

Dietro le quinte di ArchiDATA, dal computer del programmatore

Michele Lavezzi

ArchiDATA, archivio elettronico di retrodatazioni lessicali, è stato ideato e disegnato nelle sue componenti da Ludovica Maconi, curato nella realizzazione informatica da Michele Lavezzi (cfr. www.archidata.info). In queste pagine si descrive la progettazione del sito, così come è stata eseguita dal programmatore, illustrando anche pagine non visibili all'utente, ma accessibili ai collaboratori per la pubblicazione di nuove voci.

ArchiDATA è un sito web scritto in linguaggio C# (*C sharp*) sfruttando le librerie del framework .Net. Questa scelta è stata fatta perché tale linguaggio della famiglia .Net è robusto, pur avendo una sintassi semplice e concisa. Il framework .Net contiene inoltre numerose librerie già pronte per essere utilizzate nello sviluppo di applicazioni web. Da rilevare è inoltre la pressoché totale compatibilità con ogni servizio di *hosting* (non sempre garantito se si scelgono soluzioni più di nicchia).

Panoramica del sito.

Quattro sono le macro-componenti del sito:

1. interfaccia utente (o pagina di consultazione);
2. interfaccia di *backend* (o pannello di amministrazione);
3. motore di ricerca;
4. pagine di contenuti.

Interfaccia utente

L'interfaccia utente permette ai visitatori di cercare le parole retrodate e archiviate e di visualizzarne la voce completa, con fonti e commenti.

Dal punto di vista funzionale, la pagina di consultazione è suddivisa in tre zone: una maschera di ricerca dalla quale l'utente può impostare i criteri con cui accedere alle informazioni in archivio (fig. 1, sezione 1), una colonna laterale (*sidebar*) dove vengono raccolti in elenco i risultati della ricerca (fig. 1, sezione 2), un corpo centrale dove viene visualizzata la voce del lemma selezionato (fig. 1, sezione 3).

Cita come:

Michele Lavezzi, *Dietro le quinte di ArchiDATA, dal computer del programmatore*, "Italiano digitale", VII, 2018/4, pp. 159-164.

Copyright 2017 Accademia della Crusca

Publicato con licenza creative commons [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Fig. 1. Maschera di consultazione di ArchiDATA.

Per quanto riguarda la maschera di ricerca, si rimanda al paragrafo *Motore di ricerca*, che fornisce informazioni dettagliate.

La colonna laterale contiene i lemmi elencati in ordine alfabetico, subito affiancati dalla data di prima attestazione (indicata e inserita dagli autori dei contenuti del sito, linguisti e lessicografi). In caso di parole con più significati, comparirà in questa sezione solo la datazione più antica; aprendo la voce, si avranno invece le datazioni delle singole accezioni e si scoprirà a quale significato si riferisce la più antica datazione. Talvolta, se non è stato possibile risalire a una datazione precisa, la parola nel lemmario è affiancata dall'indicazione generica del secolo di prima attestazione, oppure dall'indicazione 'avanti *anno*' (questi casi si presentano in particolare per il lessico antico). Il campo datazione è marcato, ed è dunque possibile interrogare l'archivio tramite ricerca cronologica, da anno X ad anno Y.

Selezionando uno dei lemmi dalla colonna laterale, si apre il corpo centrale contenente la voce lessicografica: definizione, elenco dei significati, da quello attestato in epoca più antica al più recente, fonti della datazione, citazione o riproduzione fotografica del passo di riferimento (in questo caso, una differente grafica del titolo, in blu e in carattere sottolineato, sul quale è possibile cliccare, indica la presenza di immagini allegate), eventuale *link* esterno verso altri siti.

Interfaccia di backend

L'interfaccia di *backend* (in italiano "pannello di amministrazione") è una sezione visibile solo ai collaboratori autorizzati e registrati. Permette di inserire i lemmi e di compilare la voce lessicografica. Si tratta di un'interfaccia creata quasi esclusivamente via *javascript*. Le informazioni inserite sono salvate tramite POST Ajax, in modo che l'interfaccia si aggiorni senza dover ricaricare ogni volta l'intero contenuto della pagina. In questo modo si velocizza notevolmente il lavoro di inserimento dati.

L'interfaccia permette di inserire per ogni lemma uno o più significati; per ogni significato una o più fonti; per ogni fonte una o più immagini. Le informazioni salvate formano così una sorta di struttura ad albero, che parte dalla radice (il lemma) e si ramifica via via fino ad arrivare alle immagini.

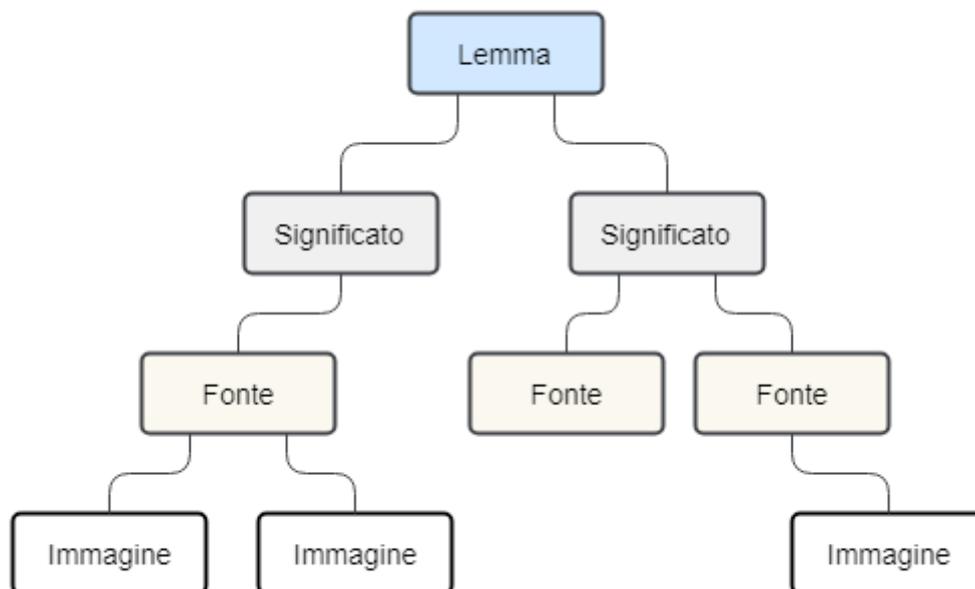


Fig. 2. Schema ad albero dell'interfaccia *backend*.

Il collaboratore, autore di voci da inserire nel sito, ha accesso a questa interfaccia di immissione dei dati. Questa sezione del sito è stata suddivisa in colonne, una per ogni livello dell'albero di fig. 2. Da sinistra verso destra, i campi da completare sono: lemma, significati, fonti e immagini (fig. 3). I dati sono modificabili solamente dal collaboratore che ha inserito e firmato la voce (e dall'amministratore), non sono accessibili agli altri collaboratori. La firma dell'autore della voce compare in automatico in calce ad ogni scheda, perché è associata all'*account* del collaboratore che immette i dati. L'amministratore del sito controlla e approva i dati inseriti dal collaboratore, marcando il lemma come "verificato" (icona a forma di triangolo sotto la colonna del lemma, cfr. fig. 3: giallo = da verificare; verde = controllato e approvato). Le voci che nella pagina di *backend* (non visibile agli utenti che siano comuni utilizzatori) sono marcate con icona di triangolo verde (= voce rivista e approvata) sono considerate definitive, e come tali non più modificabili nemmeno dall'autore che ha firmato la voce. L'amministratore, su richiesta dell'autore, potrà però sempre sbloccare la voce, riportando l'icona gialla, per permettere le modifiche o gli aggiornamenti richiesti dal collaboratore.

Fig. 3. Interfaccia *backend*, pannello di amministrazione, ovvero maschera di caricamento voci.

Rimozione dei doppioni

Per mantenere la struttura ad albero, se un lemma ha più significati, questi sono ramificazioni della stessa scheda madre (cfr. fig. 3, sotto colonna “Lemma”, il pulsante verde + per aggiungere schede con altri significati di quella stessa parola a lemma). Per evitare l’inserimento di voci doppie, è stato introdotto un blocco di sistema che impedisce al collaboratore, autore di voci, di inserire una nuova scheda per parole già presenti in archivio. Il blocco automatico dei doppioni è affiancato e preceduto da un suggerimento a video, attivo nel momento della digitazione del lemma, all’apertura della nuova scheda: una colonna laterale mostra infatti i lemmi già inseriti che iniziano con quelle stesse lettere, nella sequenza digitata. Se tale avviso non dovesse essere sufficiente per accorgersi della presenza in archivio della scheda che si sta per inserire, il sistema invierà un messaggio di errore e impedirà l’apertura della voce-doppione.

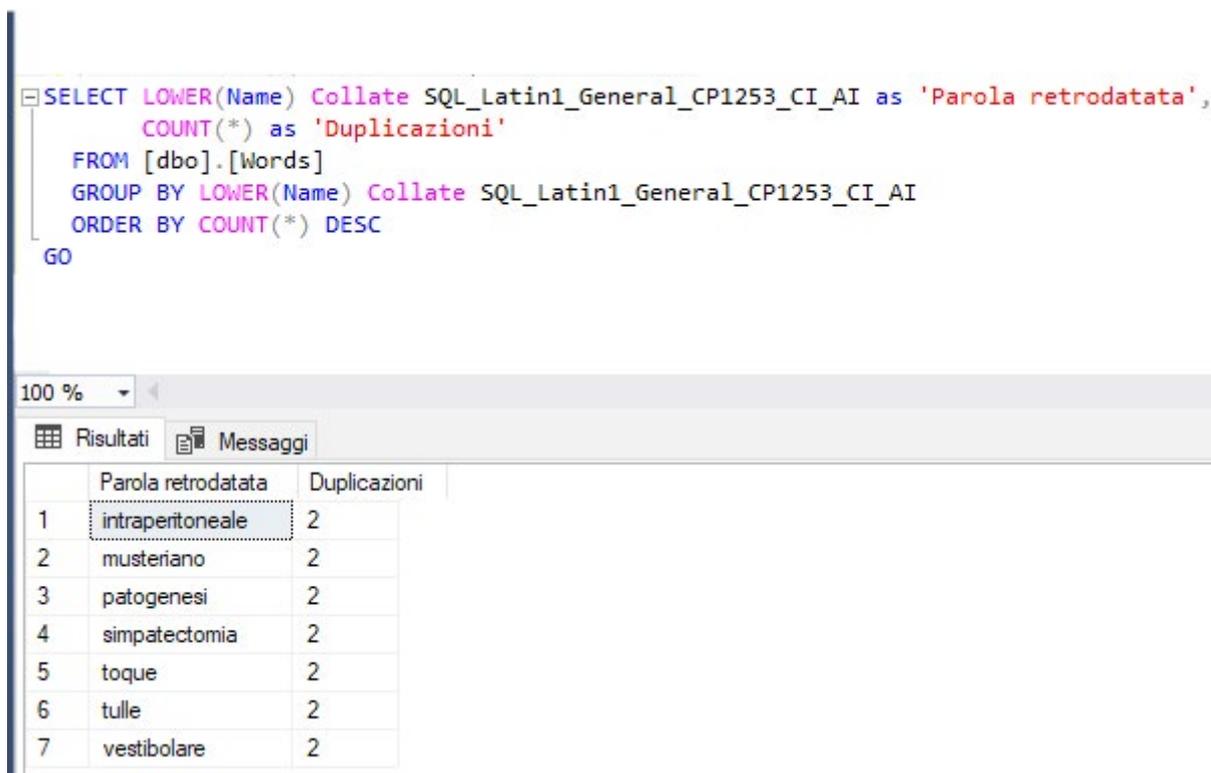
Poiché questo sistema è stato introdotto in un secondo momento, quando ormai i lemmi inseriti erano già qualche migliaio, è stato necessario un controllo su *database* dei lemmi pubblicati. Con una semplice *query* (o interrogazione) sono state trovate ed eliminate una ventina di voci erroneamente duplicate (cfr. fig. 4). Le fonti e le informazioni recuperate, se utili, sono ovviamente state accorpate sotto un unico lemma prima della cancellazione.

Per controllare le duplicazioni sono state utilizzate regole di confronto che non tenessero conto delle maiuscole, minuscole e degli accenti.

```

SELECT LOWER(Name) Collate SQL_Latin1_General_CP1253_CI_AI as 'Parola retrodatata',
       COUNT(*) as 'Duplicazioni'
FROM [dbo].[Words]
GROUP BY LOWER(Name) Collate SQL_Latin1_General_CP1253_CI_AI
ORDER BY COUNT(*) DESC
GO

```



	Parola retrodatata	Duplicazioni
1	intrapertoneale	2
2	musteriano	2
3	patogenesi	2
4	simpatectomia	2
5	toque	2
6	tulle	2
7	vestibolare	2

Fig. 4. Eliminazione dei doppiami tramite ricerca da database.

Motore di ricerca

La ricerca di base si effettua dal campo di *input* contrassegnato dalla *label* “Ricerca”. Tale campo è dotato della funzionalità di autocompletamento. Basta quindi digitare un minimo di 3 lettere per innestare una ricerca con *wildcard* che restituirà tutti i lemmi che iniziano con la sequenza di lettere inserita. L'autocompletamento sfrutta gli indici *full-text search*. In questo modo la restituzione dei dati cercati è molto veloce.

È stata disabilitata la distinzione dei segni diacritici: lettere accentate, caratteri speciali e dieresi non influiscono sui risultati della ricerca. Per esempio, cercando la parola *wurstel* scritta senza dieresi, il motore di ricerca restituisce la voce *würstel*, così come inserita dal lessicografo nel lemmario di ArchiDATA. Lo stesso vale per parole francesi, che vengono estratte anche se cercate senza gli accenti canonici.

Gli indici *full-text* permettono di avere anche il suggerimento delle polirematiche presenti in archivio. Cercando la parola *candela*, il programma segnala la presenza in archivio della locuzione *il gioco non vale la candela*. La funzionalità delle *wildcard* (simbolo ***) può essere sfruttata utilizzando il simbolo all'inizio (**ato* restituirà *altolocato*, *blefarostrato*, ecc.) o alla fine del lemma desiderato (*abb** restituirà *abbacchiaro*, *abbassalingua*, ecc.).

Ricerca avanzata

Per raffinare la ricerca sono stati introdotti campi di *input* che permettono di filtrare i risultati per data, autore, ambito d'uso, tipo di forestierismo, regionalismo, ovviamente con possibilità di cumulo dei vari criteri di ricerca.

Il campo autore funziona in modo simile al campo di ricerca base, consentendo anche l'uso di caratteri jolly (***). Le lettere inserite verranno ricercate all'interno dei cognomi degli autori opportuna-

mente marcati in fase di inserimento utilizzando i simboli *underscore* (sono marcati e quindi estraibili solo i cognomi, non i nomi di battesimo).

Il motore di ricerca sfrutta le potenzialità di Linq con Entity framework per comporre dinamicamente delle *query* a db utilizzando un linguaggio di alto livello quale C#. In questo modo le *query* si possono comporre a *runtime*, a seconda dei campi di *input* compilati dall'utente. Alla *query* così composta saranno inoltre aggiunti i parametri di paginazione, in modo da ridurre la quantità di dati scambiati fra *server* e *client*. Un utente visualizzerà così solo 10 risultati per volta.

Pagine di contenuti

Sotto questa dicitura sono raggruppate tutte le pagine del sito che racchiudono informazioni sul progetto ArchiDATA: storia, descrizione dei metodi di raccolta del materiale, monitoraggio dell'avanzamento del lavoro, bibliografia di riferimento, alla quale si rimanda in chiave all'interno delle voci lessicografiche, elenco di autori e collaboratori, contatti. Tutte queste pagine sono state curate da Ludovica Maconi. Sono raggiungibili utilizzando il menù presente nell'intestazione del sito.

In un'altra area del sito, visibile solamente agli amministratori, è possibile ampliare e aggiornare la Bibliografia. La maschera di inserimento permette di scegliere la sezione da aggiornare (bibliografia, studi, fonti) e di inserire i nuovi dati, cliccando sull'icona verde + e usando le codifiche html per la formattazione. Il riordinamento alfabetico delle aggiunte è poi realizzato in modo automatico dal sistema, dopo il salvataggio dei dati. Modifiche al testo precedentemente inserito possono essere introdotte usando l'icona della matita che affianca ogni riferimento bibliografico, come in fig. 5 (mentre l'icona del cestino serve, come ovvio, per cancellare il riferimento).



Fig. 5 Ritagli dalla maschera di aggiornamento della Bibliografia.