式会社インプレスR&D（株式会社インプレスから分割）が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン　バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

http://i.impressRD.jp/bn

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容（技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど）は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者（執筆者、写真の撮影者、イラストの作成者、編集部など）が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元（株式会社インプレスR&D）、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D（当時は株式会社インプレス）と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先
株式会社インプレスR&D
All-in-One INTERNET magazine 編集部
im-info@impress.co.jp

Copyright © 1994-2007 Impress R&D, an Impress Group company. All rights reserved.