Ridire 2.0

La nuova *release* del web Corpus promosso dalla SILFI

RIDIRE.it (Risorsa Dinamica Italiana di Rete) è un web corpus che documenta l'italiano utilizzato in rete nei domini più rappresentativi della lingua e della cultura italiana, frutto di un progetto promosso dalla SILFI, in collaborazione con un consorzio di università italiane,[1] e finanziato dal Fondo Italiano per la Ricerca di Base (FIRB).

RIDIRE è stato rilasciato per la prima volta dieci anni fa, in un'infrastruttura di rete dotata di funzioni computazionali di ricerca proprie della linguistica dei corpora necessarie a evidenziare le particolarità dell'uso linguistico e della fraseologia italiana, sia a fini di ricerca e come strumento didattico sia per il consolidamento del possesso dell'italiano da parte degli apprendenti.

Ridire 2.0 ripropone il web corpus in una nuova *release*, in cui il rumore causato dall'estrazione automatica dei testi dal web (duplicazioni dei dati, problemi di formato, *boilerplate*) è stato drasticamente ridotto e l'infrastruttura computazionale di ricerca, aggiornata agli standard contemporanei, consente oggi un accesso ai dati più semplice e veloce.

I documenti raccolti in RIDIRE costituiscono una fonte rilevante per lo studio della fraseologia italiana, sia generale sia nei suoi ambiti settoriali. La base dati interrogabile ha una dimensione di circa 1.3 miliardi di parole e è strutturata in domini divisi in due tipologie

Domini Semantici: i campi dell'eccellenza italiana in cui le scelte linguistiche sono dettate dal contenuto	***Domini Funzionali****:* contesti d'uso in cui la lingua ha convenzioni legate alla funzione sociale
• **Letteratura** • **Moda** • **Design-architettura** • **Cucina** • **Sport** • **Religione** • **Arti figurative** • **Cinema** • **Musica**	• **Informazione** • **Economia e affari** • **Amministrazione e legislazione**

Il *design* di RIDIRE

Ridire 2.0 è mantenuto in rete dal Laboratorio di Informatica Umanistica del Dipartimento di Lettere e Filosofia dell'Università di Firenze (LIU) e è liberamente disponibile all'indirizzo

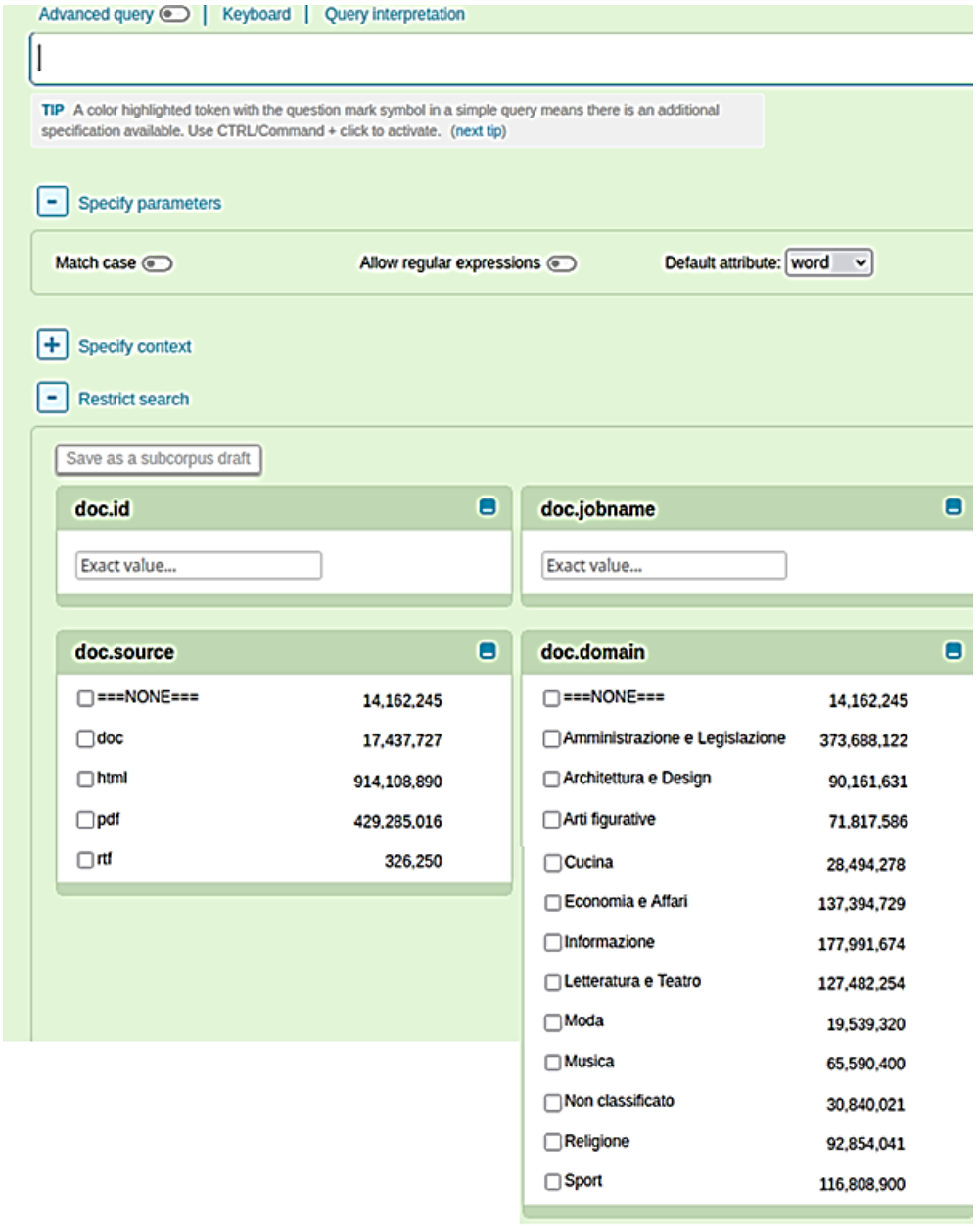
http://corpora.dilef.unifi.it/query?corpname=ridire

[1] Università di Firenze (LABLITA e Dipartimento Sistemi e Informatica), l'Università di Torino (Dipartimento Scienze Letterarie e Filologiche), l'Università di Roma3 (Dipartimento di Italianistica), l'Università di Napoli (Dipartimento di Filologia Moderna) e l'Università di Siena (Dipartimento di Economia).

Le funzioni di ricerca di Ridire 2.0

La disponibilità di strumenti computazionali per l'estrazione di informazioni vive da grandi corpora rappresentativi dell'uso è un'importante opportunità di conoscenza e costituisce il portato naturale dell'evoluzione tecnologica.

Le occorrenze del corpus RIDIRE sono lemmatizzate e annotate per Parte del Discorso (PoS). Le funzioni di ricerca sono implementate nell'infrastruttura KonText[2], che permette di effettuare ricerche sull'intero corpus o sui suoi sotto-corpora, estraendo Liste di frequenza, Concordanze, Collocazioni e *Colligation*.



La schermata di KonText per la scelta dei corpora su cui effettuare le ricerche

Una videoguida all'utilizzo delle funzioni di ricerca implementate in KonTexte realizzata da LIU e è disponibile qui http://lablita.eu/downloads/userguide_ridire2.0.mp4

[2] Machálek, T. (2020). KonText: Advanced and flexible corpus query interface. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 7003-7008).