

初代鉄人 関裕司が教える



プロフィール
昭和38年生まれ、獅子座のA型。静岡県出身。
(株)Z会・増進会出版社の情報企画部に勤務しながら、昨春秋に「第1回検索の鉄人コンテスト」で見事優勝した。自分のホームページは
URL <http://www.shikenchou.com/>

目指せ！ 検索の鉄人

「検索の鉄人」大会もいよいよ大詰め、8月30日は決勝大会です。皆さんのチャレンジの結果はいかがだったでしょうか。さて、今回はいつもと趣向を変えて、gooを使って「インターネットはこんなに便利だ!」と思えるようなサイトを検索してみましょう。有名なサイトはたくさんありますが、ちょっとめずらしいものを紹介したいと思います。

URL <http://www.goo.ne.jp>

構成 / 関裕司

Illustration: Sasaki Kazuyoshi

活用 1

芥川龍之介の作品「羅生門」の全文を、インターネットを使って入手してください。



ポイント

オンライン読書には欠かせないサイトが見つかるぞ!

解説

情報化社会が進むにつれて最近クローズアップされてきた言葉が「著作権」。あらゆる作品には著作権が存在し、著作者の権利を保護しています。しかし、著作権の消滅した作品もあるのです。それは著作者の没後50年以上たった作品です。そのような文学作品の多くが研究者やボランティアの手によってテキスト化され、WWW上に公開されています。

では、「羅生門」を探してみましょう。

キーワードを「芥川龍之介 羅生門」、検索設定を「すべての語を含む」で検索 156件

リストの上位に「作家別本のリスト」(http://www.voyager.co.jp/aozora/sakka_a.html)というページがありますね。このページには芥川龍之介の作品が多く並んでいます。探している「羅生門」もありました。

「芥川龍之介『羅生門』」というリンクをたどると、「図書カード」(<http://www.voyager.co.jp/aozora/cards/akutagawa/rasyoumon.html>)というページがあります。

ここには羅生門の出典や入力者の氏名、内容や解説、著者についての解説があり、「テキストファイル ルビなし」、「テキストファイル ルビあり」、「HTML版」、「エキスパンドブック版」の4種類のファイルをダウンロードすることができます。画面上でそのまま読むのならHTML版でもいいのですが、あとでじっくりと読みたいという方は「テキストファイル」をダウンロードするのがよいでしょう。書店や図書館に行かなくても自宅でこんなに簡単に手に入るので、便利な世の中になったものです。しかも無料です!

では、もう少しこの便利なサイトの解説をしておきましょう。サイト全体の名称は「青空文庫」(<http://www.voyager.co.jp/aozora/>)です。文学に親しむ人にとっての定番といえるこのサイトには、著作者の没後50年以上を経て著作権が消滅した作品と、著作権者が公開に同意した作品を数多く収録しています。主に明治、大正、昭和初期の作品が多いようです。「著作権が消滅した作家一覧」もありますので、興味のある方はボランティアでこのサイトの運営に協力することもできます。インターネットだからこそできたすばらしい企画ですね。

さて、近代の作品ばかりでなく古典文学にも親しみたいものです。実は古典を集めたサイトも青空文庫からリンクされています。例えば「古事記」本文のリンク先は次のサイトです。「日本文学等テキストファイル」(<http://kuzan.fukui-u.ac.jp/bungaku.htm>)

福井大学教育学部の岡島昭浩さんの運営するサイトですが、有名な古典文学はほとんど網羅しているのではないのでしょうか。一般の人はもちろん、特に高等学校の国語の先生に見ていただきたいサイトです。



活用 2

懐かしいホラー映画の定番「オーメン」。
この映画の中で数字の「666」が悪魔の印として登場しましたが、
この数字は新約聖書のどこからの引用でしょう。



ポイント

まず聖書の検索サイトをチェック!

解説

もう随分前の映画ですが、30代以上の方はよく知っているのではないのでしょうか。

では、聖書の用語が検索できるようなサイトを探してみましょう。さらに聖書全文を見ることができるサイトもあるでしょうか。

キーワードを「聖書 用語 検索」、検索設定を「すべての語を含む」で検索 226件

リストのトップの「Bible Word Search」(<http://www.yone.ac.jp/takao/biblesearch.html>) がうってつけです。旧約、新訳の聖書の全文検索ができるサイトなのです。

「新訳」にチェックを入れて、キーワード欄に「666」と入力して検索すると...

「rev:13:18 ここに知恵がある、思慮ある者はその獣の数字を数えなさい。その数字は人間をさしているからである。その数字は【666】である。」

見事に「666」を含む節が検索されました。

しかし「rev」とは？ 最初のページに「聖書分類表」というのがあり、ここを見ても、「Revelation Rev ヨハネの黙示録」と記されています。つまり先ほどの文章は「ヨハネの黙示録 13章18節」ということになるのです。

では「聖書分類表」のとなりにある「聖書通読表」とは？ ここは旧約、新約聖書の日本語、英語の全文を見ることができるページです。

このサイトは山形県立米沢女子短期大学社会情報学科の高尾哲康さんが運営するものです。宗教学の研究には欠かせないサイトでしょう。



聖書は引用文の宝庫。
小説や映画にたくさん登場しているぞ。
ほかの言葉も探してみよう!



検索活用術

ウェブサイトを活用しよう!

活用 3

無料で新聞記事のデータベースを検索できるような
サイトを探してください。



ポイント

記事検索サイトを検索しよう!

解説

文学や宗教もいいのですが、やはり日頃のニュースは気になります。過去にさかのぼって新聞記事を検索できるようなサイトはあるでしょうか。大手の新聞社はすべて新聞記事のデータベースを持っていて、記事の全文検索サービスもありますが、そのほとんどは有料です。無料で記事検索ができれば大変便利なのですが...

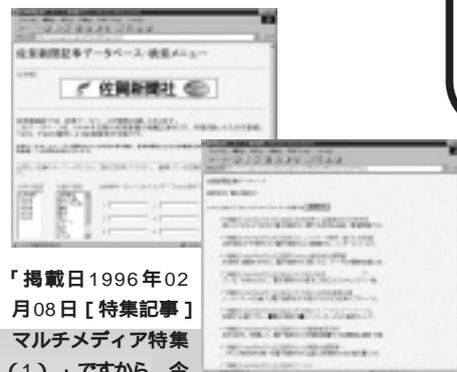
キーワードを「新聞記事 検索 無料」、検索設定を「すべての語を含む」で検索 470件

これもリストのトップでOKです。

「佐賀新聞記事データベース・検索条件入力」(<http://www.saga-s.co.jp/pubt/ShinDB/search.html>)

このサイトの説明によると、1994年以降の佐賀新聞の掲載記事のうち、外電を除いたものをデータベースとして無料公開しているとのこと。地方新聞とはいえ、これでもずいぶん役に立ちそうです。ためにしに最近話題の「電子商取引」を検索してみましょう。

検索の設定を「全年指定」「全面指定」としてキーワードを「電子商取引」で検索 51件
結構ありますね。ちなみに一番古い記事は



「掲載日1996年02月08日 [特集記事] マルチメディア特集(1)」ですから、今から2年半前にさかのぼります。そういえばインターネット上で商用サイトが急激に増えだしたのがちょうどこの頃ではなかったでしょうか。最近「EC」というと「エレクトロニック・コマース：電子商取引」のことを指すそうで、誰も「ヨーロッパ共同体」を想像しなくなりましたね。時代の流れを感じます。

バリューサーチとは？

記事情報の無料サービスといえば、忘れてならないのがgooの「バリューサーチ」です。7月末現在では、アクセス頻度で絞り込んだ300万URLにのぼる主要なWWWデータからの検索のほか、BizTech（日経BP社）のニュース・雑誌記事からの全文検索、紀伊國屋書店BookWebからの検索、帝国データバンクとアスキー企業URLからの企業情報の検索が無料で行えます。



活用 4

バリューサーチを使った検索をしてみよう!



ポイント

絞り込みが簡単!

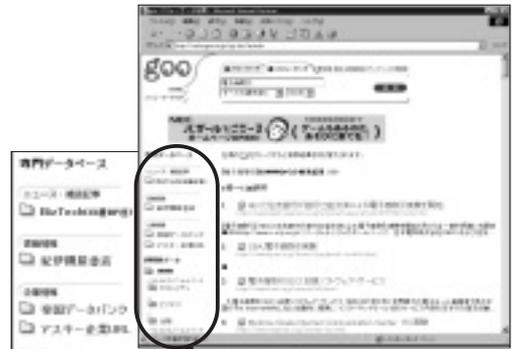
解説

それでは実際に「電子商取引」というキーワードで検索してみましょう。検索結果件数はすべて7月末時点での数字です。

キーワードを「電子商取引」で検索 WWWからの検索結果 268件

ここから各ページを見にいってもいいのですが、バリューサーチの特徴として「カテゴリによる絞り込み」があります。ページの左側にセキュリティ、ビジネス、金融、イントラネット

.....といったカテゴリが見られますね。これらを使って「電子商取引」をさらに細かく分類しているのです。たとえば「セキュリティ」というカテゴリならば、電子商取引における個人認証や暗号化についてのWebページを一覧で出してくれるというわけです。パワーサーチの場合は膨大な検索結果を絞り込むときの適切なキーワードを自分で考えなくてはいいませんが、バリューサーチの場合は「カテゴリ」という形で絞り込みの案を提示してくれるというわけです。ちょっとした調べものでしたら、パワーサーチよりも便利な機能ですね。



ウェブサイトを以外の商用データベースから有効に検索できるぞ!



ポイント

ウェブ検索以外のデータベースも使える!

解説

では「電子商取引」の記事情報はどうでしょう

「[電子商取引]のBizTech（日経BP社）からの検索結果 157件」

さすがにBizTechです。電子商取引に関する記事情報は豊富ですね。通常のWebページと違い、日々更新、蓄積される記事情報からの検索ですから、常に新鮮な情報が得られるわけです。

さて、さらに深い情報収集となると書籍は欠かせません。インターネットでの情報収集にも限界がありますから、書籍の果たす役割は大です。そこで便利なのが「紀伊國屋書店BookWeb」からの検索です。

「[電子商取引]の紀伊國屋書店BookWebからの検索結果 5件」

ここで書籍の題名や価格のほかに簡単な内

容まで確認できます。さらに紀伊國屋書店BookWebの会員ならば、その場で注文もOK。数日後には自宅までその書籍が届けられるというわけです。

今回は「電子商取引」というキーワードであったため、帝国データバンク、アスキー企業URLからの検索はできませんが、企業名検索

の場合はこれらのデータベースも活躍しそうです。

以上のように、バリューサーチの使い方も習得しておくくと、強力な情報武装ができそうですね。



検索活用術

バリューサーチを使ってみよう!

Tips!

このコーナーではgooの「くせ」や検索の際に知っておくと便利な情報を紹介します。

検索エンジンを意識したWebページ作り

皆さんはgooの「使い方」のページをご覧になったことはありますか。gooについての詳しい情報が満載ですから、ぜひ一度目を通すことをお勧めします。特にWebページを自分で作っている人には必見の、興味深い情報をいくつか紹介しましょう。

Webサイトを運営している方なら「検索結果の上位にランク付けされたい」と、常に思っていることでしょう。gooは検索語に対して各Webページにスコア(%表示)を付けていますが、そのスコアには次の事柄が関係しているようです。

- 1.ドキュメントにおける語の頻度：検索語の頻度が多ければスコアが高くなる。
 - 2.タイトル内の検索語：ページのタイトル部分(ソースの<title> ~ </title>の部分)に検索語があるとスコアが高くなる。
 - 3.キーワード内の検索語：これはメタタグ「<META NAME="keyword" CONTENT="キーワード">」の「キーワード」の部分に検索語があるとスコアが高くなる。
- この3点に注意してページ作りをすれば検索結果リストの上位に来るはず。デザインを意識しすぎて見出しのような重要な文字情報も画像を使っているようではいけませんね。

【ロボットの行動を制御する方法】

検索エンジンに検索されることによって、そのサイトのアクセス数は増加するわけですが、逆に検索されてほしくないページもあるかもしれません。たとえばプライベートな情報やCGI関係のファイルを置いているディレクトリーを直接覗かれるのは困りますし、サイトの構造上、必ずトップページから入ってほしいという場合もあると思います。そこで、ロボットがサイトヘッロー(情報収集)に来たとき、その行動を制御する方法があります。サーバーを直接管理できる立場の人なら次の方法が有効です。

・ robots.txt という名のテキストファイルを作成する。

たとえばhttp://www.a.co.jp/というサイトならばhttp://www.a.co.jp/robots.txtとなる場所にrobots.txtを作成します。その内容は、「User-agent:」によって対象とするロボットを指定し、「Disallow:」でクローリングを禁止するディレクトリーを指定するのです。すべてのロボットに対して、「/cgi-bin/」、「/tmp/」というディレクトリーをクローリングさせないためには

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
とすればいいでしょう。
```

さて、多くの人はプロバイダーの無料Webスペースサービスを利用してサイトを運営しているわけですから、上記の「robots.txt」の方法は使えません。次、メタタグによる制御はできません。

・メタタグによる制御
HTMLのソースの<HEAD> ~ </HEAD>の中に次

のようなメタタグを記述しましょう。

```
<META NAME="robots" CONTENT="(1),(2)">
```

(1)の部分はそのページをインデックスさせる(gooのデータベースに登録させる)かどうかの指定です。

index : そのページをインデックスさせる

noindex : そのページをインデックスさせない

(2)の部分はそのページからリンクされたページをインデックスさせるかどうかの指定です。

follow : そのページからリンクされたページをインデックスさせる

nofollow : そのページからリンクされたページをインデックスさせない

例えば、サイトのトップページだけが検索エンジンに検索され、他のページは検索されないようにするには、トップページに

```
<META NAME="robots" CONTENT="index, nofollow">
```

というメタタグを記述し、他のページは

```
<META NAME="robots" CONTENT="noindex, nofollow">
```

としておけばいいわけです。

詳しくは以下のページをご覧ください。

ロボットにクローリングさせないこと一般(英語)

<http://info.webcrawler.com/mak/projects/robots/exclusion.html>

メタタグについて(英語)

<http://info.webcrawler.com/mak/projects/robots/meta-user.html>

ロボットの種類(英語)

<http://info.webcrawler.com/mak/projects/robots/active.html>

今回の検索結果は98年8月1日現在のものです。

宿題と解答

【先月の宿題】文部省の調査で、大学(短大を除く)への女子の進学率が初めて20%を超えたのは平成何年のことでしょうか。

【解答】

政府の発表した資料からの検索ですから、パワーサーチの追加機能の「検索先設定」をうまく使いました。なお、日本政府に共通するドメインは「go.jp」です。

さて、目的の文部省の資料ですが、数年間の大学の進学率が一覧になったものがあるはずです。例えば「進学率の推移」といった言葉が使われているのではないのでしょうか。

キーワードを「大学 進学率 推移」、検索設定を「すべての語を含む」、追加機能の検

索先設定のサーバロケーションの設定を「go.jp」として検索 38件

検索結果リストの中の「統計資料」(<http://www.sorifu.go.jp/danjo/statistics2/index.html>)というページを見てみましょう。「第10表 進学率の推移」(<http://www.sorifu.go.jp/danjo/statistics2/10.html>)というページへのリンクが設定されています。このページには大学学部への男女別の進学率が一覧にありました。

この表によると平成5年は19.0%、平成6年は21.0%というのが女子の進学率ですね。

【正解】平成6年当選者は福山誠さん、長沢圭介さん、中村正裕さん、鬼頭幸広さん、原井彰弘さんです。

今月の宿題

毎月、正解者のなかから5名に特製テレホンカードをプレゼント!

問題

日本国憲法の中に、「【学問の自由】学問の自由は、これを保障する。」という条項があります。さてこれは第何条?

解答と、答えがあったサイトのURLをメールに書いて送ってください。メールのタイトルを「TETSUJIN」にしたうえで、ip-box@impress.co.jpまで。9月11日24:00まで到着分有効。正解と当選者は来月のこのページで。



[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の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