Metodologie di analisi e caratterizzazione della classe di utenza del servizio di Emeroteca Virtuale al CASPUR*

Marco Scarnò (mscarno@caspur.it)
Donatella Sforzini (d.sforzini@caspur.it)

Bibliocom 2004
Roma, 27 Ottobre

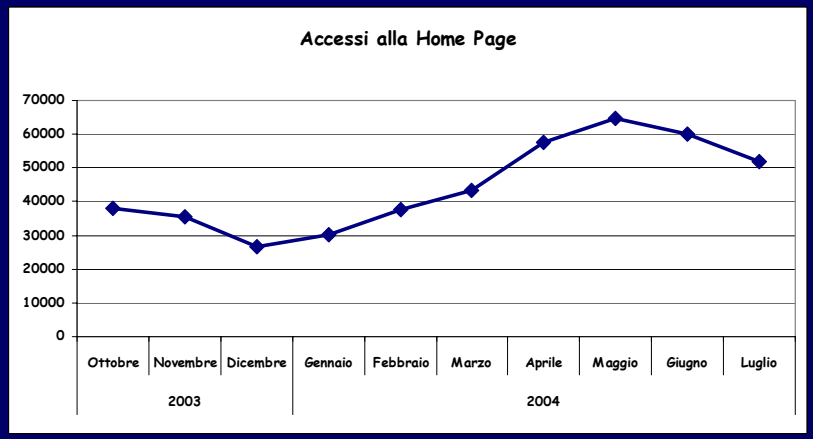
*Consorzio interuniversitario CASPUR, Università di Roma *La Sapienza*

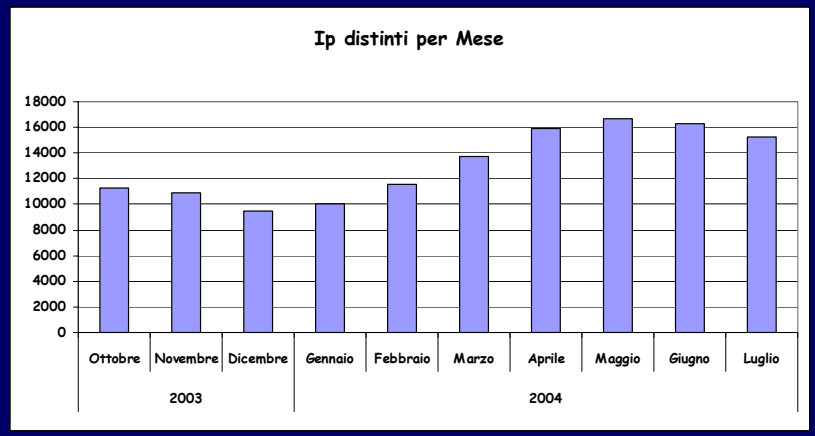
CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

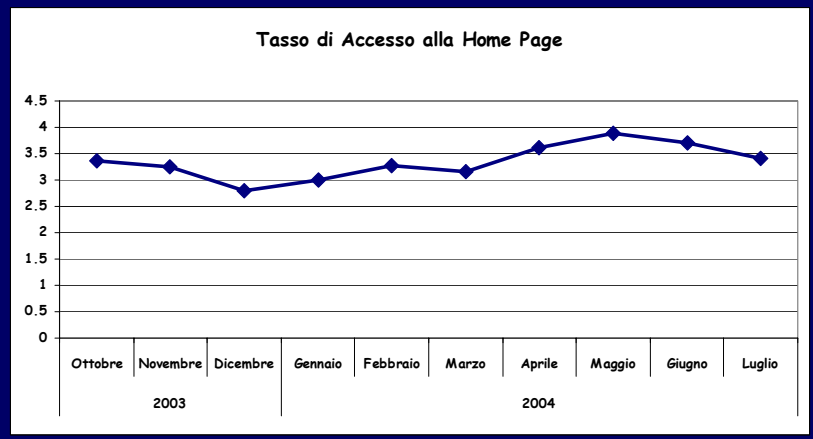
Ambito della Ricerca

Gli accessi al sito…

..da Ottobre 2003 a Luglio 2004:



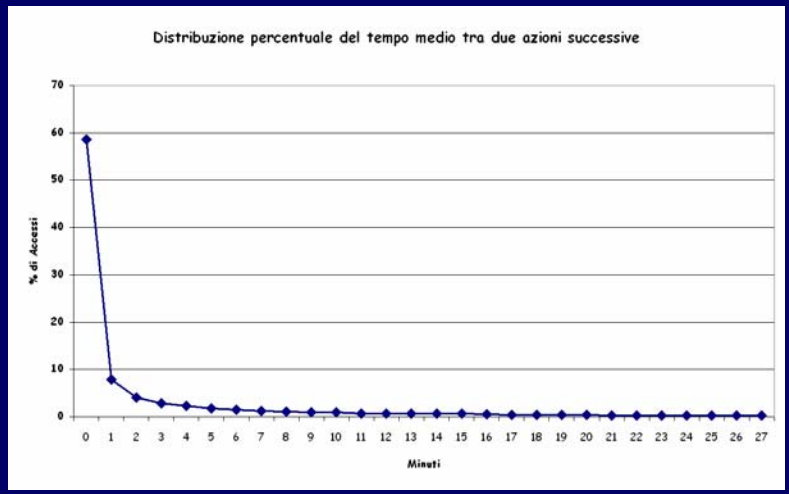


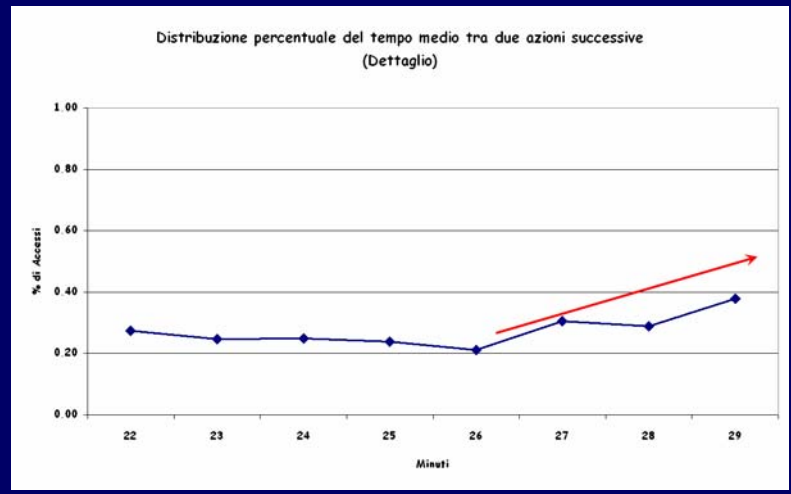


Ma.. quando si è entrati nel sito?

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

..accade che l'IP (il computer) compia più azioni successive…





Lasciando questa *traccia…*

19x.16y.10z.23t - - [01/Nov/2003:07:30:28 +0100] "GET /cgi-bin/search.pl?Database=journals&search_field=oncogenes+and+bladder+cancer&GetSearchResults=Search&fields=Any&canned_search= HTTP/1.1" 200 42145

Definizioni:

- *Contenuti:* insieme delle pagine che il server restituisce all'utente; possono essere o l'indice di una rivista, o l'indice dell'issue, o un abstract o un full text (articolo di una rivista);

- *Attività di Browsing:* insieme delle pagine web di ricerca degli articoli basate sulla visualizzazione degli elenchi delle riviste per lettera alfabetica, per argomento o per editore;

- *Attività di Ricerca:* insieme delle pagine web di ricerca che consentono all'utente di trovare un articolo tramite delle parole chiave;

- *Sessione di Ricerca:* la sessione durante la quale un utente cerca un articolo (si è identifica una nuova sessione ogni qual volta l'utente rimane inattivo per 27 minuti);

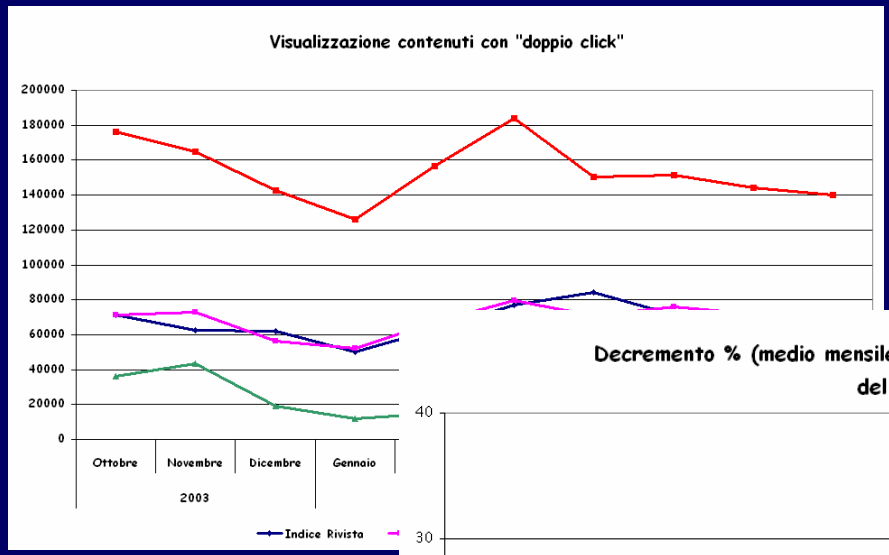
Problema: esiste un metodo *ufficiale* per trattare i dati?

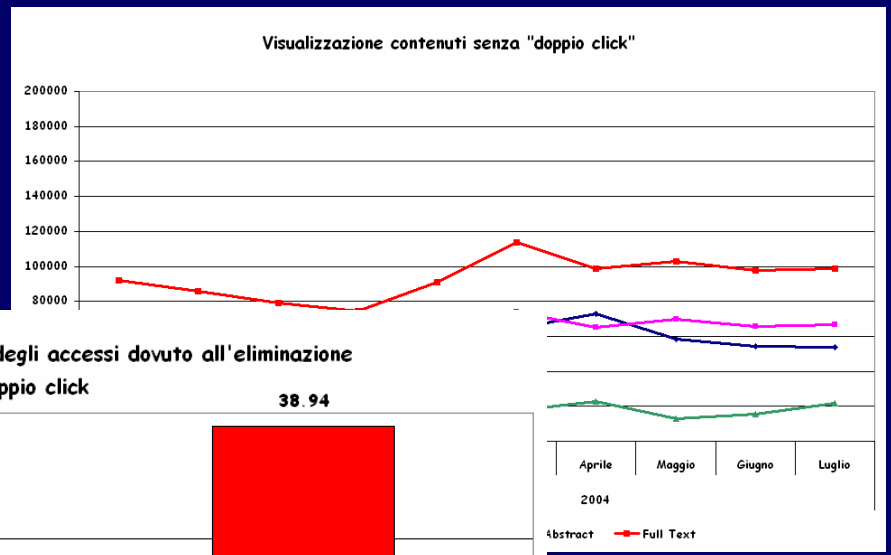
Obiettivo: eliminare i dati non validi

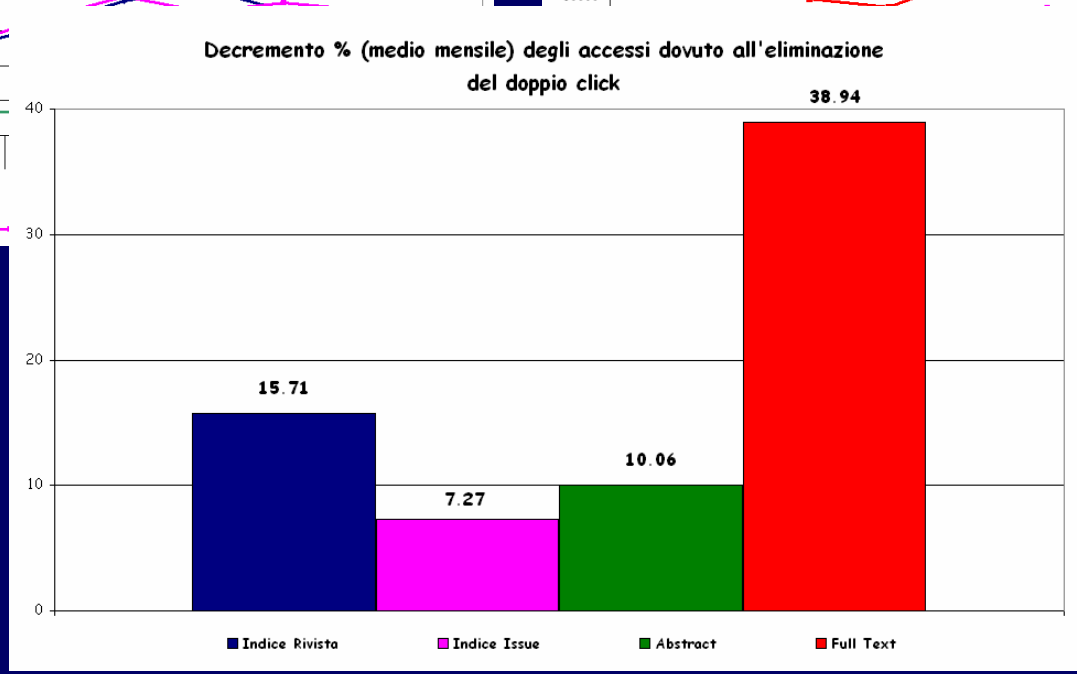
- Eliminazione dei doppi *click;*
- Eliminazione dei *comportamenti anomali*;

Fornire rapporti e analisi non falsati da valori anomali

Pre-trattamento dei dati: il problema del doppio click





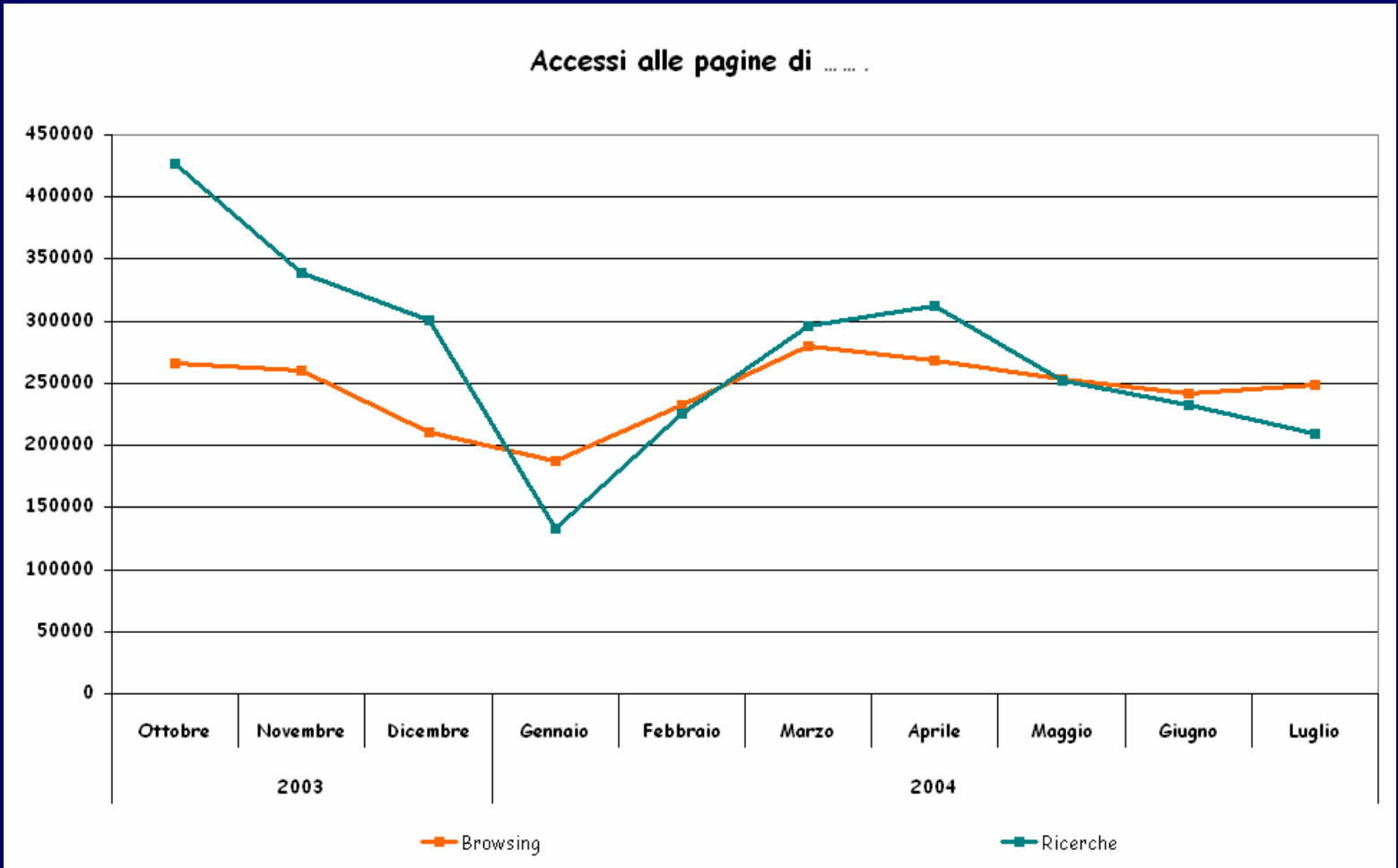


CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

Le attività possibili (e realmente effettuate) sui dati:

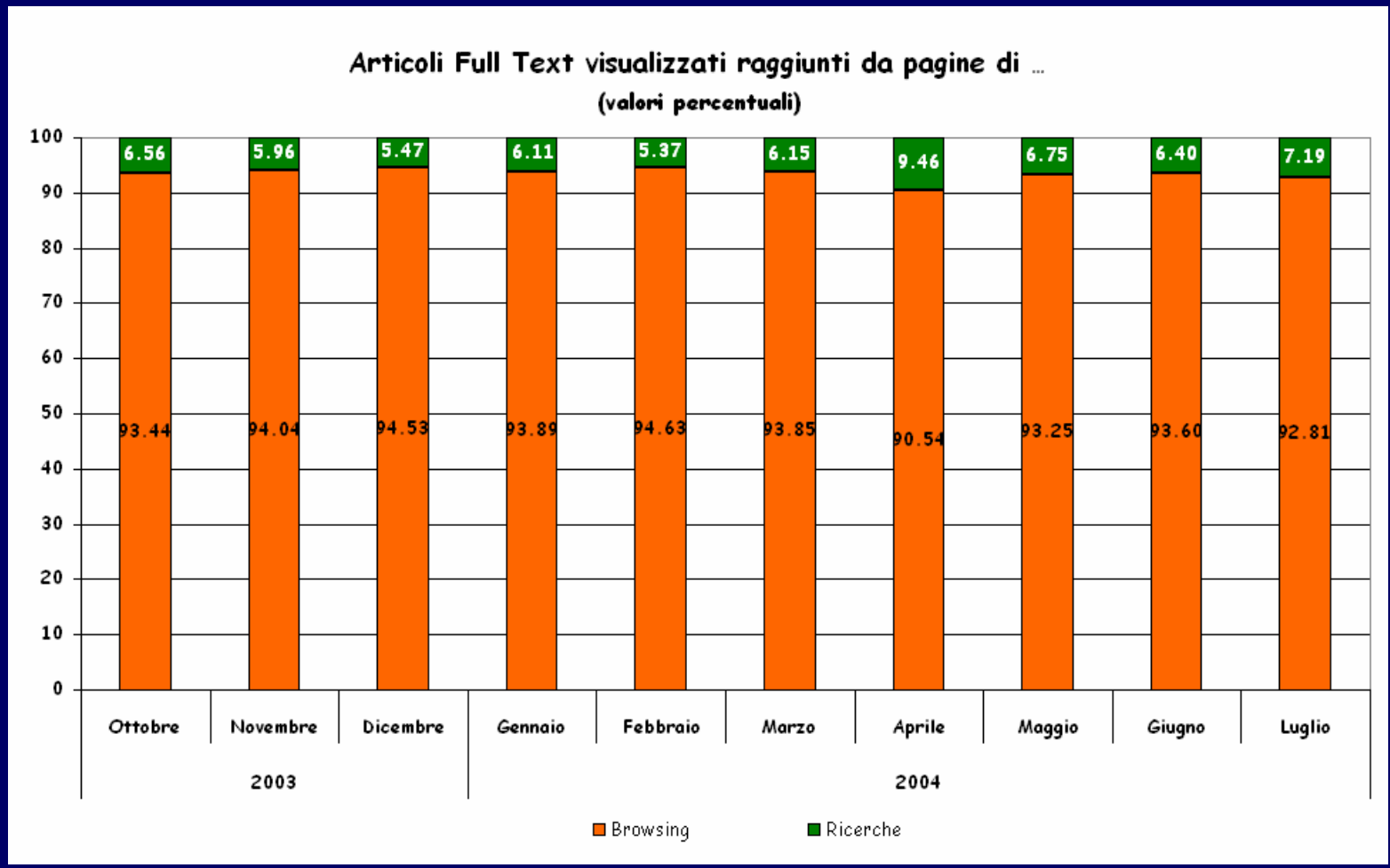
- Istituzionale (rapporti sulle consultazioni ad uso degli *utenti-responsabili*)
- Descrittiva:
  - Attività di browsing
  - Attività di ricerca tramite parole chiave
  - Visualizzazione di Full Text

Come cerco gli articoli che mi interessano? Browsing o Ricerche?



CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

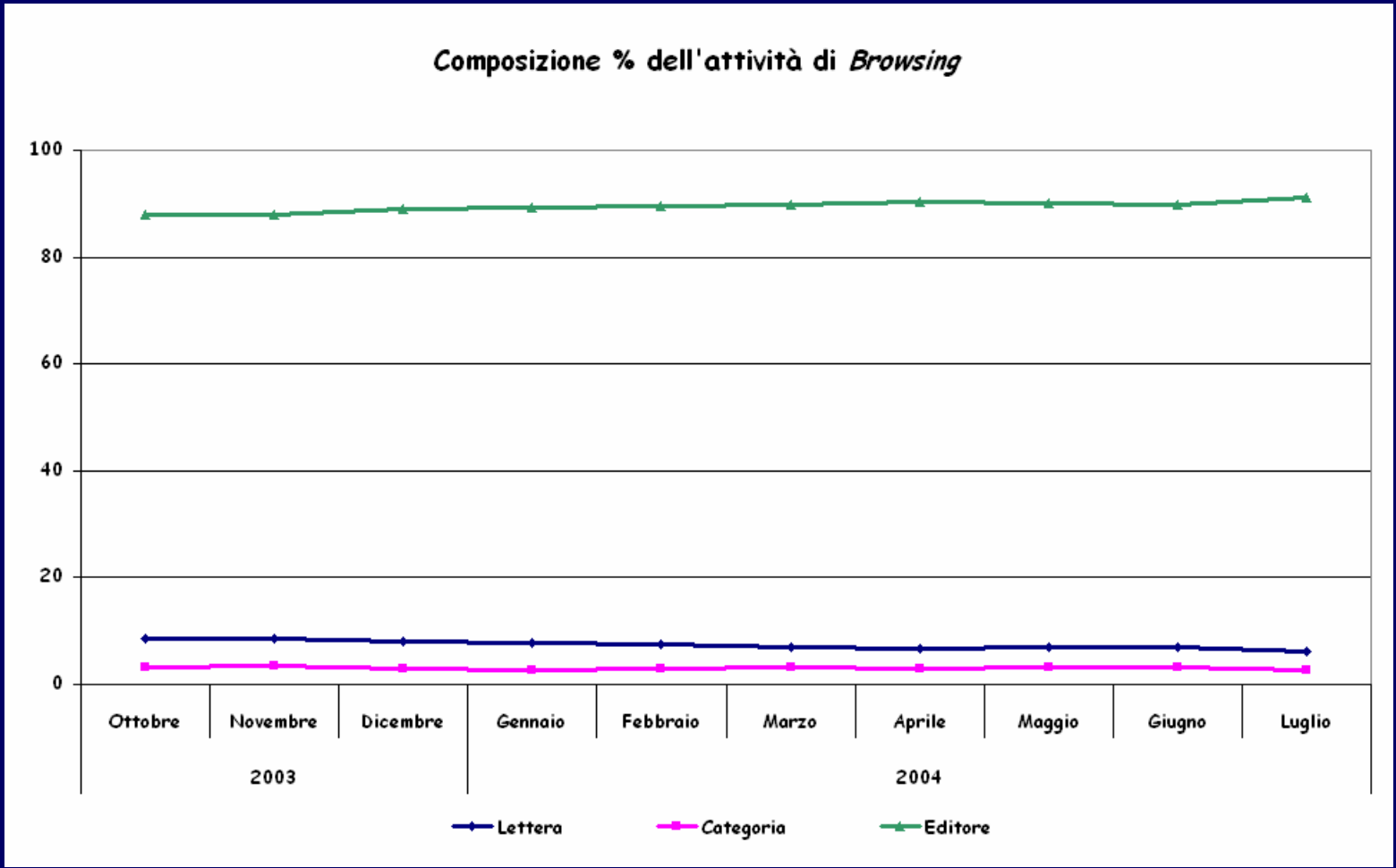
…e quale “strategia” mi aiuta di più a visualizzare gli articoli che cerco?



CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

In particolare nel browsing:

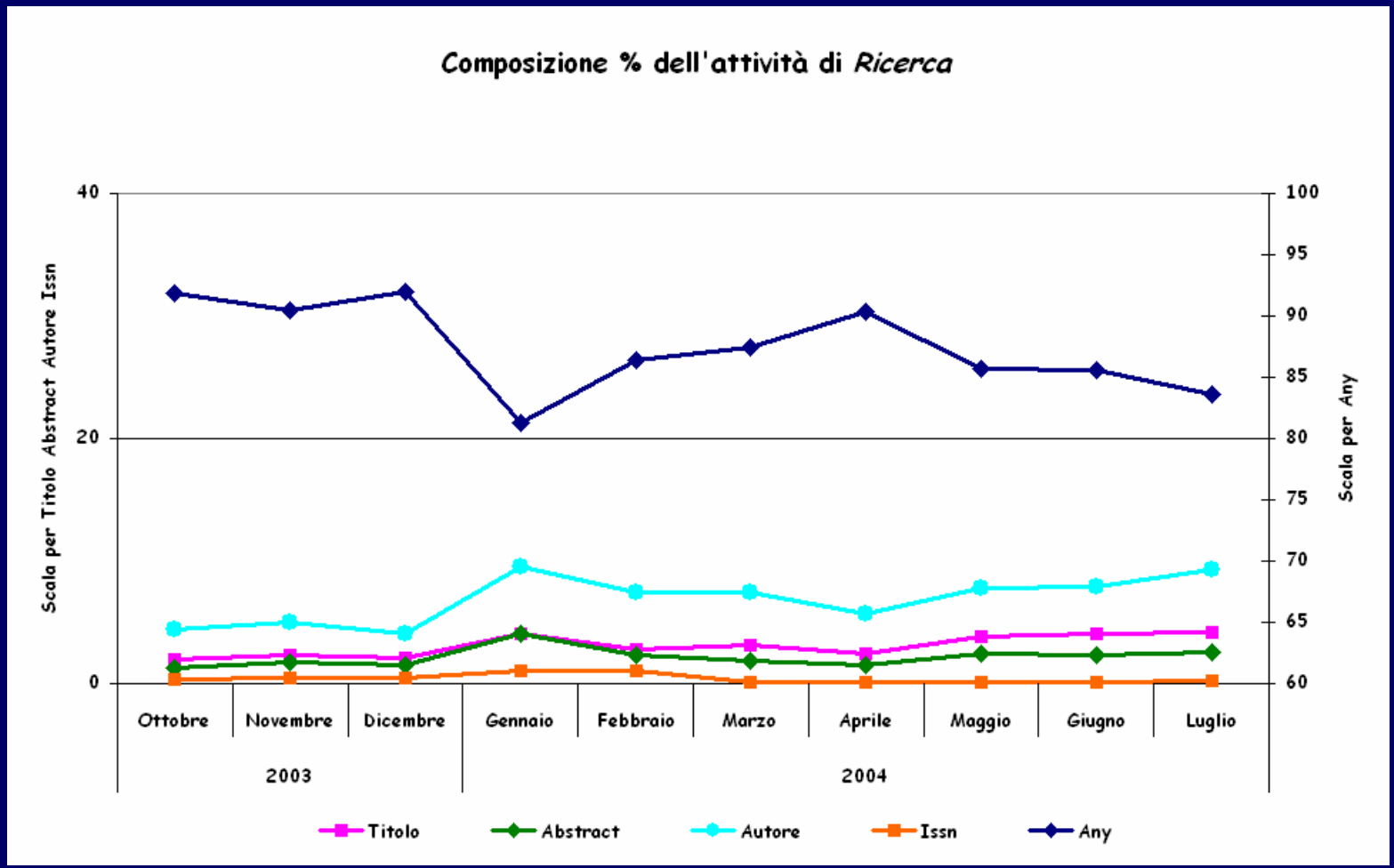
(solo una ricerca su 4 cominciata tramite browsing prosegue con il browsing stesso)



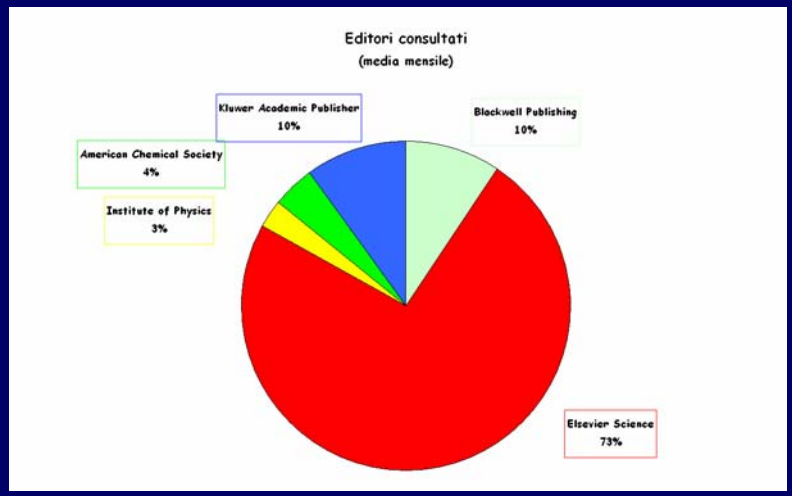
CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

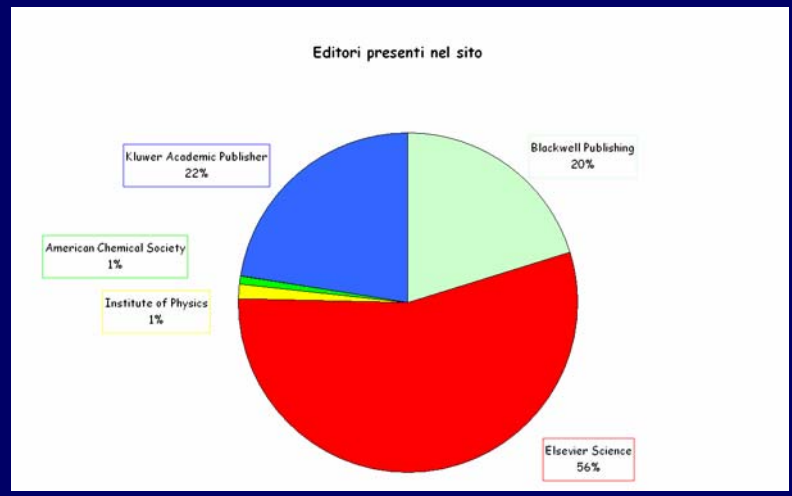
… e con le ricerche per parola chiave:

(quasi tutte le ricerche per chiave proseguono con…… *ricerche per chiave*)



Problema metodologico:
le analisi possono essere riferite o al valore assoluto delle azioni effettuate o alle possibilità che l'emeroteca offre..
…esempio..





Elsevier Science o *American Chemical Society?*

Base: attività di browsing per editore nel periodo

Curiosità..

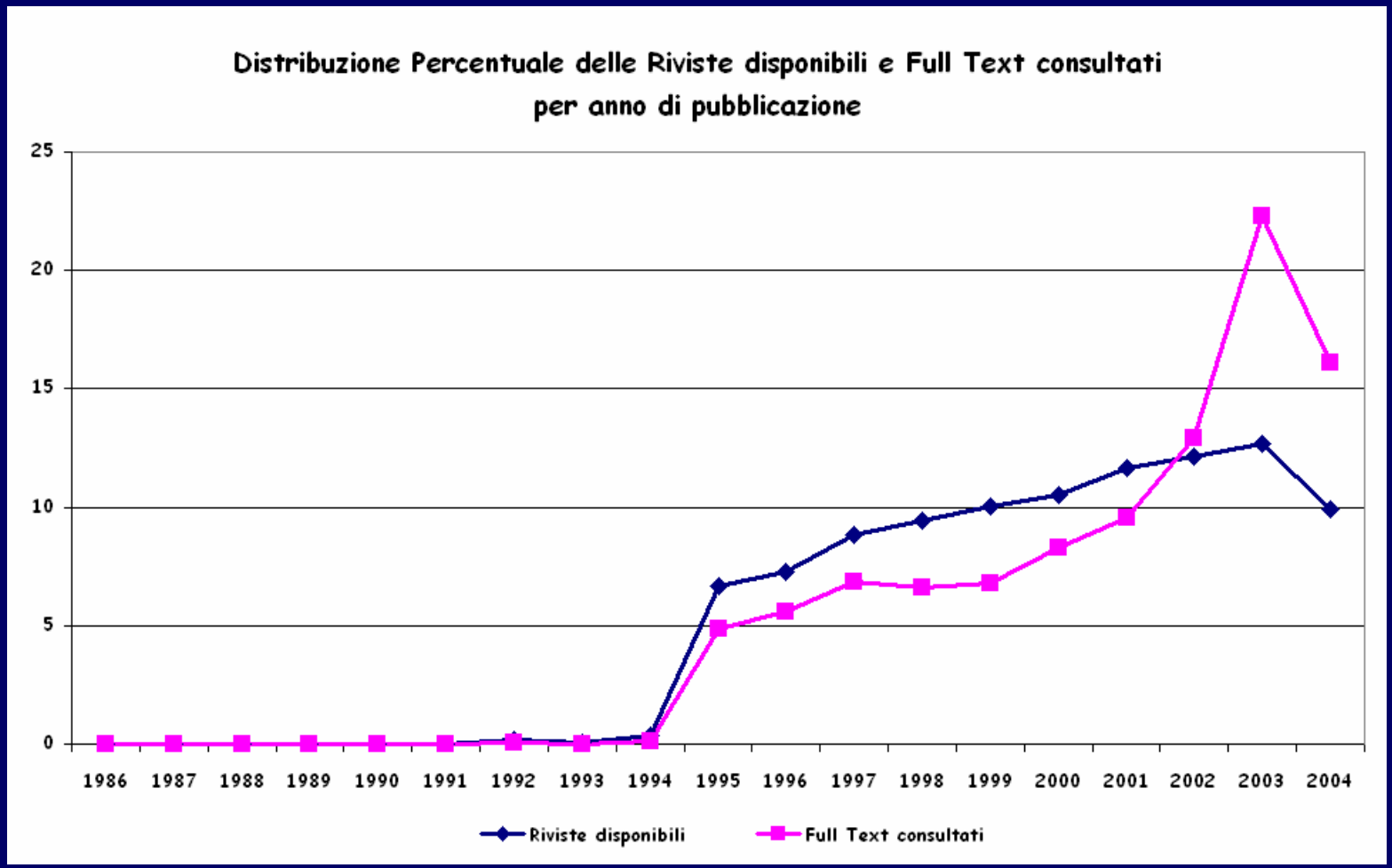
2003		
Ottobre	Novembre	Dicembre
"TOURISM"	"HEAVY OIL"	"PLAUSIBILITY"
"STYRENE AND BUTADIENE"	"PRIME NUMBER"	"SUBMODULAR"
"EBOLA"	"GEOGRAPHY"	"TESI "
"TABU SEARCH"	"MAYERS"	"DEPRESSION AND MIGRAINE"
"JOURNAL OF THERMAL ANALYSIS"	"HEAT ISLAND"	"MONTECUCCO"
"MORTALITY"	"ADOPT"	"CU-AL-NI"
"MERCURY POLLUTION"	"HEALTH SERVICE RESEARCH"	"ISLAM"
"BOUCHAUD"	"HASSANEIN"	AUTHOR=(WARD)
"BITE"	AUTHOR=(HENSEL)	"SYRIA"
"MANHATTAN"	"DOPING"	"BREAKWATER"

2004						
Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio
"TRENDS IN IMMUNOLOGY"	"SCHIZOPHRENIA RESEARCH"	"BLINDNESS"	"USABILITY"	"VERTEBRAL OSTEOMYELITIS"	"BURNOUT"	"RORSCHACH"
"THERMAL DEGRADATION"	"E-COMMERCE "	"BECK"	"CORPORATE GOVERNANCE	"LANGEREIS"	"IMPACT FACTORS"	"HYDROGEL CONTACT LENSES"
"APPENDECTOMY"	"GAS HYDRATES "	"TURNER"	"BIOELECTROCHEMISTRY"	"CHLAMYDIA PNEUMONIAE"	"PROCEEDINGS OF THE SP	"JOURNAL OF ANIMAL ECOLOGY"
CELL	"CHLAMYDIA"	"CARBON/CARBON"	"GIORDANO S"	"HEPATOCELLULAR CARCINOMA"	"EMDR"	AUTHOR=("VALSECCHI")
"OSMOSIS"	"AMERICAN ECONOMIC REVIEW"	"ONCOGENE"	"PEGYLATION"	"INDIAN J. CHEM."	"REED"	"BREAST UNIT"
	"EMAIL"	"NEUROLOGY"	"MIKLAVCIC"	"BLOOM SYNDROME"	"DONATI I."	"PNAS"
	"CAPSICUM"	"ATP"	"MONTELEONE F"	"GDNF"	"ROWING"	"NURSING"
	"ALCHEMY"	"CANCER EPIDEMIOLOGY"	"CITATION INDEX"	"DARK MATTER"	"ATKINSON"	"ENDORIBONUCLEASES"
	"CERAMIC GLASS"	"ASTHMA"	"CIRCULATION"	"NEPHRON"	"THE USE OF TRANSFORM	"X-RAY FLUORESCENCE SPECTROSCOPY"
	"NEWTON"	"CURRENT ACCOUNT"	"ANTICANCER"	"MINERVA"	"DIAPHRAGM WALLS"	

10 parole più cercate
(sono state digitate più di 10,000 volte in un mese)

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

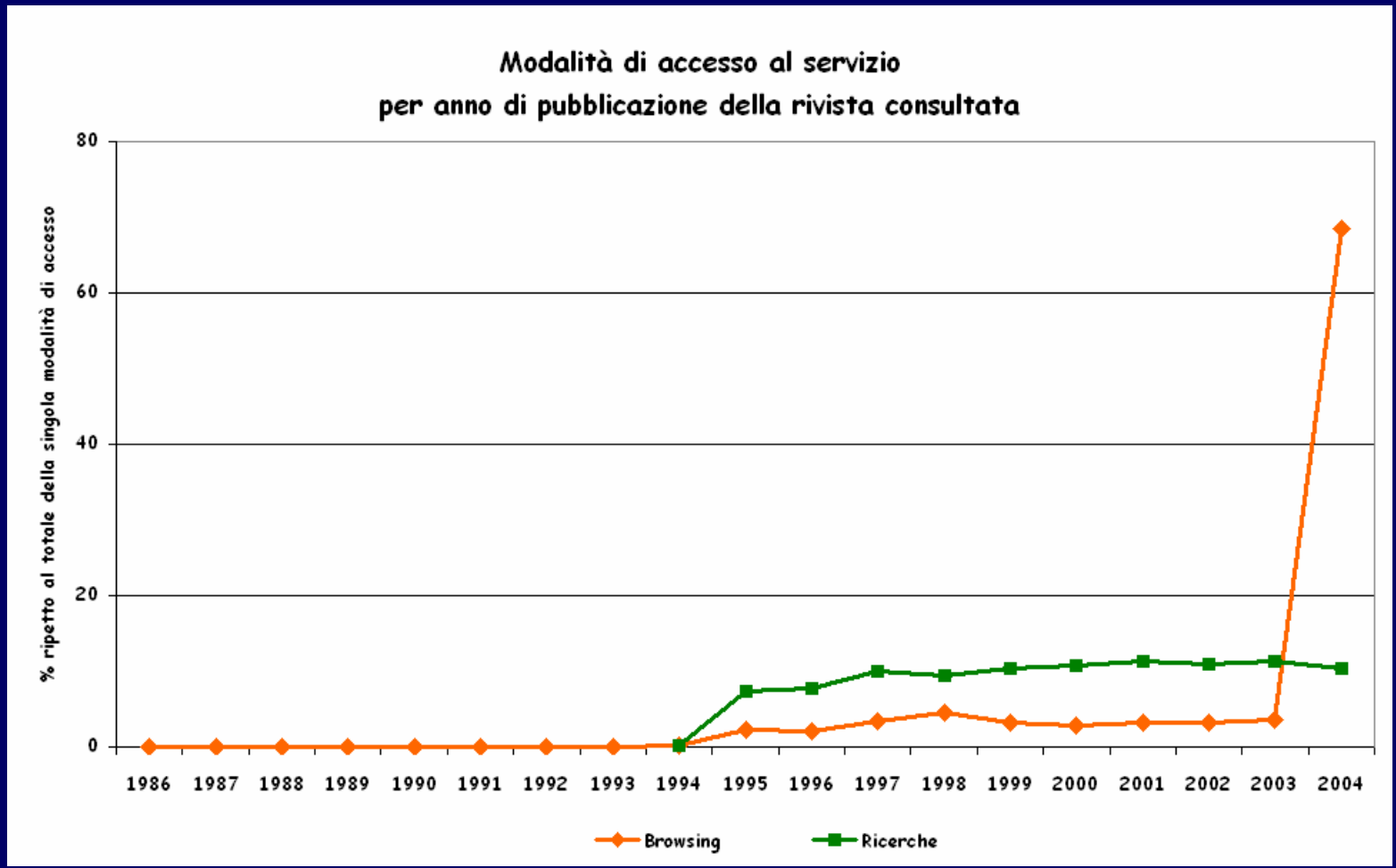
A questo punto si possono analizzare i comportamenti… e verificare se gli utenti si tengono aggiornati o se cercano rispetto interessi specifici



Periodo: Ottobre 2003- Luglio 2004

CASPUR

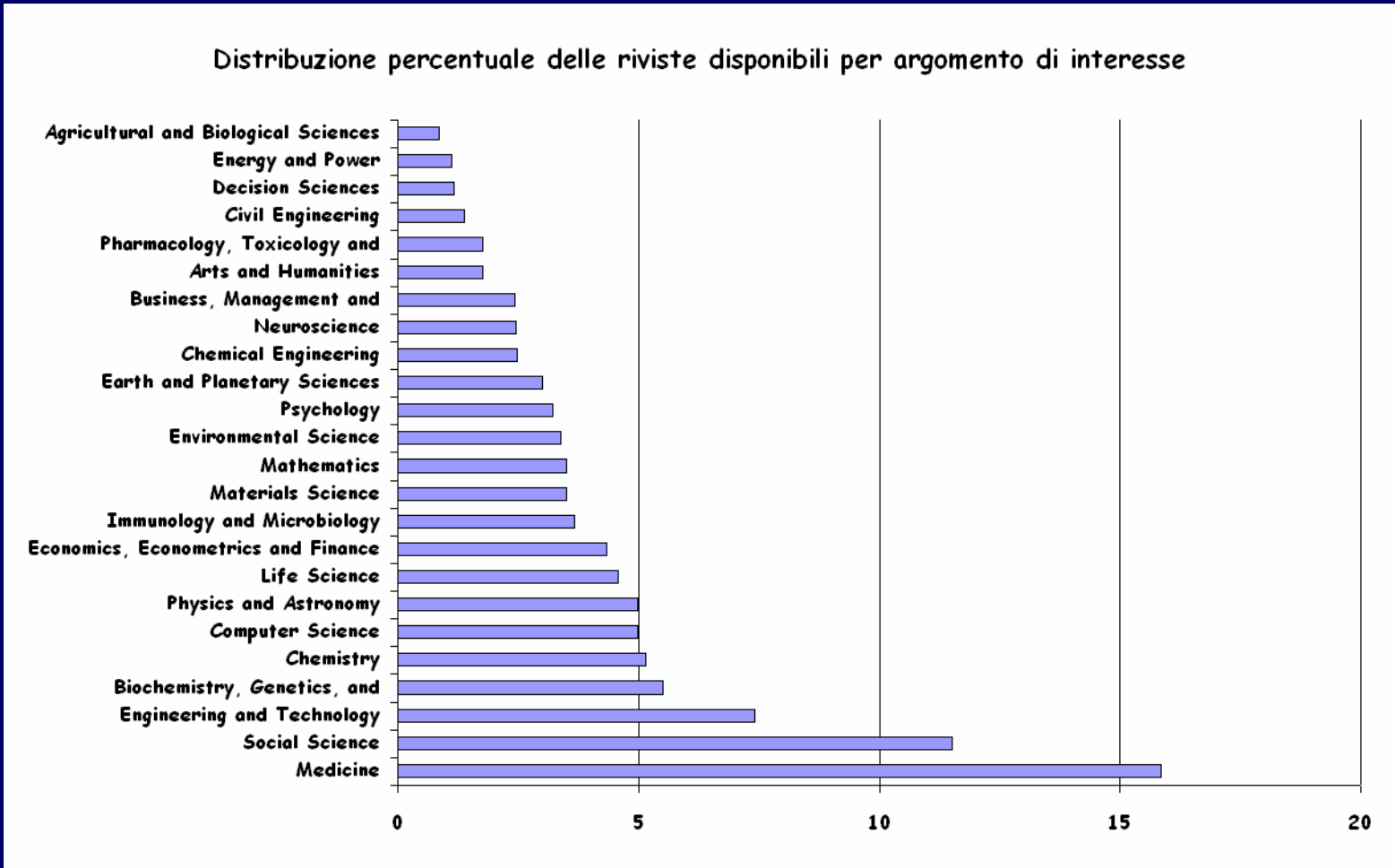
… anche rispetto alla modalità di ricerca…



Periodo: Ottobre 2003- Luglio 2004

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

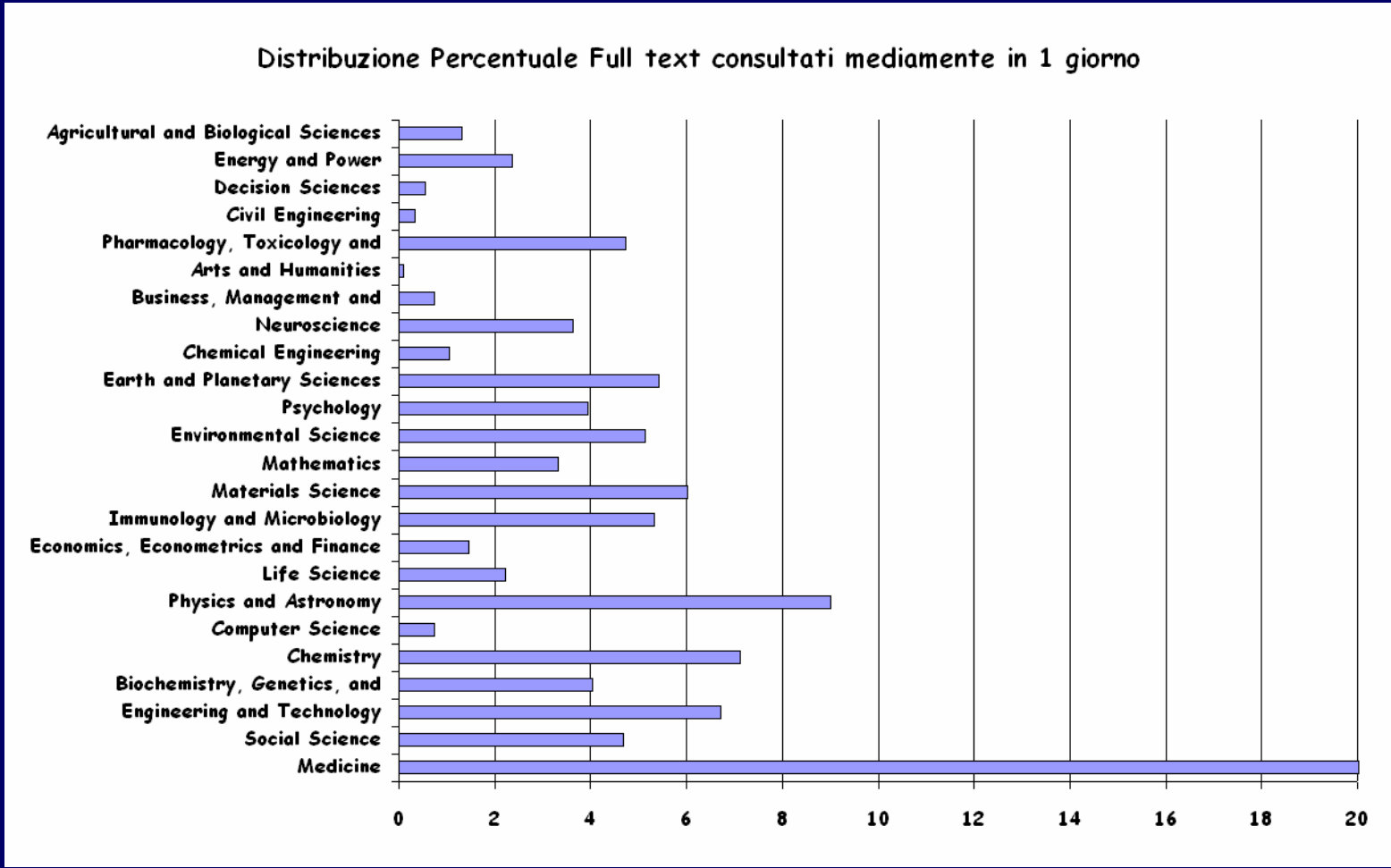
Ancora una volta il problema metodologico: riviste disponibili



Periodo: Ottobre 2003- Luglio 2004

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

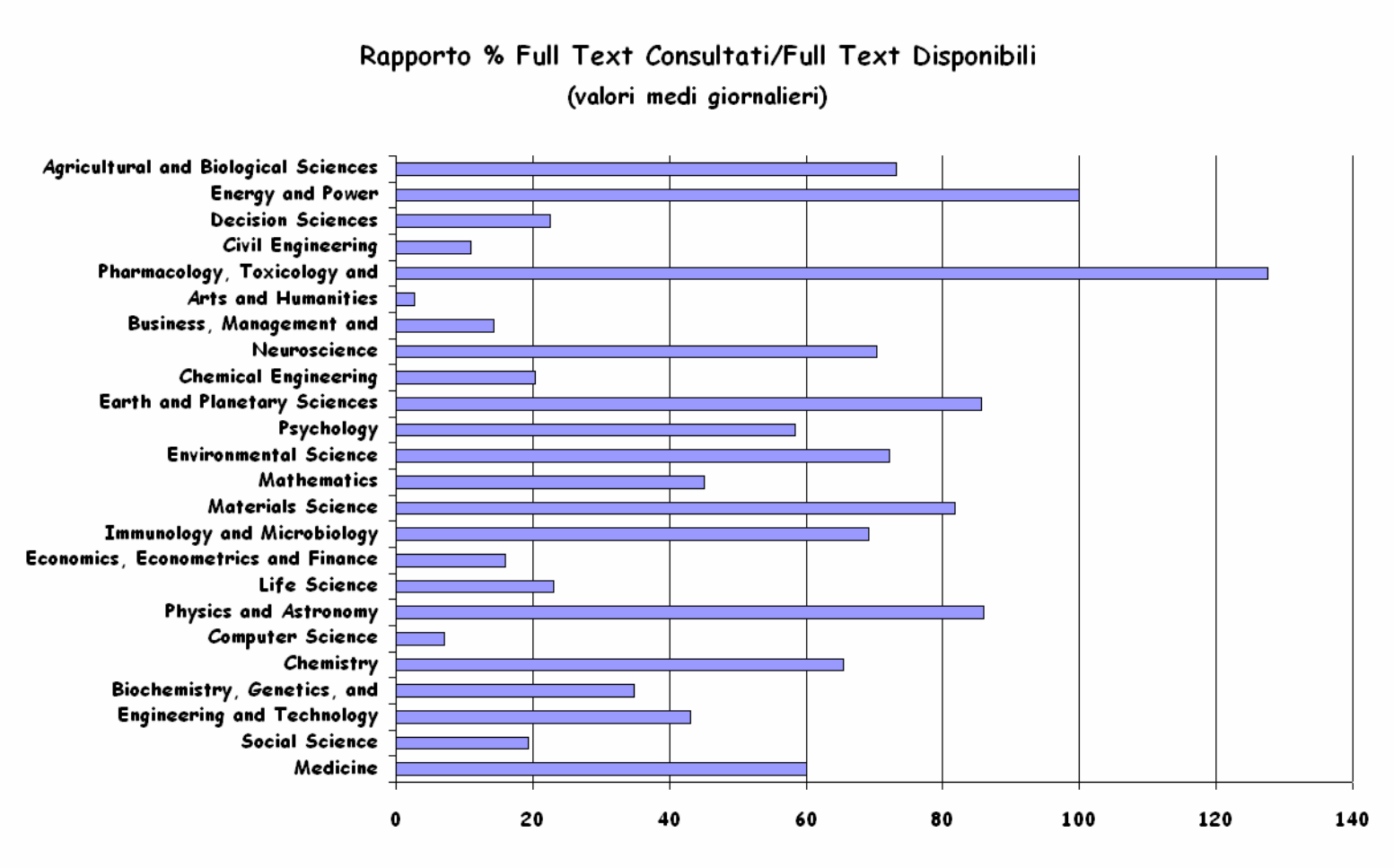
… che vengono così consultate…



Periodo: Ottobre 2003- Luglio 2004

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

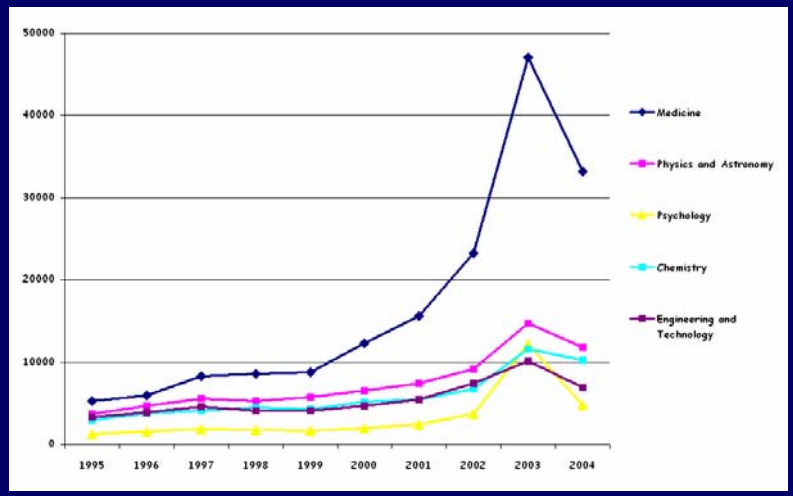
… consultazioni che potrebbero essere così sintetizzate:

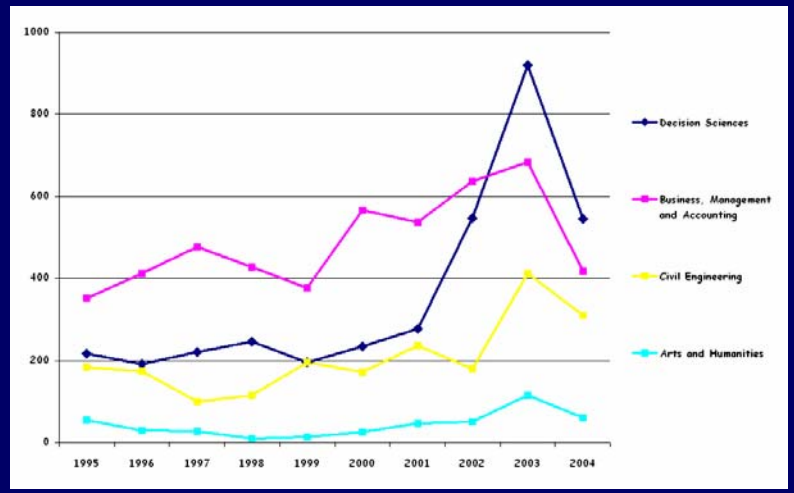


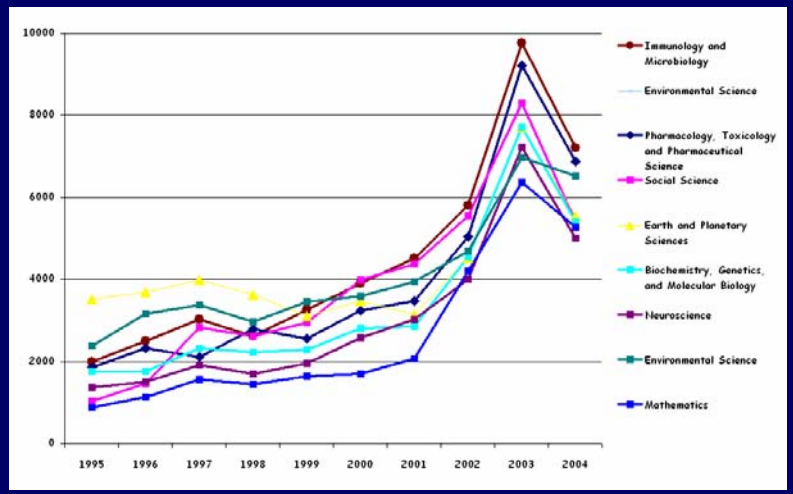
Periodo: Ottobre 2003- Luglio 2004

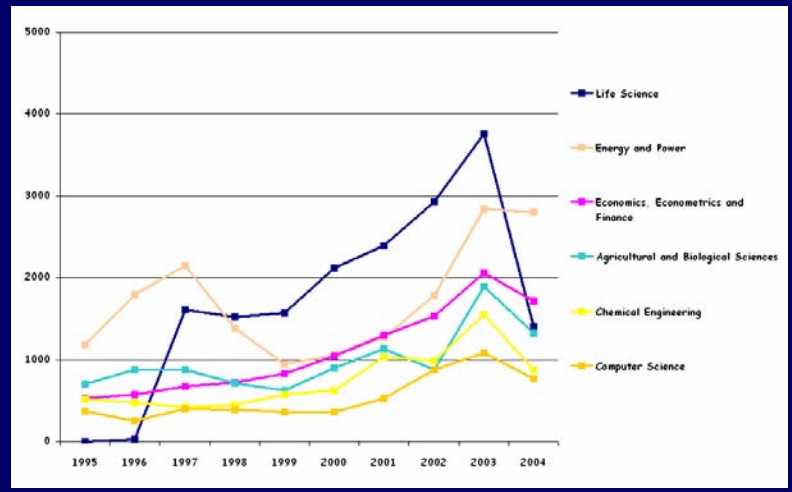
CASPUR

Inoltre come si comporta l'utente per categoria e anno di pubblicazione della rivista?









CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca

Conclusioni

- Necessità di una *metodologia standard* di analisi dei dati:

  - Pre-trattamento
  - Sintesi e normalizzazione dei risultati

- Necessità di definire la *sessione di ricerca*