

## *News letter del 30 giugno 2008*

### Agli utenti di Taltac

Dal 1 luglio è disponibile una nuova versione di Taltac2 (T2): la release **2.8.0**

Questa versione richiede una licenza diversa da quella predisposta per la versione 2.5.x.

Con la **versione 2.8.0** si introduce un'importante innovazione nella gestione della licenza, in quanto - oltre alla tradizionale **licenza stand alone**, legata fisicamente ad una sola macchina - si prevede anche una **licenza network**, che consente l'utilizzo in remoto della licenza da macchine diverse, purché visibili nella stessa *intranet*, nelle quali sia stato installato previamente il programma. Questo tipo di licenza può essere richiesta anche nella forma multi-utente, per l'uso dipartimentale di gruppi di lavoro di ricercatori o per aule didattiche informatiche.

Questa innovazione è introdotta al fine di

i) rispondere all'esigenza di una migliore gestione dei laboratori didattici con esercitazioni informatiche; non è più necessario avere privilegi particolari per l'utilizzo di T2 (mentre l'installazione va sempre eseguita come Administrator), poiché ora i file di lavoro dell'utente vengono salvati in una cartella diversa da quella di installazione, tuttavia in fase di installazione sarà necessario rendere accessibile in scrittura la cartella "**Taltac2\db**" a tutti gli utenti autorizzati ad utilizzare T2;

ii) consentire il lavoro in reti intranet: si può ora disporre di una licenza network grazie alla quale T2 è utilizzabile anche da postazioni diverse, una volta installata una **licenza mono-utente** su una macchina prescelta e installato il programma su ciascuna delle postazioni da cui si prevede di lavorare. Nella modalità network permane la possibilità di aprire più istanze di T2 simultaneamente sulla stessa macchina.

Quando un Dipartimento dispone di una **licenza multi-utente**, la licenza è posta su di un server e le singole utilizzazioni del programma vengono attivate in remoto, fino alla concorrenza del numero massimo di istanze disponibili: un messaggio avvertirà l'utente del raggiungimento del limite di utilizzo simultaneo da macchine diverse.

AVVERTENZA: se T2 si chiude irregolarmente a causa di un crash provocato da errore o altro, potrebbe accadere che al gestore della licenza risulti impropriamente ancora aperta quella istanza. In tal caso, dallo stesso PC basta aprire e richiudere il programma per permettere al sistema di controllo centralizzato di ricalcolare correttamente il numero di licenze in uso.

Contestualmente al rilascio della nuova release sono comunque messe a disposizione sul sito istruzioni dettagliate su come attivare e gestire le nuove modalità d'uso della licenza con la descrizione degli scenari più tipici.

## Novità della versione 2.8.0:

Oltre alla correzione di alcuni bugs segnalati dagli utenti e al miglioramento dell'interfaccia nell'attivazione di funzioni base, si segnalano:

- L'**ampliamento di risorse statistico-linguistiche**: il lessico di frequenza di Repubblica (Rep90) è stato quasi raddoppiato (oltre 100.000 entrate), il dizionario positivo-negativo è stato revisionato, sono ora disponibili altri dizionari tematici: figure parentali, cronomimi, Nazionalità/Etnie, e-something; alcuni di questi sono anche in forma di query nel file query predefinite che trovate nel sito nell'area utenti)

A settembre 2008 sarà disponibile il file di normalizzazione con l'aggiornamento di polirematiche, locuzioni e nomi di personaggi, enti e società al 2008 e altre risorse.

- L'**ampliamento** della funzione di **calcolo delle co-occorrenze**. Le co-occorrenze vengono calcolate nell'ambito di un intorno di  $n$  parole che l'analista è libero di impostare. Taltac calcola le co-occorrenze esistenti nel corpus tra le forme (o classi semantiche) indicate in una lista predefinita o fra le parole con almeno "x" occorrenze. Inoltre permette di calcolare anche le **collocazioni** di tali co-occorrenze, ovvero la distribuzione della posizione (precedente/successiva) di ogni parola co-occorrente rispetto al termine pivot. In particolare, i risultati sono costituiti da:

- la matrice delle co-occorrenze, una matrice quadrata di tipo *Forme x Forme* contenente le cooccorrenze rilevate tra ciascuna coppia di forme;
- la lista delle co-occorrenze e delle collocazioni maggiormente significative, dimensionabile a piacere;
- l'inventario delle poli-cooccorrenze, ovvero delle co-occorrenze che si realizzano, nell'intorno definito, tra tre o più forme.

- La scrittura delle **specificità** anche su un file esterno in formato txt, che facilita la fruibilità dei risultati e ne ottimizza la presentazione nei report di ricerca.

- La funzione del calcolo dell'**indice TFIDF** si è arricchita di un'ulteriore opzione che permette di ottenere risultati migliori in presenza di frammenti di lunghezza eterogenea.

- La funzione **Ricerca Entità** è stata razionalizzata ed ampliata. E' ora più immediato cogliere l'oggetto, l'ambito e i risultati della ricerca. Tra questi ultimi si segnala la possibilità di creare una nuova variabile contenente, per ogni frammento che soddisfa la richiesta, le  $n$  parole precedenti e/o successive all'entità ricercata. Tale funzione risulta utile, ad esempio, quando si vogliono raccogliere risposte fornite a domande presenti nei frammenti. Inoltre tutte le ricerche di entità, e le relative impostazioni, possono essere salvate in piani di lavoro testuali (**Metaquery**) allo scopo di essere riutilizzate in un secondo momento o su altri corpus. In questa prospettiva, ad esempio, l'intero processo di creazione e popolamento di variabili strutturate, create a partire dai dati testuali non strutturati, può essere replicato su un nuovo corpus con un unico comando, rilanciando una metaquery.

- La lista delle entità fornita dalla funzione Ricerca Entità mostra adesso anche la **descrizione del frammento** in cui ogni risultato è stato rinvenuto, al fine di migliorare l'interpretazione dei risultati.
- Nella finestra **Esplora Corpus** la lista dei frammenti è ora **ordinabile** in base alla descrizione del frammento o, in alternativa, all'ordine di apparizione dei frammenti nel corpus (id\_frammento), con lo scopo di facilitare la ricerca di ogni singolo frammento. E' inoltre possibile esportare subito i frammenti estratti a partire da una ricerca di entità, anche in sottoinsiemi attraverso la gestione di filtri sulle variabili categoriali.
- La **selezione di un subcorpus** è completamente ridisegnata, con la possibilità di **estrarre solo alcuni frammenti o alcune sezioni o una loro combinazione**, non solo a partire da variabili categoriali ma anche da liste di frammenti, frutto di precedenti ricerche.
- La funzione di **ricostruzione del corpus** è stata ridisegnata ed ampliata. È ora possibile riprodurre il corpus con un numero ed un assortimento di opzioni maggiore rispetto al passato, consentendo anche la **modifica del formato del corpus**, ad esempio da "collezione di file" in "file strutturato in campi".
- Il **miglioramento dell'integrazione con il software TreeTagger**. Taltac consente di importare file risultanti da un'elaborazione effettuata in una delle 5 lingue previste con il lemmatizzatore gratuito TreeTagger, mantenendo e gestendo le informazioni grammaticali fornite da quest'ultimo. Inoltre è possibile generare corpus parzialmente etichettati con Taltac nel formato richiesto da TreeTagger, al fine di migliorare le prestazioni e i risultati di quest'ultimo.

## DISPONIBILITA'

Per tutti gli utenti che hanno acquistato Taltac nel 2007 e 2008 la versione 2.8.0 è scaricabile gratuitamente, come licenza *stand alone*.

Per i possessori di versioni Taltac 2.0 o Taltac 2.5 acquistate nel 2006 o precedentemente, la nuova versione è disponibile come upgrade secondo le modalità specificate sul sito.

Il nostro sito può ospitare alla pagina "Materiali" contributi, relazioni, lucidi e papers degli utenti Taltac che testimoniano gli usi del software o forniscano esempi applicativi o metodologici da condividere nella comunità scientifica dell'area Taltac. Siete quindi invitati a spedirci lavori (liberi da copyright) o link che intendete mettere online, per ricevere commenti e reazioni: per questo scrivete a [sergio.bolasco@uniroma1.it](mailto:sergio.bolasco@uniroma1.it) e inserite allegati in formato pdf.