

AquaBrowser Library

search, discover, refine.

Breve descrizione del prodotto

1.	Il software AquaBrowser Library	2
1.1	Le funzioni di <i>Search – Discover – Refine</i>	3
1.2	<i>Search</i> : la funzione di ricerca e di presentazione dei risultati	3
1.2.1	La configurazione della sezione <i>Search</i> : alcuni esempi	3
1.3	<i>Discover</i> : il suggerimento di nuovi percorsi di ricerca.....	3
1.3.1	La configurazione della sezione <i>Discover</i> : alcuni esempi	4
1.4	<i>Refine</i> : il raffinamento dei risultati ottenuti	4
1.4.1	La configurazione della sezione <i>Refine</i> : alcuni esempi	4
1.5	La sezione amministrativa e le statistiche d'uso.....	5
2.	Caratteristiche dell'applicativo <i>AquaBrowser Library</i>	5
2.1	L'architettura tecnologica di <i>AquaBrowser Library</i>	5
2.2	I differenti sistemi di indicizzazione e trattamento dei dati in <i>AquaBrowser Library</i>	6
2.2.1	Indicizzazione standard	6
2.2.2	Web spider o web crawler	7
2.2.3	Federated search (o search on-the-fly).....	7
2.3	L'applicazione di <i>AquaBrowser Library</i> in ambienti <i>ibridi</i>	7
2.4	<i>AquaBrowser Library</i> : caratteristiche funzionali per l'utenza	8

Il documento intende descrivere brevemente le caratteristiche tecniche e funzionali di *AquaBrowser Library* e il possibile impiego nell'ambiente *ibrido* costituito dall'insieme di fonti informative, in formati e supporti differenti, prodotti o gestiti direttamente dalla Biblioteca.

1. Il software AquaBrowser Library

AquaBrowser Library, prodotto dalla Medialab Solutions, è un potente motore di ricerca ed un visualizzatore di informazioni che consente il recupero di dati su fonti differenti (catalogo bibliografico, banche dati, risorse elettroniche, documenti digitali...) in una modalità semplice, intelligente ed assolutamente innovativa. La ricerca su *AquaBrowser Library* è supportata da una rete di associazioni semantiche (associazioni tra termini, varianti ortografiche, termini provenienti da thesauri e vocabolari controllati, traduzioni in altre lingue) prodotte nella fase di configurazione ed indicizzazione dei dati sul database del sistema (si veda la descrizione dell'architettura di *AquaBrowser Library*, § 2.1 del presente documento).

AquaBrowser Library nasce da uno studio approfondito condotta sulla *psicologia del ricercatore* di fronte alle tecniche (o abitudini) di ricerca su fonti sempre più vaste ed eterogenee. L'approccio più diffuso tra chi operi ricerche in tali ambiti vasti ed eterogenei è quello che viene ormai definito "Google-like" intendendo con tale espressione l'attitudine a ricerche semplici, su fonti non necessariamente "strutturate" e senza una necessaria esperienza e competenza tecnologica nell'uso di risorse differenti.

AquaBrowser Library si propone di salvare e rendere disponibile tale logica di approccio alla ricerca su fonti eterogenee, integrando ad esso le tecniche di ricerca su cataloghi selezionati (ricerca in OPAC) e superando, così, entrambe gli approcci attraverso una serie di funzionalità avanzate, poi di seguito illustrate.

A seguire due tabelle utili a riassumere brevemente alcune delle caratteristiche funzionali proprie di *AquaBrowser Library* sia rispetto ai tradizionali motori di ricerca (qui ci si riferisce a Google solo a titolo esemplificativo) sia rispetto agli OPAC:

Confronto Google – AquaBrowser Library

Caratteristiche funzionali	Google	ABL
Concetto di rilevanza basato su metadati catalografici e non su principi "commerciali"	No	Si
Ricerca simultanea su fonti eterogenee	Si	Si
Ricerca simultanea su fonti eterogenee selezionate dalla biblioteca	No	Si
Suggerimento di termini associati per esplorare nuovi percorsi di ricerca	No	Si
Integrazione con strumenti (in modalità visibile e non) quali thesaurus, dizionari etc. selezionati dalla biblioteca	No	Si
Estrazione automatica di filtri dai risultati ottenuti	No	Si
Feedback amichevoli e "contestuali"	No	Si

Confronto OPAC tradizionali – AquaBrowser Library

Caratteristiche funzionali	OPAC	ABL
Ricerca per singolo documento	Si	Si
Esplorazione libera su termini, parole, soggetti etc.	No	Si
Presentazione di associazioni e varianti ortografiche	No	Si
Ricerca sul catalogo	Si	Si
Ricerca simultanea su catalogo e fonti diverse, eterogenee	No	Si
Arricchimento di query attraverso uso di metadati	No	Si
Estrazione automatica di filtri dai risultati ottenuti	No	Si
Feedback amichevoli	No	Si
Completamente aderente agli standard di accessibilità	No	Si

1.1 Le funzioni di *Search – Discover – Refine*

L'interfaccia di *AquaBrowser Library*, estremamente amichevole ed intuitiva, esprime tre differenti logiche di approccio alla ricerca e di presentazione dei risultati, che possono essere riassunti nelle tre funzioni di *Search – Discover – Refine*.

Di seguito una più dettagliata descrizione delle funzioni di base di *AquaBrowser Library*.

1.2 *Search*: la funzione di ricerca e di presentazione dei risultati

AquaBrowser Library consente agli utenti di cercare informazioni sulle fonti consultate, selezionate dalla biblioteca o ente, in una modalità semplice e intuitiva, e presenta i risultati di tale ricerca in una lista, personalizzabile, ordinata per *rilevanza* rispetto alla query stessa. Altri criteri di ordinamento (per titolo, per data, per autore etc.) sono disponibili per una differente presentazione dei medesimi risultati, in relazione alle specifiche esigenze di ricerca dell'utente. Dalla schermata di presentazione dei risultati è possibile, attraverso una semplice funzione di navigazione, accedere direttamente all'OPAC della biblioteca (oppure ad un qualsiasi visualizzatore di dati messo a disposizione per ciascuna fonte consultata) per fruire dei servizi a cui l'utente è abilitato (informazioni sulle copie, prestito, prenotazione, document delivery etc.).

1.2.1 La configurazione della sezione *Search*: alcuni esempi

La sezione centrale dell'interfaccia di *AquaBrowser Library* è quella della presentazione dei risultati di una ricerca effettuata sulle fonti messe a disposizione dalla biblioteca o dall'ente, secondo il criterio (di default) di *rilevanza* o secondo altri criteri selezionati dall'utente. L'amministratore di *AquaBrowser Library*, nella fase di configurazione del sistema o in momenti successivi può modificare alcuni parametri di base, in relazione alle specifiche esigenze di presentazione dei risultati e di integrazione con i visualizzatori di dati esistenti. Di seguito alcuni esempi di configurazione sulla sezione *Search*:

- selezione del set di metadati che si intenda mostrare ai ricercatori (esempio: Titolo del documento, anno di pubblicazione, Occorrenze dei termini ricercati etc.)
- adattamento della logica di navigazione (*hyperlink* su altri elementi di descrizione o altre pagine interne al sito)
- integrazione, anche grafica, con l'OPAC (o altro visualizzatore) delle fonti ricercare (*deeplinking*)
- modifica della *logica di rilevanza*: la possibilità di definire *su quali campi del record* elaborare la logica di rilevanza (esempio: i termini ricercati devono essere sull'intero record, oppure sul campo titolo, oppure sul campo titolo e il campo soggetto etc.) e definizione di *gruppi di metadati* da considerare in fase di estrazione dei dati, come da esempio:
 - 1° gruppo: termini trovati nei campi Titolo del record o del documento
 - 2° gruppo: termini trovati nei campi Soggetto
 - 3° gruppo: termini trovati nei campi Nota
 - 4° gruppo: termini in campi Autore
 - 5° gruppo: termini trovati in campo Serie
- inserimento di icone ed altri identificati grafici per la rappresentazione di tipologie differenti di materiale o di presenza di documenti digitali connessi alla descrizione del record

1.3 *Discover*: il suggerimento di nuovi percorsi di ricerca

La eccezionale funzione di navigazione delle informazioni attraverso un *grafo (word-cloud)* costituito da termini relazionati in vario modo al termine ricercato. Ponendo al centro del grafo il termine (o i termini) ricercato, il sistema recupera e presenta, come "suggerimenti" di esplorazione delle fonti consultate, una serie di termini relazionati al termine ricercato per associazioni (i termini più utilizzati nei record in associazione a quello ricercato), per traduzioni (le traduzioni linguistiche del termine

ricercato), per sinonimi, per varianti ortografiche, per gerarchia thesaurali. Il grafo di termini si "rigenera" ad ogni nuova ricerca e ad ogni nuova esplorazione del catalogo.

I *termini associati* sono estratti dalle fonti consultate attraverso una serie di algoritmi parametrizzabili dall'amministratore di AquaBrowser Library (selezione dei campi del record, numero di occorrenze e vicinanza al termine ricercato etc.).

Per le *traduzioni dei termini* AquaBrowser utilizza un dizionario di lingue europee che riconduce ciascun termine dall'esperanto e dall'esperanto riparte per proporre la traduzione del termine stesso.

Per le *relazioni gerarchiche* AquaBrowser Library utilizza strutture semantiche (thesaurus, soggettario, vocabolario controllato...) integrabile nel sistema in un qualsiasi formato elettronico (file di authority, file XML etc.)

Si vedano, di seguito, alcune possibilità di configurazione da parte dell'amministratore del sistema, della logica di creazione del grafo (*word-cloud*) nella funzione *Discover* di *AquaBrowser Library*.

1.3.1 La configurazione della sezione *Discover*: alcuni esempi

La funzione *Discover* di *AquaBrowser Library* intende rendere disponibile al ricercatore una serie di supporti alla ricerca che vanno dal suggerimento di nuovi percorsi di ricerca (attraverso i termini associati), alla integrazione di strumenti costruiti ad hoc (thesauri, soggettari, authority file etc.) da personale qualificato ma spesso di difficile comprensione ed utilizzo per gli utenti finali.

Rientra nella sezione *Discover* di *AquaBrowser Library* anche la presentazione gerarchica (ad albero) di strumenti quali thesauri, soggettari, sistemi di classificazione etc.

Di seguito alcuni esempi di configurazione della sezione *Discover*:

- definizione dei campi del record (o di altri metadati descrittivi) per l'applicazione degli algoritmi di creazione delle relazione di *associazione* (esempio: applicazione degli algoritmi al solo campo Titolo; applicazione ai campi Titolo e Soggetto; applicazione al campo Titolo ed abstract etc.)
- attivazione delle *varianti ortografiche* nel solo caso di risultati della ricerca = 0 (restrizione della funzione di offerta della variante ortografica solo nel caso di ricerca priva di risultati)
- integrazione di dizionari specifici e/o sostituzione del dizionario multilingue (esperanto)
- integrazione di strumenti semantici quali tesauri, authority file, mappe topiche etc.

1.4 *Refine*: il raffinamento dei risultati ottenuti

La sezione di *Refine* di *AquaBrowser Library* consente all'utente di filtrare la propria ricerca, partendo dal risultato di una query, al fine di recuperare con estrema precisione quanto di proprio interesse. Le categorie di informazioni che il sistema utilizza come filtri della propria ricerca sono personalizzabili dall'amministratore del sistema ed includono, in genere, nomi, titoli, tipologie di materiale, date di pubblicazione, serie editoriali, lingue del documento, soggetti di vario genere etc.

1.4.1 La configurazione della sezione *Refine*: alcuni esempi

Ogni biblioteca/ente, in relazione alle abitudini ed attitudini della propria utenza, decide quali e quante categorie di dati mettere a disposizione nella sezione di *Refine* come filtri per successive ricerche.

Ogni dato strutturato del record può diventare "filtro"; il numero dei filtri, così come l'ordine di presentazione degli stessi, sono strettamente relazionati alla ricchezza dei metadati descrittivi e all'uso che degli stessi filtri fa l'utente finale (uso documentato su una delle pagine di Statistiche di *AquaBrowser Library*, nella sezione di Amministrazione).

Strutture di metadati costruite ad hoc per rispondere a specifiche esigenze di ricerca (esempio: strutture riprodotte la localizzazione geografica di differenti biblioteche appartenenti ad una rete, oppure strutture costruite sulla base della pertinenza disciplinare etc.) possono essere integrate in *AquaBrowser Library* e diventare filtri per il raffinamento delle ricerche nella sezione *Refine*.

- **Data adapters (o Data connectors):** si tratta dello strato che connette AquaBrowser Library con i dati sorgente. Questa componente trasforma i dati in input (dati originali) nell'ultraflessibile formato di dati proprio di AquaBrowser Library. I *data adapters* sono stati sviluppati per informazioni destrutturate (documenti in formato Word e Excel, pagine HTML e siti web), ma anche (e meglio) per dati strutturati, come quelli di database relazionali, XML, file di testo strutturati etc.
- **Knowledge Builder:** effettua un'analisi completa dei dati ricevuti in input. Il Knowledge Builder analizza i dati sulla base della frequenza dei termini, di co-occorrenze e secondo altri criteri al fine di creare una rete semantica che supporti i processi di ricerca e, successivamente, di raffinamento dei risultati ottenuti dalla ricerca stessa.
- **Igor:** è il motore di ricerca di AquaBrowser Library. Si tratta di un motore di ricerca full-text estremamente potente e flessibile, sviluppato dalla Medialab Solutions nel corso degli ultimi 10 anni e integrato in AquaBrowser Library.
- **Context Bulder:** un insieme di componenti finalizzato ad offrire una serie di supporti all'utente finale. Il *context builder* consente di arricchire le query iniziali, fornisce suggerimenti alla ricerca ed altri tipi di utili feedback all'utente. Il *context builder* è anche il componente che compila i risultati delle ricerche e le associazioni prodotte dal motore di ricerca, oltre a fornire la funzione di raffinamento delle ricerche ottenute.
- **User interface:** è l'interfaccia grafica, estremamente amichevole, di AquaBrowser Library. E' un'interfaccia tripartita in cui i risultati delle ricerche con tutti i suggerimenti e le indicazioni utili vengono fornite all'utente finale. Ha la caratteristica di offrire un ambiente estremamente amichevole sia per la formulazione delle query che per la presentazione dei risultati e di rendere così il processo di ricerca particolarmente semplice, veloce ed efficiente.

2.2 I differenti sistemi di indicizzazione e trattamento dei dati in *AquaBrowser Library*

All'interno dell'architettura tecnologica di *AquaBrowser Library* sopra illustrata sono previsti differenti processi finalizzati al trattamento dei dati, in relazione alla natura dei dati di input, alla "proprietà" degli stessi, alla possibilità di operazioni di estrazione ecc.

Di seguito si riportano i tre principali possibili scenari di trattamento dei dati previsti dal sistema AquaBrowser Library.

2.2.1 Indicizzazione standard

Per *indicizzazione standard* di *AquaBrowser Library* si intende il metodo di trattamento dei dati che, partendo da un export in qualsiasi formato di un database esistente o di raccolte di documenti elettronici etc, converte in un formato proprietario ed indicizza i dati stessi sul database Igor, creando contestualmente una rete semantica successivamente utilizzata nelle differenti funzioni di recupero e presentazione delle informazioni. L'indicizzazione standard è, tra quelle previste dal sistema, sicuramente la più completa, veloce ed efficace. Questa modalità di indicizzazione, per le ragioni che poi nel seguito si spiegheranno, è quella che utilizza tecnologie di ricerca allo stato dell'arte producendo i migliori risultati nelle tre macrofunzioni (*Search – Discover – Refine*) previste dal sistema stesso.

2.2.2 Web spider o web crawler

Si intende la tecnica utilizzata dai motori di ricerca per l'indicizzazione di pagine web; il web crawler può riguardare tutte le pagine di un sito (*crawler general purpose*) o una selezione mirata dello stesso (*focused crawler*). *AquaBrowser Library* utilizza entrambe le modalità di accesso a pagine web o documenti in formati differenti (tipicamente documenti in .pdf) in relazione alla specifica esigenza (e configurazione) della biblioteca. I dati, così acceduti da *AquaBrowser Library*, sono poi comunque indicizzati sul database Igor. Questa tecnica di accesso ai dati ed indicizzazione è solitamente e preferibilmente utilizzata, in ambienti ibridi e congiuntamente alla prima modalità di indicizzazione, in presenza di dati "non strutturati" o meglio non descritti in set di metadati (file html, documenti in formati diversi etc.) più raramente per l'indicizzazione di OPAC (costituiti tradizionalmente da metadati dettagliatamente strutturati).

2.2.3 Federated search (o search on-the-fly)

La tecnologia che consente di aggregare fonti differenti di informazioni in un'unica ricerca. Le tecnologie di "federated search" consentono dunque all'utente di ricercare simultaneamente e attraverso l'uso di un'unica interfaccia più (ed eterogenee) fonti di informazione basate su web (cataloghi elettronici, banche dati, riviste elettroniche etc.), senza che alcuna indicizzazione preventiva sia avvenuta di queste differenti fonti (ricerca *on-the-fly*). Questo tipo di indicizzazione è preferibilmente utilizzato per arricchire "virtualmente" il proprio patrimonio di dati con fonti esterne, non trattabili né con la prima né con la seconda modalità di indicizzazione previste da *AquaBrowser*, ma accedibili come "fonti esterne".

2.3 L'applicazione di *AquaBrowser Library* in ambienti ibridi

Per rispondere alle esigenze di gestione di materiale sempre più eterogeneo (materiale a stampa, risorse elettroniche, informazioni on line etc.) e distante dai modelli tradizionali (evoluzione dalla biblioteca tradizionale alla biblioteca digitale) *AquaBrowser Library* offre una serie di opzioni tecnologiche e funzionali finalizzate a garantire l'accesso integrato, semplificato ma efficace, a risorse eterogenee.

Di seguito alcune caratteristiche tecniche utilizzate da *AquaBrowser Library* che rendono possibile tale accesso integrato a dati eterogenei:

- ricerca simultanea in ogni database selezionato dalla biblioteca/ente (esempio: archivi in OAI, database SQL);
- connessione a tutti i database "interni" in MS Access (esempio: lista di siti selezionati o raccomandati dalla biblioteca);
- ricerca in ogni sito o collezione di documenti (in formati diversi, come .PDF, .doc, .ppt etc.) attraverso tecniche di web spider;
- ricerca federata (*federated search*) in ogni database SRU/SRW
- ricerca federata (*federated search*) in database *proprietary* compatibili (o da rendere compatibili) con WebFeat;
- Feed RSS¹ per l'aggiornamento on line degli utenti sui contenuti di proprio interesse.

AquaBrowser Library costituisce al momento uno dei tool più potenti e diffusi al mondo per la gestione di ambienti informativi ibridi, consentendo all'utente finale di accedere attraverso un'unica interfaccia ed in modalità estremamente semplificata ed intuitiva a risorse interne ed esterne alla biblioteca, in formati e supporti eterogenei.

¹ RSS è lo standard più diffuso per l'esportazione di contenuti Web. RSS è l'acronimo per *RDF Site Summary* ed anche di *Really Simple Syndication* ed è diventato uno dei più popolari formati per la distribuzione di contenuti web.

2.4 AquaBrowser Library: caratteristiche funzionali per l'utenza

Potendo offrire in altri punti del documento la descrizione generale delle funzionalità di *AquaBrowser Library* si preferisce riportare qui elencate alcune delle caratteristiche salienti del prodotto stesso in relazione alla differente utenza che utilizzi il prodotto.

1. Caratteristiche di AquaBrowser Library per l'utenza finale

- Ricerca semplice, amichevole ma estremamente potente
- Opzioni per la ricerca esatta o avanzata
- Possibilità di estendere i propri percorsi di ricerca seguendo i suggerimenti offerti dal grafo (*word-cloud*) o di restringere i risultati della propria ricerca utilizzando i filtri (zoom sui dettagli attraverso la funzione di *Refine*)
- Correzioni ortografiche generate automaticamente
- Associazioni tra termini generate automaticamente
- Traduzioni e sinonimi generati semi-automaticamente
- Interfaccia per palmari e Smartphone

2. Caratteristiche di AquaBrowser Library per i bibliotecari

- Integrazione di strumenti quali tesauri, soggetti, sistemi di classificazione, authority file
- Front end personalizzabile
- Valorizzazione delle proprietà del catalogo
- Interfaccia unica per accesso alle fonti interne ed esterne
- Integrazioni di dizionari in differenti lingue
- Supporto di differenti set di caratteri

3. Caratteristiche di AquaBrowser Library per gli amministratori del sistema

- Completamente aderente agli standard nazionali ed internazionali di accessibilità del web
- Integrazione semplice di fonti informative eterogenee
- Possibilità di categorizzare in modo semplice e potente fonti esterne ed interne
- Interfaccia multilingue
- Piattaforme client differenti: PC, Mac e Linux
- Installazione e configurazione calcolabile in giorni e non in mesi