

# ***Dieci anni di informazione sanitaria*** ***Analisi di tre quotidiani a tiratura nazionale***

*Manuela Sassi, Alessandra Cinini*



Istituto di Linguistica  
Computazionale  
C.N.R.

## ***Dieci anni di informazione sanitaria***

### ***Analisi di tre quotidiani a tiratura nazionale***

#### **Introduzione**

*L'Osservatorio della Comunicazione Sanitaria* ha sede nel Dipartimento di Biologia dell'Università di Pisa e la responsabile è la Prof.ssa Annalaura Carducci (Professore associato di Igiene, presso il Dipartimento di Biologia). L'ILC collabora con l'OSC per il trattamento dei testi raccolti e le strategie di estrazione dei contenuti.

Quotidianamente, nella sede dell'OSC, vengono raccolti e memorizzati gli articoli su temi sanitari pubblicati da quotidiani nazionali. Rileva con continuità la quantità e la qualità dei messaggi diffusi sul tema della salute ed il loro impatto sulla popolazione. I dati rilevati dagli articoli e gli articoli stessi vengono organizzati in un data-base contenente, oltre al testo, le seguenti informazioni: data, pagina, tipologia, tematica, luogo, persone/enti, tipo, autore, sezione, occhiello, titolo, sommario, descrittori, valutazione del messaggio.

In collaborazione con l'OCS l'Istituto di Linguistica Computazionale (ILC-CNR), ha realizzato una Banca Dati Testuale composta dagli articoli che

trattano Informazione Sanitaria, pubblicati a partire dall'anno 1999, su tre quotidiani a diffusione nazionale: Corriere della Sera (C), Repubblica (R) e Stampa (S).

In questo aggiornamento vengono forniti i nuovi quadri di riepilogo dei dati presenti nel corpus, che ha ormai doppiato il decimo anno; si può quindi definire come il quadro decennale dell'informazione sui giornali di temi della salute.

### **Realizzazione degli archivi**

Gli articoli dei tre quotidiani raccolti presso l'Osservatorio della Comunicazione Sanitaria sono stati forniti inizialmente su CD, mentre il proseguimento di routine avviene tramite mail.

I testi vengono elaborati e codificati per il successivo inserimento nel DBT (Data Base Testuale, Brevetto CNR di Picchi E.).

Il DBT è un sistema di creazione, gestione e analisi di banche dati testuali. L'indicizzazione dei testi così elaborati ha generato 9 archivi, ciascuno dei quali contiene gli articoli pubblicati su uno dei tre quotidiani negli anni 1999-2000, 2001-2004 e 2005-2008; l'aggiornamento viene eseguito generalmente ogni semestre.

Tale suddivisione è dovuta principalmente all'ottimizzazione dei tempi di indicizzazione e di ricerca. Possono comunque essere adottate regole diverse di accorpamento o di suddivisione data l'organizzazione dei file in semestri.

Nella tabella 1 sono elencati gli archivi e il numero di parole corrispondente agli archivi di ciascuna testata al momento attuale.

<i>TESTATA</i>	<i>N. PAROLE</i>
Corriere	4.542.493
Repubblica	3.183.678
Stampa	4.361.003
<b>Totale complessivo</b>	<b>12.087.174</b>

Nella fase di creazione degli archivi si sono presentate diverse problematiche. Nella maggioranza dei casi gli articoli pubblicati in uno stesso giorno erano stati raggruppati in un unico documento senza apporre alcuna divisione tra l'uno e l'altro, questo ha reso impossibile l'applicazione di una procedura automatica per marcare l'inizio di ogni articolo.

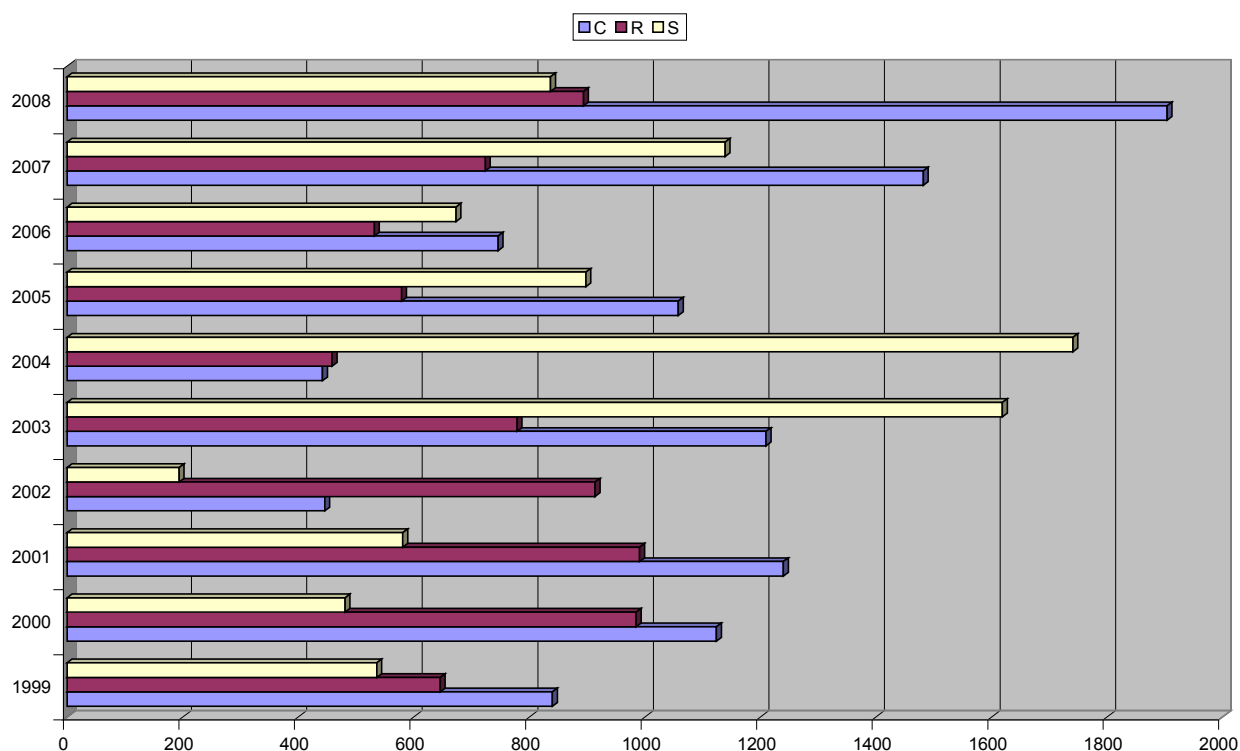
Ogni unità logica degli archivi realizzati risultava quindi costituita dagli articoli pubblicati sul medesimo quotidiano nella stessa data. Ciò ha portato ad una revisione della codifica in modo tale da permettere di suddividere i vari temi in unità logiche distinte. Ossia il riferimento logico di detta unità attualmente risulta composto da una lettera che individua il giornale (C=Corriere, R=Repubblica, S=Stampa), da una abbreviazione che indica la sezione (Cr=Cronaca, Pp=prima pagina, ecc.), dalla data di pubblicazione (nel formato americano: AAAA-MM-GG) e da una lettera di sequenza alfabetica per ogni articolo.

Nel corso del 2006-2007, tramite alcuni stagisti assegnati al Dipartimento, è stato possibile affidare questo intervento in modo da individuare tutti gli articoli della giornata con questa nuova codifica. Questo intervento ha permesso di avere statistiche testuali aggiornate per articolo, facilitando il lavoro di interpretazione dei dati.

Nella tabella 2 viene indicato il numero degli articoli per testata e per anno, i totali vengono rappresentati graficamente nella figura successiva.

<b>Testata</b>	<b>Anno</b>	<b>Totale articoli</b>
<b><i>Corriere</i></b>	1999	840
	2000	1124
	2001	1240
	2002	446
	2003	1210
	2004	442
	2005	1058
	2006	746
	2007	1482
	2008	1904
<b>Totale</b>		<b>10492</b>
<b><i>Repubblica</i></b>	1999	646
	2000	985
	2001	991
	2002	914
	2003	779
	2004	459
	2005	579
	2006	532
	2007	724
	2008	894
<b>Totale</b>		<b>7503</b>
<b><i>Stampa</i></b>	1999	536
	2000	481
	2001	581
	2002	194
	2003	1619
	2004	1741
	2005	898
	2006	673
	2007	1139
	2008	837
<b>Totale</b>		<b>8699</b>
<b><i>Totale complessivo</i></b>		<b>26694</b>

TAB 2



Gli archivi generati sono stati organizzati in Corpus e quindi sono consultabili simultaneamente; tale corpus è identificato dalla sigla CS e comprende 26.694 articoli. Il corpus CS può essere oggetto di ricerca sia nel suo complesso sia per sottogruppi di archivi, scelti dall'utente.

## Alcuni esempi di consultazione degli archivi tramite DBT

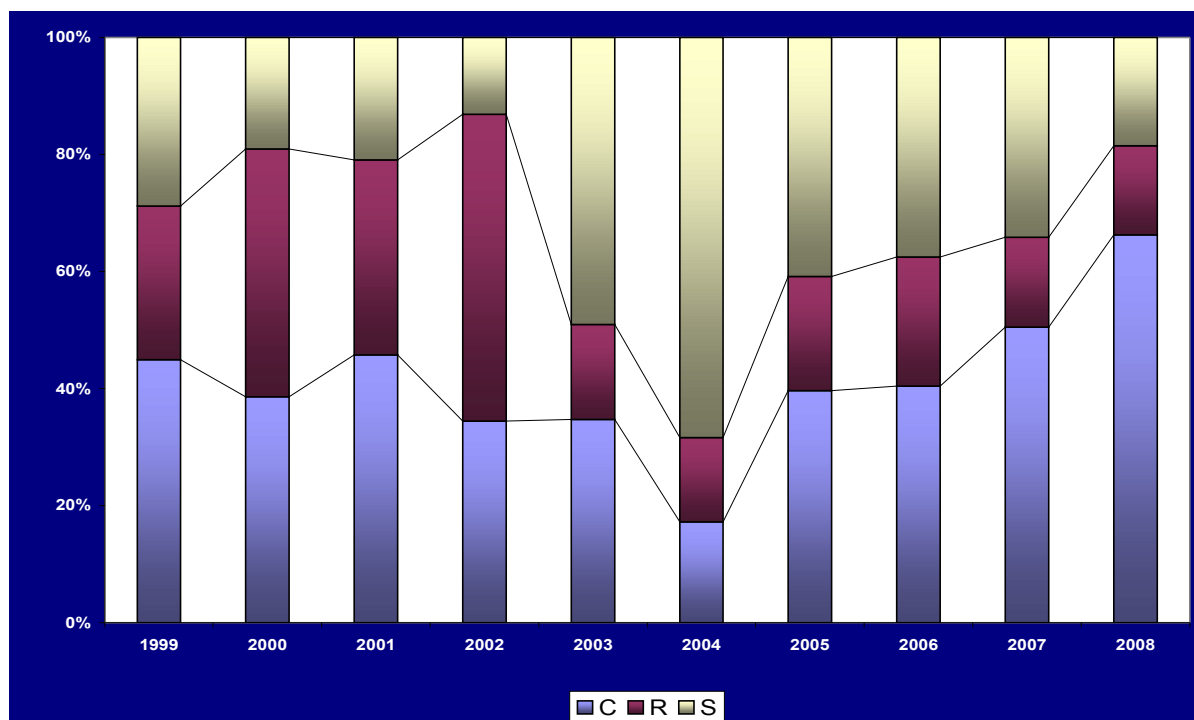
### Ricerca dei termini che condividono la sottostringa "cardio".

Sottoponendo tale richiesta al corpus, si ottiene la seguente lista di termini. Il totale delle frequenze delle parole elencate è di 3923 volte in tutto il corpus, suddivise in 1921 articoli. A seguire il grafico in percentuale suddiviso fra le tre testate nei 10 anni.

cardio	cardiopatologie
cardioanestesista	cardiopulmonare
cardioaritmologia	cardiopulmonari
cardioaritmologo	cardioprotettivi
cardioaspirina	cardioprotettivo
cardioattivi	cardioradiologia
cardiochirurgia	cardiorcolatorio
cardiochirurgo	cardiorespiratoria
cardiochirurgi	cardiorespiratori
cardiochirurgia	cardiorespiratoria
cardiochirurgiapediatrica	cardiorespiratorie
cardiochirurgica	cardiorespiratorio
cardiochirurgiche	cardiorianimatore
cardiochirurgici	cardiostim
cardiochirurgico	cardiostimolante
cardiochirurgie	cardiostimolanti
cardiochirurgo	cardiostimolazione
cardiochiurgia	cardioteam
cardiochiurgia	cardiotocografo
cardiochirurgia	cardiotonic
cardiocircolatori	cardiotonici
cardiocircolatoria	cardiotonico
cardiocircolatorie	cardiotoracica
cardiocircolatorio	cardiotossica
cardiocircovascolari	cardiotossicità
cardiociti	cardiotrapiantati
cardioembolico	cardiotrapianti
cardiofrequenzimetro	cardiovascolare
cardiogenesi	cardiovascolari
cardiogeno	cardiovascolatorie
cardiogramma	cardiovascular
cardiogrammi	cardioversione
cardiolab	cardioverter
cardiologa	cardiowest
cardiologi	ecocardio
cardiologia	ecocardiografi

cardiologica  
 cardiologiche  
 cardiologici  
 cardiologico  
 cardiologie  
 cardiologists  
 cardiologo  
 cardiology  
 cardiometabolici  
 cardiomiociti  
 cardiomiopatia  
 cardiomiopatie  
 cardiochirurgia  
 cardiopalma  
 cardiopalmo  
 cardiopatia  
 cardiopatica  
 cardiopatiche  
 cardiopatici  
 cardiopatico  
 cardiopatie

ecocardiografia  
 ecocardiografica  
 ecocardiografie  
 ecocardiografo  
 ecocardiogramma  
 ecocardiogrammi  
 elettrocardiografi  
 elettrocardiografia  
 elettrocardiografica  
 elettrocardiografico  
 elettrocardiografo  
 elettrocardiogramma  
 elettrocardiogrammi  
 microelettrocardiografo  
 miocardio  
 miocardiopatia  
 miocardiopatie  
 pericardio  
 postcardiochirurgia  
 telecardiochirurgia  
 telecardiologia





## Il linguaggio divulgativo

Il linguaggio giornalistico utilizzato per la diffusione dell'informazione sanitaria è spesso eccessivamente allarmistico.

Una testimonianza di ciò risulta dalla ricerca delle forme di *allarme* e derivati effettuata nel corpus, dove i termini recuperati compaiono in ben 5920 contesti, distribuiti in 3360 articoli, nelle seguenti forme:

- |                     |                       |
|---------------------|-----------------------|
| 1) 'allarme         | 25) allarmeranno      |
| 2) allarma          | 26) allarmi           |
| 3) allarmando       | 27) allarmiamoci      |
| 4) allarmandosi     | 28) allarmino         |
| 5) allarmano        | 29) allarmismi        |
| 6) allarmante       | 30) allarmismi'       |
| 7) allarmanti       | 31) allarmismo        |
| 8) allarmarci       | 32) allarmista        |
| 9) allarmare        | 33) allarmiste        |
| 10) allarmarla      | 34) allarmisti        |
| 11) allarmarle      | 35) allarmistica      |
| 12) allarmarli      | 36) allarmistiche     |
| 13) allarmarsi      | 37) allarmistici      |
| 14) allarmata       | 38) allarmistico      |
| 15) allarmate       | 39) assurdoallarmismo |
| 16) allarmatevi     | 40) d'allarme         |
| 17) allarmati       | 41) dall'allarme      |
| 18) allarmatissimo  | 42) èallarme          |
| 19) allarmato       | 43) l'allarme         |
| 20) allarmava       | 44) mallarmé          |
| 21) allarme         | 45) preallarme        |
| 22) allarmeaid      | 46) preallarmi        |
| 23) allarmediossina | 47) troppi allarmismi |
| 24) allarmeper      |                       |

Il grafico che segue rappresenta la suddivisione temporale e per testata dei 3360 articoli che contengono una o più delle parole precedentemente recuperate.

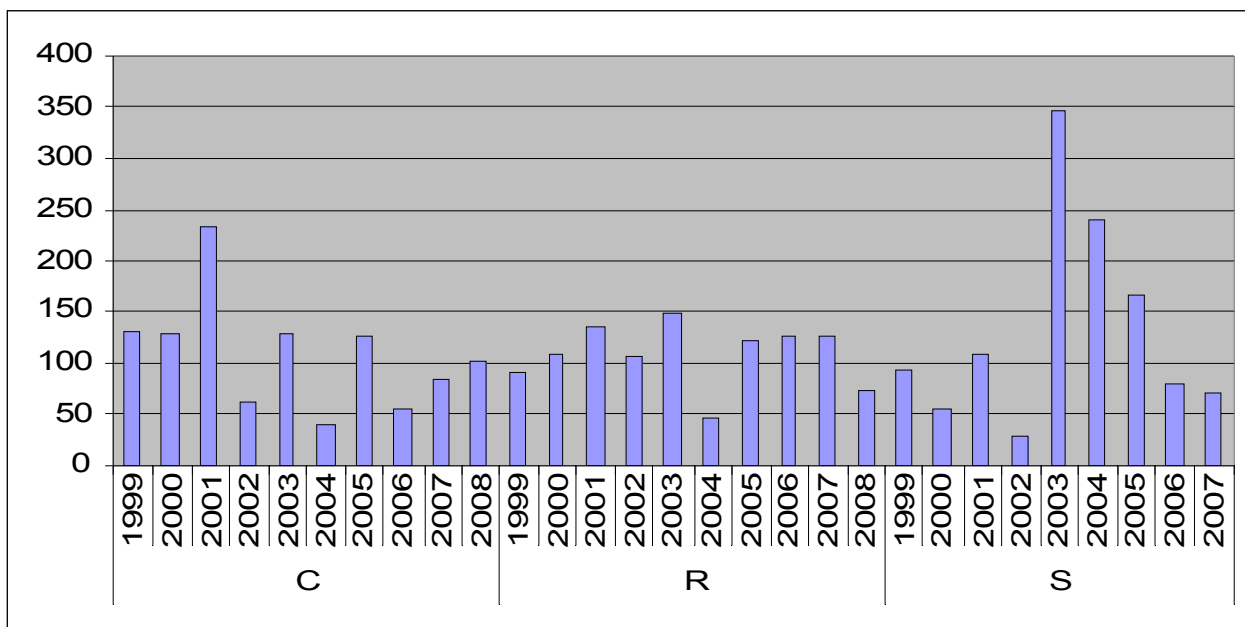


grafico 1: distribuzione degli articoli per testata ed anno

Ricerca per famiglie di parole

Il sistema di interrogazione DBT permette di definire famiglie di parole: le forme ricercate vengono combinate tra loro mediante operatori logici, ed è inoltre possibile impostare parametri per la distanza e sequenzialità tra le forme stesse. Nell'esempio che segue sono state definite 5 famiglie:

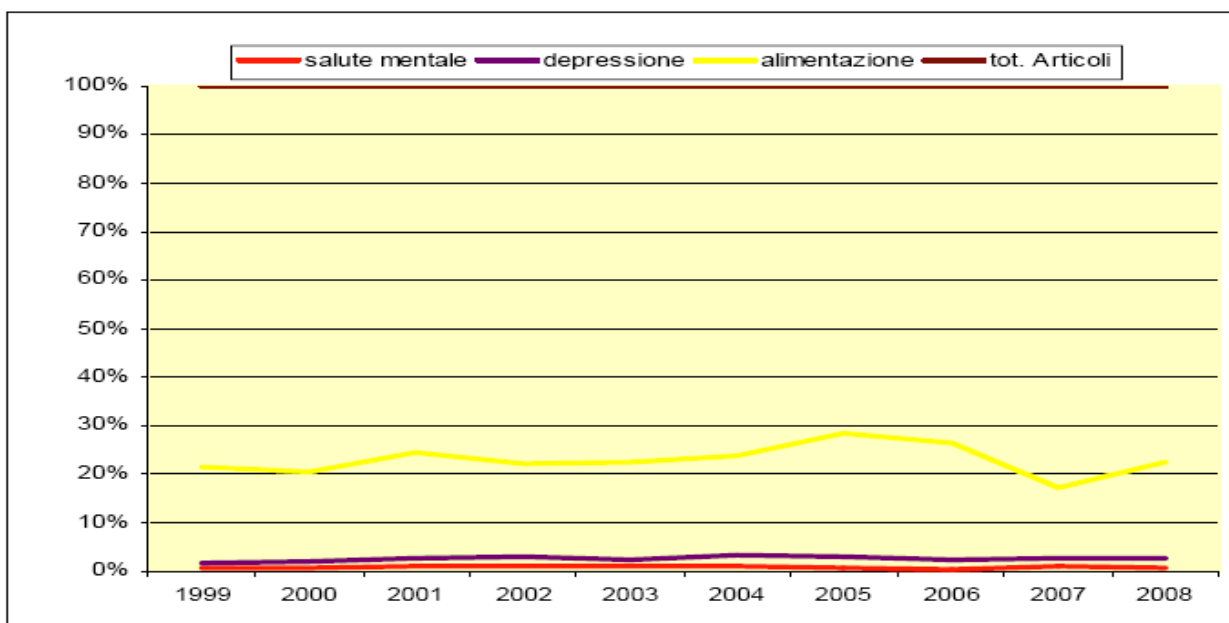
1. salute & mentale
2. malattia/e & mentale/i
3. disagio/i, & psichico/i
4. malattia/e & psichica/e

5. schizofrenia/schizofrenici

queste sono state successivamente combinate attraverso l'operatore logico "OR" per definire una nuova famiglia.

In tal modo è stato possibile recuperare tutti i contesti del Corpus CS in cui cooccorrevano le coppie di termini precedenti. I contesti generati, visualizzati in finestre DBT, sono stati salvati su di un file di testo. Utilizzando procedure automatiche è stata prodotta una lista di tutti gli articoli relativi ai contesti precedenti, con l'indicazione della frequenza. Si è rilevato che il tema della salute mentale risulta coinvolgere 269 unità, solamente l'1% del totale degli articoli della raccolta.

Tali dati sono stati confrontati con quelli di altre analisi analoghe ad esempio gli articoli sulla depressione oscillano tra il 2% e il 4% del totale quelli che riguardano le problematiche legate all'alimentazione tra il 18% e il 28% .



**Creazione di indici**

Oltre alla ricerca diretta di informazioni tramite l'interfaccia di consultazione dei testi, il sistema DBT consente la creazione di vari tipi di indici salvati in files di output riutilizzabili con un qualunque programma di videoscrittura:

- ✓ il calcolo delle *frequenze assolute* in ordine alfabetico o decrescente
- ✓ gli *index locorum*, arricchiti da riferimenti puntuali
- ✓ La lista di *forme co-occorrenti* con i relativi valori di mutual information.
- ✓ il calcolo dell' *indice di leggibilità*, basato sulle formule statistiche più accreditate.

L'Indice di leggibilità è una funzione automatica, disponibile in DBT, che applica la formula GULPEASE ad un archivio testuale; tale formula è basata sulla correlazione fra il numero dei caratteri, delle parole e delle frasi individuate nel testo indicato e in ogni sua unità logica, attraverso costanti numeriche individuate sperimentalmente:

Indice di leggibilità GULPEASE =  $89 - LP / 10 + FR \times 3$

dove LP = totale lettere  $\times$  100 / totale parole;

FR = totale frasi  $\times$  100 / totale parole

La valutazione del testo viene espressa in una scala di valori compresa tra 1 e 100, dalla maggiore alla minore difficoltà:

da 1 a 35: testo quasi incomprensibile

da 36 a 49: testo molto difficile

da 50 a 59: testo difficile

da 60 a 79: testo facile

da 80 a 100: testo facilissimo.

Questa funzione per il calcolo dell'indice di leggibilità è stata applicata ai testi che compongono il corpus CS rilevando che 66 unità logiche risultano quasi incomprensibili, 12411 molto difficili, 10925 difficili, 1384 facili e 9 molto facili.

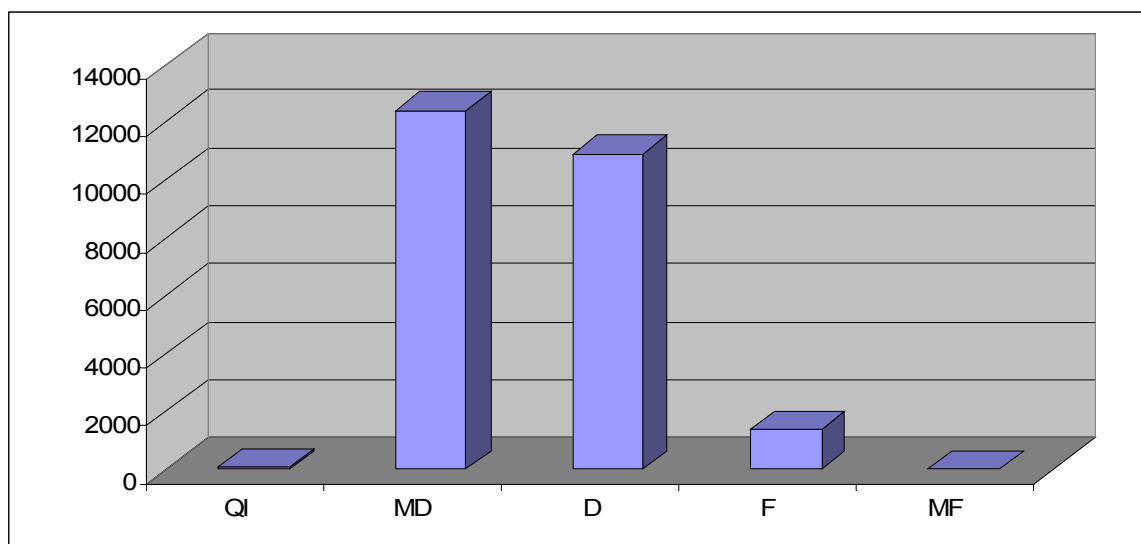


grafico 2: leggibilità

A titolo di esempio abbiamo riportato gli articoli con indice di leggibilità fra i più alti (Es. 1) e fra i più bassi (Es. 2).

ES: 1

Corriere 18 novembre 2002 cronaca

*Il medico: gli sbalzi di temperatura favoriscono le infezioni virali*

**«Anziani e malati di cuore, meglio fare il vaccino»**

*«Influenza, bronchiti e raffreddori i veri rischi. Ma coprirsi troppo è sbagliato»*

ROMA - «Il corpo umano è una macchina che si adatta benissimo ai cambi di temperatura e di umidità. Ma non sopporta gli sbalzi improvvisi. Ha bisogno di tempo». E in questi giorni il tempo per acclimatarsi non è stato davvero sufficiente. Il perché lo spiega Carlo Signorelli, professore di Igiene e sanità pubblica all'Università di Parma. **Lo scirocco ha fatto salire la temperatura anche di dieci gradi nel giro di poche ore. Ma già da oggi i**

**termometri torneranno a scendere e a scendere velocemente.**

«E il nostro corpo reagisce male, purtroppo. I cambiamenti improvvisi lo prendono di sorpresa. Sono il nostro principale punto debole, il nostro tallone d'Achille».

**Quali sono le conseguenze?**

«I medici se ne accorgono subito: aumentano le infezioni virali o batteriche delle prime vie respiratorie».

**Cioè?**

«Tracheiti, raffreddori e anche bronchiti».

**Come ci si può proteggere?**

«Prima di tutto è necessario adattare il proprio abbigliamento».

**La vecchia tecnica «a cipolla», cioè a strati.**

«Esatto. Ma bisogna vincere la pigrizia. Non va bene restare sempre con la stessa roba addosso ed è sbagliato mettere comunque il cappotto solo perché il calendario dice che siamo a novembre. La mattina bisogna coprirsi di più. A metà giornata il maglione si può togliere. La sera bisogna tornare a coprirsi. Basta ascoltare il proprio corpo: quanti di noi finiscono per sudare anche d'inverno perché sbagliano vestiti? E poi...».

**E poi?**

«C'è sempre il vaccino antinfluenzale. Per tutto il mese di novembre si è ancora in tempo per vaccinarsi. Un consiglio che vale soprattutto per chi ha più di 65 anni oppure problemi cardiaci o polmonari».

**E la pioggia, invece?**

«Il nostro corpo vive bene quando l'umidità è tra il 30 e il 70 per cento. Se piove, i valori sono molto più alti. E le vie respiratorie diventano più vulnerabili. Le malattie sono sempre le stesse: tracheiti, raffreddori e bronchiti».

**E i rimedi?**

«Non bagnarsi, se possibile, ma questo è ovvio. Se si prende la pioggia, bisogna asciugarsi subito. Può sembrare banale. Ma sa quanti raffreddori si possono evitare con questo semplice accorgimento?».

Le assicurazioni sanitarie si adeguano alle richieste di un settore «anomalo» in cui le prestazioni sempre più sono fornite contemporaneamente dal Servizio sanitario nazionale e da privati, singoli o enti. Ma quando si ferma il servizio pubblico tutti gli ulteriori interventi sono a carico dell'interessato. Nasce dunque una polizza che punta soprattutto a integrare, sia in caso di infortunio che di malattia, l'assistenza erogata dalla sanità pubblica: è «Salute Assistita» delle Assicurazioni Generali. Si tratta di una copertura assicurativa completa che comprende una serie importante di prestazioni innovative, studiate per garantire un servizio completo di assistenza che consenta al cliente di affrontare con maggiore tranquillità gli inevitabili disagi collegati al verificarsi di un infortunio o di una malattia. Quali le garanzie fornite? Indennità giornaliera di ricovero e convalescenza, rimborso delle spese di assistenza infermieristica privata post intervento chirurgico, indennità speciale erogata in caso di intervento chirurgico derivante da grave malattia, rimborso delle spese mediche sostenute nei 60 giorni precedenti il ricovero e per prestazioni sanitarie effettuate nei 90 giorni successivi alle dimissioni, assistenza medica domiciliare (un pacchetto di servizi, denominato «Medical care» - erogati tramite la Centrale Operativa di Europ Assistance, società del Gruppo Generali - che soddisfano in particolare i bisogni relativi all'assistenza domiciliare post ricovero, quali consulenze mediche, second opinion, informazioni sanitarie, trasporto dell'assicurato, assistenza domiciliare di base e integrata post ricovero, distance care). Il costo minimo annuo è di 132 euro - il premio varia in rapporto all'età - per coperture individuali di giovani fino a 18 anni (si può sottoscrivere il contratto entro i 75 anni), con durata da 1 a 5 anni. Un quarantenne assicurato per: indennità giornaliera di 75 euro, massimale di 2000 euro quale rimborso delle spese pre e post ricovero, massimale di 2000 euro quale rimborso spese per assistenza infermieristica post intervento, indennità di 3000 euro per gravi interventi chirurgici, paga di premio 251 euro all'anno, compresi i diritti fissi.

## **Creazione di archivi tematici**

In occasione della presentazione della tesi di laurea di Simona Alfani è stata sperimentata una nuova funzionalità per la generazione di archivi composti da un sottoinsieme di articoli dell'intero Corpus che soddisfano criteri prestabiliti. In particolare è stata approntata una procedura automatica di estrazione di articoli che recupera da un archivio (nel caso specifico dal corpus CS) le unità di testo il cui riferimento logico è inserito in una lista fornita in input. Le unità di testo estratte vengono riunite in un nuovo file di output, che sarà usato come base per il nuovo archivio.

Per la realizzazione della tesi era necessario costituire un archivio che si componesse esclusivamente degli articoli su temi di alimentazione. In seguito ad uno studio accurato supportato dalla consultazione in DBT dei testi del corpus, è stata creata una lista di termini in grado di individuare i testi relativi al tema dell'alimentazione.

Tramite la funzione DBT, "*ricerca per lista di parole*", sono stati generati i contesti relativi alle forme della lista predisposta. Con la funzione "*SalvaFinestra*" i contesti ottenuti sono stati salvati su file. Tramite un procedimento creato a questo scopo, si può ottenere l'elenco degli articoli da utilizzare come input per creare il nuovo archivio.

Prima di procedere alla creazione del sottoarchivio tematico, è stato fatto uno studio sulle occorrenze della lista di termini.

In base alle frequenze cumulative di tali termini per ciascun articolo è stato verificato che se tale frequenza era uguale a 1, si trattava di una presenza sporadica, tale che non denotava tale articolo come appartenente alla categoria ricercata. Pertanto dalla lista totale dei riferimenti sono stati eliminati questi ultimi.

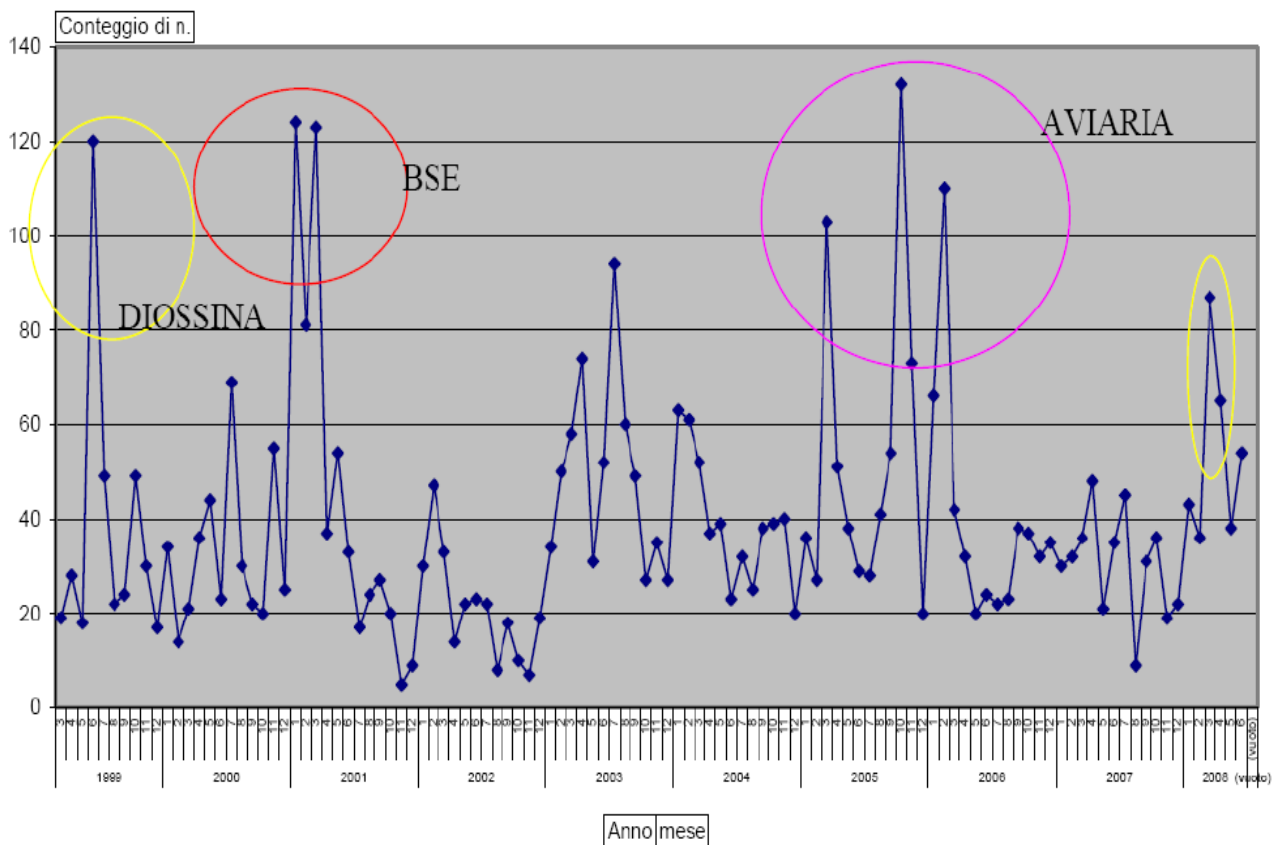
In totale sono stati quindi selezionati 4436 articoli. Su questo insieme sono stati effettuati altri studi specifici, a titolo di esempio riportiamo un grafico sintetico della distribuzione degli articoli di alimentazione per anno.



I risultati della tesi hanno evidenziato che i picchi relativi al 1999 e al 2008 riguardavano prevalentemente la tematica diossina, nel 2001 veniva trattata la problematica della BSE, nel 2005 e 2006 spiccava l'argomento influenza aviaria, mentre nel 2003 non vi era un fenomeno più rilevante rispetto agli altri (vedi grafico 4)

Con questa metodologia è possibile generare altri sotto-archivi tematici a partire dal Corpus CS.

Grafico 4: Distribuzione degli articoli di alimentazione per anno e mese (tratto dalla tesi di laurea di Simona Alfani )



## Linguistic Miner (LM)

LM è un progetto dell'Istituto di Linguistica Computazionale del CNR che ha come obiettivo la costruzione di un sistema che estragga automaticamente da testi in lingua italiana, di varia provenienza e formato, conoscenza linguistica utilizzabile per scopi molteplici: didattici, editoriali, culturali, etc.

Il progetto nasce da una considerazione preliminare a tutti i sistemi di analisi linguistica *corpus based*: una lingua, rappresentata da un insieme, il più grande possibile, di testi delle più varie tipologie è la miglior fonte di informazione linguistica, a qualunque livello di analisi la si consideri.

I testi che vengono inseriti nella miniera possono provenire dalle più svariate fonti (contenuti web e non). Il sistema prevede l'analisi del materiale già nella prima fase di recupero e l'inserimento nella grande banca-dati dei singoli elementi linguistici che costituiscono la specifica ricchezza di ogni testo, inibendone al tempo stesso la capacità di riproduzione e la lettura.

La fase di analisi dei dati costituisce il momento più importante di tutto il progetto. Gli strumenti disponibili per la fase di sfruttamento dei dati testuali che vengono stratificandosi all'interno della miniera, l'ambiente PiSystem con la sua procedura di base DBT, opportunamente integrati e riadattati, costituiscono il nucleo centrale del processo di estrazione e di sintesi delle informazioni linguistiche.

Alle funzioni già disponibili per l'analisi e la navigazione nel testo, è stato aggiunto un nuovo strumento, ossia un ambiente di lavoro per la definizione di *pattern linguistici* e successiva applicazione alla miniera e/o ad un suo sottoinsieme.

Per poter analizzare in maniera più efficace il corpus, il sistema provvede alla esecuzione, in maniera automatica, della fase di *tagging*, altrimenti detta *lemmatizzazione* (Picchi, 94). La funzione di tagging (marcatura automatica) associa il lemma, la categoria grammaticale e i codici di forma a ciascuna parola.

Ad esempio:

LEGGI= LEGGE /S/FP (equivalente a Sostantivo nella forma Femminile Plurale)

REVOCA= REVOCARE /VT/3SIP (equivalente a verbo Transitivo nella forma della Terza Persona Singolare dell'Indicativo Presente).

Tale fase viene eseguita per ciascun testo, nel momento in cui venga utilizzato per la ricerca dei pattern linguistici. Il risultato della lemmatizzazione viene poi memorizzato in modo da essere disponibile per una successiva ricerca che comprenda quel testo.

*Estrazione di pattern linguistici dal corpus CS:*

Nell'esempio 3 è illustrato il risultato della ricerca nel Data Base in esame del seguente schema linguistico:

sostantivo = *virus*, seguito da preposizione e sostantivo.

Per ogni schema recuperato è indicata la frequenza nell'intero corpus; da questo elenco mancano tutti i casi in cui *virus* è parte di una parola, come ad esempio, in *retrovirus*. 8639 è il totale delle occorrenze di *virus* nel corpus.

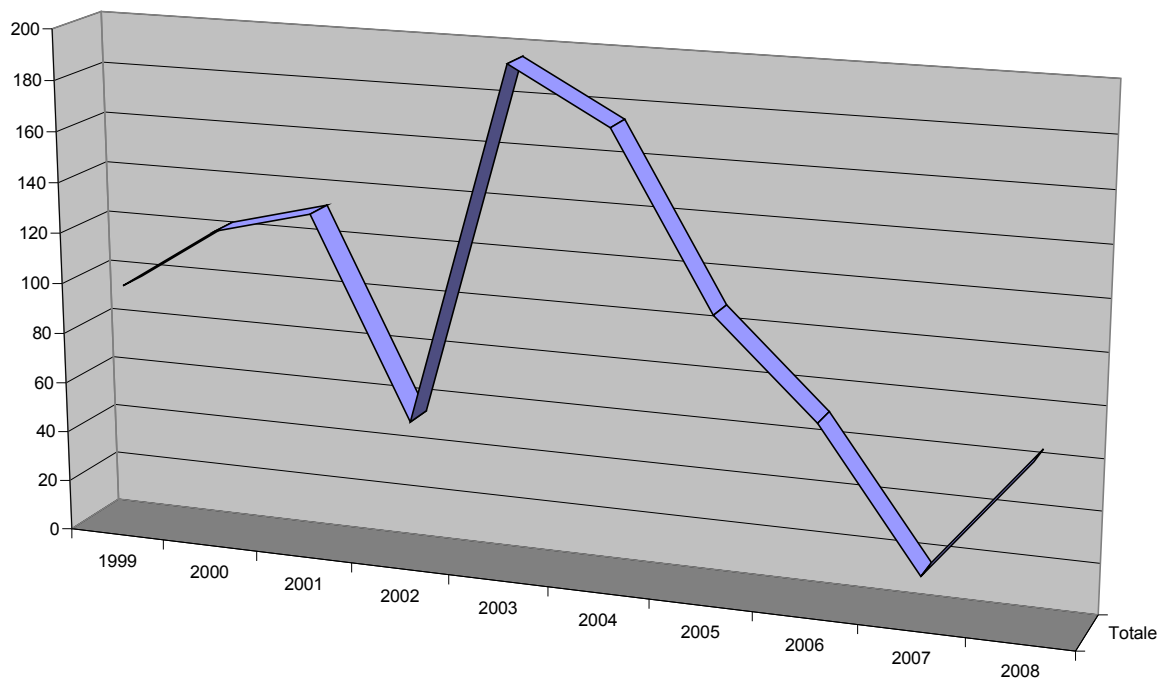
**VIRUS [1040]**

VIRUS A ANTIBIOTICO	1	VIRUS A ESSERE	2
VIRUS A ASSALTO	1	VIRUS A ESTERNO	1
VIRUS A BASE	1	VIRUS A FARMACO	1
VIRUS A BASSA	3	VIRUS A FEMMINILE	1
VIRUS A CAUSA	1	VIRUS A FETO	1
VIRUS A CELLULA	2	VIRUS A FIGLIO	1
VIRUS A CENTRAL	1	VIRUS A FRONTIERA	1
VIRUS A CENTRO	1	VIRUS A GULU	1
VIRUS A CONTATTO	1	VIRUS A H3	1
VIRUS A CORONA	3	VIRUS A H3N2	1
VIRUS A CUORE	2	VIRUS A H5	1
VIRUS A DARTMOOR	1	VIRUS A H5N1	3
VIRUS A DICEMBRE	1	VIRUS A H7N7	1
VIRUS A DNA	1	VIRUS A INTERNO	5

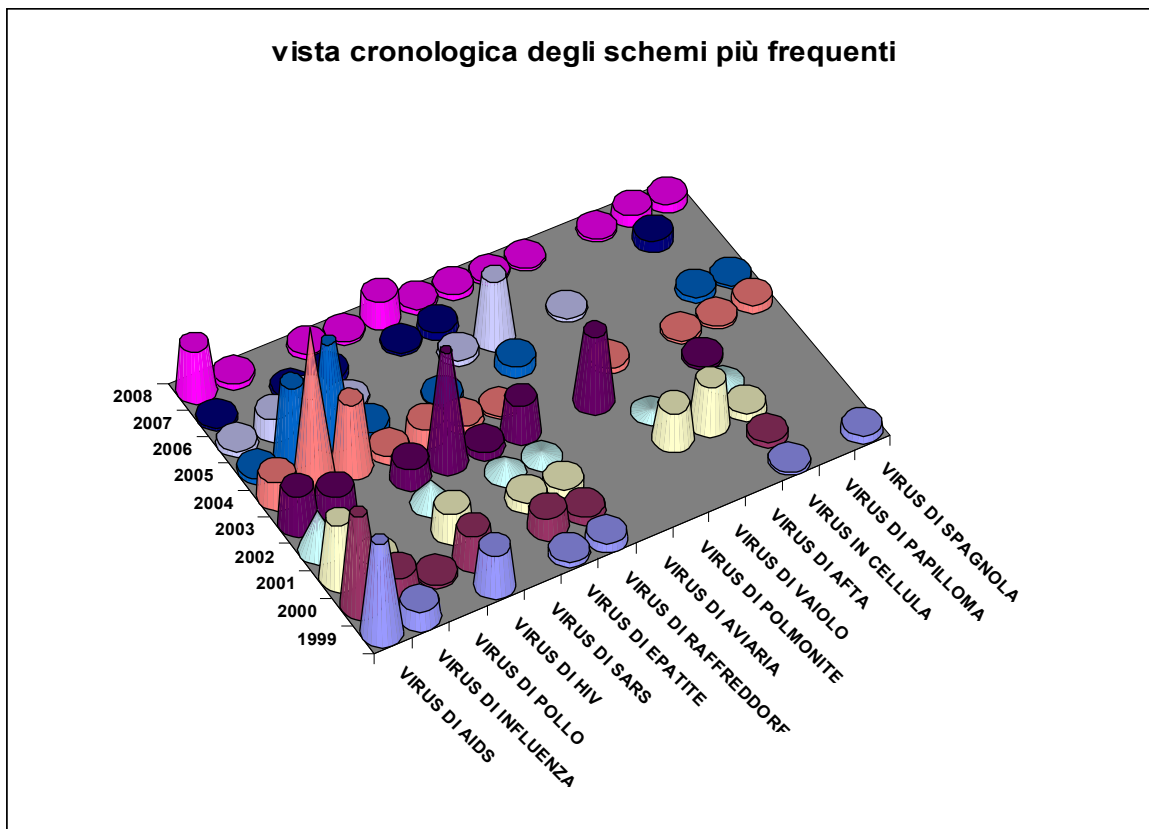
VIRUS A LA	1	VIRUS DA POTENZIALE	1
VIRUS A LABORATORIO	1	VIRUS DA PROVETTA	1
VIRUS A LAVORO	1	VIRUS DA RETROVIRUS	1
VIRUS A LIVELLO	1	VIRUS DA SCIMMIA	1
VIRUS A MILANO	1	VIRUS DA UCCELLO	3
VIRUS A MOMENTO	2	VIRUS DA VOLATILE	2
VIRUS A MORBILLO	1	VIRUS DE BAC	1
VIRUS A NATURA	1	VIRUS DE VIA	1
VIRUS A NATURA'	1	VIRUS DI AFTA	18
VIRUS A NEONATO	1	VIRUS DI AIDS	163
VIRUS A ORIGINE	2	VIRUS DI AMBIENTE	2
VIRUS A PERSONALE	2	VIRUS DI ANIMALE	1
VIRUS A POTERE	1	VIRUS DI ANNO	1
VIRUS A RAFFAELE	1	VIRUS DI AVIARIA	31
VIRUS A RNA	3	VIRUS DI BENE	1
VIRUS A STAGNO	1	VIRUS DI BLUE	1
VIRUS A TAIWAN	1	VIRUS DI BUNYA	1
VIRUS A TRASFORMISMO	1	VIRUS DI CAMBIO	1
VIRUS A TRASMISSIONE	1	VIRUS DI CANCRO	3
VIRUS A UOMO	6	VIRUS DI CEPPO	1
VIRUS A VESTITO	1	VIRUS DI CHIKUNGUNYA	1
VIRUS A WISCONSIN	1	VIRUS DI COMPUTER	2
VIRUS AD ESSERE	1	VIRUS DI COPIATURA	1
VIRUS AD ORA	2	VIRUS DI COSIDETTA	1
VIRUS AD PASSEGGERO	1	VIRUS DI CROCIERA	2
VIRUS AD PATOGENICITA'	2	VIRUS DI DENGUE	1
VIRUS AD RNA	1	VIRUS DI DESIDERIO	1
VIRUS ATTRAVERSO RAPPORTO	1	VIRUS DI DIPENDENTE	1
VIRUS ATTRAVERSO TRASFUSIONE	1	VIRUS DI EBOLA	2
VIRUS ATTRAVERSO VIAGGIO	1	VIRUS DI ENCEFALITE	2
VIRUS CON CARATTERISTICA	2	VIRUS DI ENSOLI	1
VIRUS CON EFFETTO	1	VIRUS DI EPATITE	38
VIRUS CON FARMACO	1	VIRUS DI EPATITI	1
VIRUS CON FUNZIONE	1	VIRUS DI EPIDEMIA	4
VIRUS CON GENIO	1	VIRUS DI ESSERE	1
VIRUS CON INVOLUCRO	1	VIRUS DI FAMIGLIA	8
VIRUS CON MARGINE	1	VIRUS DI FEBBRE	4
VIRUS CON RATTO	1	VIRUS DI GENE	1
VIRUS DA ANIMALE	8	VIRUS DI GENERE	1
VIRUS DA ARIA	1	VIRUS DI GRUPPO	1
VIRUS DA ASIA	1	VIRUS DI H	1
VIRUS DA BESTIA	1	VIRUS DI H5N1	2
VIRUS DA CARNE	1	VIRUS DI HERPES	4
VIRUS DA CINA	1	VIRUS DI HIV	67
VIRUS DA CONTATTO	1	VIRUS DI IMMUNODEFICIENZA	3
VIRUS DA CORRISPONDENTE	2	VIRUS DI INFANZIA	1
VIRUS DA FINE	1	VIRUS DI INFLUENZA	128
VIRUS DA GRUPPO	1	VIRUS DI MAIALA	1
VIRUS DA GUSTO	1	VIRUS DI MAIALE	5
VIRUS DA INVIATO	2	VIRUS DI MALARIA	1
VIRUS DA MAMMA	1	VIRUS DI MALATTIA	2
VIRUS DA NORTHUMBERLAND	1	VIRUS DI MENINGITE	1
VIRUS DA ORGANISMO	1	VIRUS DI MORBILLIVIRUS	1
VIRUS DA PAESE	1	VIRUS DI MORBILLO	2
VIRUS DA POLLO	1	VIRUS DI MOSAICO	1
VIRUS DA POPOLAZIONE	1	VIRUS DI MUCCA	2

VIRUS DI NILO	9	VIRUS IN CITTA'	1
VIRUS DI PANDEMIA	3	VIRUS IN COLONIA	1
VIRUS DI PANDEMIE	1	VIRUS IN CONFRONTO	1
VIRUS DI PAPILLOMA	11	VIRUS IN COREA	1
VIRUS DI PASSATO	1	VIRUS IN CORPO	4
VIRUS DI PATATA	1	VIRUS IN EMISFERO	1
VIRUS DI PAURA	2	VIRUS IN EX	1
VIRUS DI POLIO	1	VIRUS IN FARINGE	1
VIRUS DI POLIOMELITE	1	VIRUS IN FASE	1
VIRUS DI POLIOMIELITE	4	VIRUS IN FETO	1
VIRUS DI POLLO	78	VIRUS IN MALATO	1
VIRUS DI POLMONITE	29	VIRUS IN MIRINO	1
VIRUS DI PREGIUDIZIO	1	VIRUS IN MOMENTO	1
VIRUS DI PRIMATE	1	VIRUS IN MONDO	3
VIRUS DI RAFFREDDORE	33	VIRUS IN ORGANISMO	4
VIRUS DI SARCOMA	1	VIRUS IN OSPEDALE	1
VIRUS DI SARS	52	VIRUS IN PAESE	6
VIRUS DI SCIMMIA	1	VIRUS IN PANCIONE	1
VIRUS DI SECOLO	2	VIRUS IN PAZIENTE	1
VIRUS DI SINDROME	4	VIRUS IN PRELIEVO	1
VIRUS DI SPAGNOLA	10	VIRUS IN PROVINCIA	1
VIRUS DI STOMATITE	1	VIRUS IN RETINA	1
VIRUS DI TERRORE	1	VIRUS IN SANGUE	6
VIRUS DI TIPO	5	VIRUS IN SCIMPANZE'	1
VIRUS DI TUBERCOLOSO	1	VIRUS IN SUINO	2
VIRUS DI TUMORE	3	VIRUS IN TACCHINO	1
VIRUS DI UCCELLO	8	VIRUS IN TEMPO	1
VIRUS DI UOMO	1	VIRUS IN TERRITORIO	1
VIRUS DI VAIOLO	19	VIRUS IN ULTIMO	1
VIRUS DI VARIANTE	1	VIRUS IN VOLATILE	1
VIRUS DI VCJD	1	VIRUS INSIEME AGLIO	1
VIRUS DI VIA	2	VIRUS MENO AGGRESSIVO	2
VIRUS DI VOLATILE	6	VIRUS MENO CATTIVO	1
VIRUS DI ZANZARA	2	VIRUS SENZA CONFINE	1
VIRUS DURANTE ESPERIMENTO	1	VIRUS SENZA NOME	1
VIRUS IN AFRICA	1	VIRUS SOTTO ACCUSA	1
VIRUS IN ALLEVAMENTO	5	VIRUS SOTTO CONTROLLO	2
VIRUS IN AMBIENTE	4	VIRUS SU ALLEVAMENTO	1
VIRUS IN ANIMALE	1	VIRUS SU BESTIA	1
VIRUS IN ANNO	3	VIRUS SU EBOLA	1
VIRUS IN APPARTAMENTO	1	VIRUS SU ISOLA	1
VIRUS IN CAMPAGNA	1	VIRUS SU UCCELLO	1
VIRUS IN CELLULA	12	VIRUS TRA MADRE	1

Distribuzione per anno degli schemi estratti con virus



vista cronologica degli schemi più frequenti



Nell'esempio 4 è illustrato il risultato della ricerca di uno schema simile al precedente dove il sostantivo pivot è *alterazione*.

ALTERAZIONE [352]

ALTERAZIONE DI RITMO	14	ALTERAZIONE DI TESSUTO	2
ALTERAZIONE DI SISTEMA	11	ALTERAZIONE DI TIMBRO	2
ALTERAZIONE DI DNA	9	ALTERAZIONE DI TONO	2
ALTERAZIONE DI METABOLISMO	9	ALTERAZIONE DI UMORE	2
ALTERAZIONE DI COSCIENZA	8	ALTERAZIONE IN COLORE	2
ALTERAZIONE DI GENE	8	ALTERAZIONE IN GENIO	2
ALTERAZIONE DI FUNZIONE	7	ALTERAZIONE IN PRODUZIONE	2
ALTERAZIONE DI GENIO	7	ALTERAZIONE A ADOTTABILITA'	1
ALTERAZIONE DI SONNO	7	ALTERAZIONE A BASE	1
ALTERAZIONE DI SVILUPPO	7	ALTERAZIONE A CROMOSOMA	1
ALTERAZIONE DI CICLO	5	ALTERAZIONE A DENTE	1
ALTERAZIONE DI COMPORTAMENTO	5	ALTERAZIONE A DISCENDENZA	1
ALTERAZIONE DI CROMOSOMA	4	ALTERAZIONE A INTERNO	1
ALTERAZIONE DI LIVELLO	4	ALTERAZIONE A NEOPLASICA	1
ALTERAZIONE DI MECCANISMO	4	ALTERAZIONE A PARKINSON	1
ALTERAZIONE DI MOVIMENTO	4	ALTERAZIONE A SCHELETRO	1
ALTERAZIONE DI STATO	4	ALTERAZIONE DA BUCO	1
ALTERAZIONE DI ATTIVITA'	3	ALTERAZIONE DA PRO	1
ALTERAZIONE DI CELLULA	3	ALTERAZIONE DA PUNTO	1
ALTERAZIONE DI ECOSISTEMA	3	ALTERAZIONE DA SEDENTARIETA'	1
ALTERAZIONE DI EQUILIBRIO	3	ALTERAZIONE DI 22OC	1
ALTERAZIONE DI GRASSO	3	ALTERAZIONE DI 4000OC	1
ALTERAZIONE DI MEMORIA	3	ALTERAZIONE DI ABILITA'	1
ALTERAZIONE DI NERVO	3	ALTERAZIONE DI ACQUA	1
ALTERAZIONE DI PERSONALITA'	3	ALTERAZIONE DI AFFLITTO	1
ALTERAZIONE DI PROCESSO	3	ALTERAZIONE DI ALIMENTO	1
ALTERAZIONE DI PROTEINA	3	ALTERAZIONE DI ALITO	1
ALTERAZIONE DI QUALITA'	3	ALTERAZIONE DI ALVO	1
ALTERAZIONE DI STRUTTURA	3	ALTERAZIONE DI AMIGDALA	1
ALTERAZIONE IN FUNZIONAMENTO	3	ALTERAZIONE DI ANCA	1
ALTERAZIONE A CARICO	2	ALTERAZIONE DI ANGLOAMERICANI	1
ALTERAZIONE A LIVELLO	2	ALTERAZIONE DI AORTA	1
ALTERAZIONE A METABOLISMO	2	ALTERAZIONE DI ASPETTO	1
ALTERAZIONE A ORIGINE	2	ALTERAZIONE DI ASSE	1
ALTERAZIONE DI AMBIENTE	2	ALTERAZIONE DI ATMOSFERA	1
ALTERAZIONE DI APPARATO	2	ALTERAZIONE DI BIOLOGIA	1
ALTERAZIONE DI CAMPO	2	ALTERAZIONE DI CANALE	1
ALTERAZIONE DI CERVELLO	2	ALTERAZIONE DI CANCRO	1
ALTERAZIONE DI IPPOCAMPO	2	ALTERAZIONE DI CAPACITA'	1
ALTERAZIONE DI LIQUIDO	2	ALTERAZIONE DI CARICO	1
ALTERAZIONE DI MICROFLORA	2	ALTERAZIONE DI CARTELLA	1
ALTERAZIONE DI MUCOSA	2	ALTERAZIONE DI CIRCOLAZIONE	1
ALTERAZIONE DI NEURONI	2	ALTERAZIONE DI CIRCUITO	1
ALTERAZIONE DI ORECCHIO	2	ALTERAZIONE DI CLIMA	1
ALTERAZIONE DI ORGANISMO	2	ALTERAZIONE DI COAGULAZIONE	1
ALTERAZIONE DI PARTE	2	ALTERAZIONE DI COLORE	1
ALTERAZIONE DI PATRIMONIO	2	ALTERAZIONE DI CONCENTRAZIONE	1
ALTERAZIONE DI PRESSIONE	2	ALTERAZIONE DI CONNESSIONE	1
ALTERAZIONE DI RAPPORTO	2	ALTERAZIONE DI CONTATTO	1
ALTERAZIONE DI RAPPRESENTAZIONE	2	ALTERAZIONE DI CONTENUTO	1
ALTERAZIONE DI SOGGETTIVITA'	2	ALTERAZIONE DI CONTRAZIONE	1
ALTERAZIONE DI TALASSEMIA	2	ALTERAZIONE DI CONTROLLO	1

ALTERAZIONE DI CORNEA	1	ALTERAZIONE DI RIPRODUTTIVIT...	1
ALTERAZIONE DI CROMOSOMI	1	ALTERAZIONE DI RISULTATO	1
ALTERAZIONE DI CUORE	1	ALTERAZIONE DI SANGUE	1
ALTERAZIONE DI DEGLUTIZIONE	1	ALTERAZIONE DI SCHELETRO	1
ALTERAZIONE DI DIALOGO	1	ALTERAZIONE DI SEGNALE	1
ALTERAZIONE DI DOLCE	1	ALTERAZIONE DI SENO	1
ALTERAZIONE DI EMBRIONE	1	ALTERAZIONE DI SENSIBILITA'	1
ALTERAZIONE DI EMISFERO	1	ALTERAZIONE DI SESSUALITA'	1
ALTERAZIONE DI ENZIMA	1	ALTERAZIONE DI SMALTO	1
ALTERAZIONE DI FATTEZZA	1	ALTERAZIONE DI STATICA	1
ALTERAZIONE DI FIBRA	1	ALTERAZIONE DI STILE	1
ALTERAZIONE DI FLUSSO	1	ALTERAZIONE DI TERMOREGOLAZIONE	1
ALTERAZIONE DI FORMAZIONE	1	ALTERAZIONE DI TIROIDE	1
ALTERAZIONE DI FUNZIONALITA'	1	ALTERAZIONE DI VALORE	1
ALTERAZIONE DI GENOMA	1	ALTERAZIONE DI VASO	1
ALTERAZIONE DI GIUNTURA	1	ALTERAZIONE DI VENA	1
ALTERAZIONE DI GLOBULO	1	ALTERAZIONE IN CELLULA	1
ALTERAZIONE DI GUSTO	1	ALTERAZIONE IN CIRCUITO	1
ALTERAZIONE DI HABITAT	1	ALTERAZIONE IN CODICE	1
ALTERAZIONE DI IMMAGINE	1	ALTERAZIONE IN DNA	1
ALTERAZIONE DI IMPIANTO	1	ALTERAZIONE IN FETO	1
ALTERAZIONE DI IMPRINTING	1	ALTERAZIONE IN GENE	1
ALTERAZIONE DI LAMINA	1	ALTERAZIONE IN INSORGENZA	1
ALTERAZIONE DI LAVORO	1	ALTERAZIONE IN LIVELLO	1
ALTERAZIONE DI LIBIDO	1	ALTERAZIONE IN LOX1	1
ALTERAZIONE DI LINGUAGGIO	1	ALTERAZIONE IN METABOLISMO	1
ALTERAZIONE DI LIPIDICO	1	ALTERAZIONE IN MITOCONDRI	1
ALTERAZIONE DI LIPOPROTEINE	1	ALTERAZIONE IN NEURALI	1
ALTERAZIONE DI MATERIALE	1	ALTERAZIONE IN NUCLEO	1
ALTERAZIONE DI MELANOCITI	1	ALTERAZIONE IN ONCOGENI	1
ALTERAZIONE DI MICRORGANISMO	1	ALTERAZIONE IN PERCEZIONE	1
ALTERAZIONE DI MIMICA	1	ALTERAZIONE IN PESCE	1
ALTERAZIONE DI MOLECOLA	1	ALTERAZIONE IN RAGAZZO	1
ALTERAZIONE DI MOTILITA'	1	ALTERAZIONE IN SINAPSI	1
ALTERAZIONE DI MUSCOLARE	1	ALTERAZIONE IN STRUTTURA	1
ALTERAZIONE DI MUSCOLATURA	1	ALTERAZIONE IN TALAMO	1
ALTERAZIONE DI NEUROTRASMTTITORI	1	ALTERAZIONE IN UNGHIA	1
ALTERAZIONE DI NUMERO	1		
ALTERAZIONE DI OCCHIO	1		
ALTERAZIONE DI ONDA	1		
ALTERAZIONE DI OSSO	1		
ALTERAZIONE DI PARAMETRO	1		
ALTERAZIONE DI PARETE	1		
ALTERAZIONE DI PAVIMENTO	1		
ALTERAZIONE DI PELLE	1		
ALTERAZIONE DI PERCEZIONE	1		
ALTERAZIONE DI PESO	1		
ALTERAZIONE DI PRESTAZIONE	1		
ALTERAZIONE DI PROFILO	1		
ALTERAZIONE DI PROPRIETA'	1		
ALTERAZIONE DI PROTEINE	1		
ALTERAZIONE DI PSICHE	1		
ALTERAZIONE DI QUADRO	1		
ALTERAZIONE DI QUANTITA'	1		
ALTERAZIONE DI REALTA'	1		
ALTERAZIONE DI RIFRAZIONE	1		



Infine nell'esempio 5 di utilizzo delle procedure dell'ambiente di sviluppo Text Miner sono stati recuperati tutti i sostantivi legati al lemma *trapianto* dalla preposizione *di*.

TRAPIANTO [1006]	7 trapianto di faccia
1 trapianti del '97	2 trapianti del fegato
1 trapianti dell' aprile	1 trapianti di fegati
9 trapianti di arti	74 trapianti di fegato
2 trapianto degli arti	11 trapianto del fegato
7 trapianto di arti	79 trapianto di fegato
2 trapianti di articolazioni	2 trapianti di geni
1 trapianti di arto	14 trapianto di geni
5 trapianti di bari	1 trapianti del genere
2 trapianti di bergamo	1 trapianto del genere
1 trapianto di braccio	1 trapianti di genitali
2 trapianto di bulbo	1 trapianto delle gonadi
1 trapianti di capelli	2 trapianto di ieri
5 trapianto di capelli	2 trapianto di insulae
2 trapianti di cardiocirurgia	5 trapianti di intestino
1 trapianto di cartilagine	1 trapianto dell' intestino
1 trapianto della caviglia	7 trapianto di intestino
1 trapianto di caviglia	1 trapianto delle isole
1 trapianti delle cellule	1 trapianto di isole
11 trapianti di cellule	1 trapianti dell' istituto
2 trapianto delle cellule	1 trapianto dell' istituto
79 trapianto di cellule	2 trapianto di liono
2 trapianto di cervello	1 trapianto di labbra
10 trapianti di cornea	2 trapianto della laringe
3 trapianti di cornee	3 trapianto di laringe
3 trapianto della cornea	3 trapianto di lingua
15 trapianto di cornea	3 trapianto di mandibola
3 trapianto di cornee	2 trapianti della mano
2 trapianto di corpo	1 trapianti di mani
29 trapianti di cuore	11 trapianti di mano
7 trapianto del cuore	8 trapianto della mano
59 trapianto di cuore	4 trapianto delle mani
2 trapianti di cuori	10 trapianto di mani
1 trapianti di cute	42 trapianto di mano
2 trapianto di cute	1 trapianti di meo
1 trapianti di domani	20 trapianti di midollo
1 trapianti di epidermide	4 trapianto del midollo
1 trapianto della faccia	65 trapianto di midollo

1	trapianto di miociti	1	trapianto del polmone
4	trapianti di modena	1	trapianto di polmoni
1	trapianti di multiviscerale	1	trapianto di protesi
1	trapianto del neonato	1	trapianto delle regioni
3	trapianto di neuroni	1	trapianti del rene
1	trapianto di nuclei	50	trapianti di rene
1	trapianto di nuova	5	trapianto del rene
1	trapianto di nuove	49	trapianto di rene
1	trapianto dei nuovi	4	trapianti di reni
19	trapianti di organi	2	trapianto dei reni
5	trapianti di organo	10	trapianto di reni
4	trapianto degli organi	1	trapianti dei riuniti
6	trapianto dell' organo	2	trapianti di roma
22	trapianto di organi	4	trapianto di sangue
5	trapianto di organo	1	trapianto di sette
2	trapianti degli ospedali	1	trapianto dei soli
2	trapianti dell' ospedale	1	trapianto di spermatogoni
1	trapianto dell' osso	12	trapianto di staminali
1	trapianto di osso	4	trapianti di tessuti
1	trapianto delle ovaie	2	trapianto di tessuti
3	trapianto di ovaia	4	trapianto di tessuto
6	trapianto di ovaie	1	trapianti di testa
9	trapianto di ovaio	5	trapianto della testa
1	trapianti di padova	4	trapianto di testa
1	trapianto di padova	2	trapianti di torino
13	trapianti di palermo	1	trapianti di tuba
3	trapianto di pancreas	2	trapianti degli ultimi
1	trapianto di parti	1	trapianti dell' università
1	trapianto di pavia	1	trapianto di uore
4	trapianti di pelle	2	trapianto dell' utero
6	trapianto di pelle	2	trapianto di utero
2	trapianti del piemonte	1	trapianti di venezia
1	trapianti di pittsburgh	2	trapianto di viso
5	trapianti del policlinico	2	trapianti del volto
6	trapianti di polmone	1	trapianto del volto
1	trapianto dei polmoni	2	trapianto di volto

Es. 5

La procedura di ricerca di schemi linguistici fornisce per ogni archivio analizzato un file che contiene, tra le altre informazioni, il riferimento logico dell'unità cui appartiene ogni contesto recuperato. I file con i risultati sono stati fusi tramite la procedura di "merge" disponibile e successivamente

elaborati per ottenere una mappa dei sostantivi associati al lemma *trapianto* per ogni unità logica.

A titolo di esempio riportiamo in Tab. 3 una porzione dei risultati ottenuti.

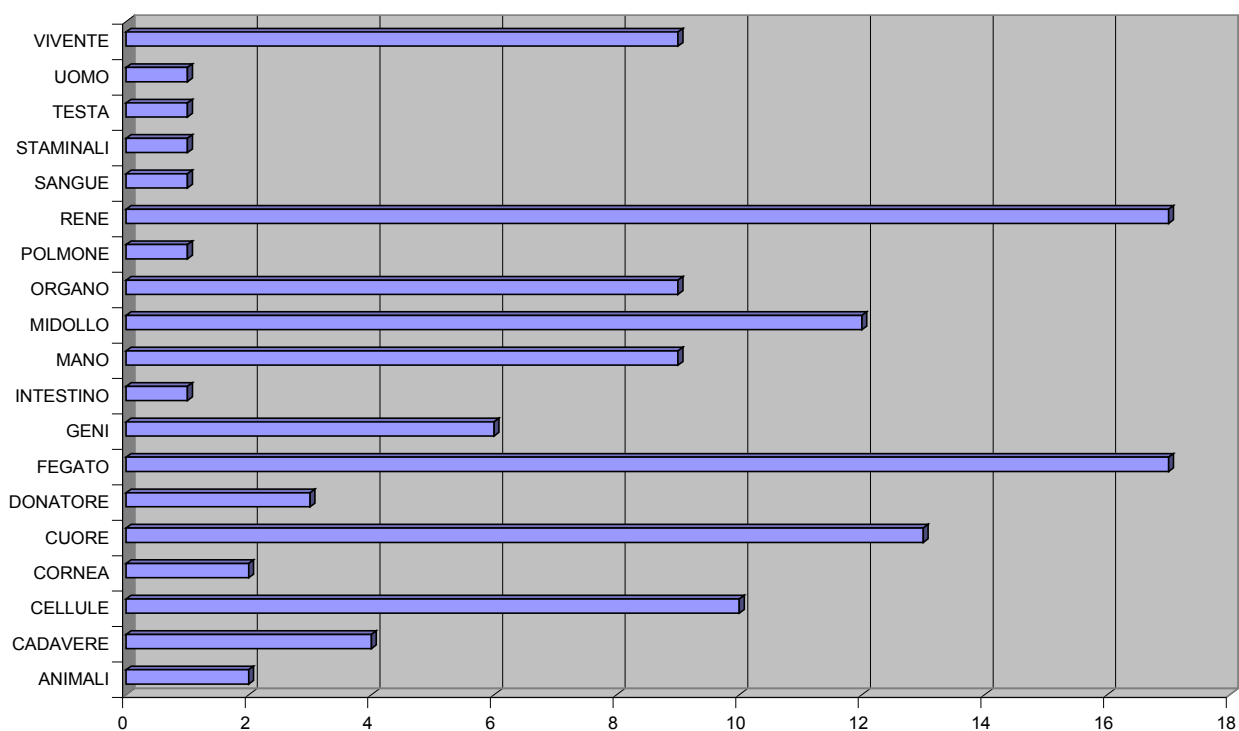
RIF	SOST2	TOT			
0C-01APRILE_00	MIDOLLO	1		MIDOLLO	2
0C-01MAGGIO_00	MANO	2		STAMINALI	1
0C-01NOVEMBRE_01	DONATORE	2		VIVENTE	1
0C-01SETTEMBRE_00	CADAVERE	1	0C-04APRILE_03	RENE	1
	FEGATO	1	0C-04LUGLIO_01	ANIMALI	1
	ORGANO	1		CUORE	2
	TESTA	1		CUORE	1
0C-01SETTEMBRE_01	CELLULE	3	0C-05APRILE_00	GENI	2
0C-02APRILE_01	CADAVERE	2	0C-05FEBBRAIO_02	MIDOLLO	1
	ORGANO	2	0C-05LUGLIO_01	CUORE	1
	VIVENTE	4	0C-05LUGLIO_03	ORGANO	1
0C-02GIUGNO_01	ORGANO	1	0C-05MARZO_00	MANO	4
0C-02GIUGNO_03	CELLULE	1	0C-05MARZO_01	FEGATO	1
0C-02LUGLIO_03	ORGANO	1		RENE	2
0C-02MARZO_03	GENI	4		VIVENTE	2
0C-02OTTOBRE_01	MANO	3		VIVENTE	2
0C-02SETTEMBRE_99	CUORE	2	0C-05NOVEMBRE_01	MIDOLLO	2
0C-03GENNAIO_00	UOMO	1		SANGUE	1
0C-03LUGLIO_03	FEGATO	1	0C-05NOVEMBRE_99	CUORE	1
0C-03MAGGIO_00	CORNEA	1		FEGATO	1
0C-03MAGGIO_03	CORNEA	1		RENE	3
	CUORE	1	0C-05OTTOBRE_00	CELLULE	1
	FEGATO	4		MIDOLLO	3
	POLMONE	1	0C-06GIUGNO_01	MIDOLLO	3
	RENE	1	0C-06GIUGNO_03	CUORE	1
0C-03SETTEMBRE_01	CUORE	2		FEGATO	6
0C-04AGOSTO_99	CELLULE	5		ORGANO	2
	RENE	1		RENE	2
	RENE	1	0C-06MARZO_01	CADAVERE	1
0C-04APRILE_02	ANIMALI	1		FEGATO	2
	CUORE	2		INTESTINO	1
	DONATORE	1		ORGANO	1
	FEGATO	1		RENE	5
				RENE	1

TA 3

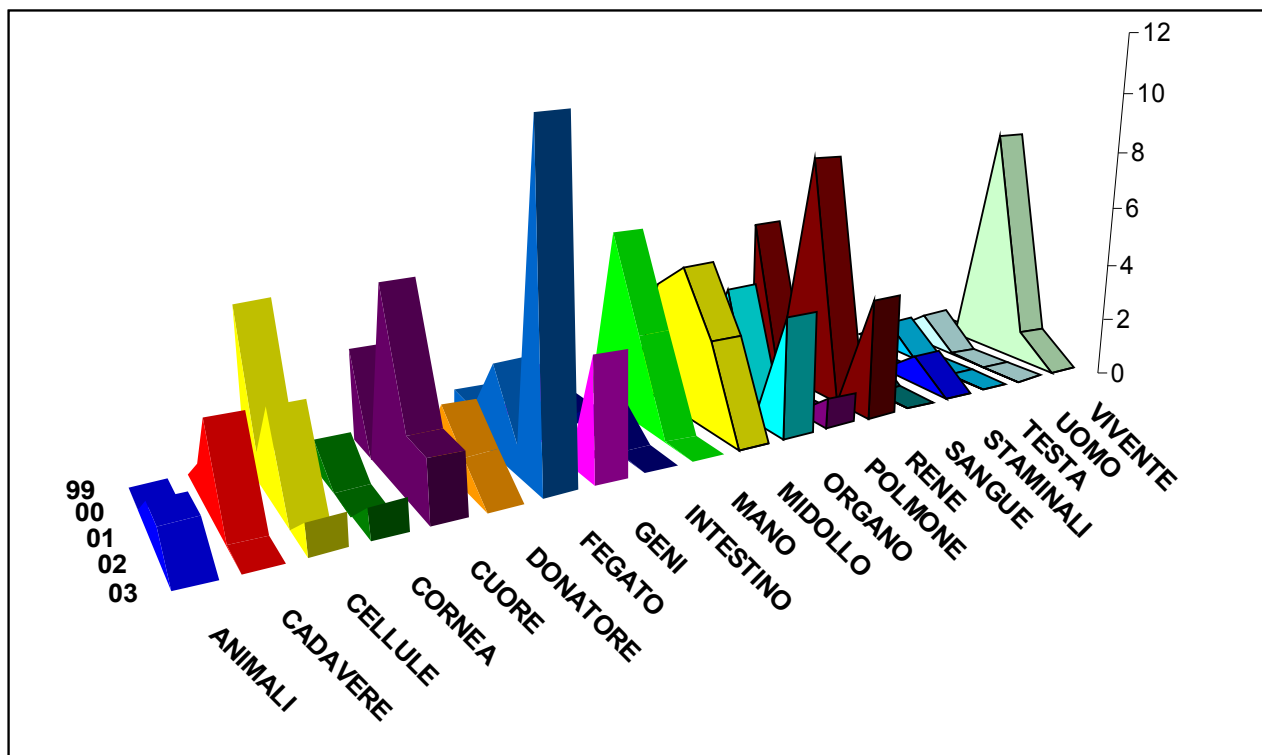
La possibilità di estrarre informazione sotto forma sintetica permette quindi anche di poter analizzare dati testuali come i dati numerici.

Ciò è reso possibile dal fatto che la lemmatizzazione e la categorizzazione permettono di unificare i termini sotto etichette e li rendono quindi trattabili numericamente e statisticamente.

Per cui la tabella precedente permette di disporre i contenuti anche graficamente, come nella figura seguente, dove viene esemplificato l'argomento "trapianto di organo" come viene trattato dal Corriere della Sera negli anni dal 1999 al 2003.



In quest'altro grafico viene aggiunto il controllo sull'anno di pubblicazione dell'articolo.



## Bibliografia

- Ceccotti, M.L. e Sassi, M. (2003), L'archivio elettronico delle opere di Carlo Emilio Gadda. Da redattori a fruitori di un data base testuale, in *Linguistica Computazionale*, Pisa-Roma, Istituti Editoriali e Poligrafici Internazionali (IEPI), pp. 221-250.
- Cinini, A. e Sassi, M. (2004), *Attività nell'ambito del WP1: Prototipo di analisi testuale a supporto del WP2*, IRSIG-ILC-CNR, Pisa.
- Hindle, D. (1994) A Parser for Text Corpora in Atkins, B.T.S., Zampolli, A. (eds.) *Computational Approaches to the Lexicon*, Clarendon Press, Oxford, pp.103-151.
- Lucisano, P. e Piemontese, E. (1988), GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana, in *Scuola e città*, n.3, Firenze, La Nuova Italia Editrice, pp. 110-124.
- Picchi, E. (1997), *DBT 3 - Data Base Testuale: Guida all'uso*, ver. 3.1. Lexis Ricerche s.r.l. su licenza del C.N.R., Roma.
- Picchi, E. (1999), *Informatica e scienze umane: Procedure di analisi testuale* in Zanetti, M.A. (a cura di), *Parola e Immagine*, Firenze, La Nuova Italia Editrice.
- Picchi, E., Ceccotti, M.L., Cucurullo, S., Sassi, M. e Sassolini, E. (2004), *Linguistic Miner: an Italian Linguistic Knowledge System* in *LREC 2004 IV International Conference, Proceedings*, Lisbona, 26-28 maggio 2004, pp.1811-14.
- Sassi, M. (2003), *La consultazione dei corpora costituzionali con il DBT*, in *Diritto alla vita e Diritto all'ambiente nel lessico costituzionale e nella dottrina giuridica. Strumenti e metodi per l'analisi linguistica-concettuale*, con CD-Rom allegato, Firenze, S.T.A.R., pp. 149-154.
- Sassi, M. e Cinini, A. (2006), *Content-analysis dei provvedimenti della sezione disciplinare del C.S.M.*, ILC-IRSIG-CNR, Pisa, S.T.A.R.
- Sezione Disciplinare del Consiglio Superiore della Magistratura, (2004) *Massimario delle decisioni*, Anni 2002-2003, Quaderno n.143 (<http://www.csm.it/pages/quaderni.html>).

## Indice

Introduzione.....	2
Realizzazione degli archivi.....	3
Alcuni esempi di consultazione degli archivi tramite DBT.....	6
<a href="#">Ricerca dei termini che condividono la sottostringa “cardio”</a> .....	6
<a href="#">Il linguaggio divulgativo</a> .....	9
Creazione di indici.....	12
Creazione di archivi tematici.....	16
Linguistic Miner (LM).....	18
Bibliografia.....	31