



Il protocollo Z39.50

Un'unica interfaccia per molti data base

N. V. Ebranati, Corso Biblioteche Digitali
Informatica Umanistica - Giugno 2004



Caratteristiche generali

Il protocollo **Z39.50**, è uno standard ANSI/NISO (American National Standards Institute, National Information Standards Organization), ed ISO 23950. E' mantenuto dalla Library of Congress (la Biblioteca Nazionale Centrale americana) e detta le regole per la ricerca ed il recupero delle informazioni (Information retrieval). La sua funzione e la sua struttura sono nate per facilitare l'**interoperabilità** tra sistemi in rete, in breve rende possibile consultare le basi dati di qualsiasi biblioteca (ma è sempre di più usato anche per basi dati non bibliografiche, quali quelle multimediali), indipendentemente da quale sia il suo sistema di database. Dalla sua prima versione del 1988 è arrivato ora alla terza versione, ampliata.

Questo protocollo, basato sull'architettura **client/server**, specifica le strutture dei dati e le regole di scambio, che permettono ad un sistema client (chiamato **Origin**) di effettuare ricerche nelle basi dati di un sistema server (chiamato **Target**) e poter consultare i records risultanti.

Le potenzialità di questo protocollo stanno nel consentire la consultazione di basi dati molto diverse tra loro **attraverso un unico linguaggio di interrogazione e con un'unica interfaccia.**

Funzionamento

Funziona così: una richiesta di ricerca (la **search query**) può essere inviata ad una base dati; i records trovati come risultato della richiesta verranno visualizzati sotto forma di un elenco (chiamato **result set**). Da questo elenco si potrà poi accedere ai records veri e propri. Il result set può anche essere oggetto di un'ulteriore ricerca, che produrrà un altro result set. All'interno della richiesta vi sono inoltre: il nome del database ed il nome degli attributi che fanno parte di un insieme particolare di definizioni registrate (**attribute-set-ID**) che servono per riconoscere l'indicizzare del database stesso. Il set di attributi usati dalle biblioteche si chiama "**bib-1**".

Il client inoltre puo' specificare in che **formato** vuole la risposta; i formati supportati sono: MARC, UNIMARC, USMARC, SUTR, GRS, e molti altri.

Certe basi di dati esportano solo in un formato (in questo caso generalmente è l'UNIMARC), e i risultati vengono visti solo in questo formato.

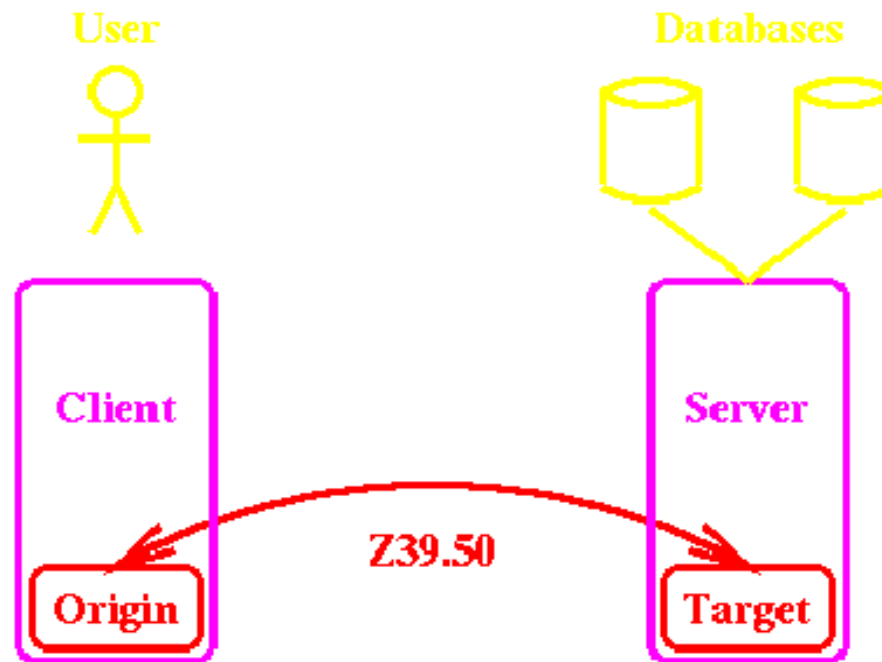
(Un formato è la **sintassi** con la quale viene strutturato e visualizzato il record).

Un'altra potenzialità dello Z39.50 è la possibilità di trasmettere contemporaneamente una richiesta a più server Z39.50. Le risposte vengono poi visualizzate divise per database. Si è pensato ad una fusione delle risposte e ad una eliminazione dei doppi, ma i problemi sono ancora molti, specie per la diversità delle descrizioni bibliografiche che potrebbero esistere per una stessa opera.

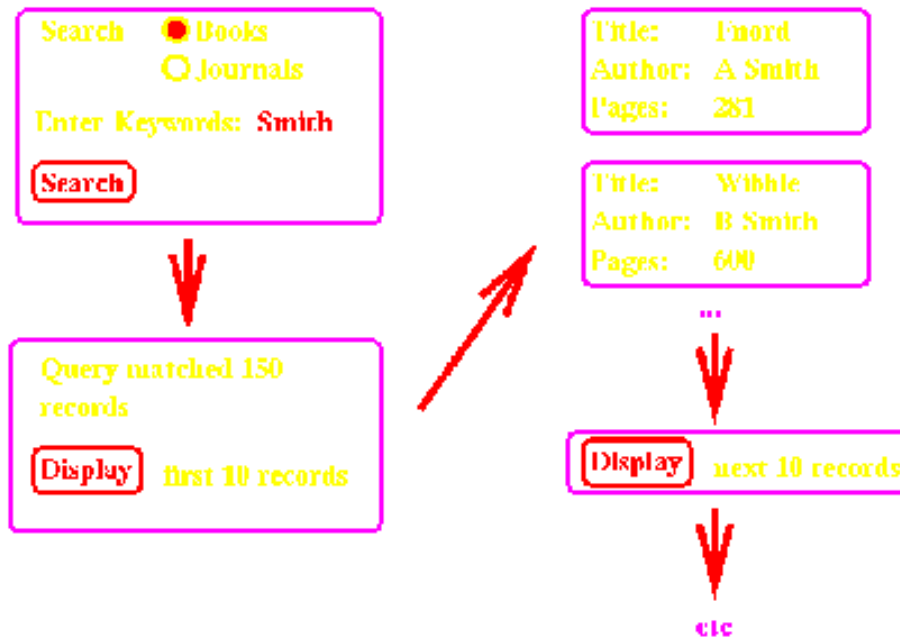
Una cosa interessante dello Z39.50 è la sua particolare architettura client/server: mentre nella classica architettura client/server del protocollo HTTP non c'è colloquio permanente tra client e server, cioè ogni coppia richiesta/risposta sta a sè, nello Z39.50 invece per ciascuna richiesta vengono creati dei gruppi di lavoro (sets) che possono essere ulteriormente elaborati. Viene tenuta perciò una **memoria** del lavoro svolto, limitata o meno da un time-out.



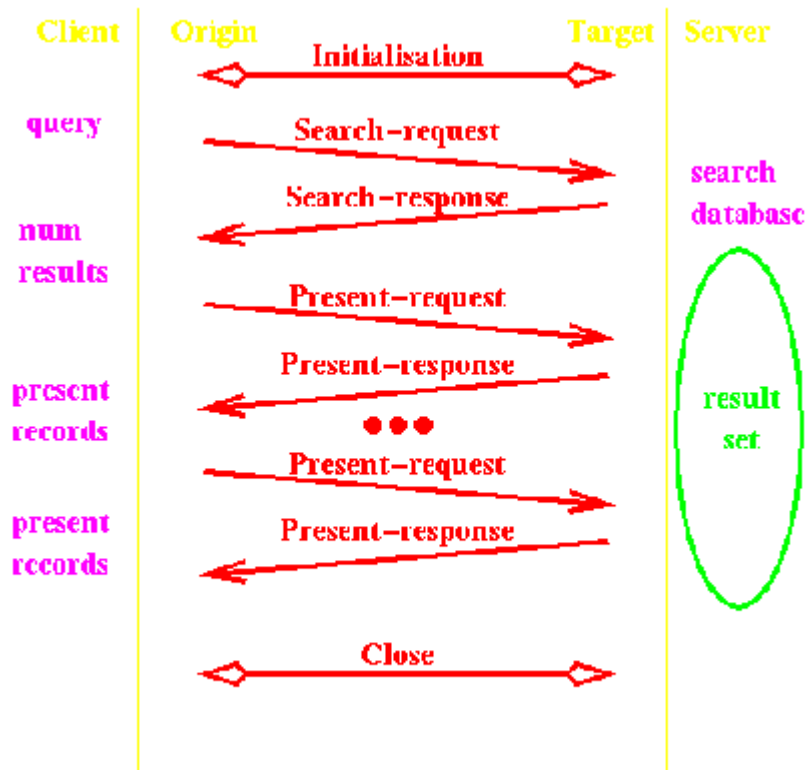
Relazione Target-Origin



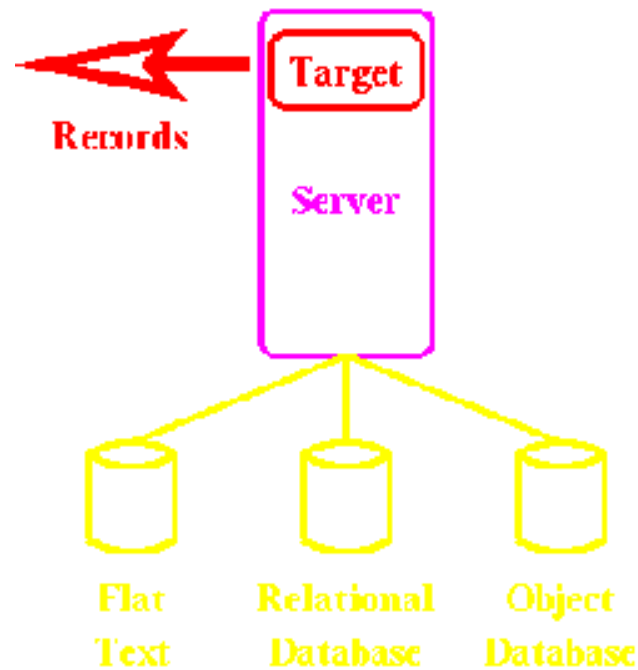
Prospettiva dell'utente



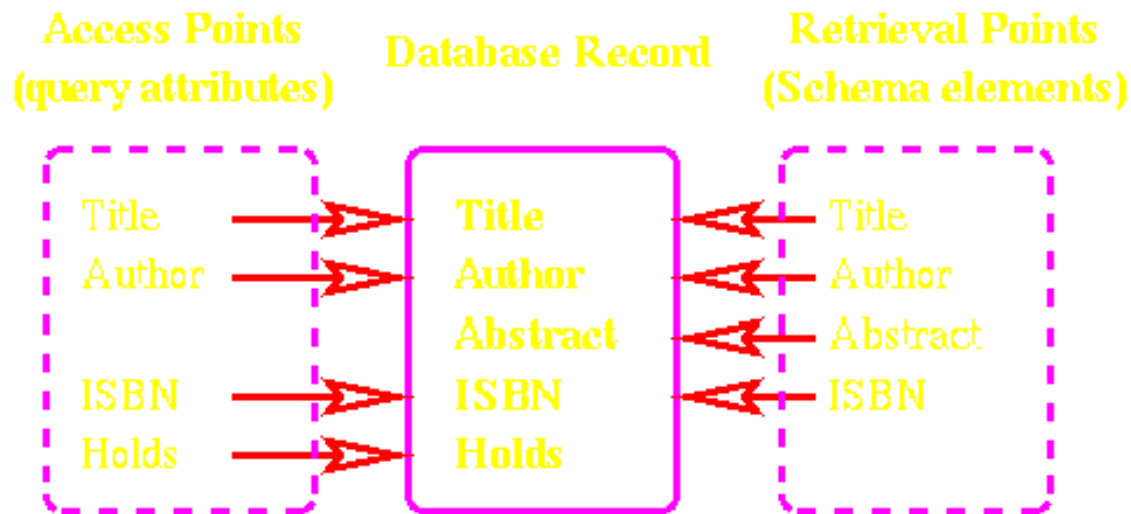
Prospettiva dell'implementazione



Rapporto server - data base



Retrieval dal data base



Due esempi

Ecco, per esempio, i dati pubblici di due basi dati, da inserire nelle configurazioni di un client Z39.50, con i quali è possibile dialogare con i relativi database:



1) The Structured Information Manager (SIM), RMIT MDS

Address: z3950.simdb.com

Port: 210

Databases: MARC (and a dynamic list of others)

Validation: not required

Version: 3

Syntaxes: SUTRS, GRS-1, USMARC (OPAC and Summary if requested)

Services and facilities: Search, Present, Delete, Scan, Resource Control, Access Control, Explain, Extended services

2) Università Cà Foscari di Venezia

Indirizzo IP: 157.138.47.40

Porta: 21210

Nome: ADVANCE

Formato dati: UNIMARC / ISO2709

Set di caratteri: ISO 5426

Versione: 3

Set attributi: Bib-1



Set di attributi Bib-1

Il set di attributi **Bib-1 (Bibliographic 1 Attribute set)** previsto dallo standard Z39.50 è adottato dal profilo SBN che vedremo di seguito. Gli attributi definiscono come il server ed il client devono colloquiare, quali campi del record possono essere utilizzati come canali di ricerca e quali criteri devono essere specificati dall'utente per svolgere un'interrogazione.

Bib-1 ha i seguenti tipi di attributi:

- **Attributi di uso:** per specificare i canali di ricerca, cioè i punti di accesso (autore, titolo, etc.) .
- **Attributi di relazione:** per specificare i rapporti tra il canale di ricerca ed il termine di ricerca (uguale a, diverso da, minore di, etc.) .
- **Attributi di posizione:** per definire la posizione del termine di ricerca nel campo o sottocampo cui appartiene .
- **Attributi di struttura:** per descrivere la tipologia del termine di ricerca (parola, frase, lista di parole, etc.) .
- **Attributi di troncamento:** per definire se il termine di ricerca deve essere considerato troncato e in che modo.
- **Attributi di completezza:** per specificare se parole aggiuntive possono apparire con il termine di ricerca in un campo o sottocampo.

Per chiarire in che modo un termine possa essere utilizzato nella ricerca si riporta la definizione degli attributi di struttura tratta da **Attribute set Bib-1 (Z39.50) Semantics:**

- **Frase (1):** Consiste di uno o più gruppi di caratteri separati da spazi vuoti. Il valore deve essere cercato esattamente come appare nel termine di ricerca rispettando l'ordine delle parole e l'adiacenza.
- **Parola (2):** Consiste in un gruppo di caratteri. Specifica che la parola è da ricercare esattamente, a meno che non sia esplicitamente troncata. Il termine parola non contiene spazi vuoti.
- **Anno (4):** Consiste in termine numerico composto da quattro cifre.

- **Data normalizzata (5):** Consiste nel giorno, mese ed anno in cui la transazione si è svolta.
- **Lista di parole (6):** Consiste in una o più parole separate da spazi vuoti. Non è considerato l'ordine delle parole. Ogni singola parola può essere troncata.
- **Stringa (108):** Consiste in un termine di ricerca trattato come stringa, una sequenza o un insieme di parole singole.

Facilities e funzioni

All'interno del protocollo Z39.50 sono definite alcune funzioni dette **facilities**. Ognuna di queste funzioni è sostenuta da uno scambio di messaggi tra l'Origin e il Target. Un messaggio può essere una **richiesta** o una **risposta**. Esaminiamo ora le principali operazioni disponibili:

- La facility **Init** definisce una serie di messaggi tra Origin e Target per stabilire una connessione, inizializzare una sessione e negoziare le aspettative e le limitazioni delle attività, come il massimo numero di records che possono essere trasferiti da un server a un client, la versione del protocollo, ecc... Solo dopo queste disposizioni il client può sottoporre la query.

- Attraverso la facility **Search** lo Z39.50 client traduce la query ad un set di database in una rappresentazione standard e la passa al server Z39.50. Il Server esegue la ricerca e crea un result set, che rappresenta l'insieme dei risultati identificati dalla query e che vengono mantenuti per una successiva richiesta.
- Attraverso la facility **Retrieval** il client può richiedere alcuni record dal result set oppure richiedere un nuovo processo di interrogazione al result set. Tale facility consiste di due servizi: il servizio **Present** permette al client di richiedere record dal result set, riferendosi alla loro posizione nel set; il servizio **Segment** permette al Target di segmentare le risposte mandando uno o più Segment richiesti prima del risultato del Present.



- La facility **Sort** permette all'Origin di richiedere al Target di scegliere e ordinare un result set secondo una specifica sequenza.
- La facility **Browse** attraverso il servizio **Scan** permette di scandire una lista ordinata di termini, cioè di punti di accesso, disponibili su un database o su un gruppo di database.
- Infine abbiamo la facility **Close** che termina la comunicazione tra Origin e Target.

Una particolare facility, detta **Explain**, rappresenta il supporto usato dal Target per la memorizzazione dei dati che viene consultato dall'Origin nelle fasi di ricerca e di recupero.

Un Target può permettere la ricerca in più database che risiedono sul server stesso. In particolare esso contiene anche un database detto **Explain Database**, che contiene tutte le informazioni riguardanti i database presenti nel Target.

L'Origin deve poter accedere al database Explain come un qualunque altro database, quindi in fase di search deve essere possibile accedere a tale database mediante gli attributi necessari.

La facility Explain permette all'Origin di ottenere dettagli relativi all'implementazione del Target, includendo i database disponibili per la ricerca e la sintassi supportata per il recupero.



Caratteristiche del profilo SBN Z39.50

Il profilo **SBN** è stato definito per le componenti Z39.50 lato server e lato client.

L'implementazione del profilo consente ad un client Z39.50 generico (anche non SBN) di effettuare ricerche e di interoperare con il server dell'Indice. Nel profilo sono definiti i servizi Z39.50 implementati sul server profopac.html e gli attributi e le combinazioni di attributi necessari per effettuare le ricerche e visualizzare i risultati (profilo Bib-1).

- Linea di comunicazione: TCP/IP
- Versione del protocollo: Z39.502,3
- Indirizzo Internet: opac.sbn.it
- Numero della porta: 2100
- Ore di disponibilità: 0-24
- Servizi: Init, Search, Present, Scan, Sort, Delete, Close.
- Base dati: nopac
- Tipo di interrogazione: 1(RPN), 101
- Insieme di attributi: BIB-1
- Sintassi dei record: SUTRS
- Insieme dei caratteri di interrogazione: ISO 646 + ISO 8859-1
- Insieme dei caratteri dei record: ISO 646 + ISO 8859-1
- Presentazione del record: Full, Brief

Caratteristiche del server SBN Z39.50

- Hardware: IBM 7013-J30 RISC 6000
- RAM: 512 MB
- Disk: 40GB
- Sistema: IBM AIX 4.3
- DBMS: BASIS Plus Versione 8.2.3

Links utili

- www.math.unipd.it/~derobbio/z3950.htm

Risorse generali

- <http://www.cenfor.net/database/menu.htm>

Repertorio degli Opac italiani Z39.50

- <http://www.informatik.tu-darmstadt.de/VS/Infos/Protocol/Z39.50/z39.50-nutshell.html#what>

Rassegna sui temi di Z39.50

- <http://www.lsbu.ac.uk/~litc/caseli b/software.html>

Per scaricare il software

