



***Content analysis dei provvedimenti
della sezione disciplinare del C.S.M.***

Manuela Sassi, Alessandra Cinini

ILC-CNR / IRSIG-CNR

Pisa 2006





***Content analysis* dei provvedimenti
della sezione disciplinare del C.S.M.**

Manuela Sassi, Alessandra Cinini

ILC-CNR / IRSIG-CNR
Pisa 2006

In copertina: elaborazione grafica della foto del Palazzo dei Marescialli,
tratta dal sito del CSM.

1. Introduzione

Nell'ambito del progetto di ricerca denominato "Tecnologie dell'informazione e della comunicazione per la giustizia" coordinato dall'Istituto di Ricerca sui Sistemi Giudiziari (IRSIG-CNR di Bologna)¹ e co-finanziato dai Fondi Integrativi Ricerca di Base (FIRB) del Ministero dell'Università e della Ricerca, l'Istituto di Linguistica Computazionale (ILC-CNR di Pisa) ha collaborato con l'IRSIG e con il Centro Studi e Ricerche sull'Ordinamento Giudiziario dell'Università di Bologna² per la realizzazione di una base dati elettronica per l'analisi dei provvedimenti della sezione disciplinare del Consiglio Superiore della Magistratura.

A questo scopo è stato bandito un Assegno di Ricerca, espletato dalla Dr.ssa Alessandra Cinini, laureata in Informatica, con esperienza nel trattamento di dati testuali.

La collaborazione all'attività di ricerca prevista per l' Assegno verteva sul tema seguente: "Individuazione ed estrazione di informazioni da banche dati giuridiche tramite lo studio e la creazione di regole testuali con parametri di tipo linguistico. Creazione e sperimentazione di protipi per l'analisi della giurisprudenza".

1 Si tratta di un progetto quadriennale, iniziato nel 2002, promosso e gestito nell'ambito della commessa dell'IRSIG-CNR: "Organizzazione, funzionamento ed innovazione tecnologica nell'amministrazione della giustizia".

2 Per l'IRSIG la collaborazione è stata prestata da Marco Fabri, coordinatore del progetto, mentre per il CeSROG da Domenico Piscitelli e da Daniela Cavallini, che ha in particolare fornito il contributo tecnico-giuridico per lo sviluppo del prototipo.

1.Introduzione

L'IRSIG e il CeSROG hanno messo a disposizione del progetto i testi dei provvedimenti emessi dalla Sezione Disciplinare del CSM.

I testi, sotto forma di archivio storico di files e in vari formati, consistevano di circa 1600 unità. Su tali materiali si è operato in modo da renderli trattabili con software di produzione ILC, procedendo alle varie fasi lavoro descritte di seguito.

Il primo prodotto ottenuto è un database consultabile, di cui verranno descritte le caratteristiche e le modalità di interrogazione. Inoltre verrà data esemplificazione delle possibilità di sfruttamento di tale database, che vanno oltre la semplice consultazione, attraverso anche fasi successive di elaborazione.

2. Schema di lavoro

Allo scopo di riassumere le fasi principali del lavoro di ricerca svolto, si elencano i seguenti passi operativi:

1. studio della tipologia dei testi e codifica degli elementi individuati come prioritari;
2. applicazione ai testi già codificati dei moduli informatici che creano il database;
3. verifica dell'implementazione effettuata attraverso consultazioni mirate e volte ad individuare la migliore codifica dei fenomeni rilevanti come, ad esempio, la struttura concettuale, le date, i nomi propri del collegio, dell'incolpato, dei testimoni, e di quant'altro ritenuto necessario;
4. sviluppo di software ad hoc finalizzato al riconoscimento o alla normalizzazione dei dati (es.: riduzione delle date di

2. Schema di lavoro

emissione ad un medesimo formato, estrazione dell'anno di inizio procedimento o di deposito, riconoscimento delle sezioni formali del documento, degli acronimi, delle sigle, dei numeri, ecc.);

5. rigenerazione degli archivi ogni qualvolta si sia reso necessario, per modifiche nella codifica o nuovi inserimenti, e aggiornamento del database.
6. strategie di consultazione del database allo scopo di produrre report di content-analysis;
7. elaborazione dei risultati delle query, sia con software ad hoc, sia con comuni prodotti di office automation, in modo da produrre la documentazione necessaria.
8. Applicazione di metodologie di POS-TAGGING per il riconoscimento morfologico-sintattico delle parti del discorso e etichettatura di tutte le parole del database per ulteriori analisi di tipo contenutistico.

3. Studio della tipologia dei testi

Con i testi informatizzati delle Sentenze e delle Ordinanze della Sezione Disciplinare del Consiglio Superiore della Magistratura è stata costituita una prima banca-dati testuale da implementare nel sistema DBT (Data Base Testuale, Brevetto CNR di E. Picchi¹). Tale raccolta è stata successivamente integrata con gli aggiornamenti recuperati fino al settembre dell'anno 2004; attualmente si opera su un insieme di 1618 provvedimenti disciplinari, di cui 148 Ordinanze e 1470 Sentenze, relative agli

1 Picchi Eugenio, *DBT - Data Base Testuale*. Bollettino CRIBECU, Pisa, Scuola Normale Superiore, I/1 (1991), pp. 135-137.

3. Studio della tipologia dei testi

anni 1990-2004 (per quanto riguarda la data di emissione, infatti sono presenti anche procedimenti iniziati negli anni '80).

Buona parte dei testi relativi ai procedimenti disciplinari dei primi anni '90 si presentavano in formato WS2, ossia Wordstar, uno dei primi programmi di elaborazione di testi (i cosiddetti wordprocessor) che si sono diffusi precedentemente a Winword, soprattutto in ambito giudiziario e scolastico. Inoltre i materiali raccolti si trovavano in un formato già semielaborato, comprendente alcuni codici di controllo e pertanto si è provveduto alla realizzazione di un apposito programma che convertisse i documenti direttamente al formato testo per procedere al successivo inserimento nella banca-dati.

Altri tipi di documenti presenti erano in WordPerfect o in RTF (Rich Text Form) oppure in Word. I testi così elaborati e ricondotti ad un unico formato-testo sono stati poi codificati per il successivo inserimento nel DBT.

4. Creazione del database

I 1618 file sono stati raggruppati per anno e indicizzati con la funzione apposita del DBT. In questa fase vengono generati gli archivi in modo tale da ottimizzare sia lo spazio di memoria occupato, sia la velocità di recupero dell'informazione. Così il testo viene “predigerito” e predisposto in modo tale che, con un semplice tocco del mouse, ogni parola diventa un link, facilitando la navigazione nei documenti e nelle loro sezioni.

Sono stati prodotti due archivi, uno per le Sentenze ed uno per le Ordinanze, in modo che possano essere consultati sia singolarmente che congiuntamente (in quest'ultimo modo è

4. Creazione del database

possibile fare una sola richiesta che produrrà risultati distinti). Per questo motivo parliamo di database, visto che l'insieme dei due archivi, per omogeneità di trattamento, può essere visto come una unica fonte di dati.

Sono state elaborate un totale di 3.608.880 parole, di cui 168.321 nei 147 procedimenti delle Ordinanze e 3.440.459 nei 1470 procedimenti delle Sentenze.

Alcuni procedimenti disciplinari risultano classificati come Sentenze-Ordinanze: il riferimento di questi è stato inserito in entrambi gli archivi, l'intero testo è presente solamente nel database delle Sentenze per evitare duplicati che potrebbero alterare i risultati e le statistiche delle analisi testuali.

5. Caratteristiche del database

L'indicizzazione ha permesso l'applicazione di funzioni di analisi testuale agli archivi ottenuti quali:

1. la ricerca libera nel testo di una o più parole o porzioni di esse, in AND e OR;
2. la ricerca per lemma che, tramite il motore morfologico, recupera tutte le forme del lemma dato e/o dei sinonimi;
3. il calcolo delle frequenze assolute in ordine alfabetico o decrescente;
4. le concordanze di tutte le parole secondo l'ordine di presenza nei testi, oppure secondo l'ordine alfabetico delle forme che nella frase seguono/precedono la parola oggetto di ricerca;
5. l'index locorum, arricchito da riferimenti puntuali per tutte le parole del testo;

5. Caratteristiche del database

6. il calcolo dell'indice di “mutual information” ossia la formula statistica che applicata ad un testo genera un elenco di coppie di parole, dette co-occorrenze, che sono collegate da uno stretto legame di correlazione nei documenti analizzati;
7. il calcolo dell'indice di leggibilità, basato sulle formule statistiche più accreditate.

Nell'esplorazione dell'archivio tramite le funzioni elencate, si sono evidenziati casi in cui era possibile intervenire con software, da sviluppare ad hoc, in modo da isolare e contrassegnare alcuni fenomeni testuali da usare in seguito per migliorare la navigazione e il recupero di informazione utile a ulteriori analisi di tipo concettuale (content-analysis). Fra questi citiamo il riconoscimento delle date, utili per analisi di tipo temporale.

Nei paragrafi successivi verranno descritti i vari fenomeni testuali che sono stati oggetto di individuazione e codifica, ossia:

- i **riferimenti**, che individuano univocamente ciascuna sentenza o ordinanza attraverso il numero e l'anno;
- le **date**, indicate dall'Estensore alla fine di ciascun procedimento;
- i **numeri**, cioè tutti i tipi presenti nei testi e che possono indicare articoli o leggi, l'ammontare di cifre, unità di misura e date (escluse quelle di cui al punto precedente);
- le **sigle**, che possono essere acronimi, titoli, enti, ecc.;
- la **struttura**, ossia le sezioni o parti in cui è organizzata la sentenza.

5. Caratteristiche del database

Nell'ultima parte verranno proposte alcune strategie di consultazione del database, con l'esemplificazione di come sia possibile ricondurre i risultati ottenuti a vere e proprie analisi di contenuto.

5.1. Il riferimento

Il *riferimento logico*, ossia l'intestazione univoca che identifica ogni singolo procedimento, è stato dedotto dal numero e dall'anno di inizio del procedimento disciplinare.

Tale dato risulta dallo stesso nome del file, così come era stato registrato nella memoria del computer, ed è stato contrassegnato, a livello di codifica-testo, con il codice %, che poi il DBT tratterà come identificativo di quel procedimento.

Ad es.:

%Sen001_90

identifica la Sentenza n. 1 del 1990.

Per ragioni di utilità nell'elencare i documenti, è stato trasformato il numero di Sentenza/Ordinanza, con l'introduzione di zeri non significativi alla sua sinistra, in modo da poter ottenere un elenco in ordine alfanumerico che rispettasse anche la sequenza cronologica.

Nel paragrafo seguente vedremo come sia stato ulteriormente arricchito il codice del riferimento.

5.2. Le date

Studiando le peculiarità dei documenti ci siamo resi conto delle varianti di uso nell'indicazione della data.

5. Caratteristiche del database

Dopo un'attenta analisi di tali varianti, è stato studiato un programma per la marcatura automatica della data di emissione del procedimento disciplinare, per poi procedere alla sua normalizzazione e renderla così utilizzabile come parametro nella fase di consultazione.

Nella tabella 1 si esemplificano alcune delle suddette annotazioni; in questa lista l'ordine è basato sull'anno e poi sul numero di sentenza; nel caso di dato mancante nella colonna "Anno" compare 0.

Come si può notare nella colonna "Data di emissione" essa viene espressa nei testi in vari formati.

Rif. Sentenza	Data di emissione	Anno
%Sen107_00	0	0
%Sen111_02	0	0
%Sen001_90	Roma, 21 settembre1990	1990
%Sen004_90	Roma, 23 novembre1990	1990
%Sen015_90	Roma, 19 ottobre1990	1990
%Sen016_90	Roma, 19.10.1990	1990
%Sen019_90	Roma, 28 settembre1990	1990
%Sen025_89	Roma, 9 novembre1990	1990
%Sen025_90	Roma, 16/11/1990	1990
%Sen033_88	Roma, 26 ottobre1990	1990
%Sen037_89	Roma, 23.11.1990	1990
%Sen037_90	Roma, 28 settembre1990	1990

Tab. 1 - Riferimenti e Date

Il riconoscimento della data ha permesso successivamente di arricchire il riferimento logico con l'inserimento dell'anno di

5. Caratteristiche del database

conclusione del procedimento disciplinare.

L'idea di introdurre questa ulteriore codifica è scaturita, come anzidetto, nella fase di consultazione di prova degli archivi; a seguito di questa integrazione il reperimento dei procedimenti conclusi in un determinato intervallo risulta più semplice e veloce.

Infatti è possibile applicare, in fase di interrogazione, la restrizione sul riferimento, limitando la ricerca a quei documenti il cui riferimento rispecchia determinati criteri.

Il nuovo riferimento logico dei documenti risulta così strutturato:

<i>Sen/Ord</i>	<i>N.Proc.</i>	<i>_Anno Inizio</i>	<i>(Anno Conclusione)</i>
SEN	045	_92	(95)
ORD	002	_93	(95)

Di seguito si propone un'analisi fatta estraendo i riferimenti con la richiesta in DBT di tutte le date di emissione contenute nell'archivio delle sentenze, ottenuta con i seguenti passi:

- Il risultato della query è stato salvato su file con la funzione *SalvaFinestra* (il sistema DBT consente il salvataggio successivo delle finestre, di volta in volta attive, su uno stesso file di salvataggio).
- Successivamente è stata recuperata la data di termine dei procedimenti componenti l'archivio dove la data è facilmente individuabile grazie all'operazione di *marcatura citata*. Si è rilevato che in 1467 Sentenze è presente la data di conclusione delle medesime, per questo motivo il calcolo della durata media dei

5. Caratteristiche del database

procedimenti nella precedente fase di lavoro è stato effettuato in base alla data di emissione piuttosto che a quella di deposito, mancante in numerosi casi.

- Confrontando il totale dei riferimenti delle unità logiche recuperate nelle due fasi di ricerca è stato possibile realizzare una lista delle Sentenze in cui non è presente il numero di Registro di Deposito.
- I dati sono stati poi trattati con un foglio elettronico che è in grado di fare i calcoli e riprodurli sotto forma di grafico.

Nella tabella 2 viene illustrato l'andamento del numero di procedimenti disciplinari iniziati e conclusi per ogni anno dal 1990 al 2004 (quest'ultimo anno non è completo).

<u>ANNO</u>	<u>N. PROC. CONCLUSI</u>	<u>N. PROC. INIZIATI</u>
1990	17	45
1991	93	85
1992	76	64
1993	59	57
1994	36	84
1995	107	105
1996	120	123
1997	160	151
1998	154	141
1999	83	97
2000	122	130
2001	138	121
2002	113	101
2003	119	92
2004	56	23

Tab. 2 - Elenco procedimenti per anno

5. Caratteristiche del database

Nella figura 1 viene rappresentato graficamente l'andamento degli stessi procedimenti.



Fig. 1 - Procedimenti iniziati/conclusi per anno

Basandosi sui dati precedenti, è stato svolto uno studio sulla durata dei procedimenti disciplinari inseriti nel database delle sentenze.

La data di inizio è stata dedotta dal numero di procedimento disciplinare, estraendo l'anno e calcolando il mese sulla base di una ipotetica media di procedimenti iniziati per ogni mese di quell'anno. Per alcuni anni tale media è stata diversamente valutata dato che questi si discostavano dalla media generale.

Il termine del procedimento, valutato in mese ed anno, è stato dedotto dalla data di emissione della sentenza.

La durata media di un procedimento disciplinare è risultata di sette mesi.

5. Caratteristiche del database

Sono state individuate cinque classi per la valutazione della durata dei procedimenti e la figura 2 illustra la quantità di procedimenti per ciascuna classe di durata.

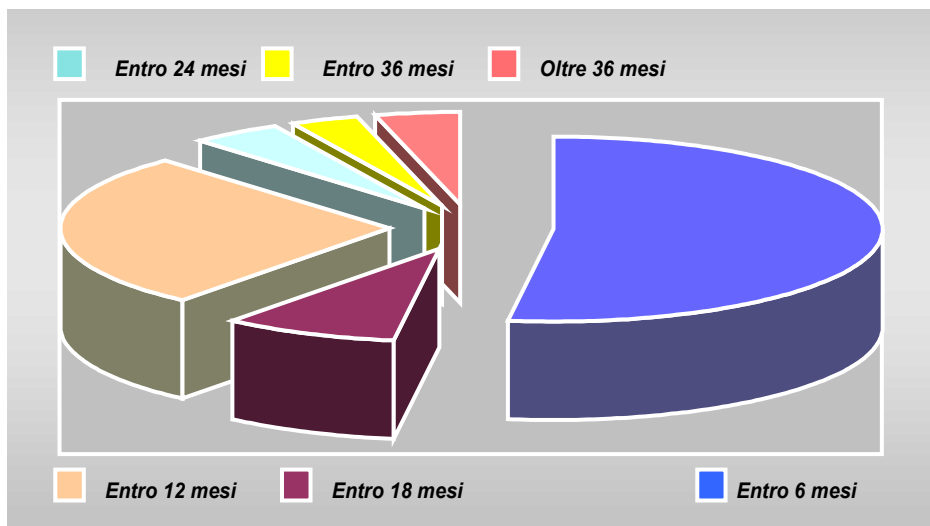


Fig. 2 - Durata dei procedimenti emessi

5.3. I numeri

Un'altra operazione che si è ritenuto utile per la consultazione è quella di marcare i numeri presenti nel testo. Questo permette di individuare, ad esempio, tutte le sentenze che fanno riferimento ad un determinato articolo di legge, oppure che fanno riferimento a determinate quantità, monetarie, di misura, di tempo, ecc. Per illustrarne l'utilità, si propone un grafico che quantifica il numero di sentenze nelle quali si citano gli articoli del R.D.L. del 31 maggio 1946.

5. Caratteristiche del database

In fase di interrogazione tramite DBT, è stata definita una famiglia di ricerca¹ che individua tutti i contesti in cui compaiono le abbreviazioni [art.] od [artt.] ed il numero [1946] e stabilendo 5 come parametro di distanza massima. In questo modo sono state recuperate tutte le citazioni degli articoli del Regio Decreto Legislativo del 31 maggio 1946.

I dati ottenuti dal DBT sono stati salvati nel formato-testo con le modalità offerte dall'interfaccia utente, che tramite il pulsante destro del mouse (oppure la barra dei menu in alto) registra il contenuto della finestra attiva su file.

Successivamente i dati salvati sono stati elaborati con prodotti di Office Automation.

Nella figura 3 sono state rappresentate le citazioni degli articoli del R.D.L. Del 1946 che compaiono nel database delle Sentenze con una frequenza maggiore di 5.

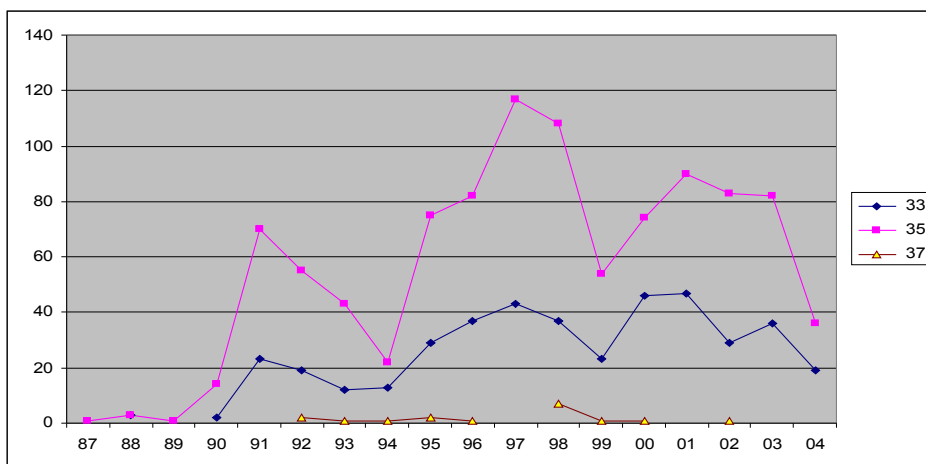


Fig. 3 - Percentuale di citazioni del RDL 1946 per anno

1 Per Famiglia di Ricerca si intende la possibilità di richiedere due o più parole compresenti nella stessa frase.

5.4. Le sigle o acronimi

Un'ulteriore trasformazione effettuata sugli archivi testuali è stata la codifica delle abbreviazioni, molte delle quali sono state utilizzate come forme di citazioni di atti normativi, e degli acronimi, che si riferiscono generalmente ad enti, istituti, etc.

L'individuazione degli acronimi e delle abbreviazioni nei documenti è stata realizzata con l'ausilio di un modulo del DBT denominato "CercaFenomeni", la codifica è stata effettuata aggiungendo una marcatura iniziale e finale alla forma presente nel documento.

L'elenco di tutte le "sigle" codificate può essere rapidamente generato in fase di consultazione degli archivi con DBT, usando il canale di ricerca apposito (esiste la voce specifica "Ricerca Sigle").

Di seguito sono illustrate quelle usate più di 50 volte nell'archivio delle Sentenze.

dott.	36682	nn.	276	disp.	109
n.	13189	cass.	274	fasc.	107
art.	8387	pen.	266	sig.ra	107
dott.ssa	4009	g.i.	262	c.	106
avv.	3364	vol.	262	d.l.	106
r.d.l.	2962	cpv.	252	p.s.	105
prof.	2891	sez.	239	reg.	104
p.m.	2172	sig.	233	fl.	103
g.i.p.	1900	s.r.l.	212	a.g.	101
r.g.	1818	cost.	211	ord.	101
c.p.p.	1784	disc.	206	att.	100
p.q.m.	1472	co.	202	cc.	99

5. Caratteristiche del database

c.p.	1327	cod.	176	segg.	98
cfr.	1286	ff.	175	ss.uu.	96
reg.dep.	1197	all.	171	s.a.s.	93
dr.	1093	mod.	167	cap.	92
artt.	1034	s.p.a.	157	pagg.	92
c.s.m.	1034	fol.	155	gen.	89
proc.	970	c.c.	153	d.m.	81
l.	824	r.d.	153	f.f.	72
p.g.	803	lett.	147	s.u.	72
dr.ssa	802	r.g.n.r.	145	dep.	66
v.	591	cit.	137	prot.	63
p.	508	g.u.p.	132	i.s.p.i.	61
r.d.lgs.	431	fg.	131	soc.	59
sent.	422	gg.	129	dott.sa	53
pag.	420	m.llo	127	giud.	51
f.	407	doc.	121	un.	51
d.p.r.	349	ing.	121	n.ri	50
c.p.c.	344	on.	114		
c.d.	316	on.le	112		

Esaminando tali sigle si rileva una difformità di uso e per ottimizzare la consultazione sarebbe opportuno uniformare le abbreviazioni e gli acronimi presenti nei testi secondo standard comuni.

Di seguito sono elencati i vari modi in cui viene citato il Regio Decreto Legislativo e in corrispondenza di ognuno è indicata la frequenza di utilizzo nell'archivio delle Sentenze.

r.d. l.vo	1	r.d.lgs	9
r.d. lgs	3	r.d.lgs.	431
r.d. lgs.	7	r.d.lgs.vo	5
r.d.l.gvo	1	r.d.ls.	1
r.d.l.vo	16	r.d.lsg.	1
r.d.leg.	2	r.d.lvo	3
r.d.lg.vo	2		

5. Caratteristiche del database

La citazione degli articoli avviene generalmente utilizzando le due abbreviazioni art. ed artt. seguite da uno spazio e dal numero.

5.5. La struttura delle sentenze

Tutti i provvedimenti sono impostati secondo una struttura fissa; allo scopo di recuperare tale struttura a livello informativo, si è proceduto alla individuazione dei segnali che indicano i livelli o sezioni di tale struttura.

In alcuni casi sono chiaramente indicati nel testo: ad esempio la sigla P.Q.M. (o p.q.m. dato che, a livello di etichetta, il DBT non differenzia i due modi) individua il dispositivo, quella parte della sentenza che contiene la decisione del caso; per identificare tale parte è stato usato il marcatore \PQM\.

Nella sezione dedicata all'esposizione dei fatti e dei motivi, vi sono diversi modi di organizzazione del testo: è stato scelto di lasciare divise le due sezioni, qualora sia indicato chiaramente nel testo, oppure di non distinguerle, nei casi in cui non vi siano segnali inequivocabili.

In mancanza di segnali non viene introdotto nessun marcatore, mentre in presenza dei seguenti segnalatori, viene inserito il marcatore corrispondente, come risulta dalla seguente tabella di corrispondenze:

SVOLGIMENTO DEL PROCESSO (e simili)	\FATTO\
MOTIVI DELLA DECISIONE (e simili)	\DIRITTO\
Entrambe le frasi	\FATTO_DIRITTO\

5. Caratteristiche del database

\FATTO\ ... In data 21.1.1993 la Procura Generale presso la Corte di Cassazione chiedeva pronuncia di non doversi procedere a carico del dott. Xxxxxx per cessata appartenenza all'Ordine giudiziario. La richiesta è accoglibile sulla scorta degli elementi di causa sopra menzionati e pertanto va pronunciata sentenza di proscioglimento del dott. Xxxxxx per tale motivo.

\PQM\ P.Q.M.

La Sezione Disciplinare del Consiglio Superiore della Magistratura;
Visto l'art. 35 del R.D.L. 31 maggio 1946, n. 511,
DICHIARA

non doversi procedere nei confronti del dott. Xxxxxx perché il procedimento disciplinare si è estinto per cessazione dell'appartenenza all'Ordine Giudiziario.

Roma, 19 febbraio 1993

L'ESTENSORE

IL PRESIDENTE

IL MAGISTRATO SEGRETARIO

6.

7. Esempi di consultazione del database

Come si è detto, una volta che i testi siano stati codificati e elaborati per la creazione degli indici, è possibile consultare il database.

L'interfaccia appare nella figura 4 ed è lo stesso per la creazione del database, per l'indicizzazione, per la consultazione e per la elaborazione di funzioni speciali che il DBT mette a disposizione con i vari menu.



Fig. 4 - Interfaccia iniziale del DBT

7. Esempi di consultazione del database

Con la voce 1 "Ricerche testo singolo" si accede alla fase di consultazione di un singolo archivio o testo o database.

Con la voce 2 "Ricerche corpus di testi" è possibile consultare un corpus, ossia un insieme di testi, preventivamente indicizzati come archivi singoli e poi riuniti sotto l'etichetta di un corpus (una indicizzazione collettiva) sempre con la voce 3. Quest'ultima opzione è utilizzabile per procedere alla consultazione degli archivi delle Sentenze e delle Ordinanze in modo congiunto.

Con la voce 3 "Preparazione nuovo testo" si procede alla fase di indicizzazione.

In figura 5 appare la schermata corrispondente alla scelta della voce 1.



Fig. 5 - Interfaccia del DBT dopo la selezione 1

7. Esempi di consultazione del database

Con la scelta 1 si può procedere all'interrogazione.

Con la scelta 2 si accede ai sotto-menu per la generazione di indici/liste di tipo statistico.

Con la scelta 3 si generano le concordanze complete o parziali dell'archivio selezionato.

Per quanto riguarda le funzioni descritte precedentemente¹, queste possono essere applicate in modo interattivo al database tramite la scelta 1, proponendo richieste puntuali. Con la scelta 2 e/o 3 è possibile operare sul database completo.

I risultati di tali funzioni, una volta visualizzati, sono salvati su file esterni, in modo da essere visionati/elaborati successivamente. Ogni qualvolta il DBT genera un file di salvataggio, il nome e il luogo in cui viene registrato è chiaramente indicato in una mascherina di dialogo che l'utente deve confermare e chiudere.

Ovviamente fra le funzioni elencate in nota, la 1 e la 2 sono utilizzabili solo nella fase di interrogazione; nella figura 6 possiamo vedere un esempio di richiesta fatta tramite la finestra di dialogo che appare automaticamente all'apertura del programma.

-
1. la ricerca libera nel testo di una o più parole o porzioni di esse, in AND e OR;
 2. la ricerca per lemma;
 3. il calcolo delle frequenze assolute in ordine alfabetico o decrescente;
 4. le concordanze di tutte le parole;
 5. l'index locorum;
 6. le co-occorrenze;
 7. il calcolo dell'indice di leggibilità.

7. Esempi di consultazione del database

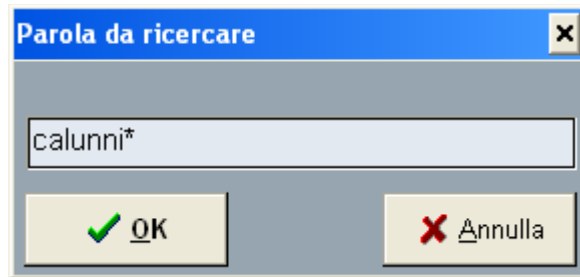


Fig. 6 - Richiesta di parola con mascheramento della finale


Nella figura 7 compare il risultato ottenuto con la precedente richiesta, cioè tutte le parole che iniziano per "calunni" dove l'asterisco indica qualunque terminazione.




A screenshot of a search results window. The title bar is blue and contains the text '[F]-calunni*' and window control buttons (minimize, maximize, close). The main area of the window contains a list of 12 items, each consisting of a number in parentheses, a word, and a frequency value to the right.

Numero	Parola	Frequenza
1)	calunnia	189
2)	calunniato	7
3)	calunniatore	8
4)	calunniatori	1
5)	calunniatorio	1
6)	calunnie	8
7)	calunniosa	6
8)	calunniosamente	2
9)	calunniose	7
10)	calunniosi	6
11)	calunniosità	6
12)	calunnioso	1

Fig. 7 - Elenco delle parole trovate e loro frequenza

Nei casi a risposta multipla, come il precedente, viene data la possibilità di procedere automaticamente alla spunta di tutte le parole elencate con il pulsante matita  e quindi alla formazione di una famiglia di ricerca in OR con l'altro pulsante

7.Esempi di consultazione del database

attivato, in modo da recuperare tutti i contesti delle parole elencate. 

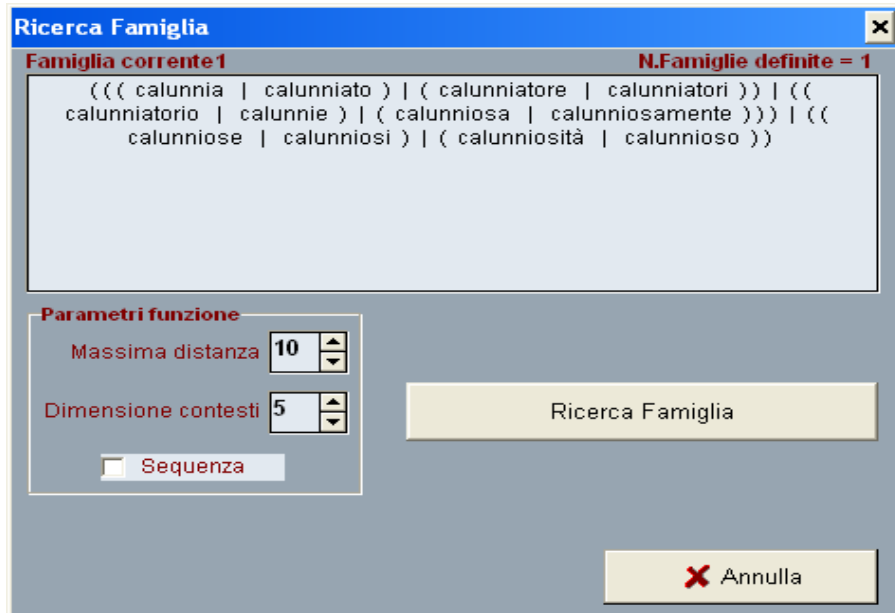


Fig. 8 - Ricerca per gruppo o famiglia di parole in OR

Nella figura 8 appare la finestra di dialogo generata dal DBT, dove si possono stabilire alcuni parametri o accettare quelli proposti:

Massima distanza

Numero massimo di elementi (parole, punteggiatura, accapo, ecc.) presenti tra due o più parole cercate; influisce solo nel caso della ricerca in AND; in caso di ricerca di parole adiacenti va indicato 1.

7.Esempi di consultazione del database

- Dimensione contesti** Numero di elementi visualizzati prima e dopo la/e parola/e richiesta/e. In questo modo si determina l'ampiezza del contesto.
- Sequenza** Rispetto o meno dell'ordine delle parole richieste; influisce solo nel caso della ricerca in AND.

Nella figura 9 appare il risultato della ricerca effettuata relativa ai primi 7 contesti sui 242 trovati.

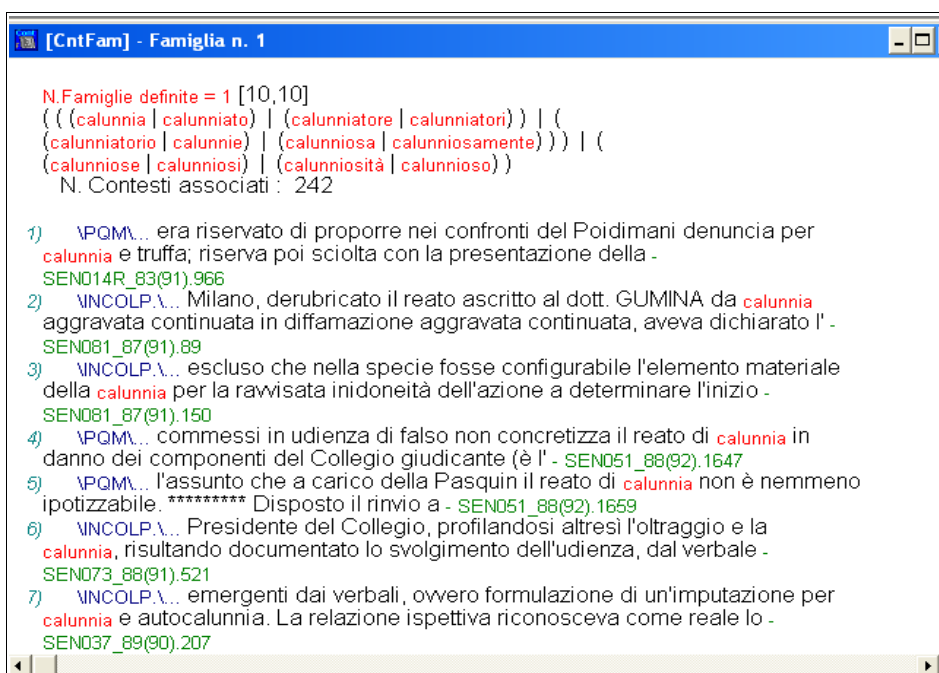



Fig. 9 - Risultato della ricerca di parole in OR

La finestra-risultato può contenere fino a circa 200 righe, per

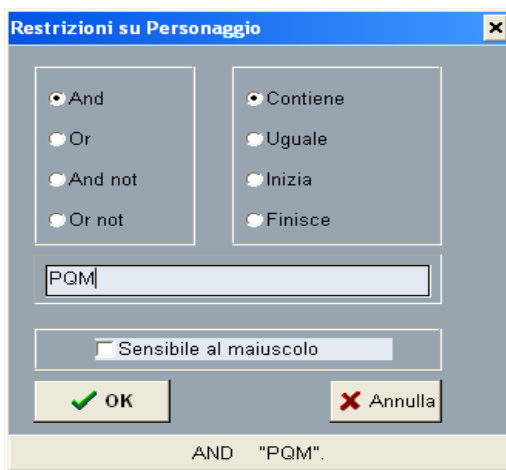
7. Esempi di consultazione del database

visionare le successive finestre è necessario scorrerle tramite il pulsante freccia  che si attiva nella barra dei pulsanti in alto .

Come si può notare nella precedente figura 9, ciascun contesto inizia con l'indicazione della sezione della sentenza da cui è stato estratto il contesto: ad es. Il contesto 1) inizia con PQM, che significa che la frase è stata estratta dal dispositivo, mentre il 2) fa parte della sezione INCOLP. Per ogni contesto estratto viene sempre indicata dal DBT la sezione di appartenenza.

La codifica della struttura permette quindi un ulteriore affinamento della ricerca, in quanto può essere utilizzata per la scelta dei contesti in base alla loro collocazione nelle varie sezioni.

Se utilizziamo il risultato dell'esempio precedente, si potrebbe ridurre il numero di contesti applicando la condizione che appartengano alla sezione PQM.



Per imporre questa regola è necessario usare il menu "Opzioni", e, al suo interno, "Restrizioni su Personaggio", che presenta la finestra di dialogo della figura 10.

Fig. 10 - Indicazione della regola di restrizione

7. Esempi di consultazione del database

In particolare si può notare che la regola viene costruita con la selezione dell'operatore voluto (*And*) e del comando appropriato (*Contiene*) con l'indicazione della stringa cercata (*PQM*). Ossia nella ricerca di contesti, successiva a questa operazione, si recupereranno solamente quelli in cui la sezione¹ contiene la stringa PQM.

Nella figura 11 compare la finestra di gestione delle regole che possono essere inserite, modificate o eliminate utilizzando i comandi che compaiono nella figura 10.

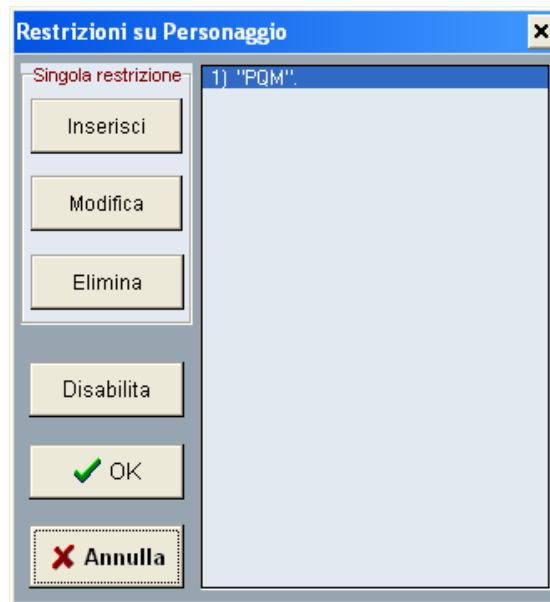



Fig. 11 - Finestra di gestione delle regole

1 Viene usata la dicitura *Personaggio* in quanto sono stati usati gli stessi codici dei testi teatrali, che permettono il recupero differenziato del testo di ciascun personaggio.

7. Esempi di consultazione del database

Anche la "Restrizione sul Riferimento" funziona nello stesso modo.

A questo punto con il pulsante  si può recuperare la ricerca precedente e si otterrà il risultato mostrato in figura 12.

In pratica solo i 3 contesti rimasti rispettano la regola data, mentre sono stati esclusi i 239 contesti che non contengono le parole date nel dispositivo.

Ogni contesto visualizzato nella finestra è a sua volta un link al testo completo della sentenza indicata nel riferimento, posto alla fine. L'ultimo numero della riga di contesto rappresenta la sequenza della riga all'interno del testo della sentenza.

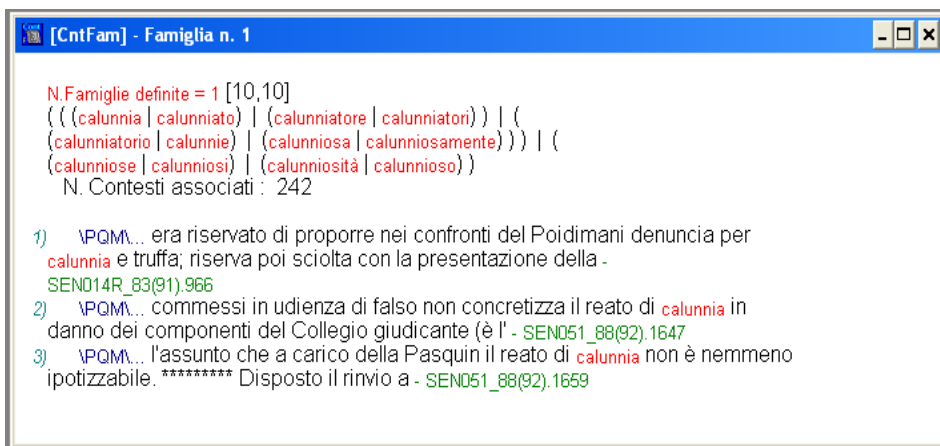
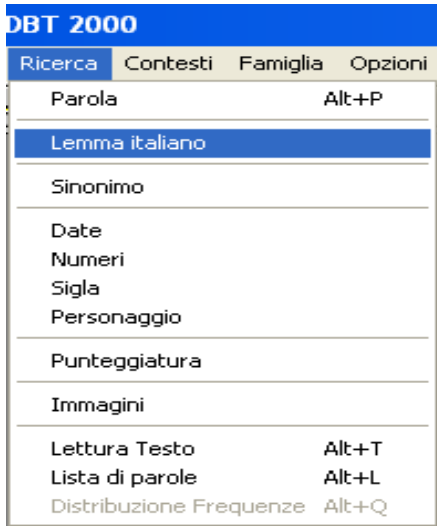


Fig. 12 - Risultato della richiesta con la Restrizione attiva

Il contenuto di ciascuna finestra può essere salvato con il click del pulsante destro del mouse posizionato all'interno della finestra medesima (vedi nella figura 15 la finestrella dei salvataggi possibili). Tra le opzioni segnaliamo il salvataggio su

7. Esempi di consultazione del database

file in Word, che viene attivato automaticamente.



Nella figura 13 sono elencate le possibilità di ricerca offerte dal DBT, fra le quali si può selezionare "Lemma italiano" per proporre, ad esempio un verbo all'infinito.

Fig. 13 - Canali di Ricerca offerti dall'interfaccia DBT

Indicando il verbo *infliggere* si ottiene la risposta mostrata nella figura 14, dove appaiono tutte le forme del verbo con le relative frequenze.

1)	infligge	366
2)	infliggendo	5
3)	infliggere	58
4)	infliggeva	57
5)	infliggevano	1
6)	inflitta	99
7)	inflitte	12
8)	inflitti	1
9)	inflitto	39

Fig. 14 - Risultato della ricerca del verbo infliggere

Utilizzando i pulsanti indicati nelle pagine precedenti, si possono ottenere tutti i contesti relativi (638).

7.1. Esempio di ricerca complessa

Con l'inserimento delle varie tipologie di codici nei testi di input, come precedentemente descritto, i dati vengono trasformati in informazioni che possono essere selezionate ed estratte tramite funzioni di ricerca del DBT; successivamente vengono elaborate e trattate come informazioni e sintetizzate con programmi di uso comune che ne permettono anche elaborazioni grafiche.

Ad esempio, al fine di individuare la tipologia delle sanzioni inflitte nell'anno 1995 e la loro frequenza di applicazione nell'anno medesimo, è stata utilizzata la seguente metodologia:

- si richiedono tutti i contesti in cui è stato usato il verbo *infliggere* (solitamente nella forma della terza persona singolare del presente indicativo);
- si imposta la restrizione sul riferimento, imponendo la condizione che l'intestazione delle unità logiche recuperate contenga la stringa alfanumerica (95), che individua i procedimenti depositati nell'anno 1995;
- si imposta la restrizione sulla sezione, limitando il recupero ai casi in cui *infliggere* sia presente nella parte di Sentenza con il marcatore PQM.

Dai contesti recuperati sono state estratte le parole che seguono il termine *infligge*, tramite l'ordinamento destro mostrato nella figura 15, ed i risultati sono stati poi organizzati nella tabella 3.

7. Esempi di consultazione del database

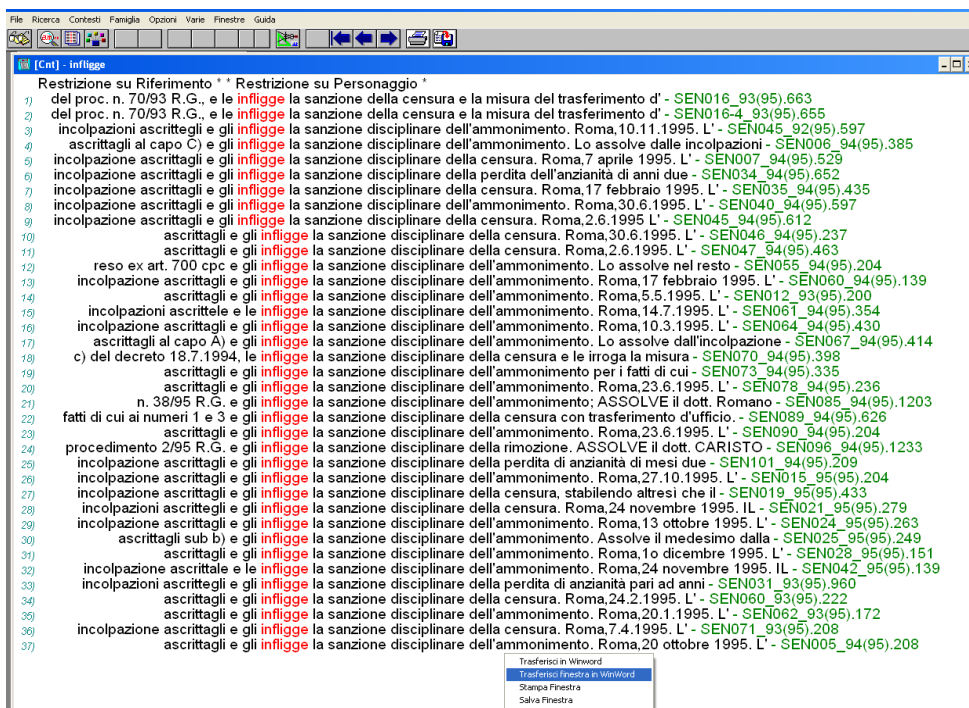


Fig. 15 - Elenco dei contesti del '95 in PQM (ordinamento destro)

CENSURA	AMMONIMENTO	PERDITA DELL'ANZIANITA'	RIMOZIONE
SEN016_93(95)	SEN073_94(95)	SEN031_93(95)	SEN078_94(95)
SEN016-4_93(95)	SEN045_92(95)	SEN101_94(95)	
SEN089_94(95)	SEN012_93(95)	SEN034_94(95)	
SEN070_94(95)	SEN062_93(95)		
SEN019_95(95)	SEN005_94(95)		
SEN060_93(95)	SEN006_94(95)		
SEN071_93(95)	SEN040_94(95)		
SEN007_94(95)	SEN055_94(95)		
SEN035_94(95)	SEN060_94(95)		
SEN045_94(95)	SEN061_94(95)		
SEN046_94(95)	SEN064_94(95)		
SEN047_94(95)	SEN067_94(95)		

7.Esempi di consultazione del database

CENSURA	AMMONIMENTO	PERDITA DELL'ANZIANITA'	RIMOZIONE
SEN021_95(95)	SEN078_94(95)		
	SEN090_94(95)		
	SEN015_95(95)		
	SEN024_95(95)		
	SEN025_95(95)		
	SEN028_95(95)		
	SEN042_95(95)		
	SEN085_94(95)		

Tab. 3 - Raggruppamento delle sentenze del '95 per contenuto

Il risultato risulta graficamente sintetizzato nella figura 16.

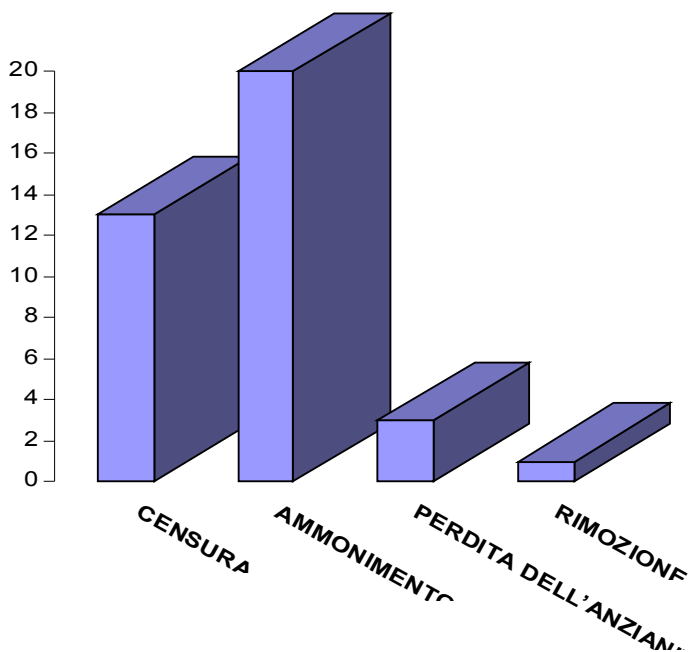




Fig. 16 - Raggruppamento delle sentenze del '95 per contenuto.

La ricerca combinata di più parole in compresenza può essere

7.Esempi di consultazione del database

ottenuta in più modi, dove in ognuno sono presenti automazioni che facilitano la richiesta manuale.

- 1) Tramite il menu Ricerca Parola, con il pulsante  nella cui finestra di dialogo è anche possibile inserire più di un elemento.
- 2) Con il menu Famiglia, pulsante , dopo aver richiesto almeno 2 parole singolarmente.
- 3) Sempre con il menu Famiglia, senza aver preventivamente richiesto nessuna parola, usando "Famiglia hard".

Il caso 1 si presenta come in figura 17, dove sono state indicate le due parole e automaticamente scatta la definizione di famiglia e la richiesta dei contesti.

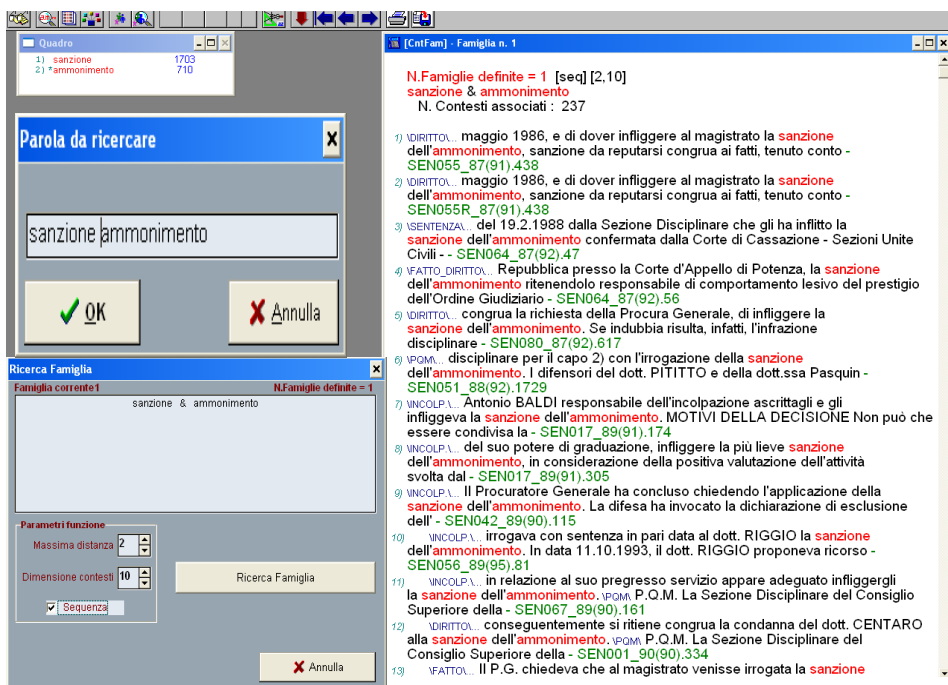


Fig. 17 - Richiesta di più parole e risultato della ricerca.

7.Esempi di consultazione del database

In questo modo si ottiene automaticamente la formazione della famiglia di ricerca ed è sufficiente indicare la distanza e spuntare la casella di rispetto della sequenza per ottenere i 237 casi in cui le due parole compaiono vicine (Distanza = 2) e nella sequenza indicata.

Nel caso 2, dopo aver richiesto separatamente le parole volute, si costruisce la famiglia di ricerca con il pulsante o con il menu "Famiglia - Definizione", che presenta l'interfaccia mostrato nella figura 18.

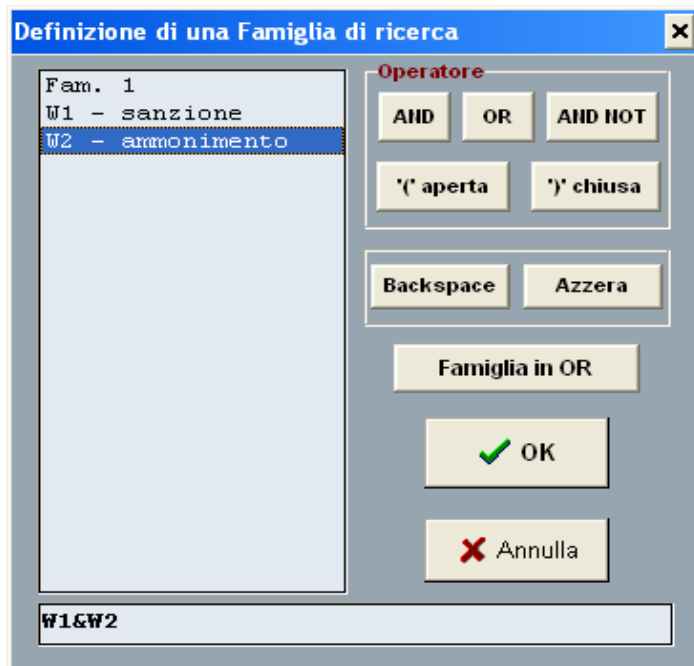


Fig. 18 - Definizione della Famiglia di Ricerca.

Nel riquadro a sinistra vengono elencate tutte le parole già richieste e con i pulsanti a destra e con il click del mouse si

7. Esempi di consultazione del database

definisce il gruppo di parole da ricercare congiuntamente, cioè W1 and W2, che vengono aggiunti automaticamente cliccando sulle parole della lista.

Nel caso 3 non c'è bisogno di fare selezioni preventive, ma si usa l'interfaccia mostrato nella figura 19 per indicare le richieste. Poiché si tratta di una funzione più complessa delle precedenti, è possibile anche combinare più richieste (tante quante sono le caselle predisposte) partendo anche dai lemmi e non solo dalle singole forme.

Definizione e ricerca di famiglia "hard" di parole

1)
L=infliggere

2)
L=sanzione

3)
L=disciplinare

4)

5)

Parametri funzione

Massima distanza 3

Dimensione contesti 10

Sequenza

OK

Annulla

Fig. 19 - Famiglia Hard: interfaccia e richieste

7.Esempi di consultazione del database

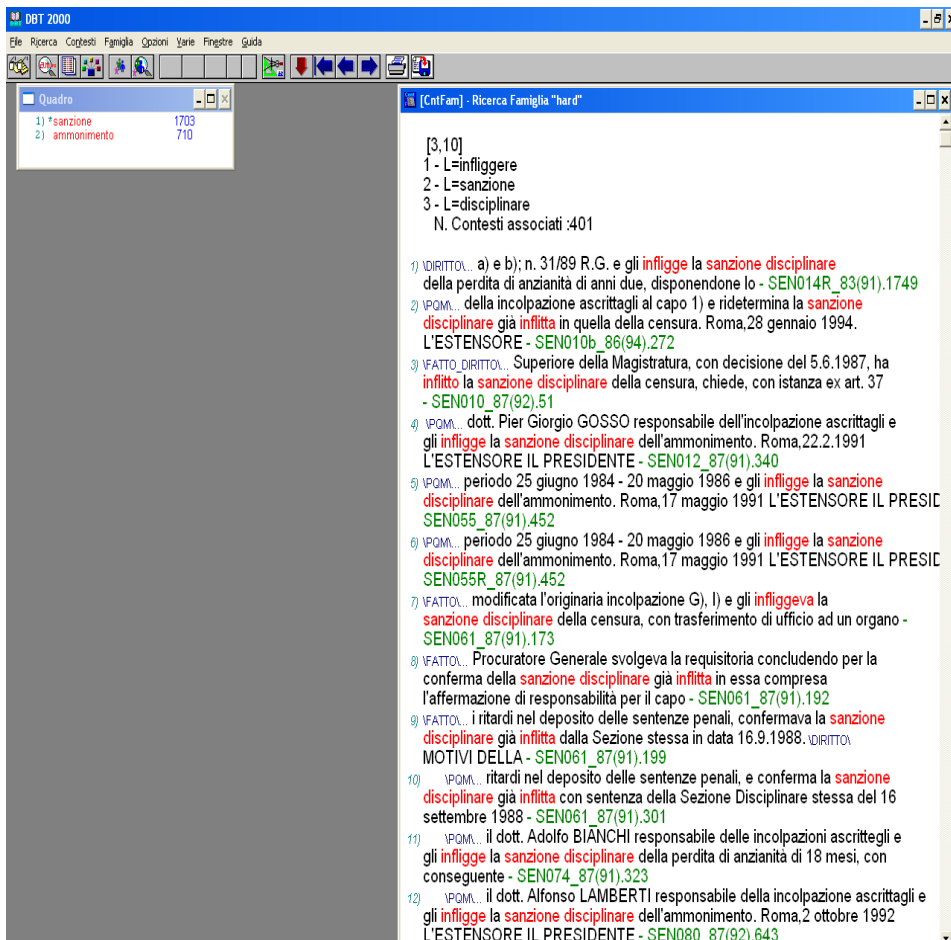


Fig. 20 - Famiglia Hard: risultati


Nella figura 20 appare il risultato della ricerca effettuata; come si può notare nel Quadro¹, non compaiono i lemmi *infliggere* o *disciplinare*, poiché non è necessario averli pre-selezionati, dato che la funzione "Famiglia hard" raggruppa in un unico

1 Il Quadro è la finestra riassuntiva di tutte le ricerche effettuate nel corso della sessione e in quanto tale si chiude solo all'uscita dal programma.

7. Esempi di consultazione del database

procedimento i comandi descritti nei casi 1 e 2. I contesti, risultato di ricerche complesse, possono essere utilizzati per ottenere analisi di tipo statistico. In Appendice si forniscono i grafici per anno e globale del periodo 1990-2004 in base all'esito.

7.2. Distribuzione delle frequenze

Per avere un quadro riassuntivo della presenza di ciascuna parola cercata, cioè presente nella finestra Quadro, è possibile selezionarla e, in presenza dell'asterisco che la segnala come parola corrente, premere il pulsante per  ottenere il risultato mostrato dalla figura 21.

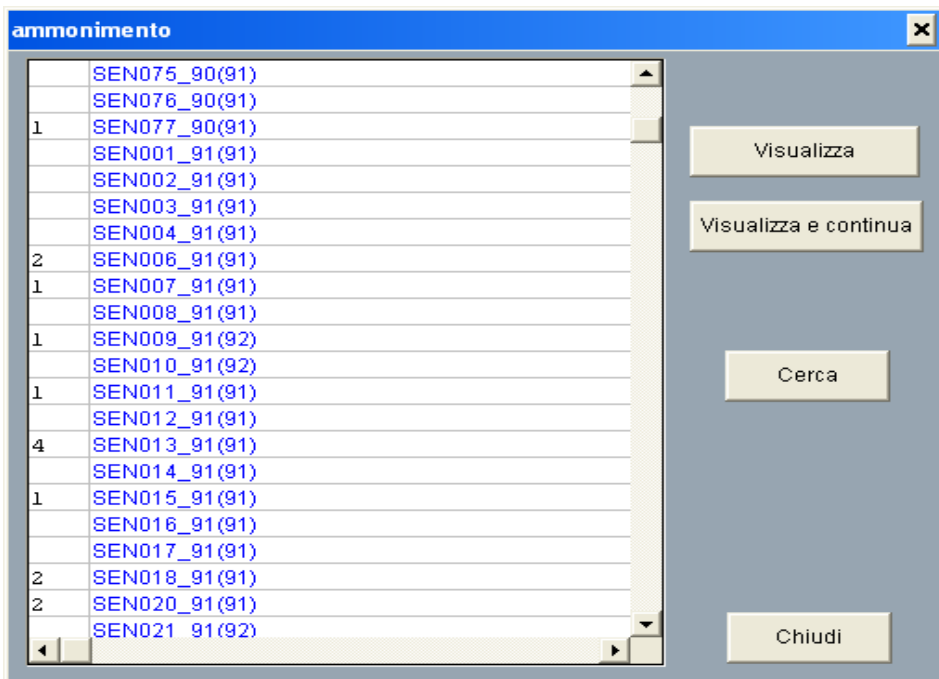


Fig. 21 - Risultato del pulsante di Distribuzione delle frequenze

7. Esempi di consultazione del database

Per visionare i contesti relativi è sufficiente utilizzare i pulsanti "Visualizza" oppure saltare alla sentenza desiderata con il pulsante "Cerca", che funziona anche per sottoinsieme di caratteri.

7.3. Cooccorrenze statistiche

Sotto il menù "Varie" si trova la funzione di "Co-Occorrenze Statistiche" che elabora una lista di parole che sono fortemente connesse ad una o più parole date.

Per illustrare tale funzione, sono state selezionate le forme del verbo *infliggere*, a cui è stata poi applicata la suddetta funzione usando i parametri che appaiono nella figura 22.

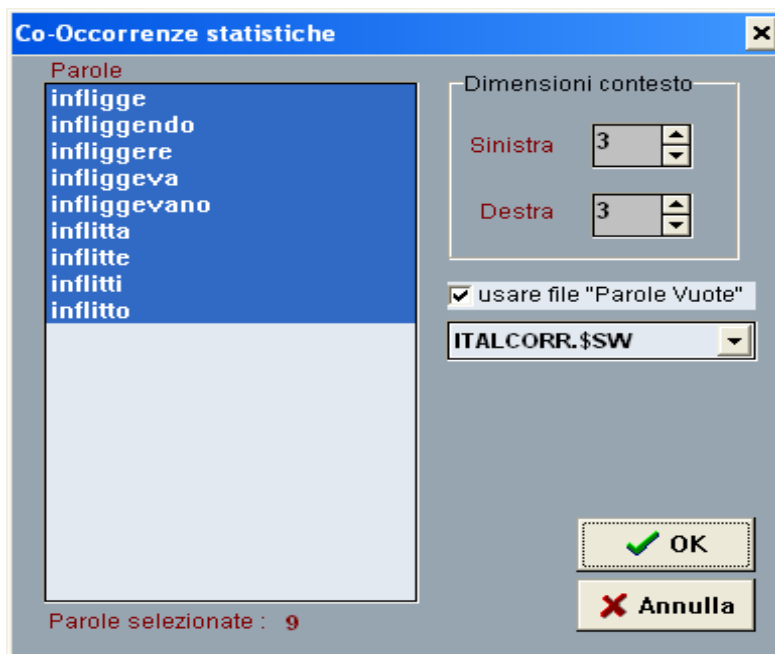


Fig. 22 - Finestra di dialogo per la selezione dei parametri

7. Esempi di consultazione del database

Nella figura 23 si può vedere il risultato di tale elaborazione. Come si può notare sono state estratte parole che hanno una connessione molto forte con quelle indicate.

Item	Frequency	Count	Word
1)	4	1	14.396
2)	9	116	8.708
3)	173	328	11.474
4)	19	33	11.600
5)	4	60	8.489
6)	46	133	10.865
7)	4	8	11.396
8)	8	8003	2.430
9)	4	623	5.113
10)	5	152	7.470
11)	4	549	5.296
12)	4	3437	2.649
13)	405	15921	7.100
14)	40	1	17.718
15)	5	341	6.305
16)	7	49	9.589
17)	8	10554	2.031
18)	8	2185	4.303
19)	4	3897	2.468
20)	6	855	5.242
21)	9	7326	2.727
22)	8	4814	3.163
23)	6	19905	0.700
24)	8	5168	3.061
25)	4	985	4.452
26)	4	125	7.431
27)	4	1311	4.040
28)	4	6005	1.844
29)	23	559	7.793
30)	5	1607	4.068
31)	18	1	16.566
32)	8	957	5.494
33)	505	1703	10.643
34)	9	79	9.263
35)	7	6129	2.622
36)	10	9600	2.489
37)	5	4901	2.489
38)	4	8318	1.374
39)	4	6687	1.689
40)	12	2120	4.931
41)	4	198	6.767

Fig. 23 - Risultato della funzione di Co-Occorrenze Statistiche

Tale elenco è stato ottenuto applicando la formula statistica, detta di *mutual information*, basata sulla frequenza assoluta nel testo delle parole e sulla loro presenza nell'immediato contesto di quelle selezionate in partenza, secondo le definizioni dei parametri iniziali, cioè che si trovino nell'ambito di 3 elementi prima e dopo. Con l'attivazione della lista di stopwords vengono

7.Esempi di consultazione del database

tolte dal computo tutte le forme grammaticali come articoli, congiunzioni, preposizioni, ecc. Ciascun elemento della lista può essere selezionato e vengono così visualizzati i relativi contesti.

Il calcolo delle Co-occorrenze può essere molto utile nella fase di recupero dei contenuti in quanto ci aiuta a individuare la mappa concettuale che sottende ai testi.

Nella figura 24 sono esemplificate le richieste per Sinonimo di 4 parole - *pena*, *ammenda*, *sanzione* e *condanna* - con i risultati ottenuti.

The screenshot shows the DBT 2000 software interface with four windows displaying synonym search results. Each window lists 14 synonyms with their corresponding frequency counts.

Word	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)	14)														
[F]-Sinonimo = pena	sanzione	1703	sanzioni	79	punizione	16	ammenda	28	sofferenza	23	sofferenze	5	preoccupazione	95	preoccupazioni	31	ansia	29	affanno	2	angoscia	4	afflizione	5	tortura	1	compassione	1
[F]-Sinonimo = sanzione	convalida	220	convalide	14	ratifica	11	approvazione	59	consensi	2	consenso	120	avallo	24	pena	559	pene	79	penale	4662	penali	1308	multa	80	ammenda	28	risarcimento	87
[F]-Sinonimo = ammenda	multa	80	penale	4662	penali	1308	contravvenzione	46	contravvenzioni	39	riparazione	19	riparazioni	7	risarcimento	87	scusa	12	scuse	8	perdono	7						
[F]-Sinonimo = condanna	pena	559	pene	79	sanzione	1703	sanzioni	79	disapprovazione	13	biasimo	6	deplorazione	1														

Fig. 24 - Risultato della Ricerca per Sinonimo

Dopo aver applicato la funzione di Co-Occorrenze Statistiche all'insieme risultato dalle precedenti ricerche, si ottiene il risultato mostrato nella figura 25.

7. Esempi di consultazione del database

The screenshot shows the DBT 2000 application window. The main area displays a table of co-occurrence statistics. The table has three columns: a number (likely a rank or ID), two columns of numbers (likely counts for different contexts), and a word. The words are color-coded: red for words with a frequency of 20 or more, and black for those with a frequency of 19 or less. The words listed include: infligge, pecuniaria, codice, pecuniarie, irrogazione, detentiva, ammonimento, accessoria, pregiudizialità, irrogazione, infliggeva, irrogare, inflitta, sanziona, sanzioni, disapprovazione, biasimo, deplorazione, contravvenzioni, multa, penale, penali, contravvenzione, riparazione, riparazioni, scusa, scuse, perdono, condizionale, inflitto, dibattimentali, settore, dibattimenti, minima, costituendo, equo, pendente, procedura, instaurato, perdita, illiceità, collegi, rimozione, iii, conclusosi, giudicato, terza, procedimento, lieve, and clanni.

The 'Co-Occorrenze statistiche' dialog box is open, showing a list of selected words: multa, ammenda, risarcimento, sanzione, sanzioni, disapprovazione, biasimo, deplorazione, contravvenzioni, multa, penale, penali, contravvenzione, contravvenzioni, riparazione, riparazioni, scusa, scuse, and perdono. The 'Dimensioni contesto' section shows 'Sinistra' and 'Destra' both set to 3. The 'usare file "Parole Vuote"' checkbox is checked. The dropdown menu shows 'ITALCORR.\$SW'. The 'OK' button is highlighted with a green checkmark, and the 'Annulla' button is highlighted with a red X. Below the list, it says 'Parole selezionate : 42'.

Fig. 25 - Risultato della funzione di Co-Occorrenze Statistiche

Da notare che il risultato è stato ridotto alle parole che avevano una frequenza di comparsa a salire da 20 volte in su con l'insieme dato, dove la somma totale delle frequenze è di 15.522.

7.4. Menu Indici vari

Oltre alla consultazione del testo, è possibile utilizzare il menù "Indici e Statistiche" che presenta le 5 voci indicate dalla figura 26. Ognuna di queste presenta una descrizione sommaria del risultato che si può ottenere. Con la voce 1, dopo la scelta del testo, viene visualizzato il menu che permette di scegliere il tipo

7. Esempi di consultazione del database



Fig. 26 - Menu "Indici Vari" del DBT

di lista, come mostrato nella figura 27. La schermata si presenta come una serie di schede dove, cliccando sull'intestazione, si ottiene la pagina per l'inserimento dei parametri richiesti. Il risultato del tipo di elaborazione richiesta viene salvato su file, di cui viene sempre indicato il nome a fine procedimento per la conferma da parte dell'utente. Segue l'elenco in ordine di frequenza decrescente, dove si può notare che in pratica le parole più frequenti fanno tutte parte del linguaggio tecnico di stesura delle sentenze. Ci sia permessa una annotazione: anche qui si rispettano le quote rosa reali, le dottoresse sono poco meno del 10% dei dottori.

36682 dott.	8155 data	6005 ordine	4677 perché
19905 magistrato	8003 aveva	5867 cassazione	4662 penale
15921 disciplinare	7969 merito	5796 degli	4499 stesso
10727 procuratore	7674 quale	5757 fatto	4492 funzioni
9600 sezione	7326 incolpato	5707 quanto	4423 atti
9584 presidente	7305 consiglio	5505 parte	4353 richiesta
8887 procedimento	7111 essere	5228 repubblica	4171 procura
8850 corte	6849 confronti	5168 magistratura	4124 fatti
8469 anche	6845 stato	5076 udienza	4048 aver
8387 art.	6794 tale	4889 giudice	4010 roma
8318 tribunale	6687 ufficio	4814 incolpazione	4009 dott.ssa
8207 come	6129 sentenza	4801 superiore	4001 carico

7. Esempi di consultazione del database

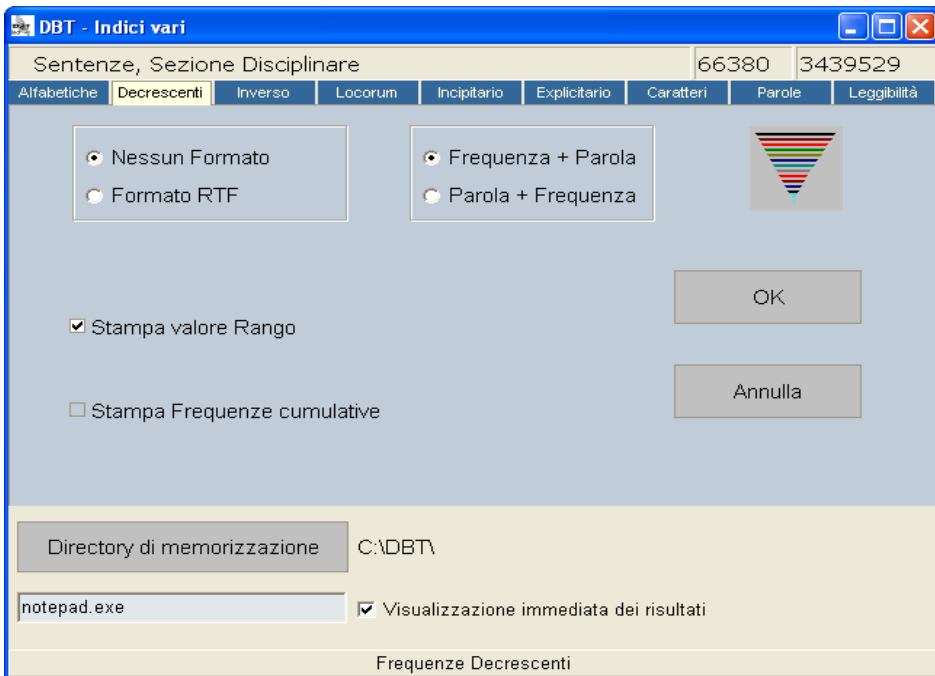


Fig. 27 - Menu "Decrescenti" all'interno di "Indici Vari"

Un ulteriore esempio di calcolo statistico è l'Indice di Leggibilità, che corrisponde all'ultima scheda a destra del menu "Indici vari" e che si visualizza nella figura 28.

Come spiegato dalla scheda, si tratta di una funzione che calcola il valore statistico di leggibilità di un testo secondo la formula di Flesh, originariamente riservata alla lingua inglese ma rivista e riadattata all'italiano.

I calcoli statistici vengono effettuati sulla base della quantità di frasi (da punto a punto) sulla quantità di parole delle stesse e sulla lunghezza delle parole usate. Si tratta quindi di un sistema puramente numerico che ha però il fondamento nella maggiore difficoltà di comprensione di frasi o parole più lunghe.

7. Esempi di consultazione del database

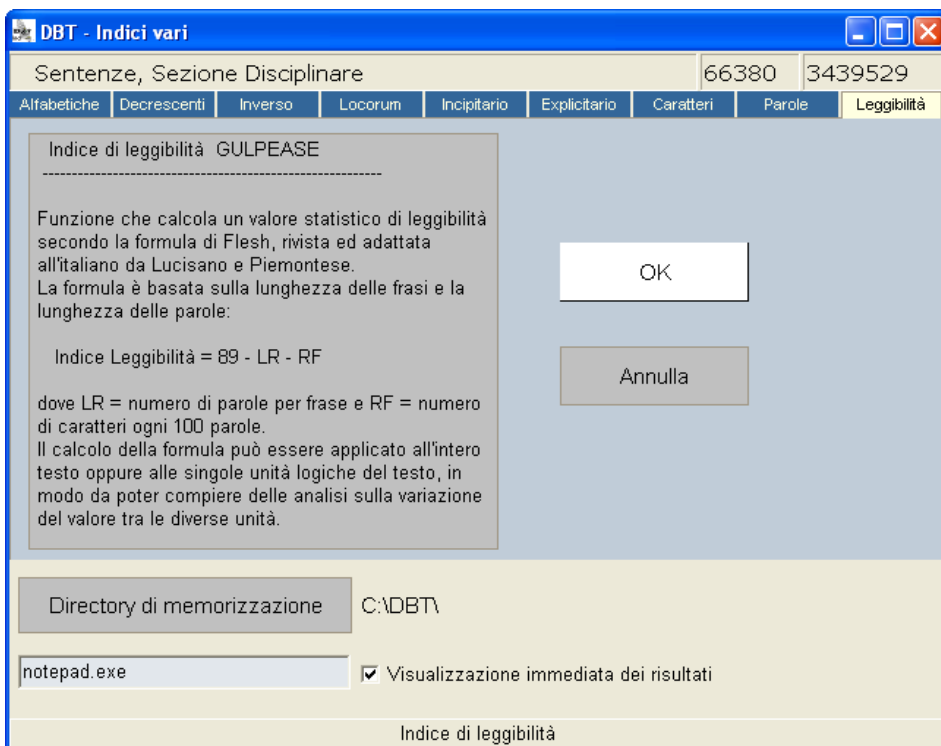


Fig. 28 - Menu "Leggibilità" all'interno di "Indici Vari"

Si elenca una parte del risultato ottenuto dall'archivio delle sentenze.

DBT - Sentenze, Sezione Disciplinare

Nr Frasi	Nr Parole	Nr Lettere	--LP--	--FR--	* Indice * *GULPEASE *	* Unità di testo *
62	2770	15832	571.552	2.238	* 38.560 *	SEN004_90 (90)
195	8958	50229	560.717	2.177	* 39.459 *	SEN005_90 (91)
83	4671	25345	542.603	1.777	* 40.070 *	SEN006_90 (91)
88	4327	23945	553.386	2.034	* 39.763 *	SEN010_90 (92)
62	2460	13518	549.512	2.520	* 41.610 *	SEN015_90 (90)
77	3879	21322	549.678	1.985	* 39.987 *	SEN016_90 (90)
59	3697	21022	568.623	1.596	* 36.925 *	SEN017_90 (91)
30	1509	8705	576.872	1.988	* 37.277 *	SEN019_90 (90)
32	1439	8162	567.199	2.224	* 38.951 *	SEN025_90 (90)
82	3663	20862	569.533	2.239	* 38.762 *	SEN026_90 (92)

7.5. Ricerca per Strutture Concettuali

Un'altra applicazione del menu "Indici e Statistiche", come riportato nella figura 26 alla scelta 5, consente una ricerca sul testo applicando delle strutture concettuali precostituite sulla base di termini significativi.

La struttura deve essere definita anteriormente all'uso del DBT, tramite un generico programma di gestione testi, compreso il blocco-note, purché recante come estensione la sigla RSI che permette al DBT di caricarla durante la funzione "Ricerca Strutture Concettuali". La sintassi per la stesura di tale struttura prevede alcune regole per la definizione dei livelli e l'inserimento delle regole. Trattandosi di questioni abbastanza complesse, dipendendo altresì dalla stretta collaborazione interdisciplinare fra esperti di entrambi i settori, ci si riserva di approfondire l'argomento, ritenendo importante di segnalarne le potenzialità.

Lo schema della struttura concettuale qui presentata è stato costruito traendo spunto dalle voci di classificazione utilizzate nel Quaderno N.143¹. La sottostruttura dettagliata è stata inserita solo in merito ad alcune voci, in quanto di più facile ed immediata realizzazione per non esperti della materia.

```
(1) --- VOCI DI CLASSIFICAZIONE
(2) ----- AZIONE DISCIPLINARE
(3) ----- CONDIZIONE
(4) ----- [*]APPARTENENZA DELL'INCOLPATO ALL'ORDINE GIUDIZIARIO {344}
                CESSAZIONE & APPARTENENZA {344}
(4) ----- ESERCIZIO (MODI E TEMPI) IN GENERALE
(4) ----- (IN)SINDACABILITA' DEI PROVVEDIMENTI GIUDIZIARI
(4) ----- LIMITI
(2) ----- CONDOTTA
(3) ----- DOVERI DEL MAGISTRATO CORRETTEZZA
(4) ----- IN GENERALE
```

1 Quaderni del C.S.M. (www.csm.it/pages/quaderni.html).

7.Esempi di consultazione del database

```
(4)----- CORRETTEZZA NEI CONFRONTI DEI COLLEGHI
(4)----- CORRETTEZZA NEI CONFRONTI DEGLI ORGANI ISTITUZIONALI
(4)----- CORRETTEZZA NEI CONFRONTI DEI PRIVATI
(4)----- CONDOTTA PRIVATA E SOCIALE
(3)----- DOVERI DEL MAGISTRATO DILIGENZA
(4)----- IN GENERALE
(4)----- [*]VIGILANZA {205}
          VIGILANZA {205}
(4)----- INOSSERVANZA DI NORME DI PROCEDURA
(3)----- DOVERI DEL MAGISTRATO - IMPARZIALITA'
(4)----- ASTENSIONE
(4)----- EQUIDISTANZA DEL GIUDICE DALLE PARTI
(3)----- DOVERI DEL MAGISTRATO - OPEROSITA'
(4)----- [*]RITARDO O OMISSIONE NELL'ADOZIONE DI PROVVEDIMENTI {305}
          RITARDO & DEPOSITO {305}
(4)----- GIUDICE CIVILE
(4)----- GIUDICE PENALE
(4)----- PUBBLICO MINISTERO
(3)----- DOVERI DEL MAGISTRATO - RISERBO
(2)----- PROCEDIMENTO DISCIPLINARE
(3)----- IN GENERALE
(3)----- RAPPORTI CON PRECEDENTE PROCEDIMENTO DISCIPLINARE
(3)----- SENTENZA DI NON FARSI LUOGO A RINVIO A DIBATTIMENTO
(3)----- GIUDIZIO DI RINVIO
(2)----- [*]SANZIONI {1296}
          FUNZIONI & STIPENDIO {181}
          PERDITA & ANZIANITA {123}
          SANZIONE & AMMONIMENTO {566}
          SANZIONE & CENSURA {368}
          RIMOZIONE {58}
(2)----- SOSPENSIONE CAUTELARE
(3)----- [*]APPLICAZIONE {4}
          ACCOLTA & SOSPENSIONE {4}
(3)----- [*]REVOCA {54}
          REVOCA & SOSPENSIONE {54}
(2)----- REVISIONE DEL GIUDIZIO DISCIPLINARE
```

Nello schema è stato definito uno specifico tipo di *Sanzioni* mediante coppie di parole co-occorrenti, secondo criteri di massima distanza e tra parentesi graffe viene indicato il numero dei contesti associati. Allo stesso modo sotto Doveri - Ritardo il concetto è stato definito tramite la coppia RITARDO & DEPOSITO, che risulta avere la frequenza di 305 volte. In questo modo tutte le forme rintracciate saranno evidenziate dal sistema e sarà possibile risalire direttamente al contesto. I risultati possono anche essere stampati su file di output.

E' possibile recuperare più parole cooccorrenti utilizzando il mascheramento per individuare sia le forme plurali che singolari di una parola così come le forme di un verbo. Ad esempio con la notazione DEPOSIT* & RITARD* si reperiscono i contesti di *depositato, depositati, depositare, etc.* e *ritardo, ritardare,*

7.Esempi di consultazione del database

ritardato, etc. Questa opzione può essere utilizzata ad esempio per rilevare forme alternative che esprimono uno stesso concetto, tuttavia il numero dei contesti generati in output può aumentare notevolmente. Alcuni esempi dei contesti recuperati sono illustrati in figura 29, in cui ognuno è un link al testo.

- 187) \INCOLP.\... del resto, scorrendo l'elenco delle sentenze **depositate** in **ritardo**, dalla sommaria descrizione dell'oggetto della causa si rileva - [SEN013 93\(93\).138](#)
- 188) \INCOLP.\... Maria FOSCHINI, incorsi in gravi e reiterati **ritardi** nel **deposito** delle sentenze civili. In particolare al dott. TOTI veniva - [SEN050 93\(94\).77](#)
- 189) \INCOLP.\... che rispecchiavano l'interesse vivo delle parti, **ritardando** il **deposito** di provvedimenti relativi a cessazione della materia del contendere, - [SEN050 93\(94\).171](#)
- 190) \INCOLP.\... in sede disciplinare degli oggettivi ed innegabili gravi **ritardi** nel **deposito** dei provvedimenti giudiziari a lui affidati, sia nella Sezione - [SEN050 93\(94\).258](#)
- 191) \INCOLP.\... già sottoposto a procedimento disciplinare per gravissimi **ritardi** nel **deposito** di sentenze ed ordinanze (con riferimento alla data del - [SEN007 94\(95\).48](#)
- 192) \INCOLP.\... Federico GOVERNATORI, che da anni presentava gravi **ritardi** nel **deposito** di numerose sentenza e ordinanze, nonostante i reiterati e - [SEN007 94\(95\).58](#)
- 193) \INCOLP.\... già sottoposto a procedimento disciplinare per gravissimi **ritardi** nel **deposito** di sentenze e ordinanze (con riferimento alla data del - [SEN007 94\(95\).65](#)
- 194) \INCOLP.\... incolpato lo ha onestamente riconosciuto, che i **ritardi** nel **deposito** delle sentenze e delle ordinanze elencate negli allegati A e - [SEN007 94\(95\).179](#)
- 195) \INCOLP.\... su diversi fronti, che ha certamente accumulato **ritardi** nel **deposito** delle sentenze, ma non a causa di disimpegno o - [SEN007 94\(95\).244](#)

7. Esempi di consultazione del database

7.6. Esempi di consultazione del Corpus

Gli esempi finora mostrati riguardano solo l'archivio delle Sentenze; infatti è stato creato anche l'archivio delle Ordinanze con le stesse modalità. Come abbiamo già detto, è stato creato un Corpus con entrambi gli archivi, in modo da poter essere consultati contemporaneamente con una unica richiesta. Nella figura 4 a pag. 21 si può vedere come nella stessa videata iniziale sia compresa la scelta 2 per la consultazione del Corpus. Dopo la scelta del Corpus, l'interfaccia nella figura 29 è simile a quello dell'interrogazione del testo singolo.

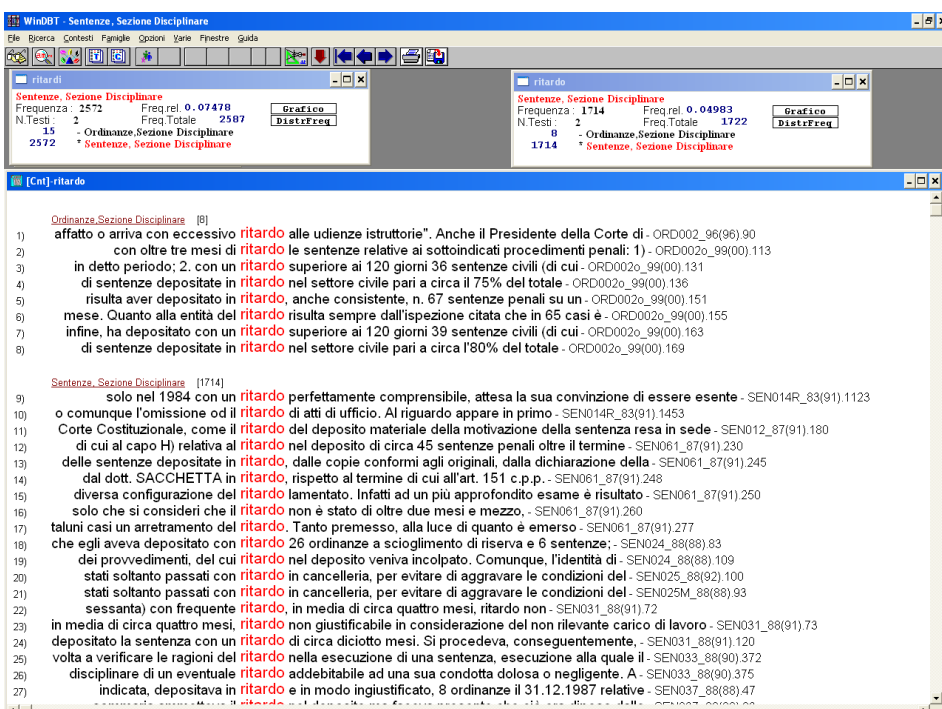



Fig. 30 - Consultazione del Corpus delle Sentenze e delle Ordinanze

Nell'esempio precedente si può notare un pulsante in più 

7. Esempi di consultazione del database

che serve per avere la mappa della presenza della parola suddivisa fra i testi che compongono il corpus; nelle 2 finestre piccole in alto nella figura 30 si può vedere il risultato della richiesta delle due parole, *ritardo* e *ritardi*, con le relative frequenze nelle Ordinanze e nelle Sentenze. Da queste è possibile accedere alla funzione "Grafico" che visualizza graficamente le percentuali di presenza delle parole selezionate in relazione all'estensione dei testi, calcolata in numero di parole, come mostrato nella figura 31.

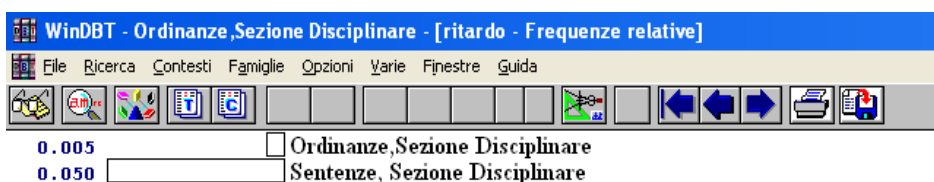


Fig. 31 - Consultazione Corpus: Funzione Grafico

A parte questa opzione in più e la possibilità di avere i contesti da più archivi contemporaneamente, le altre funzioni hanno le stesse caratteristiche già presentate nelle pagine anteriori.

8. Altri tipi di elaborazioni

Nei paragrafi precedenti sono state descritte ed esemplificate le funzioni del DBT, così come sono disponibili nel pacchetto che viene distribuito nella sua ultima versione standard.

Poiché l'ILC è un istituto di ricerca, vi sono molte altre funzioni a corollario che ci permettono di utilizzare la galassia DBT e il PiSystem¹ come componenti in fieri e quindi in fase di sviluppo.

1 <http://www.ilc.cnr.it/pisystem/procedure/index.html>.

8. Altri tipi di elaborazioni

Si cita ad esempio il Lemmatizzatore o Tagger, che è in parte integrato nel sistema di interrogazione, per quanto riguarda la possibilità di sviluppare e riconoscere, a livello morfologico, le forme dei lemmi.

E' inoltre possibile l'applicazione all'archivio in DBT della fase di riconoscimento delle funzioni svolte nella frase dalle parole (pos-tagging). In questa ottica sono state utilizzate le procedure, tuttora in fase di sviluppo e/o raffinamento, del *Linguistic Miner*¹ (LM), che è un sistema integrato per la creazione di grandi archivi testuali di riferimento per la lingua italiana e per l'estrazione sempre più veloce mirante alla sintesi qualificante di conoscenza linguistica da grandi quantità di dati.

8.1. Ricerca di schemi linguistici

Nella sperimentazione del LM sono stati utilizzati i testi indicizzati tramite la funzione di tagging automatico che assegna a ciascuna parola il suo lemma, la categoria grammaticale e la sua funzione.

A seguito della lemmatizzazione sono state applicate ai testi alcune regole di ricerca di schemi linguistici. Praticamente sono state estratte le coppie di sostantivi collegate da preposizione che seguono tutte le forme verbo *violare* e/o quelle del sostantivo *violazione*.

Lo studio è stato effettuato sia sull'archivio delle ordinanze che su quello delle sentenze ma ha prodotto risultati significativi

1 Picchi, E., Ceccotti, M.L., Cucurullo, S., Fiorentini, G., Sassi, M., Sassolini, E., (2003) *Linguistic Miner*, in Atti del Congresso Annuale AICA, Trento, 15-17 Settembre 2003.

8. Altri tipi di elaborazioni

solo nel secondo caso. Dai risultati ottenuti sono state generate due tabelle. Nella prima tabella sono elencati per ogni sentenza i 2 sostantivi che seguono le forme di *violare* e/o di *violazione*; nella seconda tabella le righe sono raggruppate secondo il primo lemma, con il sottototale per ciascun gruppo.

SENTENZA	LEMMA_UNO	LEMMA_DUE	tot
OSEN001_01	DOVERE	LABORIOSITA'	3
OSEN001_P_98	DOVERE	CORRETTEZZA	1
OSEN002_00	DOVERE	LABORIOSITA'	1
OSEN002_03	DOVERE	CORRETTEZZA	1
OSEN002_04	DOVERE	DILIGENZA	2
OSEN002_92	DOVERE	UFFICIO	1
OSEN002_97	DOVERE	CORRETTEZZA	2
		LEALTA'	1
OSEN003_00	DOVERE	LABORIOSITA'	1
OSEN003_99E9E053_00	DOVERE	CORRETTEZZA	6
OSEN004_01	DOVERE	DILIGENZA	1
OSEN004_03	DOVERE	CORRETTEZZA	1
OSEN004_93	DOVERE	DILIGENZA	2
OSEN004ROI_00	DOVERE	LABORIOSITA'	1
OSEN005_00	DOVERE	LABORIOSITA'	2
OSEN005_01	DOVERE	LABORIOSITA'	1
OSEN005_04	DOVERE	CORRETTEZZA	2
	OBBLIGO	DILIGENZA	1
OSEN005_97	DOVERE	CORRETTEZZA	2
OSEN005_98	DOVERE	CORRETTEZZA	2
OSEN005E22_03	DOVERE	DILIGENZA	2
OSEN006_00	DOVERE	LABORIOSITA'	1
OSEN006_02NCNC	PARAMETRO	LABORIOSITA'	1
OSEN006_94	DIRITTO	DIFESA	2
	DOVERE	DILIGENZA	1

<i>Lemma_uno</i>	<i>Lemma_due</i>	<i>Totale</i>
DOVERE	ASTENSIONE	16
	AVVEDUTEZZA	1
	COLLABORAZIONE	2
	COMPORTEMENTO	2
	COMPOSTEZZA	1

8. Altri tipi di elaborazioni

<i>Lemma_uno</i>	<i>Lemma_due</i>	<i>Totale</i>
	CONFRONTO	1
	CONTINENZA	1
	CORRETTEZZA	297
	CUSTODIA	1
	DILIGENZA	258
	DIRIGENZA	2
	DISTRIBUZIONE	1
	EQUILIBRIO	2
	FUNZIONE	3
	GIUDICE	1
	IMPARZIALITA'	22
	INFORMAZIONE	1
	LABORIOSITA'	66
	LEALTA'	20
	LESIONE	1
	MAGISTRATO	16
	MINISTERO	2
	NEGHISSITA'	1
	OPEROSITA'	15
	OSSERVANZA	1
	PARTE	2
	PARTECIPAZIONE	1
	PRECEDENZA	1
	PROBITA'	1
	PRUDENZA	3
	PUNTUALITA'	1
	RIPARTIZIONE	1
	RISERBO	28
	RISERVATEZZA	31
	RISPETTO	2
	SEGRETEZZA	3
	SOGGEZIONE	9
	STATO	3
	STRETTA	1
	UFFICIO	38
	VERIDICITA'	1
	VIGILANZA	5
DOVERE Totale		872
OBBLIGO	ASSISTENZA	1
	ASTENSIONE	7
	CORRETTEZZA	8
	DECORO	1
	DENUNZIA	1

8. Altri tipi di elaborazioni

<i>Lemma_uno</i>	<i>Lemma_due</i>	<i>Totale</i>
	DEPOSITO	4
	DILIGENZA	17
	LEGGE	2
	PARTICOLARE	1
	PERMANENZA	1
	RESIDENZA	6
	RISERBO	1
	RISERVATEZZA	1
	RISPETTO	2
OBBLIGO Totale		53

Nella figura 32 viene mostrata una sintesi grafica della presenza nel database delle Sentenze della *violazione del dovere di ...*

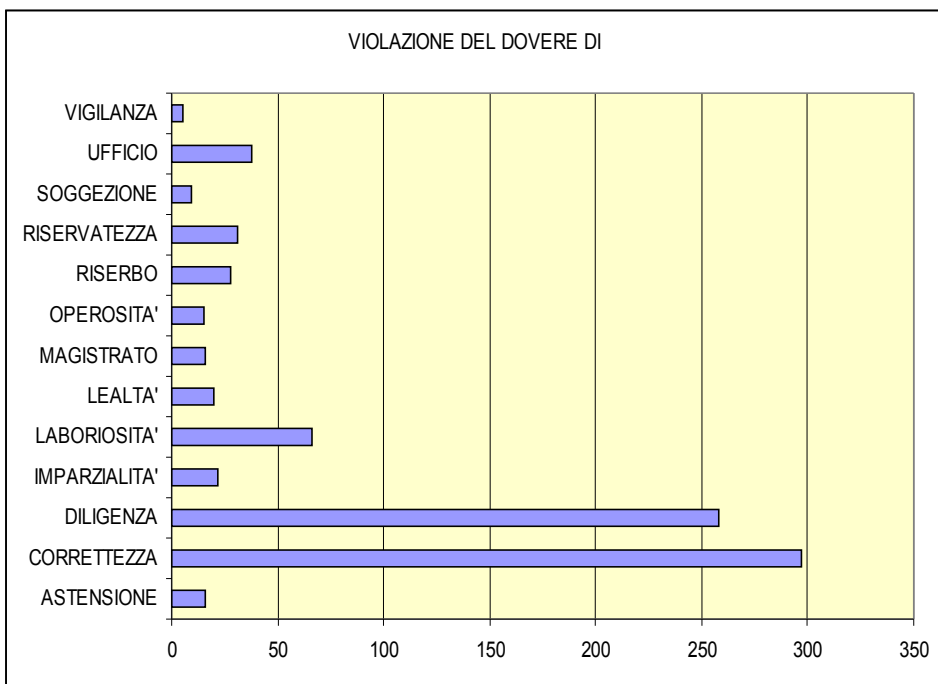


Fig. 32 - Grafico sulla violazione del dovere di ...

8.2. Sperimentazione dell'anonimazione di testi

E' stata sperimentata l'anonimizzazione dei testi dell'archivio delle Ordinanze della Sezione Disciplinare del CSM, in modo da rendere disponibile il database privato dei dati personali.

Il riconoscimento dei Nomi Propri nei testi avviene tramite l'utilizzo di un componente del progetto Linguistic Miner, descritto nel paragrafo 7.

In particolare ne fa parte il programma *CercaFenomeni* che, lavorando sull'archivio, individua nel contesto i Nomi Propri.

Questa operazione comporta varie difficoltà poiché tali entità possono essere costituite dal solo nome, dal solo cognome o da entrambi in vario ordine o abbinati alle sole iniziali. Una volta individuate tali entità in base anche ad un database esistente, che vengono considerate come un tutto unico, sono memorizzate con alcune informazioni aggiuntive in un file di output.

E' possibile intervenire manualmente sulla lista generata per apportare modifiche come inserimento o esclusione di elementi. Successivamente viene effettuata la riconversione dei documenti generando nuovi archivi, in cui ciascun elemento appartenente alla lista precedente viene marcato all'interno del testo. La fase finale prevede l'anonimizzazione dei nomi codificati.

Nel caso specifico dei procedimenti disciplinari le entità marcate sono sostituite da una stringa alfanumerica composta di 2 caratteri tra *, in modo tale che nel contesto di una stessa Ordinanza ogni occorrenza del nome e/o cognome della medesima persona sia sostituita con uguale sigla. Il software che effettua la riconversione dei nomi genera contemporaneamente un file di output utilizzabile in fase di controllo, in esso per ogni

8. Altri tipi di elaborazioni

Ordinanza sono elencate tutte le occorrenze dei Nomi Propri associate alla stringa che li sostituisce nel testo anonimato.

Di seguito riportiamo a titolo di esempio il testo di un Ordinanza della Sezione disciplinare del CSM in cui è stata realizzata l'anonimazione dei Nomi Propri.

La verificata dei risultati della prova effettuata sull'archivio delle Ordinanze ha evidenziato alcune problematiche: difficoltà nel distinguere i nomi di luoghi da quelli di persona ad esempio nel contesto di una determinata Ordinanza il nome di regione composto Friuli Venezia Giulia non viene riconosciuto come tale e pertanto convertito in Friuli Venezia *B1*. Una soluzione a tale inconveniente è essere quella di utilizzare un dizionario dei nomi geografici, che è in fase di costruzione all'interno del progetto LM. Tuttavia in alcuni casi permane l'ambiguità, ad esempio Rossano si riferisce talvolta ad una località, talvolta costituisce nome di persona non associato al relativo cognome o nome e tali casi possono essere risolti solamente valutando il contesto.

ORD3_82-21_83(96)

N. 69/96 Reg. Dep.

REPUBBLICA ITALIANA

LA SEZIONE DISCIPLINARE

del Consiglio Superiore della Magistratura

Composta dai Signori:

Prof. *A1* -Vice Presidente del Consiglio

Superiore della Magistratura

PRESIDENTE

Avv. *A2* -Componente eletto dal Parlamento

Prof. *A3* -Componente eletto dal Parlamento

dott. *A4* -Magistrato di Cassazione

dott. *A5* -Magistrato di merito

dott. *A6* -Magistrato di merito

dott. *A7* -Magistrato di merito

dott. *A8* -Magistrato di merito

dott. *A9* -Magistrato di merito

COMPONENTI

ha pronunciato in Camera di Consiglio la seguente

\ORDINANZA\ ORDINANZA

nei procedimenti disciplinari nn. 382 e 21/83 del registro generale a carico

del dott. *B1*,

8. Altri tipi di elaborazioni

(nato a Potenza il 3.11.1927)

già presidente di sezione del Tribunale di Santa Maria Capua Vetere.

Con istanza in data 23 giugno 1995 il dott. *B1*, già presidente di sezione del Tribunale di S. Maria Capua Vetere ha chiesto al Consiglio Superiore della Magistratura: a) di annullare ex art. 4 comma 2 legge sul contenzioso amministrativo le sanzioni disciplinari adottate nei suoi confronti con la sentenza disciplinare del 13.4.1984 e precisamente quelle adottate con D.P.R. 29.10.1985 registr. il 6.2.1987 , e con decreto di trasferimento d'ufficio 21.12.1987 e 29.5.1989 registr. il 15.11.1989); b) annullare il decreto di nomina del dott. *B2* a Presidente del Tribunale di S. Maria Capua Vetere del maggio 1989 ed il decreto di decadenza dall'impiego 7.8.1989 registr. il 30.8.1990 .

Il Procuratore Generale presso la Corte di Cassazione ha presentato le sue conclusioni in merito alla suindicata istanza, chiedendo che la stessa, se interpretata come istanza di revisione del procedimento disciplinare, sia dichiarata inammissibile; se interpretata come istanza di condono sia rigettata. Successivamente (7 e 10 novembre 1995) il dott. *B1* ha chiesto l'applicazione del condono previsto dalla legge 20.5.1986 n. 198 alla sanzione disciplinare dell'ammonizione inflittagli, in via definitiva, con sentenza 15.7.1978 .

In ordine a tale istanza il Procuratore Generale in data 15.12.1995 ha concluso, chiedendone l'inammissibilità - sia perché il rapporto di servizio del dott. *B1*, fondamento della giurisdizione della Sezione Disciplinare, è cessato dal 7 febbraio 1989 , sia perché una precedente istanza del dott. *B1* - di identico contenuto - è stata rigettata dalla Sezione Disciplinare con ordinanza del 30.9.1988 , non impugnata per Cassazione.

Con telegrammi in data 21.12.1995 e 10.1.1996 il *B1* ha insistito per l'annullamento delle sanzioni disciplinari e con memoria del 21.2.1996 ha sottolineato come erroneamente la seconda commissione referente, non avesse subito trasmesso alla Sezione Disciplinare la sua istanza di annullamento delle sanzioni disciplinari, presentata unitamente a quella di annullamento della decadenza dall'impiego dichiarata con D.P.R. 7.8.1989 , pur essendo la prima questione preliminare rispetto alla seconda.

Infine con memoria 11.5.1996 il *B1* ha insistito per l'accoglimento delle istanze presentate.

Tutto ciò premesso, va preliminarmente osservato che, in ordine alla richiesta di annullamento del decreto presidenziale con il quale il dott. *B1* è stato dichiarato decaduto dall'impiego, il C.S.M. , in conformità alla proposta formulata dalla seconda commissione referente, con delibera in data 25 maggio 1995 , ha rigettato la suddetta istanza, affermando che l'annullamento d'ufficio di atti amministrativi illegittimi è provvedimento essenzialmente discrezionale, dovendosi ponderare le ragioni di pubblico interesse, diverse dal mero ripristino della legalità, e che nella specie non vi fossero motivi validi per l'esercizio del potere di autotutela, tenuto conto del ridotto periodo di servizio per il quale potrebbe essere utilizzato il dott. *B1* e dei precedenti disciplinari dello stesso.

Sicché sul punto il Consiglio Superiore ha deliberato e qualsiasi rimostranza al riguardo può essere fatta valere solo con gli ordinari mezzi di impugnazione previsti per legge.

Passando all'esame dell'istanza di annullamento delle sanzioni disciplinari, si deve anzitutto rilevare che erroneamente l'istante richiama

8. Altri tipi di elaborazioni

l' art. 4 sul contenzioso amministrativo, posto che per i provvedimenti in materia disciplinare l' art. 17 legge n. 195/1958 è previsto unicamente il ricorso alle Sezioni Unite della Corte di Cassazione, stante la natura giurisdizionale delle decisioni adottate dalla sezione disciplinare del C.S.M. (cfr. Sez. Un. Ci v. 1.7.1992 n. 8083).

Per completezza si deve precisare che la sentenza disciplinare 13 aprile 1984 è passata in giudicato, in quanto confermata dalle Sezioni Unite Civili con sentenza n. 4822 del 28 luglio 1986 .

Al riguardo va, altresì, osservato che il passaggio in giudicato della sentenza impugnata in Cassazione coincide con la data di deposito della stessa e non, come erroneamente ritenuto dall'istante, con quella di deposito della sentenza pronunciata sul ricorso per revocazione.

Siffatta tesi interpretativa trova conforto non solo nella dottrina più attenta, ma è consacrata ora legislativamente dall' art. 391 bis terzo comma c.p.p. ("la pendenza del termine per la revocazione della sentenza della Corte di Cassazione non impedisce il passaggio in giudicato della sentenza impugnata con ricorso per Cassazione respinto").

Va, infine, osservato che anche il ricorso per revocazione è stato dichiarato inammissibile dalle Sezioni Unite Civili con sentenza 7 luglio 1988 n. 4507 e che soltanto dopo tale provvedimento giurisdizionale è stata chiesta il 16.9.1988 l'esecuzione della sanzione disciplinare da parte del Presidente della Corte d'Appello di Napoli (cfr. sentenza TAR Campania 10/7 - 24.11.1991).

Sulla base di tutti questi rilievi, non può non concludersi per l'assoluta inammissibilità dell'istanza relativa al richiesto "annullamento delle sanzioni disciplinari", sia perché trattasi di provvedimento non di competenza di questa Sezione disciplinare, sia perché non può neppure essere interpretata come istanza di revisione del procedimento disciplinare, non essendo stati allegati elementi nuovi, siccome disposto dall' art. 37 R.D.Lgs. 31.5.1946 n. 511 ,sia perché non può essere interpretata come richiesta di condono, trattandosi di infrazioni disciplinari riguardanti fatti del giugno 1982 , e quindi successivi al 31.12.1979 , data ultima prevista dalla legge 20.5.1986 n. 198 sul condono.

Del pari è inammissibile l'istanza del dott. *B1* di applicazione del condono alla sanzione disciplinare dell'ammonizione inflitta con sentenza 15.7.1978 e ciò in quanto una precedente istanza, di identico contenuto, è stata rigettata dalla Sezione disciplinare con ordinanza 30.9.1988 , divenuta definitiva, non essendo stata impugnata per Cassazione.

\PQM\ P.Q.M.

La Sezione Disciplinare del Consiglio Superiore della Magistratura;
Su conformi conclusioni del Procuratore Generale presso la Corte di Cassazione;

DICHIARA

inammissibili le istanze presentate dal dott. *B1*.

Roma, 17.5.1996

L'ESTENSORE

IL PRESIDENTE

IL MAGISTRATO SEGRETARIO

9. Presentazione del Database

Per la diffusione sul piano divulgativo dei risultati prodotti con il lavoro svolto nella collaborazione all'attività di ricerca nell'ambito del progetto "Tecnologie dell'informazione e della comunicazione per la giustizia" è stata realizzata una presentazione utilizzando il software applicativo Microsoft Power Point. La presentazione risulta composta da una serie di diapositive nelle quali si affrontano in maniera concisa le seguenti tematiche:

- La fase di codifica dei testi che se correttamente e adeguatamente effettuata permette l'utilizzo di specifiche funzionalità del sistema DBT in aggiunta a quelle applicabili in seguito alla codifica base
- L'indicizzazione dei file trattati che ha prodotto i due archivi delle Sentenze e delle Ordinanze della Sezione Disciplinare del CSM
- La consultazione dei documenti inseriti nei data base testuali attraverso l'interfaccia DBT illustrata tramite alcuni esempi significativi
- La possibilità di creare indici statistici all'interno di files di output utilizzabili con un qualunque programma di videoscrittura (cooccorrenze statistiche, leggibilità)
- Il calcolo della durata media dei procedimenti disciplinari componenti l'archivio delle Sentenze della Sezione Disciplinare del CSM
- La fase di lemmatizzazione dei testi degli archivi

9. Presentazione del Database

indicizzati in seguito alla quale è possibile applicare regole di ricerca linguistica

- Le problematiche inerenti l'anonimizzazione dei procedimenti disciplinari e le prove effettuate in merito a questa tematica sull'archivio delle Ordinanze.

10. Sviluppi futuri

L'archivio dei provvedimenti descritto è di per sé una banca dati con cui monitorare più di un decennio di lavoro della sezione disciplinare. Nello stesso tempo costituisce un nucleo da integrare sia con gli anni precedenti, dove possono essere d'aiuto le pubblicazioni già effettuate, sia con l'incremento dei provvedimenti che man mano vengono emessi. Il lavoro di analisi dei dati è già fatto in buona parte e si tratta solo di ingegnerizzare i nuovi inserimenti.

Inoltre una valutazione importante da fare è che trattandosi di dati testuali già codificati, nei fenomeni più rilevanti, già strutturati organicamente, è possibile passare ad altre codifiche in modo da poter trasmigrare i dati ad altre applicazioni e/o scambiarli con altre istituzioni.

11. Appendice

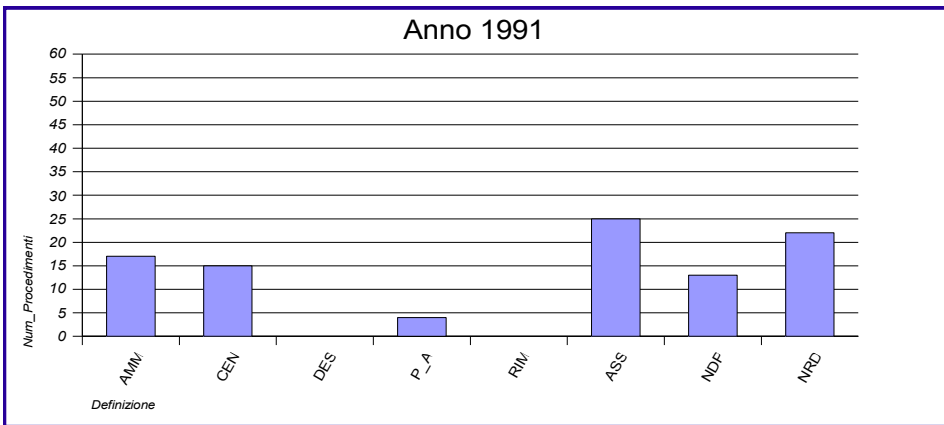
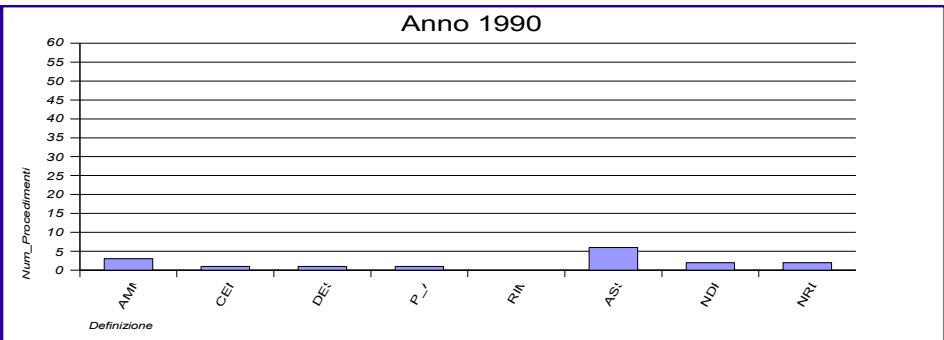
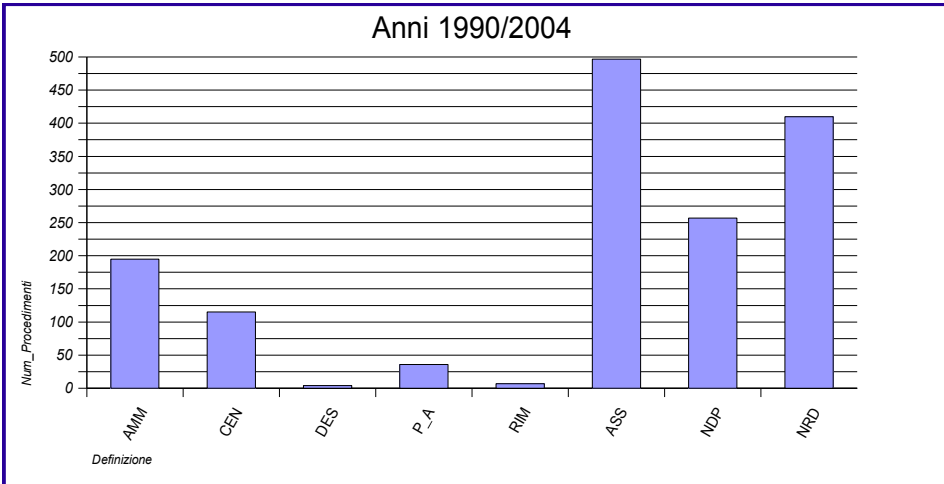
11. Appendice

I risultati salvati durante le richieste in DBT sono stati organizzati in fogli di calcolo. Utilizzando la data di emissione della sentenza e le frequenze delle parole-chiave che ne individuano l'esito nel dispositivo, sono stati generati i seguenti grafici. Le sigle indicate hanno il seguente significato:

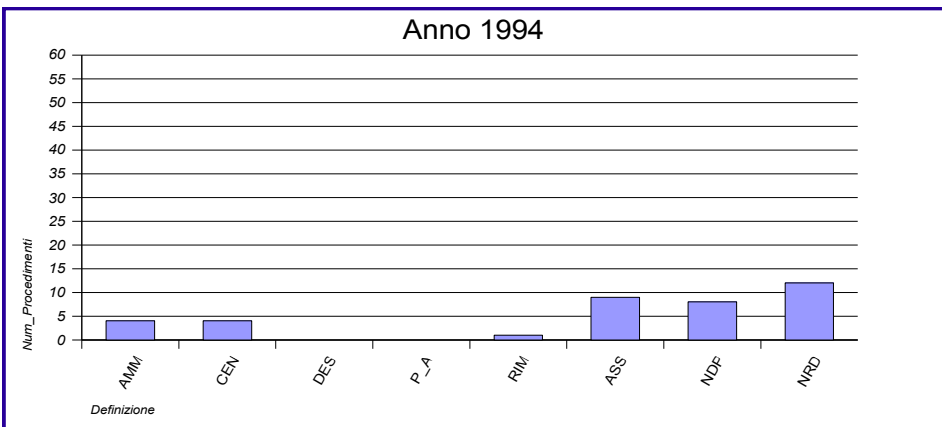
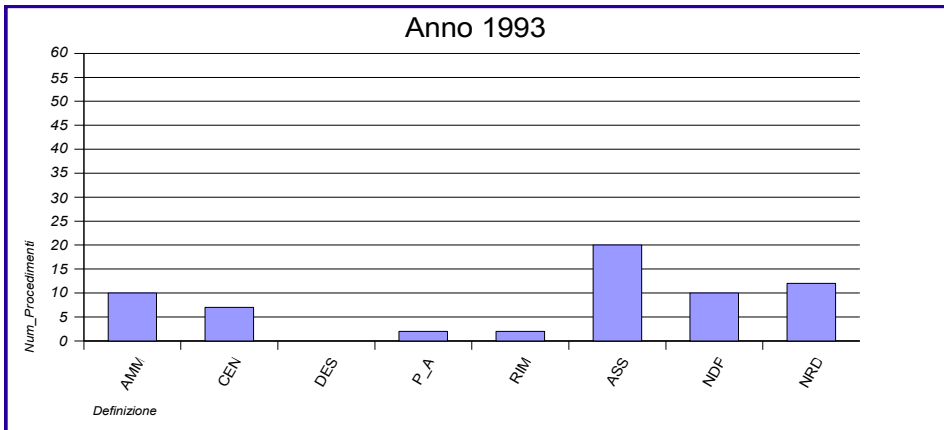
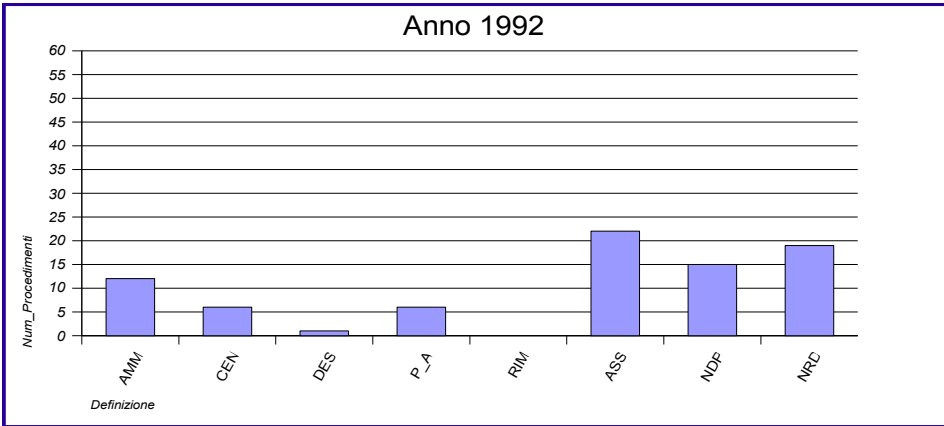
AMM	Condanna con sanzione dell'ammonimento	
CEN	"	" della censura
DES	"	" della destituzione
P_A	"	" di perdita di anzianità
RIM	"	" della rimozione
ASS	Assoluzione	
NDP	Non doversi procedere / Non farsi luogo a procedere	
NRD	Non rinvio a dibattimento	

DEF/AN	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	TOT
AMM	3	17	12	10	4	20	12	20	15	8	15	18	17	18	6	197
CEN	1	15	6	7	4	13	20	17	8	2	7	4	5	5	1	115
P_A	1	4	6	2		3	4	2	4		3		2	5		36
RIM				2	1	1			1		1		1			7
DES	1		1										1		1	4
TOT CON	6	36	25	21	9	37	36	39	28	10	26	22	26	28	8	359
ASSOLVE	6	25	22	20	9	34	25	60	41	30	49	57	54	42	23	500
NRD	2	22	19	12	12	28	36	43	37	24	46	46	29	36	18	412
NDP	2	13	15	10	8	15	30	36	49	19	9	17	9	18	7	258
TOTNCO	10	60	56	42	29	77	91	139	127	73	104	120	92	96	48	1170
TOTALE	16	96	81	63	38	114	127	178	155	83	130	142	118	124	56	1529

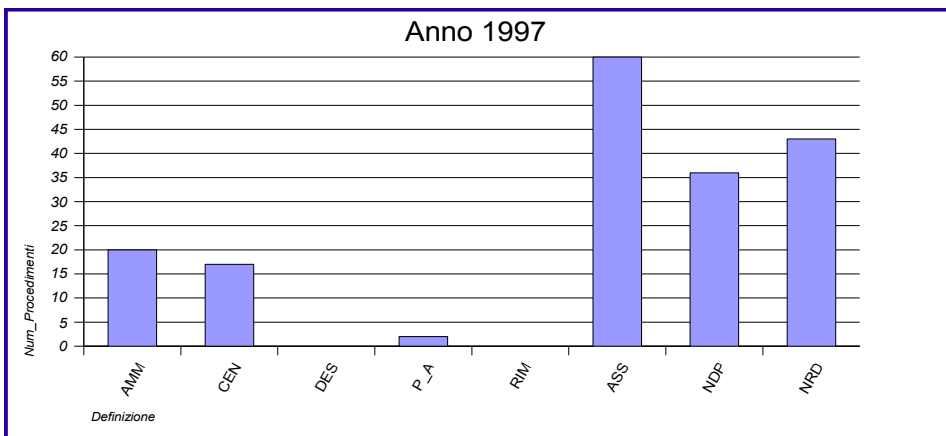
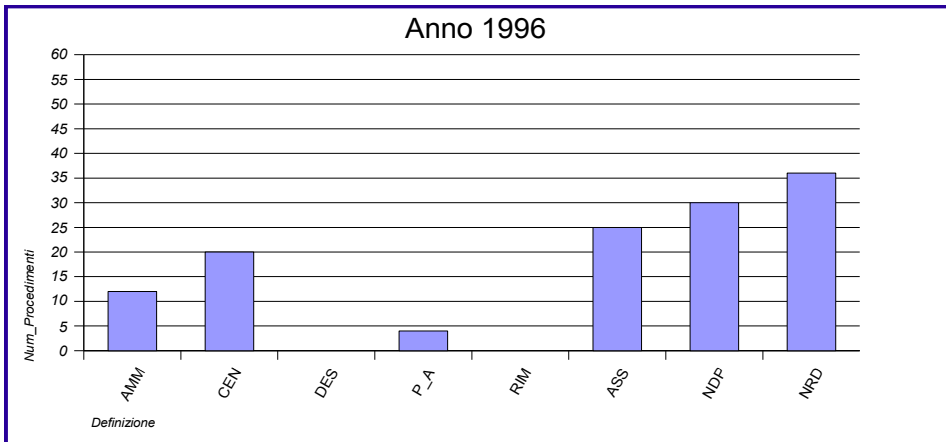
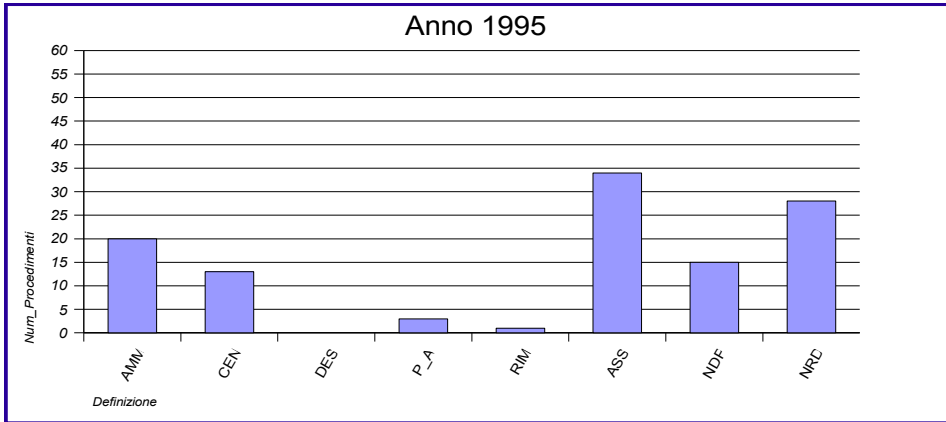
11.Appendice



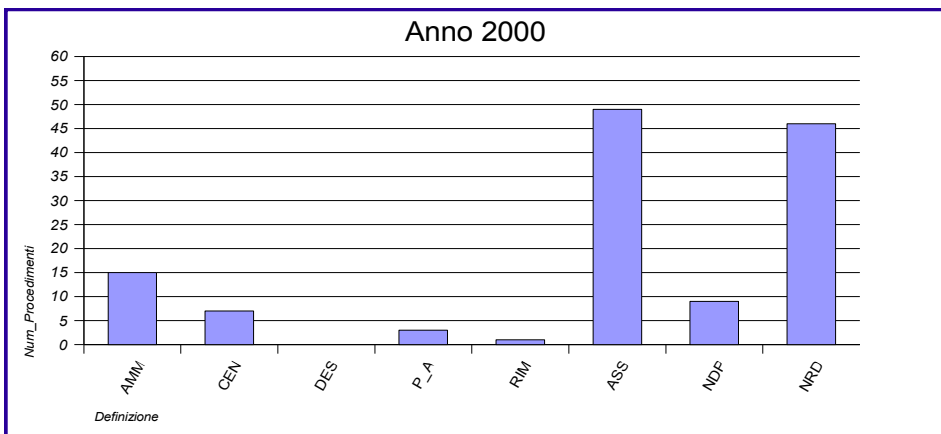
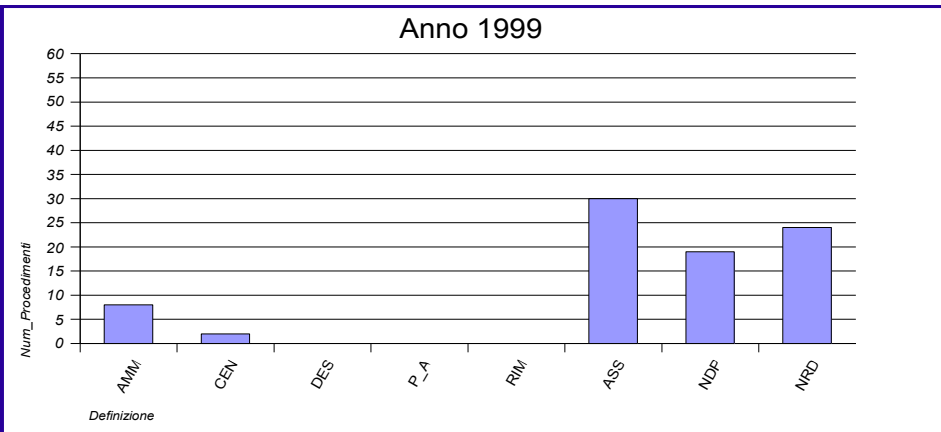
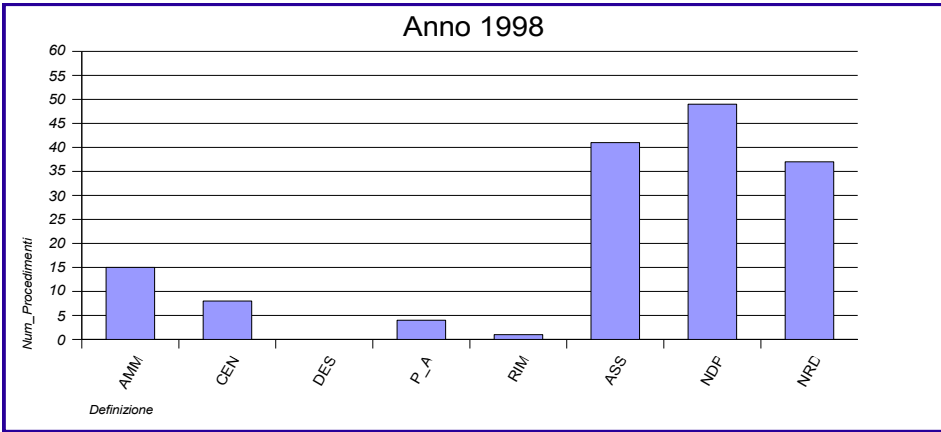
11.Appendice



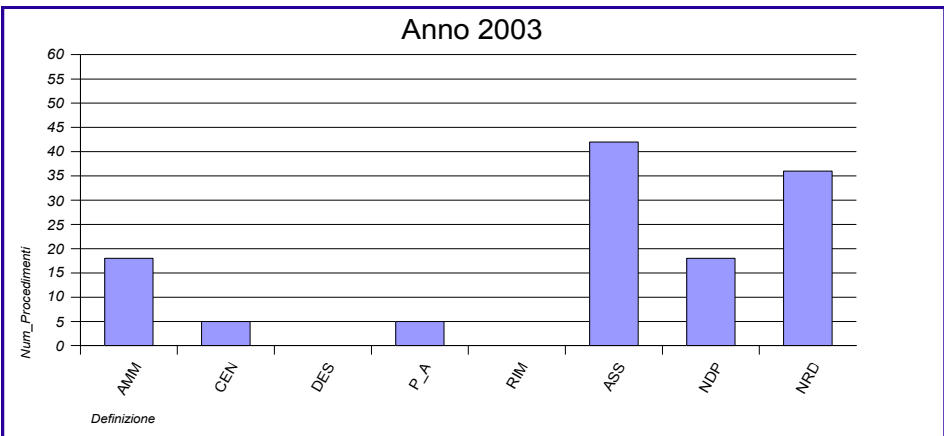
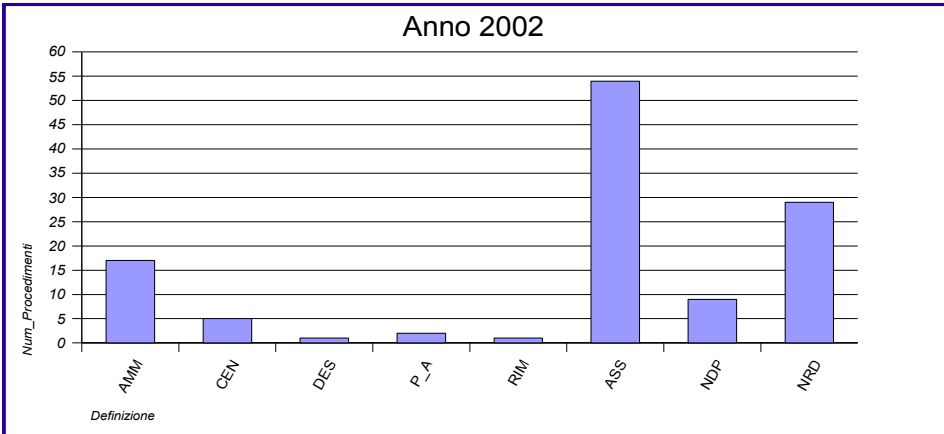
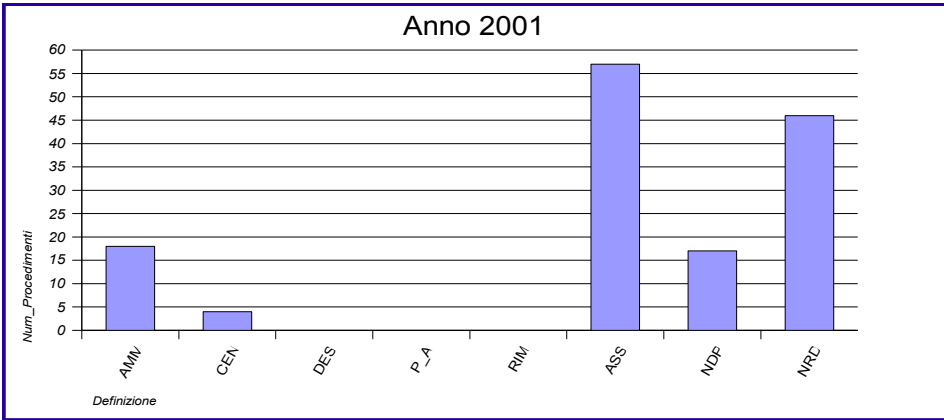
11.Appendice



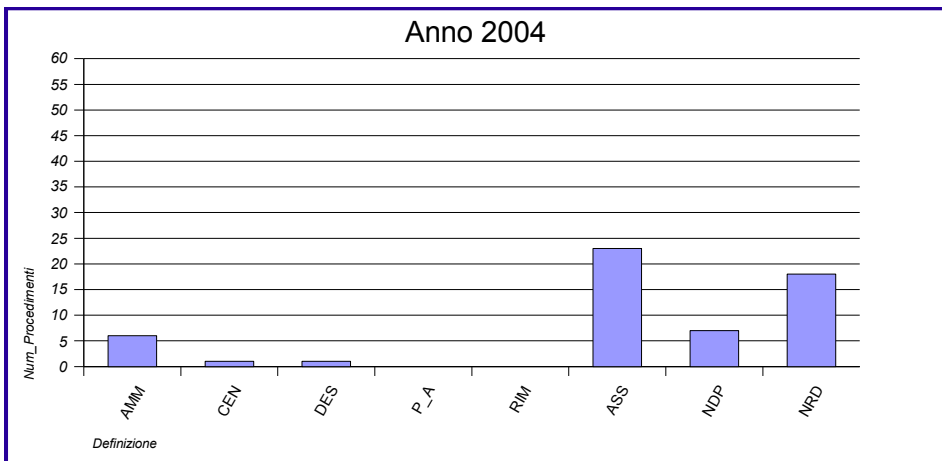
11.Appendice



11.Appendice



11.Appendice



Indice

1. Introduzione.....	3
2. Schema di lavoro.....	4
3. Studio della tipologia dei testi.....	5
4. Creazione del database.....	6
5. Caratteristiche del database.....	7
5.1. Il riferimento.....	9
5.2. Le date.....	9
5.3. I numeri.....	14
5.4. Le sigle o acronimi.....	16
5.5. La struttura delle sentenze	18
6. Esempi di consultazione del database.....	21
6.1. Esempio di ricerca complessa.....	31
6.2. Distribuzione delle frequenze.....	38
6.3. Cooccorrenze statistiche.....	39
6.4. Menu Indici vari.....	42
6.5. Ricerca per Strutture Concettuali.....	46
6.6. Esempi di consultazione del Corpus.....	49
7. Altri tipi di elaborazioni.....	50
7.1. Ricerca di schemi linguistici.....	51
7.2. Sperimentazione dell'anonimazione di testi.....	55
8. Presentazione del Database	59
9. Sviluppi futuri.....	60
10. Appendice.....	61