

第8章 ウェブページの活用

- ✓ 自分の知りたい情報のあるウェブページを探す
- ✓ 検索エンジンを使う
- ✓ ウェブページを印刷する

1. 検索エンジン(サーチエンジン)とは

データベースやウェブページの検索では、見出しや著者など文書に含まれる特定の要素のみを対象に検索する場合と、文書に含まれるすべての情報を対象に検索する場合がある。後者を全文検索と言う。検索範囲としては最も広く、最も「漏れ」の少ない検索方法であるが、逆に細かい条件を指定しにくく絞り込みにくい方法でもある。

検索エンジンの役割は、

- ①クローラー(ロボット)はウェブの各ページを自動的に巡回し情報を収集。
- ②キーワードを抽出し、整理してインデックス(索引)データベース作っておく。
- ③ユーザーが探している情報を入力すると、検索語に分解¹して内容を把握。
- ④ランキングアルゴリズムに基づき、インデックスされているウェブページをスコアリングする。
- ⑤点数の高い順にウェブのアドレス(URL)のリストを返す。

2. 検索エンジンの例

ロボット検索エンジンの代表的なものは以下である。

(1) Yahoo! JAPAN yahoo.co.jp

(2) Google google.co.jp

そのほか、ウェブページだけでなく、画像、地図、ニュース等に的を絞った検索ができ、「Fresheye」は画像や動画の検索を得意としている。それぞれのエンジンが得意としている分野があるので、色々使ってみよう。ニュース、海外など分野別にエンジンの一覧(リンク集)を載せている「検索デスク」<https://www.searchdesk.com/>というサイトもある。

¹ 形態素解析と言い、自然言語処理の手法の一つで、文章を単語に分解して内容を判断するために使う。

「Yahoo!JAPAN」のように、インターネットには、その入り口として機能する巨大なウェブサイトがいくつか存在する。そういったサイトのことを**ポータルサイト**と呼ぶ。ニュースや株価などの情報提供サービス、ウェブメールサービス、電子掲示板、チャットなど、ユーザーがインターネットで必要とする機能が用意されている。

3. キーワード検索

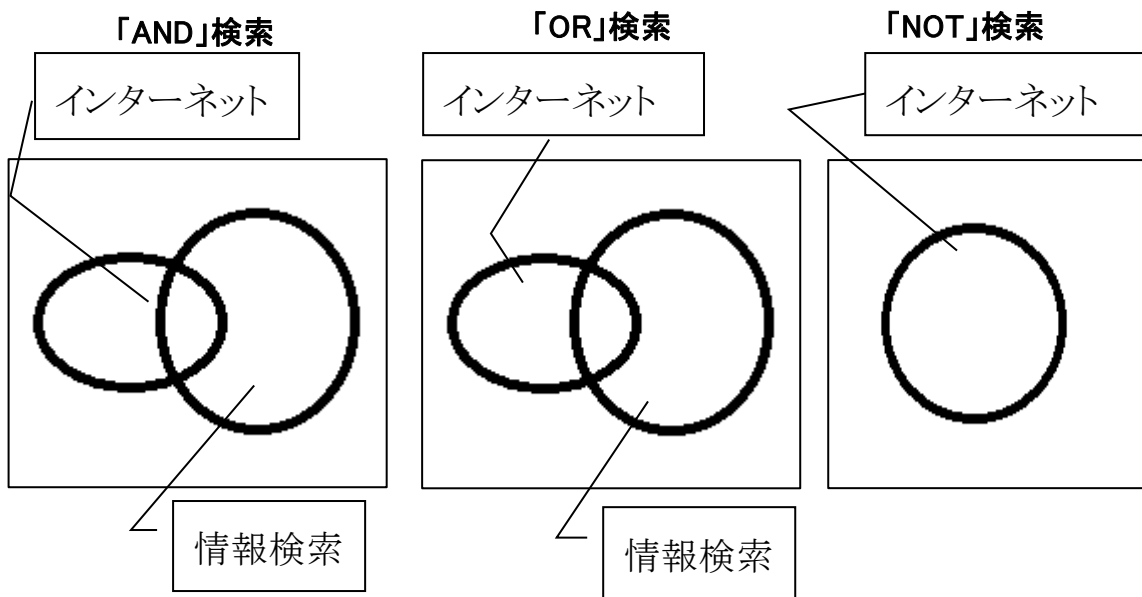
調べたい事柄に関係のある言葉(キーワード)を使って探す(☞ **演習1~5**)。

- キーワード検索の窓にキーワードを入力し、[検索]をクリックする。
- 該当が多すぎたときは先に入力したキーワードの後にスペースを入れ、さらに別のキーワードを追加し、[検索]をクリックする。
- 出てきたホームページ一覧から、目的とするタイトルをクリックする。

4. 「AND」と「OR」を用いた検索

(1) 「AND」 「OR」 「NOT」

「AND」検索は、その両方が含まれた結果が返され、「OR」検索はその少なくともどちらか一方が含まれた結果が返される。



検索結果の集合である。「OR」を用いた検索結果のほうが「AND」を用いた検索結果よりも数が多くなる(☞ **演習4**)。

「NOT」を使った検索では、キーワード以外の結果が返される。

(2) AND 検索

ここでは図書館のOPAC (蔵書検索システム)を用いて検索をする。

図書館トップ²/資料を探す/本学蔵書検索(OPAC)をたどる。

検索条件の枠に「インターネット」と「情報検索」を空白で区切って入力し、

[検索]ボタンを

クリックしよう。

すると、「イン

ターネット」と

「情報検索」の

両方含まれる

本が図書館の

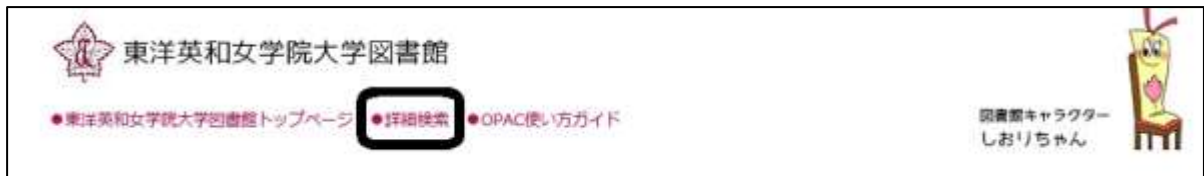
蔵書目録の中

から探し出される。



(3) OR 検索

「詳細検索」をクリックする。



キーワードの欄に「インターネット」と「情報検索」を空白で区切って入力し、

「いずれかを含む」を選択する。すると、「インターネット」と「情報検索」のい

れかが含まれる

本が図書館の

蔵書目録の中

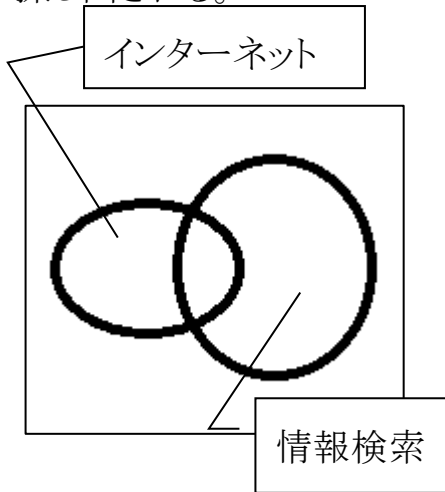
から探し出される。



²本学図書館のウェブページのURLは <http://www.toyoeiwa.ac.jp/daigaku/shisetsu/library/>

(4) [発展] NOT 検索

「詳細検索」のページで、キーワードの欄に「インターネット」を入力し、「いずれかを含む」を選択する。タイトルの欄に「情報検索」を入力すると、「インターネット」の中で「タイトルに情報検索」を含まないものが蔵書目録の中から探し出される。



5. クッキー

(1) クッキーとは

自分のユーザーアカウントを入力しなくても自動的に表示される。以前訪ねたことがあるウェブサイトで2度目には「こんにちは〇〇さん」と表示される。しかし、そのウェブサイトを閲覧したことのない別のパソコン(例えば、大学)で同じウェブサイトを見ても、「こんにちは〇〇さん」と表示してくれるわけではない。



ウェブアクセスの基本的な仕組みである HTTP は、アクセスを1度きりの単位として扱うので、複数のウェブアクセスを関連づけて管理することはできない。そこで、この弱点を補うために開発されたのがクッキー(Cookie)である。ウェブページにアクセスするときにやり取りされる小さなテキストデータである。

(2) クッキーの仕組み

- 《ユーザーが初めてウェブサイトを訪問したとき》

ブラウザ側ではウェブページを表示するとき、そのウェブサイトからクッキーを受け取る。受け取ったクッキーをパソコン内に保持しておく。

- 《ユーザーが再び同じウェブサイトを訪問したとき》

ブラウザ側では保持していたクッキーをそのウェブサイトに送り返す。ウェブサイト側はそのクッキーを参照し、前と同じユーザーであることを確認する。

(3) クッキーの中身

「名前とその値」、「有効期限」、「適用範囲」が書かれている。これにより、発行元と違うウェブサイトにクッキーを送らないようにしたり、不正なウェブサイトへクッキーを送信してしまう危険性を回避することができる。

(4) クッキーの危険性

クッキーを通信途中で横取りされ³、横取りしたクッキーをサーバーに提示することで、本来のユーザーになりすまして、そのユーザーのクレジットカードや個人情報などを盗み取ることができてしまう。このような危険を避けるためにも、金融サイトやショッピングサイトなどでは、必要な作業が終了したら、**手動でログオフ**するようにしよう。

クッキーはウェブサイトのサーバーが作成・発行し、ブラウザは受け取ったクッキーの内容に合わせて動作するだけである。クッキーにどのような情報を入れるかは発行するサーバー次第であり、クッキーに個人情報を入れてやりとりすることもでき、悪意を持ったサイトが怪しいクッキーを作成することも可能である。

(5) クッキーの危険性と安全対策

それぞれのウェブサイトごとに、そのクッキーを使うかどうかを判断し、**おかしい、怪しいと思ったサイトのクッキーは使わない**ようにする。

日頃訪れるウェブサイトのクッキーを把握し、それ以外は削除してしまう方法もある。例えば、I E ではVer6 から、クッキーの拒否設定を簡単に行うことができるようになっているので、**不要なクッキーは削除**する。

重大な情報を扱うサイトでは、クッキーの有効期限がセッション限りであることを確認する。

上記のようなサイトにログイン中は、決して他のサイトを見に行かない。

上記のようなサイトでは、サービス利用後に「ログオフ」したり、すべてのブラウザを閉じる。


クッキー自体は便利な機能を提供してくれる。しかし、ユーザーのあずかり知らないところで、情報のやり取りが行われている可能性があること、クッキー

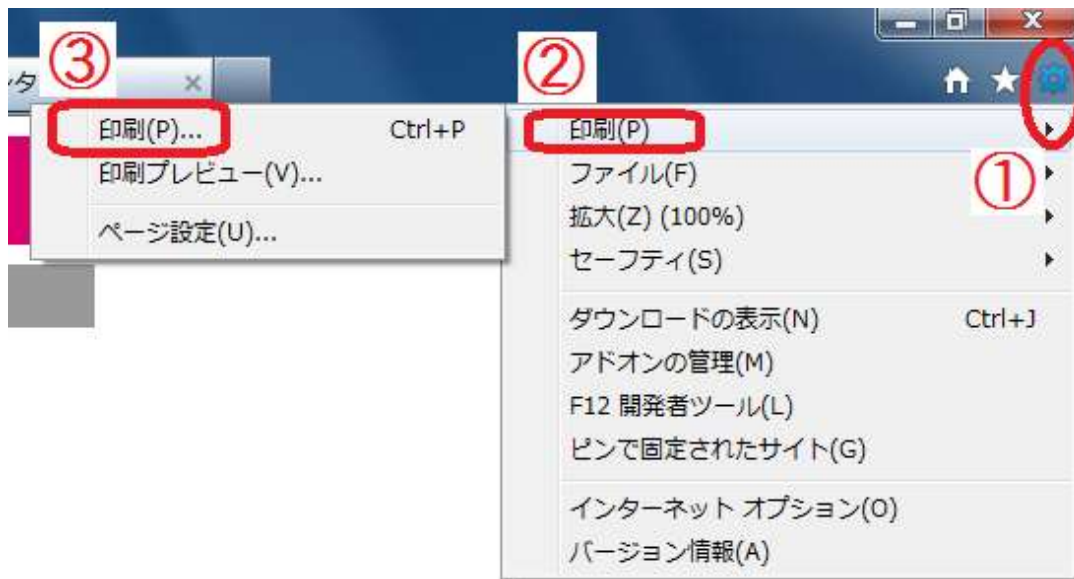
³ 「クロスサイトスクリプティング」という手法

を利用したユーザーの追跡や情報収集が行われていることは、知っておいた方がよい。

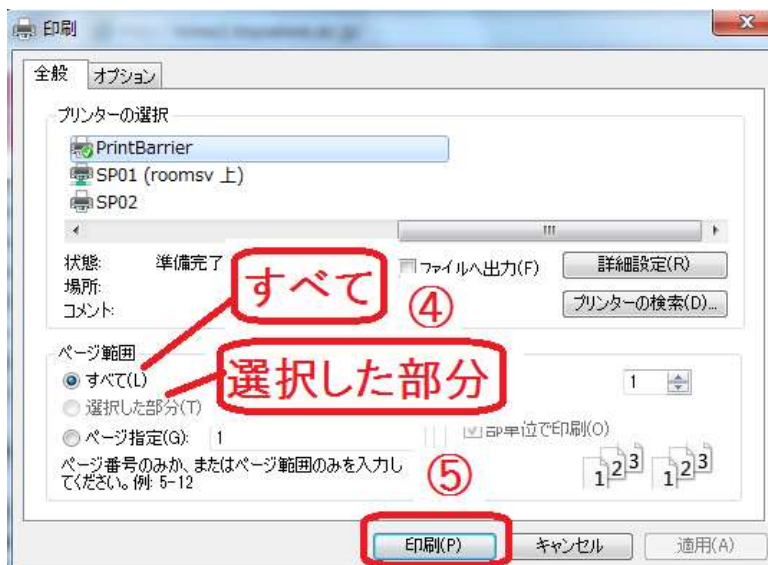
6. ウェブページの印刷

Internet Explorer では、印刷プレビューを確認して、印刷範囲をページで指定したり、ブラウザ上で選択した範囲のみを印刷したりすることができる。

- ① ツールボタン 
- ② 印刷を選択。
- ③ 印刷を選択。



- ④ 印刷する範囲を指定して
- ⑤ 印刷ボタンをクリック。



コラム 圧縮と解凍

インターネットでダウンロードしたファイルや、電子メールで送られてきたファイルが読めなくて困ったことはありませんか。普通、ファイルはそれを作ったソフトウェアで読みやすい形で保存されています。しかし、ファイルの転送時間をより短くするために、圧縮して、小さいサイズのファイルを作ることができます。画像や音楽のファイルは文字のファイルに比べ非常に大きいので、よく圧縮してあります。この圧縮してあるファイルは、そのままでは使えず、**解凍**してソフトウェアで使えるように戻さなければなりません。

自己解凍形式のファイル(拡張子.sea)は、ファイルを実行(例えばデスクトップに置いてダブルクリック)すれば、ファイルが出現し使えるようになります。圧縮されているかどうかは、拡張子(☞**第4章 5 節**)を見ればわかります。日本で広く使われている「L H A 形式」(拡張子.lzh)と世界中で広く使われている「ZIP形式」(拡張子.zip)があります。

また、+Lhaca、LHMelt、Lhaz のような圧縮・解凍のフリーソフトウェアがあります。

デジカメで取った写真をウェブサイトには、「JPEG形式」(拡張子.jpg)を使い、転送時間短縮のため、人間には画像の劣化がほとんどわからないようにして、画像を圧縮しファイルサイズを小さくします。ただし、一度圧縮したら元には戻らない不可逆圧縮です。

ウェブサイトの画像に「PNG形式」(拡張子.png)と呼ばれる新時代の国際標準画像ファイル形式が誕生しました。きれいな画質が保たれ、可逆性ですが、ファイルサイズが JPEG 形式より大きく、ブラウザやアニメーション機能には対応していません。

