



CBT - Servizi di ricerca

Versione 1.0

Indice

1 Introduzione.....	3
2 Composizione della query.....	3
2.1 Indirizzo.....	3
2.2 Parametri.....	3
2.3 Stringa di ricerca.....	4
2.3.1 Operatore booleano implicito.....	4
2.3.2 Normalizzazione diacritici.....	4
2.3.3 Minuscole e maiuscole.....	5
3 Tipologie di ricerche (qt).....	5
3.1 Any keyword.....	5
3.2 Titolo.....	5
3.3 Autore (Nome).....	5
3.4 Editore.....	6
3.5 Soggetti.....	6
3.6 Note.....	6
3.7 Ricerca avanzata.....	6
3.8 Dettaglio record con dati di copia.....	7
4 Ordinamento (sort).....	7
5 Filtri (fq).....	8
6 Faceted search.....	9
7 Formati di risposta (wt).....	12
7.1 Links utili.....	13

Indice delle tabelle

Tabella 1: Parametri di ricerca	4
Tabella 2: Tipologie di ricerche disponibili.....	5
Tabella 3: Attributi di ricerca avanzata.....	7
Tabella 4: Criteri di ordinamento.....	8
Tabella 5: Faccette.....	11
Tabella 6: Formati di risposta disponibili.....	12

1 Introduzione

Il Catalogo Bibliografico Trentino (CBT) espone una serie di servizi di ricerca attivabili tramite utilizzo del protocollo HTTP (modo GET).

Ciò significa che, fatte salve le questioni relative all'autorizzazione ed autenticazione che vanno al di là dello scopo di questo documento, un client HTTP può utilizzare tali servizi per eseguire delle ricerche all'interno del sistema e avere in risposta un set di dati in formato XML o JSON da utilizzare per diversi scopi (e.g. app mobile, applicazione standalone, web application).

Nelle sezioni successive vengono dettagliati i servizi in oggetto, fornendo informazioni sui parametri che tali servizi accettano, sulle differenti modalità e possibilità di ricerca che è possibile eseguire ed infine sui formati in risposta.

2 Composizione della query

2.1 Indirizzo

L'indirizzo del servizio presenta la seguente nomenclatura:

```
<schema>://<host>:<porta>/solr/main/select
```

dove

- schema: http o https a seconda di come il servizio risulta configurato;
- host: il nome host o l'indirizzo IP dove il servizio è in ascolto;
- porta: la porta dove il servizio è in ascolto; nel caso in cui il servizio risponda sulle porte di default (80 per http e 443 per https) questa informazione può essere omessa.

2.2 Parametri

La seguente tabella riporta i parametri che è possibile specificare attraverso la URL http. Alcuni di essi (principalmente qt e wt) vengono descritti in apposite sezioni.

- P = Nome parametro
- O = Obbligatorio
- R = Ripetibile (e.g. &a=v1&a=v2&a=v3)

P	Descrizione	O	R	Default
q	Indica la query da seguire, in accordo a quanto specificato nella sezione successiva.	S	N	N.A.
fq	Indica un filtro, applicata dopo aver eseguito la query specificata nel parametro q . La sintassi è <nome campo>: <valore>	N	S	N.A.
qt	Indica la tipologia di ricerca che si intende effettuare. Viene dettagliato nella sezione successiva.	N	N	any_bc
wt	Indica il formato desiderato nella risposta. Possibili valori sono json oppure xml	N	N	xml
sort	Indica il criterio di ordinamento (vedere sezione successiva)	N	N	N.A.
rows	Indica la dimensione massima di ogni pagina di risultati	N	N	10
start	Indica l'indice del primo risultato. Viene utilizzato per effettuare la paginazione tra i risultati.	N	N	0
facet	Attiva o disattiva le faccette (i.e. i raggruppamenti) all'interno dei risultati di ricerca	N	N	true
facet.field	Indica il nome di un attributo per il quale si vuole attivare i raggruppamenti (i.e. faccette) nei risultati di ricerca.	N	S	???

Tabella 1: Parametri di ricerca

2.3 Stringa di ricerca

Tutte le ricerche descritte nelle sezioni successive, fatta eccezione per le ultime due (dettaglio con dati di copia e ricerca avanzata) sono ricerche dove il valore dell'attributo "q" deve contenere solamente uno o più termini di ricerca, senza operatori o attributi.

2.3.1 Operatore booleano implicito

I termini di ricerca si intendono implicitamente legati tra di loro con l'operatore booleano AND, quindi la stringa di ricerca

storia repubblica italia

verrà logicamente tradotta dal sistema nella ricerca

storia AND repubblica AND italia

2.3.2 Normalizzazione diacritici

I termini di ricerca possono contenere o meno diacritici: in ogni caso il

sistema provvederà alla sostituzione di ogni carattere diacritico con la corrispondente versione “non diacritica”. Ad esempio le seguenti stringhe di ricerca sono considerate identiche:

- Széchenyi Lánchíd
- Szechenyi Lanchid

2.3.3 Minuscole e maiuscole

I termini di ricerca vengono normalizzati in minuscolo pertanto la ricerca, dal punto di vista del richiedente, è da considerarsi case-insensitive. Ad esempio le seguenti stringhe di ricerca sono considerate identiche:

- Dante Alighieri
- DANTE ALIGHIERI
- dante alighieri

3 Tipologie di ricerche (qt)

Il parametro qt indica la tipologia di ricerca che è possibile effettuare. Ogni tipologia di ricerca ha un determinato ambito di competenza.

La tabella seguente riporta i valori ammessi per questo parametro:

Valore	Descrizione
any_bc (default)	Ricerca dovunque
title_bc	Ricerca per titolo
author_bc	Ricerca per autore
publisher_bc	Ricerca per editore
sub_bc	Ricerca per soggetto
note_bc	Ricerca all'interno delle note
adv	Ricerca avanzata

Tabella 2: Tipologie di ricerche disponibili

3.1 Any keyword

Il termine o i termini inseriti vengono ricercati all'interno di tutti gli attributi che compongono l'indice del sistema.

3.2 Titolo

Il termine o i termini inseriti vengono ricercati all'interno delle diverse tipologie di titoli appartenenti a ciascun record.

3.3 Autore (Nome)

Il termine o i termini inseriti vengono ricercati nei nomi o autori associati ad ogni record.

3.4 Editore

Il termine o i termini inseriti utilizzati per effettuare una ricerca tra i nomi degli editori ed i luoghi di pubblicazione.

3.5 Soggetti

Il termine o i termini inseriti utilizzati per effettuare una ricerca tra i soggetti associati ad i record.

3.6 Note

Il termine o i termini inseriti utilizzati per effettuare una ricerca tra le note associate ai record.

3.7 Ricerca avanzata

Le tipologie di ricerca sino ad ora descritte vengono spesso classificate “ricerche semplici”, in quanto il richiedente indica solamente il termine o i termini, delegando al sistema il pieno controllo della ricerca che verrà eseguita.

Nella ricerca avanzata, il richiedente ha un maggiore controllo sulla query prodotta e quindi sulla ricerca che viene eseguita. Importante sottolineare che l'unica differenza rispetto a quanto descritto nel resto del documento risiede nel valore del parametro q; il resto (e.g. faccette, filtri) risulta completamente valido anche per questa tipologia di ricerca.

*Come già detto in precedenza, per attivare questa ricerca, è necessario che il parametro **qt** sia impostato con il valore **adv***

La query che è possibile specificare come valore del parametro q è in sostanza una espressione nel linguaggio di interrogazione del motore fulltext sottostante e pertanto deve rispettare la seguente nomenclatura (il valore deve essere tutto su una riga, viene diviso in righe nell'esempio per comodità di lettura):

```
_query_:
"{!edismax qf=$<attributo1>_query pf=$<attributo1>_phrase}<valore1>"
AND | OR | ANDNOT
"{!edismax qf=$<attributo2>_query pf=$<attributo2>_phrase}<valore2>"
...
```

al posto di <attributoN> è possibile specificare uno dei valori presenti nella seguente tabella:

Valore	Descrizione
any_bc	Ovunque (Any keyword)
title_bc	Titolo
author_bc	Autore
pub_bc	Editore
sub_bc	Soggetto
classification	Classificazione Decimale Dewey
isbn_or_issn	ISBN o ISSN
an	Amicus Number (identificativo del record)
note_bc	Note
publ_place	Luogo di pubblicazione
collocation	Collocazione (codice)
barcode	Codice a barre
holding_note	Note di copia

Tabella 3: Attributi di ricerca avanzata

3.8 Dettaglio record con dati di copia

Oltre alle ricerche precedentemente descritte, esiste una ulteriore possibilità, ossia il recupero delle informazioni di un record di cui si conosce l'identificativo (i.e. Amicus number).

In questo caso la URL del servizio cambia leggermente dalle precedenti, nel senso che

- il valore del parametro q, ossia il termine di ricerca, deve essere un Amicus Number valido;
- non è necessario specificare alcun parametro qt;
- è invece necessario sostituire /select con /doc

Quindi il servizio presenta la seguente nomenclatura:

```
<schema>://<host>:<porta>/solr/main/doc
```

Quindi per fare un esempio:

```
http://host:8292/solr/main/doc?q=832833
```

4 Ordinamento (sort)

Il valore del parametro sort è una lista di attributi, separati da virgola, che specificano i criteri di ordinamento in ordine di priorità, seguiti ciascuno da "asc" o "desc" a seconda che si voglia applicare un ordinamento (per

quello specifico attributo) ascendente o discendente. Quindi, ad esempio

```
sort=a1 desc, a2 desc, a3 asc
```

La seguente tabella riepiloga gli attributi che possono essere dichiarati in questo parametro:

Nome	Descrizione
score	Ranking attributo al risultato di ricerca.
title_sort	Titolo
author_sort	Autore
publication_date_sort	Data di pubblicazione

Tabella 4: Criteri di ordinamento

L'attributo è opzionale e se non presente, il valore di default applicato è

```
score desc, title_sort asc, author_sort asc, publication_date_sort desc
```

5 Filtri (fq)

Il parametro fq può essere utilizzato, ripetutamente, per applicare dei filtri ai risultati di ricerca ottenuti da una delle tipologie di ricerca descritte di sopra. E' importante sottolineare che il criterio di filtro non viene applicato contestualmente alla ricerca, bensì dopo aver prodotto i risultati derivanti dalla stessa. Chiaramente per il richiedente tutto questo è trasparente;

La nomenclatura del valore ammesso per il parametro fq è la seguente

```
fq=<nome attributo 1>:<valore 1> AND | OR <nome attributo 2>:<valore 2>
```

Alcune precisazioni:

- qualora il valore secondo cui effettuare il filtro contenga degli spazi, è consigliabile racchiuderlo tra virgolette (e.g. fq=a1="codice")
- se vengono specificati 2 parametri fq, questi si intendono in AND ed hanno lo stesso significato di un solo parametro fq con due espressioni legate da AND. Tuttavia, in questi casi è preferibile specificare due differenti parametri fq per fare un uso migliore dei meccanismi di cache del sistema. Cioè ad esempio

Invece di

fq=n1:v1 AND n2:v2

è preferibile

fq=n1:v1&fq=n2:v2

- se due criteri devono essere legati in OR, allora è necessario specificare un solo parametro fq con due espressioni. Ad esempio

fq=n1:v1 **OR** n2:v2

I nomi e i valori degli attributi sono gli stessi elencati nella sezione "Faceted Search". Laddove l'attributo non possiede un dominio di codici, ma può avere qualsiasi valore (e.g. nome dell'autore) il criterio utilizzato per il filtro deve avere esatta corrispondenza col valore, possibilmente utilizzando doppi apici per racchiudere il valore.

6 Faceted search

I risultati di ricerca, oltre a contenere gli item che hanno trovato una corrispondenza nell'indice rispetto ai termini inseriti, contengono una sezione aggiuntiva denominata facets, consistenti in una serie di aggregazioni, effettuate su una serie di attributi, rispetto ai risultati di ricerca stessi.

Per ogni criterio di aggregazione (i.e. per ogni attributo) vengono forniti i valori più rappresentativi con associato il conteggio degli item che posseggono quel valore.

Le faccette, questo il nome attribuito a tali aggregazioni, restituite dal sistema assieme ai risultati di ricerca sono riassunte nella seguente tabella.

Nome	Descrizione
format	<p>Formato. E' costituito da un codice avente uno dei seguenti valori:</p> <p>mas=Materiale a stampa re=Risorsa elettronica vc=Videocassetta vdvd=Video DVD vr=Videoregistrazione cdm=CD musicale lpm=Long Playing mus=Musica alccass=Audiolibro con cassetta alccd=Audiolibro con CD rs=Registrazione sonora mfilm=Microfilm pamus=Partitura musicale mcs=Materiale cartografico a stampa mcm=Materiale cartografico manoscritto manos=Manoscritto sw=Software mm=Materiale misto m=Monografia sr=Periodico a=Spoglio mling=Materiale linguistico msom=Spartito musicale mcsoim=Carta geografica mcmim=Materiale cartografico, manoscritto mdpro=Materiale da proiettare dvd=DVD vhs=VHS rsmusic=Registrazione musicale rsnomusic=Audiolibro rgrafbi=Grafica soso=Servizio on line file=File computer kit=Kit mlingman=Manoscritto o Tesi manutri=Oggetto docebook=E-book cddvdrom =CD/DVD-ROM docel=Elettronico doccar=Cartaceo website=Sito internet compprog=Programma informatico intermult=Multimedia interattivo compfile=Computer file</p>

publication_date_interval	Intervallo di pubblicazione. Può assumere uno dei seguenti valori: aty=Anno in corso bl2y=Ultimi 2 anni cl5y=Ultimi 5 anni dl10y=Ultimi 10 anni el50y=Ultimi 50 anni fmt50y=Oltre ultimi 50 anni cl3y=Ultimi 3 anni dl4y=Ultimi 4 anni el5y=Ultimi 5 anni fmt5y=Oltre 5 anni
language	Lingua di pubblicazione del record. E' costituito da un codice di 3 lettere ed è in sostanza il valore presente nel record MARC originale.
name	Nome di autore o persona (e.g. illustratore)
subject	Soggetto
index_term_facet	Parola chiave
dewey_1digit	Prima cifra della Classificazione Decimale Dewey
target_audience	Codice che può assumere uno dei seguenti valori: a=Sconosciuto b=Sconosciuto c=Sconosciuto d=Sconosciuto e=Adulto f=Sconosciuto g=Sconosciuto j=Giovanile u=Sconosciuto
user_category_ad	Classificazione all'interno dei record che appartengono alla vista adulti (e.g. "Opera per giovani adulti")
user_category_rag	Classificazione all'interno dei record che appartengono alla vista ragazzi (e.g. "Opera per bambini", "Opera ragazzi")
collection_code	Collezione di provenienza del record
series_statement_facet	Collana
genre_form_subject	Genere
library	Codice / Simbolo della biblioteca (Main Library). I codici vengono forniti in un allegato a parte.

Tabella 5: Faccette

7 Formati di risposta (wt)

I formati di risposta previsti dal protocollo sono 3: XML (default), JSON e MARCXML. La seguente tabella mostra pertanto i valori ammessi per il parametro wt.

Valore	Descrizione
xml (default)	Formato XML
json	Formato JSON
marc_xml	MARC XML

Tabella 6: Formati di risposta disponibili

Da precisare che il formato MARCXML si intende applicato al singolo risultato di ricerca: l'interno flusso XML non è conforme allo standard in quanto include alcune informazioni come numero totale di risultati, indice del primo record restituito, etc.

Per alcuni esempi di risultati fare riferimento al link

<http://docs.lucidworks.com/display/lweug/Query+and+Response+Examples>

7.1 Links utili

Apache SOLR e Lucene

<http://lucene.apache.org/solr>

<https://wiki.apache.org/solr/XMLResponseFormat>

<https://wiki.apache.org/solr/SolJSON>

<http://docs.lucidworks.com/display/lweug/Query+and+Response+Examples>

Codici MARC per le lingue di pubblicazione

<http://www.loc.gov/marc/languages>

CBT - Catalogo Bibliografico Trentino

<http://www.cbt.biblioteche.provincia.tn.it>