

4

Appunti di Laboratorio di Informatica

La ricerca di informazioni in rete

COSA SI IMPARA IN QUESTO MODULO:

- ✓ Quali tipi di risorse si possono trovare su Internet
 - ✓ Come si sfruttano le pagine-indice, i servizi di ricerca, i servizi bibliografici e i documenti scaricabili
 - ✓ Quali astuzie conviene adottare per rendere efficiente il lavoro di ricerca
-

Ricerca di informazioni in rete

Internet comprende moltissime sorgenti di informazioni. Il problema tipicamente consiste nel selezionare fra una enorme numero di sorgenti, ciascuna di grandi dimensioni.

Esistono diversi servizi che consentono l'accesso alle informazioni disponibili in rete.

Per quello che riguarda la ricerca scientifica, essi sono principalmente:

- le **pagine-indice** delle varie organizzazioni che fanno o finanziano la ricerca (Università, Dipartimenti, gruppi universitari di ricerca, gruppi e consorzi di studio, singoli studiosi, enti di ricerca non universitari, Ministeri, la Comunità Europea...)
- i servizi di accesso in rete ai **cataloghi** delle biblioteche
- i **motori di ricerca (search engines)**, che mantengono enormi indici di pagine su cui svolgono ricerche su richiesta degli utilizzatori.
- le **directory**, che catalogano grosse quantità di pagine per argomenti e sotto-argomenti
- i **documenti elettronici scaricabili**, che si possono leggere sul monitor o stampare su carta
- i **corsi in linea**, vere e proprie lezioni in rete

Una pagina-indice (**index page**) contiene link ad altre pagine su un dato argomento

Oltre a conoscere l'uso di ciascun servizio, l'utilizzatore esperto conosce anche gli accorgimenti per raggiungere in breve le informazioni utili e filtrare quelle inutili. Alcuni suggerimenti verranno dati qui, altre tecniche si imparano dall'esperienza.

Pagine-indice

I centri di ricerca, i singoli studiosi di una determinata disciplina, e in generale chiunque sfrutti a fondo il WWW, fanno spesso il lavoro di raccogliere, catalogare e mettere a disposizione i link in rete più utili secondo la loro esperienza.

Una pagina WWW che contiene solo o prevalentemente indirizzi di altre pagine è detta una **pagina-indice (index page)**.

Per sfruttare le pagine indice occorre conoscere l'indirizzo di rete della pagina dell'ente, gruppo o persona che mantiene l'indice.

Per esempio, se io so che il Prof. Toluene dell'Università di Smarburg in Germania è uno studioso di microspore, per trovare un elenco di siti che trattano le microspore mi procurerò l'indirizzo della pagina personale del prof. Toluene:

<http://www.unisma.de/~toluene>

e verificherò se egli mette a disposizione un elenco di link che, nel suo lavoro, ha trovato utili. (**Nota: nome e indirizzo sono di fantasia**)

SE NON SO L'INDIRIZZO?

Le università italiane hanno indirizzi WWW della forma <http://www.uniXX.it> (sostituire a XX la sigla della città). I politecnici italiani hanno indirizzi WWW della forma <http://www.poliXX.it> (idem).

Esempio: <http://www.unige.it> Università di Genova

Esempio: <http://www.unimi.it> Università di Milano

Anche in molti Paesi stranieri esistono convenzioni analoghe.

Gli indirizzi accademici USA sono del tipo www.NOMEISTITUZIONE.edu ; quelli britannici sono del tipo www.NOMEISTITUZIONE.ac.uk ; quelli tedeschi sono del tipo www.uni-NOMECITTÀ.de .

I siti personali dei singoli sono spesso accessibili passando (cliccando successivamente sui link presenti sulle pagine):

- dal sito dell'istituzione
- poi da quello del dipartimento
- poi da quello del gruppo di ricerca a cui appartengono.

Queste regole consentono di "**tirare a indovinare**" un indirizzo non conosciuto, con buona probabilità di successo.

Se uno o due tentativi non dovessero andare a segno, si consiglia di ricorrere al sito del MURST (ministero dell'università), www.murst.it .

Dalla home page del sito MURST, cliccando su "Università", si trovano un paio di link a pagine-indice che elencano i siti di moltissime università in tutto il mondo (alla voce intitolata "atenei").



Collegarsi al sito <http://www.unipv.it/webbio/fastlink.htm> . È una pagina-indice di risorse utili agli studiosi di biologia animale, messa a disposizione dal Dipartimento di Biologia Animale dell'Università di Pavia.

Collegarsi poi a <http://www-phys.science.unitn.it/useful-links.html> . È una pagina-indice di risorse utili ai fisici, messa a disposizione dal Dipartimento di Fisica dell'Università di Trento.

Cercare in rete altre pagine-indice su temi specifici, utilizzando i suggerimenti forniti sopra.

Quello che si è detto finora vale se si sa già dove cominciare a cercare: p.es., se si conosce l'indirizzo WWW di uno studioso che ha una pagina in rete; p.es., se si conosce il nome di una istituzione che si occupa della materia di interesse.

E se non si sa da dove iniziare?

Motori di ricerca e altri servizi in rete

MOTORI DI RICERCA

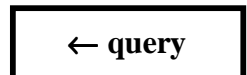
Un **motore di ricerca** è un servizio in rete che ha la funzione di indicizzare **automaticamente** le pagine del World-Wide Web.

A un motore di ricerca si accede attraverso una pagina WWW con un proprio indirizzo.

Su questa pagina si trovano principalmente 2 elementi:

1. Un gran numero di **pubblicità**, fonte di finanziamento del servizio (che per l'utilizzatore è di solito **gratuito**)
2. Una zona per effettuare la ricerca, comprendente una casellina per introdurre la **query**, il tasto **Search** (o simili), ed eventuali altre opzioni.

Nel gergo delle basi di dati, si chiama **query** l'espressione di una richiesta di dati che soddisfino un dato insieme di condizioni.



Nel caso dei motori di ricerca, la query più semplice è costituita da un semplice termine da cercare.

Premuto il tasto Search (o quel che è), si ottiene una o più paginate di risposte. Ogni risposta comprende:

- il titolo di una pagina WWW;
- l'indirizzo (cliccabile) della pagina;
- un estratto della pagina, per avere un'idea del contenuto;

- eventuali altre informazioni, come p.es. la dimensione in byte della pagina, la data di creazione o ultimo aggiornamento, ecc.

Alcuni popolari motori di ricerca:

<http://www.altavista.com>

<http://www.excite.com>

<http://www.google.com>

<http://www.lycos.it>

<http://www.yahoo.com>

← Indirizzi di motori
di ricerca

METARICERCA

Un tipo interessante di ricerca negli indici è la "metaricerca". Alcuni siti forniscono il servizio di ricerca non nel proprio indice, ma in quelli di numerosi altri motori di ricerca. In questo modo si ha a disposizione una collezione di indici, anziché un indice unico.

Una simile raccolta di motori di ricerca si trova p.es. all'indirizzo

<http://cui.unige.ch/meta-index.html>

(ATTENZIONE: qui unige sta per Università di Ginevra, non Genova: notare che il dominio nazionale è **.ch**, non **.it**).

Si può anche provare a chiedere alla mamma...

<http://www.mamma.com>

DIRECTORY E ALTRI SERVIZI DI RICERCA IN RETE

La parola **directory**, sul dizionario, indica un **elenco del telefono**, come p.es. le pagine gialle.

Un sito definito **directory** ha la stessa struttura delle pagine gialle, ossia si consulta per **argomento** e all'interno di un argomento si consulta per **nominativo**.

Una possibile differenza tra la versione di carta e la versione in rete è nel fatto che in una directory in rete esistono **argomenti**, **sotto-argomenti**, **sotto-sotto-argomenti**, ecc.

Un'altra differenza è che i nominativi dei siti (titolo o indirizzo) possono essere organizzati sia **alfabeticamente**, sia secondo **altri criteri** (p.es. per popolarità).

In sostanza, una directory non è altro che una collezione di numerose **pagine indice**, relative a molti argomenti diversi e organizzate in modo logico per facilitare l'accesso.

I principali portali e motori di ricerca hanno sia funzione di motore di ricerca, sia funzione di directory.

Essi hanno tipicamente una home page del tipo:

← motore di
ricerca + directory

— Pagina principale —

Cerca:

- Argomenti
- Altri argomenti
- Ancora argomenti
- Cose utili
- Cose inutili

Accanto al comando “cerca” c’è una casella dove inserire la propria query.

Nel resto della pagina ci sono tutti gli argomenti principali della directory.

Sparse per tutta la pagina (e qui non rappresentate) si trovano infine numerose pubblicità.

Indirizzi di alcune directory:

← Indirizzi directory

<http://dmoz.org/>

<http://www.tc.cornell.edu/Edu/MathSciGateway>

<http://directory.google.com/Top/Science/>

<http://directory.netscape.com/Science/index.html>

http://www.lycos.it/dir/Scienza_e_Tecnologia/index.html

<http://listoflists.com/Top/Science/>

Come strutturare una query

A questo punto, anche se l’argomento non è completato, si può accedere già ad un enorme quantità di informazioni.

Per sfruttare bene gli strumenti presentati conviene quindi capire qualcosa in più su **come è fatta una query**.

Tipicamente, le parole di una query vengono cercate all’interno dei documenti in qualunque ordine e non necessariamente tutte insieme.

Inoltre esse vengono filtrate per eliminare parole troppo comuni (p.es. gli articoli). Questa operazione è detta **stoplisting** perché la lista delle parole da “non far passare”, o stopwords, è detta stoplist.

← stoplisting

Esse vengono anche filtrate per eliminare le varianti (p.es. desinenze dei verbi, plurali). Questa operazione è detta **stemming**, perché, privata della desinenza, una parola si riduce alla sola radice (stem).

← stemming

Gli indirizzi che vengono riportati corrispondono quindi a pagine che contengono:

- Tutte o alcune delle parole della query
- Parole simili a quelle della query (p.es. un'altra voce dello stesso verbo)
- A volte, parole con significato correlato a quello delle parole della query

← Ricerca di tipo
“information retrieval”

“Analogica”

Cerca tutti i documenti
potenzialmente interessanti

Gli indirizzi vengono inoltre ordinati per **rilevanza**.

Un motore di ricerca ritiene più rilevante un documento in base a criteri diversi (infatti motori diversi danno ordinamenti diversi).

Ci si può aspettare che un documento sia più rilevante, rispetto alla query, se:

- Contiene più parole della query
- Contiene le parole della query vicine fra loro
- Contiene le parole della query nel titolo

Inoltre i motori di ricerca tentano anche di valutare l'**affidabilità** di un documento, ritenendo meno affidabili:

- I documenti che nel testo **non** hanno le parole del titolo
- I documenti che presentano **liste di parole-chiave nascoste**
- Documenti per i quali hanno ricevuto **segnalazioni** di insoddisfazione da parte di utilizzatori delusi.

Documenti inaffidabili saranno in coda alla lista presentata.

Segue una piccola guida su come strutturare una query.

SPECIFICITÀ

Mettere dei termini di troppo non esclude i documenti che non li contengono. Quindi conviene mettere tutti i termini utili.

Inoltre, poiché esistono i filtri che eliminano le parole inutili, è possibile anche inserire come query **una domanda vera e propria**. P.es.:

`does my IBM computer work with Windows 2000?`

COME OBBLIGARE L'INCLUSIONE DI UN TERMINE

Se si vuole che un termine sia considerato obbligatorio anziché preferenziale si può usare il segno + davanti al termine. P.es. con

`IBM Windows`

troverò molti documenti che parlano di calcolatori IBM oppure del sistema operativo Windows. Invece, con

```
+IBM +Windows
```

troverò solo i documenti che parlano di Windows su un calcolatore IBM.

Questo accorgimento serve a ridurre le “inondazioni di risultati” e ad evitare di annegarci dentro.

COME OBBLIGARE L'ESCLUSIONE DI UN TERMINE

Se si vuole evitare di ottenere tutte le pagine che riguardano un “secondo significato” dei termini della query, si può aggiungere un termine caratteristico solo del secondo significato, e imporre che esso NON sia presente aggiungendo il segno – davanti. P.es. con

```
medusa -giasone
```

otterrò tutti i documenti che contengono il termine medusa, escludendo però tutti quelli che si riferiscono al mito greco in cui compare anche il personaggio chiamato Giasone (e quindi, presumibilmente, selezionando quelli che si riferiscono al significato biologico del nome)..

COME OBBLIGARE L'ADIACENZA DI DUE TERMINI

Sono interessato a Windows 2000, ma con la query

```
Windows 2000
```

ottego anche tante pagine sugli eventi dell'anno 2000 e tanti cataloghi di installatori di finestre.

Posso allora richiedere che il motore mi cerchi proprio l'espressione “Windows 2000”. Per fare questo, racchiudo i termini dell'espressione fra virgolette.

```
"Windows 2000"
```

Attenzione: per fare questo, occorre che io sappia con ragionevole certezza che i termini sono **sempre usati in quell'ordine**. Se invece so che sono possibili varianti, **devo aggiungerle** come se fossero termini in più, distinti. Per esempio:

```
"camera dei deputati" "camera deputati"
```

(Questo perché più è precisa la mia richiesta, più selettiva sarà la ricerca).

Questi suggerimenti valgono con tutti i principali motori di ricerca. Conviene consultare la pagina di **help** se in qualche caso non dovessero funzionare (magari è diverso il modo di impostare le query, anche se le operazioni sono le stesse).

Il sito <http://www.searchenginewatch.com> fornisce altri utili suggerimenti per sfruttare al meglio i motori di ricerca e per comporre query più specifiche e più selettive.

Anche le pagine di aiuto dei singoli motori di ricerca forniscono preziose informazioni.



IN PRATICA

*Scegliere un tema di ricerca. Individuare in rete i siti più importanti relativamente a tale tema, utilizzando query appropriate da sottomettere a vari motori di ricerca
Riepilogare i risultati ottenuti in un rapportino che elenca i siti trovati, attribuisce loro un giudizio di effettiva rilevanza, riassume il tipo di materiale trovato ed indica le query e i motori di ricerca utilizzati.*

Un'altra categoria di informazioni reperibili in rete sono i lavori scientifici.

Una delle motivazioni originarie per la creazione di Internet era proprio lo scambio di informazioni fra istituzioni di ricerca.

Ovviamente la circolazione dei lavori **pubblicati** è vincolata dalle leggi sul copyright. Quindi un lavoro pubblicato su una rivista scientifica o negli atti di un congresso, di norma, **non sarà disponibile in rete**.

L'argomento che segue riguarda **come rintracciare in rete le informazioni bibliografiche sui lavori pubblicati**. Sarà mio compito, successivamente, recuperare in biblioteca tali lavori.

Successivamente, vedremo **come recuperare in rete il testo completo di lavori non pubblicati** oppure eventualmente **l'abstract di lavori pubblicati**.

Bibliografie e cataloghi in linea

Ho intenzione di iniziare un lavoro di ricerca scientifica.

La primissima cosa da fare è sapere qual è lo "stato dell'arte" sull'argomento di mio interesse. Ovviamente, se il lavoro che intendo fare è già stato svolto e pubblicato da qualcuno, riprenderlo da capo è uno spreco di tempo e risorse. Quindi devo fare una ricerca bibliografica per sapere quali sono le pubblicazioni sull'argomento. Perciò vado in biblioteca e comincio a scartabellare.

Qual è il difetto di questa tecnica?

Se mi interessasse sapere **se esiste almeno una** pubblicazione, potrei smettere di cercare appena la trovo. Purtroppo, io vorrei conoscere **tutte** le pubblicazioni sull'argomento, quindi devo fare una ricerca **esaustiva** in tutta la biblioteca.

← Vorrei evitare le ricerche esaustive

(Tutte significa tutte le principali; ovviamente non è possibile conoscere tutto.)

E se la mia biblioteca non tiene una rivista importante?

L'unico modo è rendere la ricerca più sistematica. I concetti di ricerca in rete che abbiamo imparato fin qui possono essere sfruttati nei modi spiegati di seguito.

RICERCA PER TEMI: PAGINE DI BIBLIOGRAFIA IN RETE

Si tratta di identificare un certo numero di **siti www** di persone che si occupano del tema di nostro interesse.

Esempio:

<http://www.sunysb.edu/sciencelibs/bioref.htm>

Come abbiamo visto, l'uso è di mettere in rete delle pagine-indice con gli indirizzi di interesse per il proprio specifico ambito di ricerca.

Allo stesso modo, frequentemente viene messa in rete anche una **bibliografia** tematica, ossia i riferimenti bibliografici dei lavori più importanti pubblicati sul tema.

Si tratta poi di annotarsi i riferimenti più interessanti e andare in biblioteca a "fare la spesa".

Suggerimento: i riferimenti bibliografici sono spesso molto sintetici: conviene copiarli per esteso, in modo da non tralasciare dati potenzialmente necessari (p.es. l'anno di pubblicazione).

Altro suggerimento: non è detto che nella mia biblioteca ci sia tutto, quindi, anche se mi sembra che due lavori parlino della stessa cosa, conviene che me li annoti tutti e due: così, se non trovo l'uno, posso ancora trovare l'altro.

Commento: il caso più frequente è quello di bibliografie tematiche, piuttosto che bibliografie per grandi discipline.

P.es. troverò più facilmente una lista bibliografica sui poriferi o sulla biologia marina del mediterraneo meridionale, piuttosto che una grossa lista sulla biologia animale.

Altro commento: nel caso specifico delle discipline biologiche, a differenza di altre discipline, si può osservare che in rete sono presenti numerose risorse orientate all'uso da parte delle **scuole elementari o medie**. Questo può rendere necessario sfrondare i risultati di una ricerca da quelli che non risultano di livello sufficientemente approfondito.

Siti con bibliografie di interesse biologico:

<http://www.stfx.ca/people/gbayesp/marinebioreference.htm> (St. George Bay)

<http://accstr.ufl.edu/biblio.html> (sea turtles)

http://www.sasked.gov.sk.ca/curr_inst/iru/bibs/secsci/index.html
(didattica scuola superiore)

← indirizzi di bibliografie

<http://version0.neonet.nl/ceos-idn/index.html> (vari temi)
<http://www.collegeboard.org/ap/biology/tc/biblio.html> (didattica)
http://www.savetheredwoods.org/r_res/rrbibl.htm (sequoie)
<http://www.soc.surrey.ac.uk/~scs1ec/evolbio.html> (evoluzionismo)
http://www.nal.usda.gov/pgdic/Probe/v1n3_4/bib.html (Computational Molecular Biology = bioinformatica e modelli matematici)
<http://www.bioanth.org/pacific/humanbio.htm> (biologia umana e antropologia nel Pacifico)
<http://www.lavc.cc.ca.us/Library/bib-biology.htm> (dal sito di una biblioteca)
[http://www.amicus.nlc-bnc.ca/wbin/resanet/resultsm/s=b/n=SU/l=1/d=1/r=1/e=0/h=10/t=Life+\(Biology\)--Bibliography--Catalogues](http://www.amicus.nlc-bnc.ca/wbin/resanet/resultsm/s=b/n=SU/l=1/d=1/r=1/e=0/h=10/t=Life+(Biology)--Bibliography--Catalogues) (Biblioteca Naz. Canada)
<http://www.sti.nasa.gov/Pubs/Aeromed/Aeromed.html> (questa è una **pubblicazione in linea** – vedere più avanti – edita dalla NASA, il cui scopo è formare una bibliografia in continuo aggiornamento sul tema della biologia e medicina aerospaziale)

RICERCA PER PAROLE CHIAVE: BASI DI DATI BIBLIOGRAFICI

Una “base di dati” o **database** è, nella sua forma più fondamentale, un elenco di schede (dette i **record**). Ogni record è composto da un insieme prefissato di spazi, ciascuno dei quali accoglie una specifica informazione, chiamati i **campi** del record.

← un database è composto di record che contengono campi

Questa struttura è esattamente analoga allo schedario che accoglie il catalogo di una biblioteca: per ogni libro o rivista esisterà una scheda (un record), e ogni record conterrà spazi (campi) dedicati al **titolo**, all’**autore**, all’**editore**, all’**anno di pubblicazione**, all’**argomento**, alla **collocazione**, e così via.

Esempio:

Scheda n.	4		
Scheda n.	3		
Scheda n.	2		
Scheda n.	1		
<i>Titolo</i>			
<i>Autore</i>			
<i>Editore</i>			
<i>Anno</i>			
<i>Argomento</i>			
<i>Collocaz</i>			

Le **biblioteche universitarie** hanno normalmente un catalogo informatizzato con questa struttura. Esso è accessibile da un calcolatore presente in biblioteca, oppure, sempre più comunemente, direttamente **dalla rete**.

Per il **Sistema Bibliotecario di Ateneo** dell'Università di Genova, l'accesso è dall'indirizzo:

<http://www.sba.unige.it>

e mette a disposizione l'accesso al **catalogo unico di ateneo** (ma si può anche restringere la ricerca ai cataloghi dei singoli Centri di servizio bibliotecario, cioè delle biblioteche dei singoli raggruppamenti disciplinari, p.es. il CSB-BSTM, per Biologia e Scienze della Terra e del Mare).

I record del catalogo di una biblioteca normalmente contengono solo informazioni del genere di quelle indicate sopra. Perciò non è possibile fare una ricerca sul **contenuto** di una pubblicazione, ma solo sul titolo, sull'argomento generale, sull'autore, ecc.

Esistono altri tipi di basi di dati bibliografiche (o **banche dati**, termine che suggerisce il loro scopo commerciale). Esse raccolgono grandissime quantità di pubblicazioni periodiche, per moltissime annate, e vengono regolarmente aggiornate. Anche se i periodici che mi interessano sono presenti nel catalogo della biblioteca, le banche dati presentano vantaggi in quanto di solito includono un numero di campi superiore, tra cui spesso anche un campo **Abstract**.

L'abstract è un riassunto del testo di un articolo, che dovrebbe dare un'idea piuttosto precisa sul contenuto limitandosi a 10-30 righe. Eseguendo la ricerca sull'abstract ho moltissime probabilità in più di trovare quello che cerco.

Queste basi di dati sono normalmente limitate a uno specifico ambito di conoscenze. Occorre quindi scegliere quella di interesse.

Le biblioteche hanno spesso tali basi di dati su CDROM, accessibili al pubblico da un calcolatore dedicato.

Dalla pagina del SBA dell'Università di Genova si può accedere a un certo numero di basi di dati di questo tipo. L'indirizzo diretto è

<http://basidati.csita.unige.it>

Però per alcune l'accesso è dietro username/password, che occorre possedere.

Le modalità di accesso dipendono dagli accordi tra il fornitore della banca dati e l'Università; perciò per ogni banca dati occorre leggere le informazioni fornite sulla pagina di accesso.

(Per alcune banche dati, i docenti sono autorizzati all'accesso dietro semplice richiesta; quindi, se mi occorre un accesso nel corso di un lavoro svolto in collaborazione con un docente, p.es. la tesi, posso chiedere a lui di assistermi.)

← accesso al catalogo
unificato delle
biblioteche di ateneo

← banche dati
bibliografiche

← abstract

← accesso alle banche dati
del servizio biblioteche
di ateneo

Nomi di banche dati, disponibili all'indirizzo

<http://basidati.csita.unige.it> ,

di possibile interesse per le discipline biologiche:

- **Current contents**
- **Geobase**
- **MathSci**
- **SIGLE**

← banche dati di
interesse biologico

Siti che forniscono l'accesso a servizi bibliografici (**non verificati – potrebbero avere limitazioni all'accesso**):

http://www.imbc.gr/biblio_serv/

<http://griffin.wsu.edu/>

<http://igm.nlm.nih.gov/>

<http://www.fiu.edu/~library/internet/subjects/health.html>

<http://www.dimdi.de/engl/hoste/dbkurze.html>

← siti bibliografici di
interesse biologico

SUGGERIMENTI PER LA RICERCA BOOLEANA

I programmi di accesso ad una base di dati consentono di effettuare ricerche di tipo simile a quelle viste per i motori di ricerca.

Ossia, ci sarà uno spazio da riempire, in cui scriverò le parole chiave che mi interessano, e ci sarà un comando **cerca**.

Tuttavia c'è **una importante differenza**: i motori di ricerca valutano il **grado di rilevanza** di un documento come un numero. Esistono documenti più rilevanti e documenti meno rilevanti.

I programmi di ricerca nelle basi di dati valutano solo **se un documento risponde o no** alla query. Quindi, in questo caso, non ci sono gradi di rilevanza: un documento o risponde alla query, oppure no; in tal caso non viene mostrato.

Questo tipo di ricerca è detto **ricerca booleana**. L'algebra booleana è quell'algebra che si basa solo sui due valori **vero** e **falso**. Quindi nella ricerca booleana il fatto che un documento sia interessante può essere solo **vero** o **falso**: non ci sono sfumature intermedie.

Questo significa anche che non esiste il concetto di sinonimo: nella query devo mettere esattamente i termini che occorre trovare.

Tutte le banche dati indicate sopra sono accessibili con il programma di interfaccia ERL, che poi si chiama WebSPIRS (vedi istruzioni su pagina www). I suggerimenti che seguono si riferiscono genericamente a tale interfaccia, anche se – come al solito – i concetti base sono generali, ed è solo il dettaglio di forma che cambia da un caso all'altro.

Per cercare tutte le pubblicazioni la cui scheda contenga la parola **miceti** userò la query semplice:

miceti

← Ricerca di tipo
“accesso a database”

“Booleana”

Cerca tutti e solo i
documenti che soddisfano
un criterio rigoroso

Se però sono in dubbio che sia usato il termine **funghi**, dovrò indicare l'alternativa:

miceti OR funghi

Otterrò tutti i documenti la cui scheda contiene almeno uno dei due termini.

Se invece mi interessano tutte le pubblicazioni che riguardano la raccolta dei funghi, potrò indicare:

raccolta AND funghi

Otterrò tutti i documenti la cui scheda contiene entrambi i termini, in qualunque posizione. Invece la query

raccolta funghi

troverà solo i documenti la cui scheda contiene l'espressione "raccolta funghi", con i due termini esattamente in quest'ordine e adiacenti.

Se mi interessa che le parole della query siano esclusivamente nel titolo (perché p.es. un autore molto prolifico ha un cognome che si scrive esattamente come la mia parola chiave) posso limitare la ricerca alle schede che hanno, p.es. **funghi** nel titolo. Posso farlo con un apposito comando dell'interfaccia, oppure scrivere

TI(funghi)

Se voglio che la ricerca sia solo nell'abstract e nel titolo, scriverò

TI(funghi) OR AB(funghi)

Una ricerca più complessa è:

AU(pippo) AND (TI(funghi) OR AB(funghi))

Questa query serve a trovare tutti i lavori di cui è autore pippo, e che contengono la parola funghi nel titolo oppure nell'abstract. Notare l'uso delle parentesi per indicare la "sotto-ricerca".

Ogni base di dati avrà un certo insieme di campi, con nomi non necessariamente uguali a quelli che abbiamo visto. Per sapere quali sono i campi disponibili, c'è un pulsante "database information" sulla destra in basso sulla finestra di interfaccia "WebSPIRS"



IN PRATICA

Scegliere un tema di ricerca. Individuare in rete i lavori più importanti pubblicati su tale tema, utilizzando database bibliografici delle biblioteche o disponibili su Internet. Riepilogare i risultati ottenuti in una bibliografia scritta.

Documenti scaricabili in linea

Se sono fortunato, dalla mia ricerca in rete otterrò non solo un riferimento bibliografico, ma il testo completo del lavoro che mi interessa.

Troverò cioè un file che contiene tutto l'articolo che cercavo.

Tale file sarà "in linea" (**on line**), cioè accessibile dalla rete Internet, e sarà "scaricabile", cioè potrò copiarlo (**download**) sul mio calcolatore e leggerlo sullo schermo, o stamparlo.

TIPI DI DOCUMENTI OTTENIBILI IN LINEA

I lavori che si trovano più facilmente in linea sono i **rapporti tecnici** (technical reports).

← **Rapporto tecnico**

Un rapporto tecnico è un lavoro pubblicato internamente da un'università o ente di ricerca, disponibile autonomamente (cioè non all'interno di una rivista o libro o antologia o raccolta...).

Un rapporto tecnico, se sufficientemente curato, può essere inviato ad una rivista per la pubblicazione vera e propria.

Per questo motivo, un rapporto tecnico può essere quasi equivalente a un articolo su rivista, ma con i seguenti notevoli vantaggi:

- Il copyright non è stato trasferito a un editore, ma è ancora dell'autore che può farne quello che vuole – p.es. lo distribuisce in rete
- Per pubblicare un lavoro su rivista ci vuole tipicamente un minimo di sei mesi (facilmente ancora di più), quindi lo stesso lavoro in forma di rapporto tecnico è disponibile con sei mesi–un anno di anticipo.

Lo svantaggio principale è il seguente:

- Il lavoro non è stato sottoposto a revisione, quindi è possibile che contenga delle sciocchezze.

A questo svantaggio si deve rimediare con un po' di buon senso da parte del lettore, e verificando l'attendibilità scientifica dell'autore.

La morale è che i rapporti tecnici, se usati con cautela, sono una preziosissima fonte di aggiornamento e studio scientifico.

I rapporti tecnici sono tipicamente messi a disposizione dalle **università, dipartimenti o enti di ricerca**, piuttosto che dai singoli autori. Vanno quindi cercati non sulle pagine personali, ma sui siti delle rispettive istituzioni.

Un altro tipo di documento è l'**abstract**. Come nelle basi di dati bibliografiche, è possibile trovare in rete gli abstract dei lavori pubblicati. Infatti l'editore non consente la diffusione di un lavoro di cui detiene il copyright, ma consente (e incoraggia) la diffusione del suo abstract.

← **Abstract**

Gli abstract si trovano tipicamente tra le **pagine personali** degli autori, che li mettono personalmente a disposizione.

Alcune riviste mettono in linea il **testo completo (full text)** degli articoli che pubblicano.

← Testo completo
di articoli

Normalmente ciò avviene dietro pagamento di una quota di **abbonamento** con caratteristiche analoghe all'abbonamento alla rivista convenzionale su carta.

In alcuni casi, un abbonamento è valido per **tutte le persone che si collegano da un dato indirizzo**. Per esempio, potrebbe esistere una rivista che consente a tutti i calcolatori che possiedono un indirizzo che termina per **.unige.it** di scaricare gratuitamente gli articoli. Questa possibilità va indagata chiedendo p.es. al personale delle biblioteche (eventualmente per e-mail).

In altri casi, le riviste rendono disponibili gratuitamente **articoli scelti a campione**, per dare la possibilità di valutare un eventuale abbonamento. Di solito questi articoli non sono interessantissimi, soprattutto perché sono pochi.

In tutti questi casi, la ricerca va effettuata **sul sito della casa editrice della rivista** su cui è comparso l'articolo che mi interessa.

Infine, alcuni autori mettono illegalmente in linea il full-text dei propri lavori.

Illegalmente, perché, una volta pubblicato un lavoro, il diritto di diffonderlo (copyright) è trasferito all'editore, e l'autore non ha più il diritto di decidere sulla diffusione della propria opera.

È quindi possibile che **sulle pagine personali dell'autore** si trovino i suoi articoli in forma scaricabile. Ovviamente, non vanno scaricati senza il permesso dell'editore...

IL FORMATO ELETTRONICO DEI DOCUMENTI SCARICABILI

I documenti elettronici sono dei **normali file**, che posso scaricare con un programma **FTP** (vedi Modulo 3) oppure **direttamente da pagine www** (vedi sempre Modulo 3).

Il **tipo** di tali file può essere variabile.

Occasionalmente si trovano lavori nella forma di **pagine www**. In questo caso sono facilmente consultabili come tutte le altre pagine di un sito www.

Tuttavia se cerco di stamparli il risultato **non** è di solito eccellente.

Raramente posso trovare articoli nel formato proprio di qualche programma di elaborazione di testi, con cui è stato creato.

P.es.: potrei trovare un file che si chiama **article.doc** ed è un documento nel formato proprio di Microsoft Word.

Questo tipo di diffusione è raro perché, se mi viene dato il file originale, io potrei modificare il contenuto dell'articolo, o per manomissione, o per errore.

I due formati di gran lunga più frequentemente usati sono il **Postscript** e il **Portable Document Format**.

Questi formati di file garantiscono che, usando appositi **programmi di visualizzazione e stampa**, l'articolo possa:

- essere letto sullo schermo di un calcolatore qualunque
- essere stampato su una stampante qualunque
- mantenere la forma grafica e l'impaginazione decise dall'autore;
- mantenere il contenuto invariato, senza possibilità di manomissioni o errori.

Nel Modulo 7 di questo corso si vedranno questi tipi di file con un minimo di dettaglio in più.

Siti che mettono a disposizione articoli in linea:

<http://www.ces.fau.edu/library/>

<http://www.dmoz.org/Science/Biology/Publications/Journals/> (cercare i sottoargomenti di interesse)



IN PRATICA

Scegliere un tema di ricerca. Recuperare in rete i lavori più importanti disponibili in linea su tale tema. Compilare una raccolta di lavori stampabili..

© 1998 Randy Glasbergen. E-mail: randy@glasbergen.com www.glasbergen.com



“First I searched for Larry in Yahoo, then Lycos, Excite and Infoseek. Eventually, I found him in the bathroom.”

Annotazioni

