

Fondamenti di informatica

Appunti sulla ricerca di informazioni per le scienze umane

dal capitolo 8 del testo:
M. Lazzari, *Informatica umanistica*, McGraw-Hill, 2014

Marco Lazzari

Università di Bergamo, Dipartimento di Scienze umane e sociali,
anno accademico 2014-2015

Le informazioni

1. letteratura di riferimento: conoscenze esplicite (condivise) su un dominio
 2. informazioni fattuali: dati, osservazioni, eventi
- 1 e 2 registrate nei documenti:
 - contenuto intellettuale
 - supporto

I documenti

- caratteristiche dei documenti:
 - supporto
 - tipo e forma di contenuto
 - ciclo di vita

Information retrieval

- utente: percezione soggettiva di lacuna di conoscenza
- 
- esigenza informativa
- 
- ricerca informazioni

Ricerche alternative

- chiedo a un esperto (tacit knowledge retrieval)
- uso catalogo biblio, motore di ricerca, repertorio web (information retrieval)
- consulto un documento (full text retrieval)
- consulto documento pertinente per ottenere altre fonti (reference linking)

Information retrieval

- utente: per soddisfare un'esigenza informativa, ricerca informazioni
- sistema: in risposta a una domanda (query) dell'utente, fornisce una lista di possibili fonti dell'informazione

Information retrieval

- selezione di info rilevante in una collezione di documenti organizzata
- processo interattivo: richiesta utente, risposta sistema
- indicizzazione di una risorsa informativa: creazione di una sua rappresentazione ai fini del reperimento

Information retrieval

- risultato di IR: selezione di risorse adeguate alla query
- valutazione della rilevanza
- ordinamento per rilevanza (relevance **ranking**)

Motori di ricerca

- da una lista di parole chiave dell'utente a una lista di indirizzi di risorse web
- spider (crawler, bot)
- indicizzazione
- creazione dell'archivio del motore
- ricerca nell'archivio
- ranking

Motori specializzati

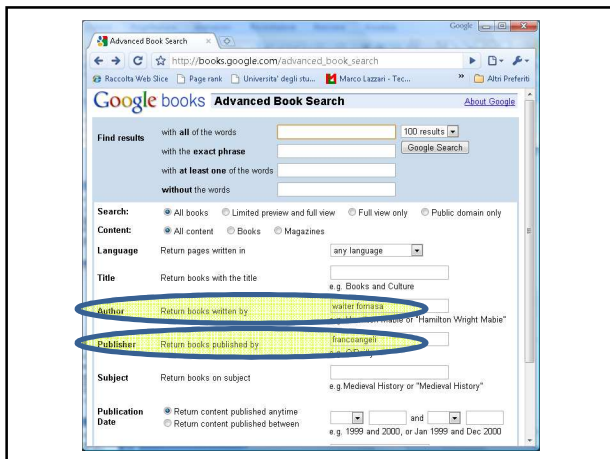
- metamotori (Metacrawler)
- plurimotori / motori paralleli (123people)
- clusterizzatori (Clusty)
- motori specialistici (Google Scholar, Google Books)

Metadati

- dialogo utente-sistema tramite stringhe di testo dette:
 - termini di ricerca (dal lato utente)
 - termini indicizzati (dal lato sistema)
- i termini possono essere:
 - generici (keyword)
 - coppie <attributo = "valore"> chiamate metadati (es. <autore = "Lazzari">)

Metadati

- attributi (strutturati) del documento
- ne promuovono la ricerca



Metadati

- vari tipi di metadati, raggruppati in schemi di metadati
- i metadati possono risiedere:
 - dentro il documento primario (es. nell'intestazione del file HTML)
 - in un surrogato (es. la scheda catalografica)
- surrogati raggruppati in indici o in database

Metadati nel doc primario

```
<meta name="author" content="Marco Lazzari" />
```

Metadati nel doc primario

```
<meta name="keywords" content="Università di Bergamo, Marco Lazzari, informatica generale, didattica, comunicazione multimediale, comunicazione ipermediale, ipermedialità, tecnologie dell'educazione, tecnologie per l'istruzione, podcasting universitario, Pluriversiradio)" />
```

Metadati nel doc primario

```
<meta name="description" content="Pagine informative per gli studenti dei corsi del prof. Marco Lazzari, Dipartimento di Scienze umane e sociali, Università di Bergamo; fondamenti di informatica, comunicazione multimediale, comunicazione ipermediale, ipermedialità; tecnologie dell'educazione, tecnologie per l'istruzione, didattica; podcasting universitario, Pluriversiradio" />
```

Metadati nel surrogato

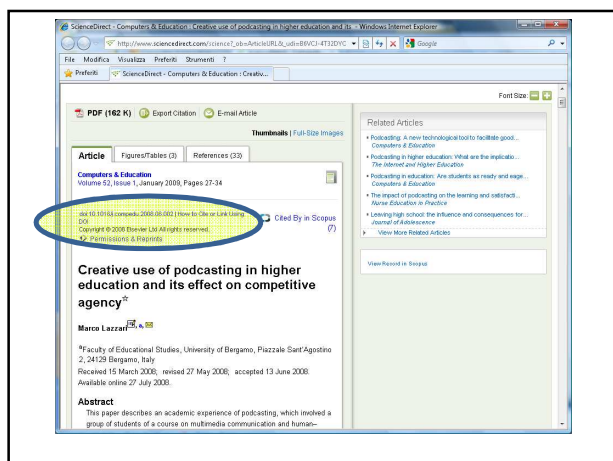
ISBN 978-88-17-02646-8
Autore Darwin, Charles
Titolo L'origine delle specie / Charles Darwin ; a cura di Giuliano Pancaldi
Pubblicazione Milano : BUR, 2009
Descrizione fisica XLIII, 565 p. ; 18 cm
CDD 576.82
In biblioteca Biblioteca umanistica

ISBN

- 978-88-386-6557-8
 - 978: libro
 - 88: Italia
 - 386: casa editrice
 - 6557: specifico testo
 - 8: codice di controllo

Digital Object Identifier

- DOI: identificatore per oggetti digitali pubblicati in Internet con diritti di proprietà intellettuale
- [doi:10.1145/2499149.2499175](https://doi.org/10.1145/2499149.2499175)



Tipi di metadati

- descrittivi: riguardano il doc in generale
 - titolo, autore, data, lingua, meta di legame (rel fra doc diversi, fra vers di un doc, fra parti logiche di un doc)
- semantici: riguardano il contenuto
 - parole chiave, abstract, codice di classific.
- localizzazione: riguardano l'esemplare fisico (copia del doc)
 - segnatura, url

La biblioteca

- mission: incontro tra esigenze informative degli utenti e documenti che possono soddisfarle
- documenti rappresentati da surrogati: schede catalografiche
- schede catalografiche raccolte nel catalogo

Schede catalografiche

- prodotte manualmente da personale specializzato
- una scheda catalografica contiene
 - metadati descrittivi (es. titolo, autore)
 - metadati semantici (es. parole chiave)
 - metadati di legame (es. relazioni con altri documenti)
 - segnatura → collocazione documento primario (es. 001.302 85 INF UMA)



La catalogazione

- metadati prodotti seguendo standard internazionali
- catalogazione descrittiva: per problemi di omonimia e sinonimia → controllo di autorità dei nomi

Record di autorità

Università di Bergamo
Servizi bibliotecari
Catalogo delle biblioteche

Rinvii - Autori - Stendhal

Sys. nr. 00007995
Intestazione: Sylva, Hieron.
Forme scartate: Sylva, Hieron.
Forme scartate: Bombet, Louis Alexandre César

Due tipi di catalogazione semantica

- soggettazione
- classificazione

Titolo: Kant e l'ornitorinco/Umberto Eco
Soggetto.....: 1. Semiotica
Classificazione.....: 121.68 (e. 21) – NATURA DELL'INTERROGAZIONE. SIGNIFICATO, INTERPRETAZIONE, ERMENEUTICA

La soggettazione

- stringa di testo → rappresenta l'argomento del documento
- termini della stringa scelti da vocabolario controllato
- soggetti raggruppati in un database, l'authority file dei soggetti

La classificazione

- codice (alfa)numerico → argomento del documento tramite
- utile per la collocazione a scaffale
- schema di classificazione più diffuso a livello mondiale: Classificazione Decimale Dewey – DDC (es. 005.276 POL CHE)

Classi Dewey

000-099 Generalità
 100-199 Filosofia e Psicologia
 200-299 Religione
 300-399 Scienze Sociali
 400-499 Linguaggio
 500-599 Scienze Naturali
 600-699 Tecnologia e scienze applicate,
 700-799 Arti
 800-899 Letteratura e Retorica
 900-999 Geografia e Storia

Classi Suddivisioni Sezioni

600 Scienze applicate
 610 Medicina
 616 Malattie
 616.1 Specifiche malattie
 616.12 Malattie del cuore
 616.123 Malattie delle coronarie
 616.123 2 Arteriosclerosi coronarica
 616.123 7 Infarto miocardico

Esempi Dewey unibg

- Informatica umanistica :
001.302 85 INF UMA (4)
- Neuropsicologia dello sviluppo:
155.413 NEU DEL (1)
- Educazione, pedagogia e scuola dall'Umanesimo al Romanticismo:
370.109 EDU PED (1)

Come cercare informazioni

1. scegliere il target informativo
2. tradurlo in un profilo di ricerca
3. inserire il profilo in uno o più sistemi di information retrieval

Come cercare informazioni

- in base a numerosità e bontà risultati:
- 4.0 rinegoziare il profilo di ricerca
 - 4.1 allargare la ricerca (strategia pearl growing)
 - 4.2 restringere la ricerca (strategia onion peeling)
 - 4.3 navigare tra le citazioni (citation search)

Come cercare informazioni

una volta raggiunto un set soddisfacente:

5. filtrarlo
6. monitorare le fonti

OPAC Online Public Access Catalog

- usare navigatore Dewey, liste di browsing e menu per esplorare l'OPAC
- sfruttare potenzialità linguaggio di interrogazione OPAC
- scorrere efficientemente lista risultati usando i vari criteri di ordinamento
- navigare tra record e contenuti aggiuntivi esterni all'OPAC
- sfruttare metaOPAC

Nel Web, oltre l'OPAC

- motori specialistici per ricerca di particolari tipologie di documenti
- banche dati bibliografiche
- directory
- servizi di social tagging per risorse informative

Barriere all'accesso

- non sempre è possibile aprire il testo pieno di un documento:
 - barriere economiche
 - abbonamento, pay per view
 - barriere legali
 - pochi utenti, Digital Rights Management (no stampa, no download)
 - barriere tecniche
 - software di lettura

Documenti liberamente accessibili in Rete

- documenti non più soggetti a copyright
→ appartenenti al dominio pubblico (PD)
- documenti distribuiti secondo forme di copyleft (es. licenze Creative Commons)

Testi pieni in Rete

- digital libraries
- archivi aperti
- riviste open access
- easy-publishing (blog, wiki)

Testi pieni in Rete

- verificare sempre quali usi dei documenti sono consentiti
- in mancanza di indicazione, valgono le leggi sul copyright del nostro Paese
- sempre e in ogni caso è necessario riconoscere la paternità intellettuale della risorsa (citare, non rubare)

Un buon sistema di IR

- fa reperire idealmente tutti e soli i documenti rilevanti per la specifica esigenza informativa
- li fa reperire rapidamente e senza fatica
- indica uno o più modi per accedere ai documenti