



Induction Team

La ricerca all'ennesima potenza

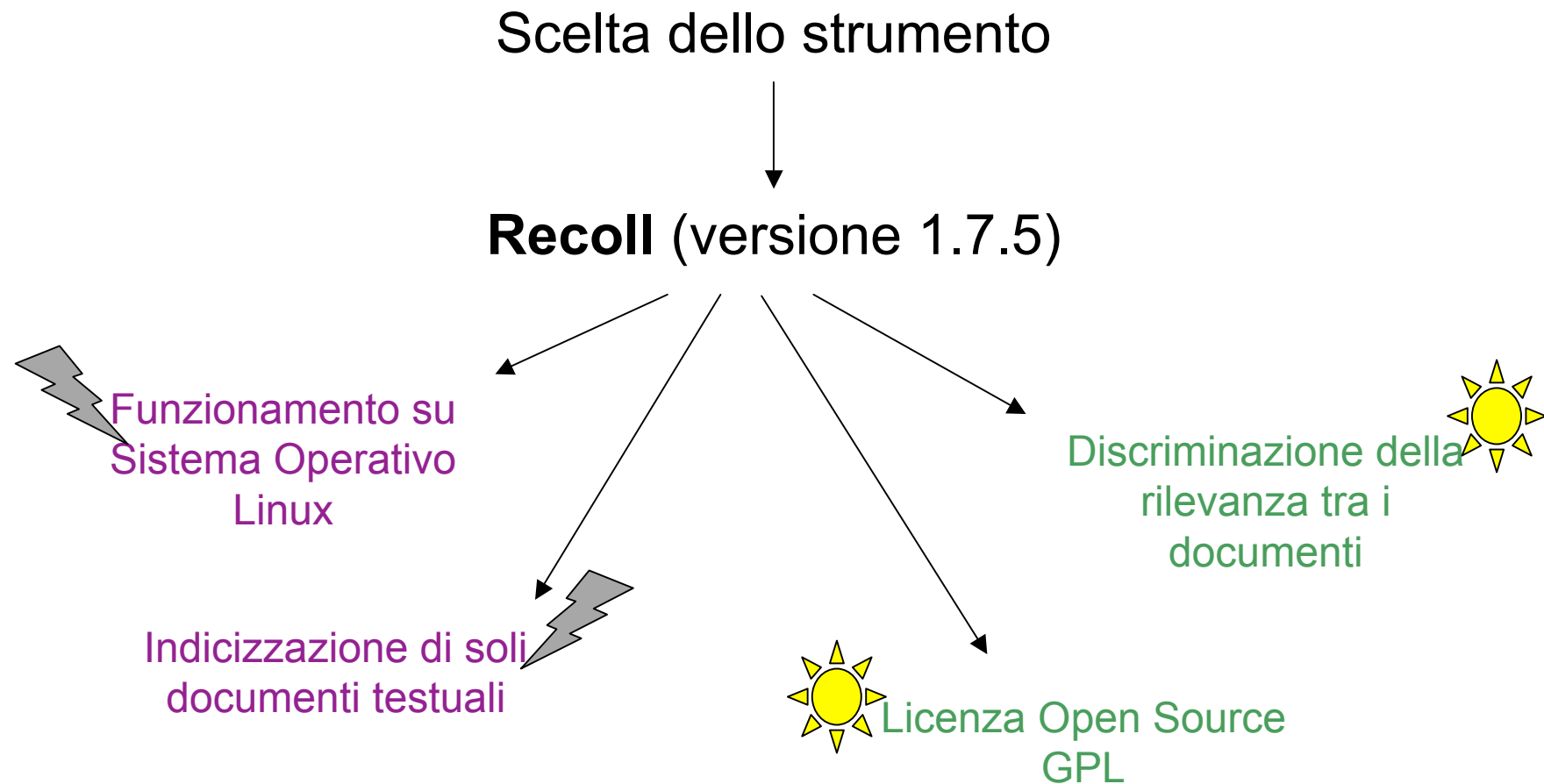
Amerigo Vespucci

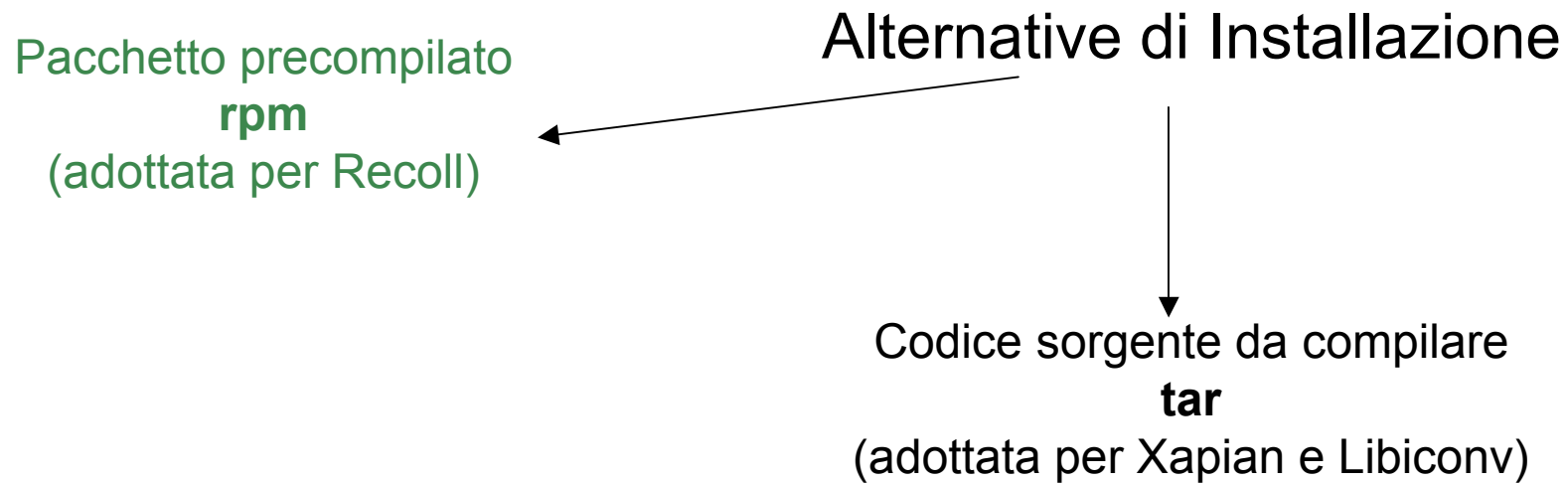
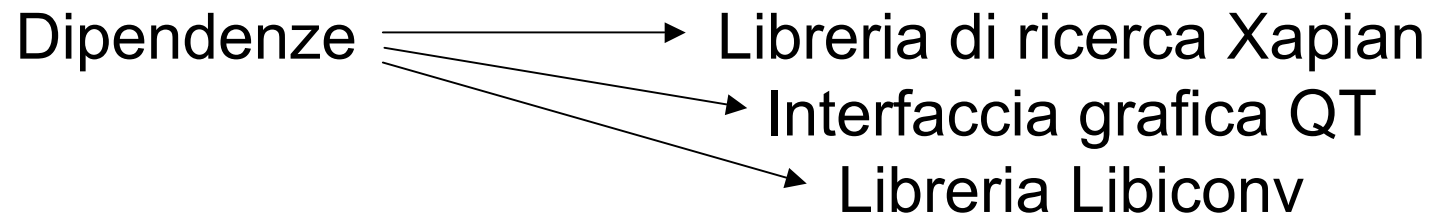
Padova, 30 Marzo 2007

Dindo Stefano	Matr.	566949-IF
Sandri Francesco	Matr.	566520-IF
Scremin Thomas	Matr.	566991-IF
Tono Raffaella	Matr.	569864-IF
Vantini Marco	Matr.	566948-IF



- ☐ Scelta del Desktop Search
- ☐ Installazione e dipendenze
- ☐ Caratteristiche principali
- ☐ Algoritmo di stemming
- ☐ Interfaccia grafica
- ☐ Analisi dei risultati
- ☐ Conclusioni







- Non gestisce eliminazione stop word
- Effettua lo stemming
- Ordina i risultati (ranking)



❖ Regola dello stemming

(Condizione) S1 -----> S2

❖ Esempio:

(m>1) EMENT ----->
REPLACEMENT -----> REPLAC

❖ Algoritmo:

Se la parola ha un numero di lettere ≤ 2 rimane invariata.

- 1) Rimozione genitivo sassone.
- 2) Elimina i suffissi e i verbi (per esempio: “-ing”, “-ed”, “-es”).
- 3) In questi tre passi vengono tolti i suffissi terminanti con doppia consonante.
- 6) rimozione della vocale e dall'ultima posizione delle parole.
- 7) rimozione delle doppie consonanti presenti al termine delle parole.

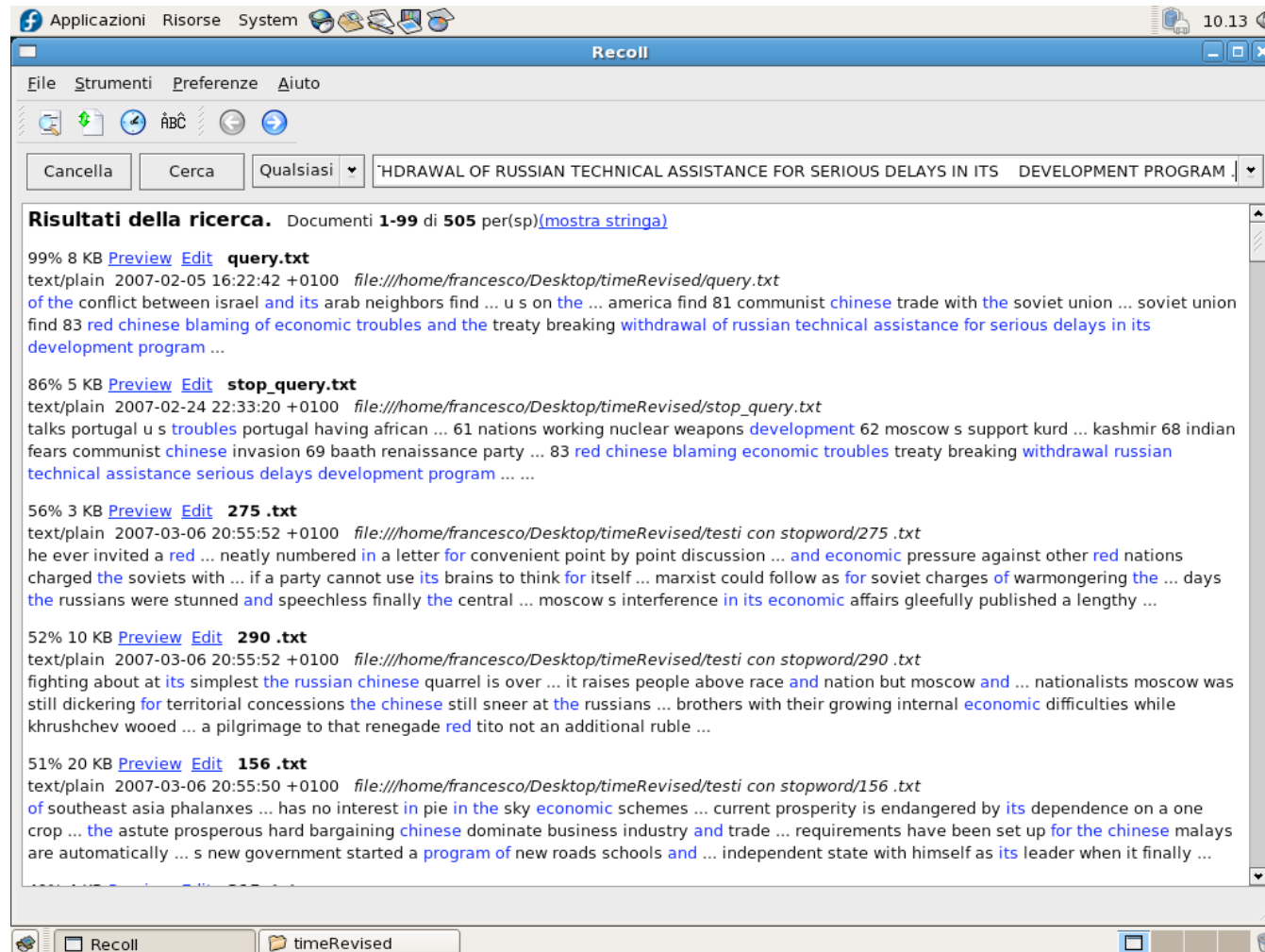


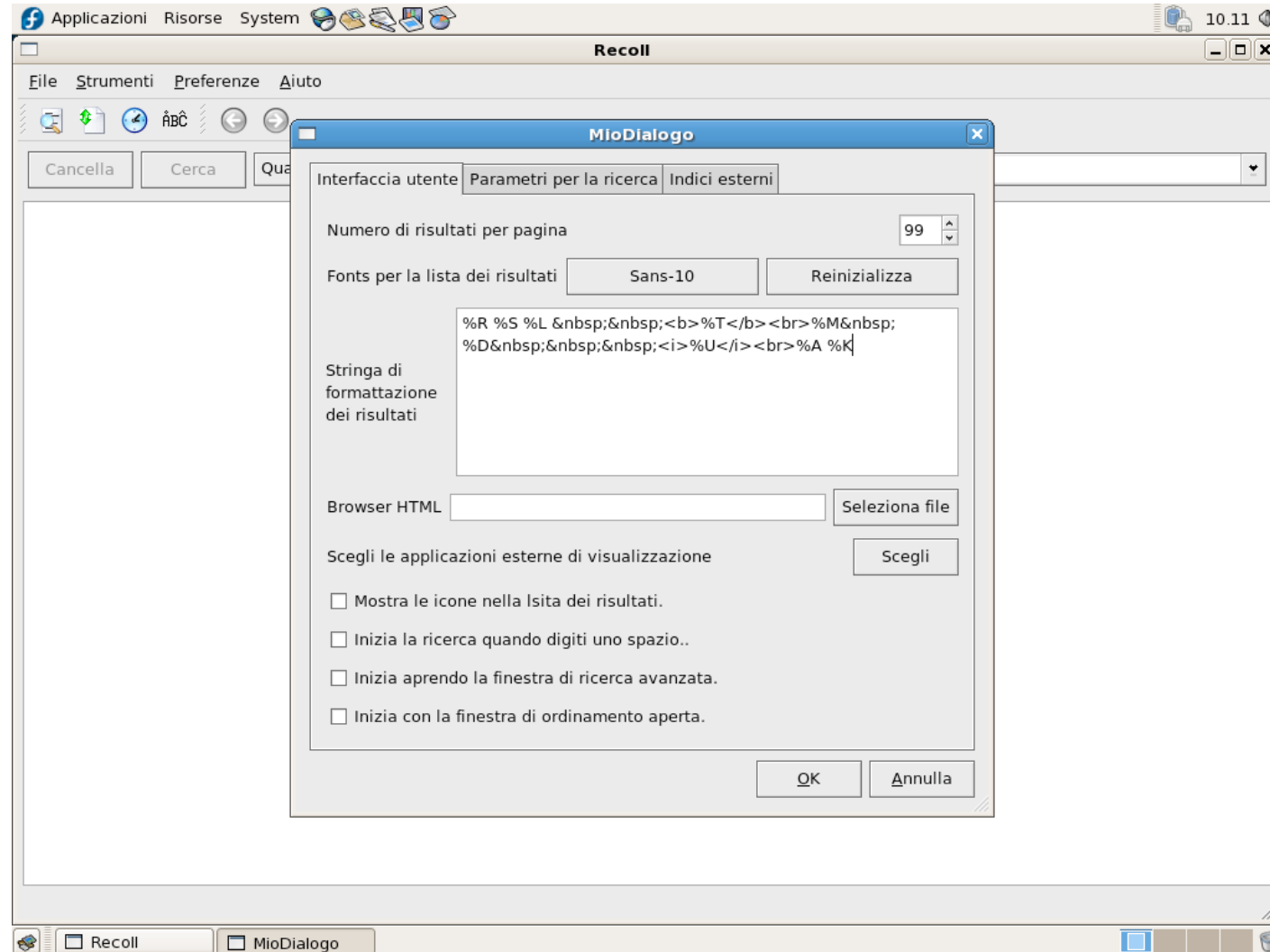
Formula di
Robertson

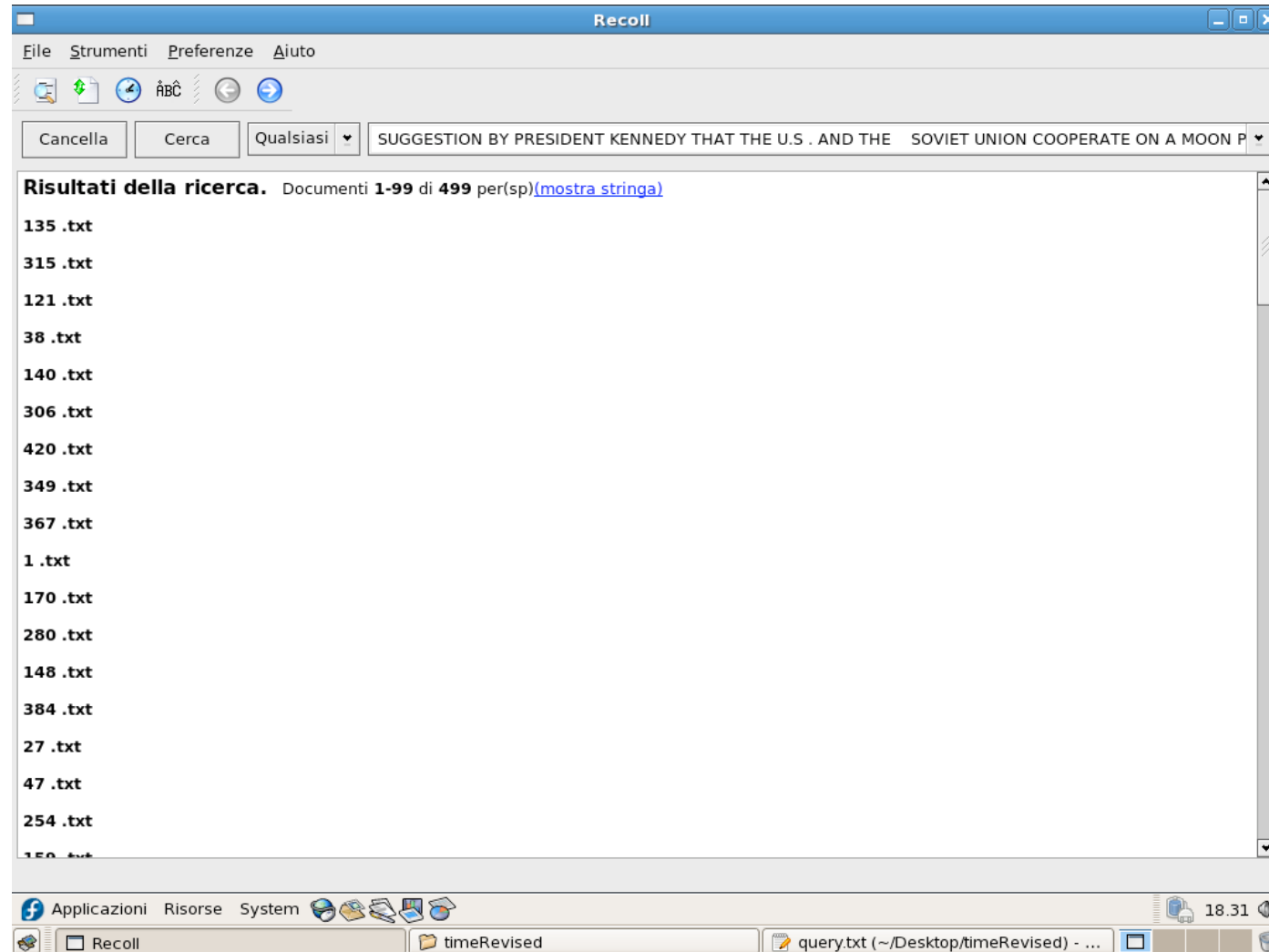
$$w(t) = \log \frac{p(1-q)}{q(1-p)}$$



$$w(t) = \log \frac{(r+h)(N-R-n+r+h)}{(R-r+h)(n-r+h)}$$









Ricerca avanzata

Cerca tra documenti che contengono::

Qualsiasi parola

Questa frase

Tutti i termini

Elimina condizione

Aggiungi condizione

Nessuna parola

Nome del file

☐ Limita il tipo di file. ☐ Per categorie

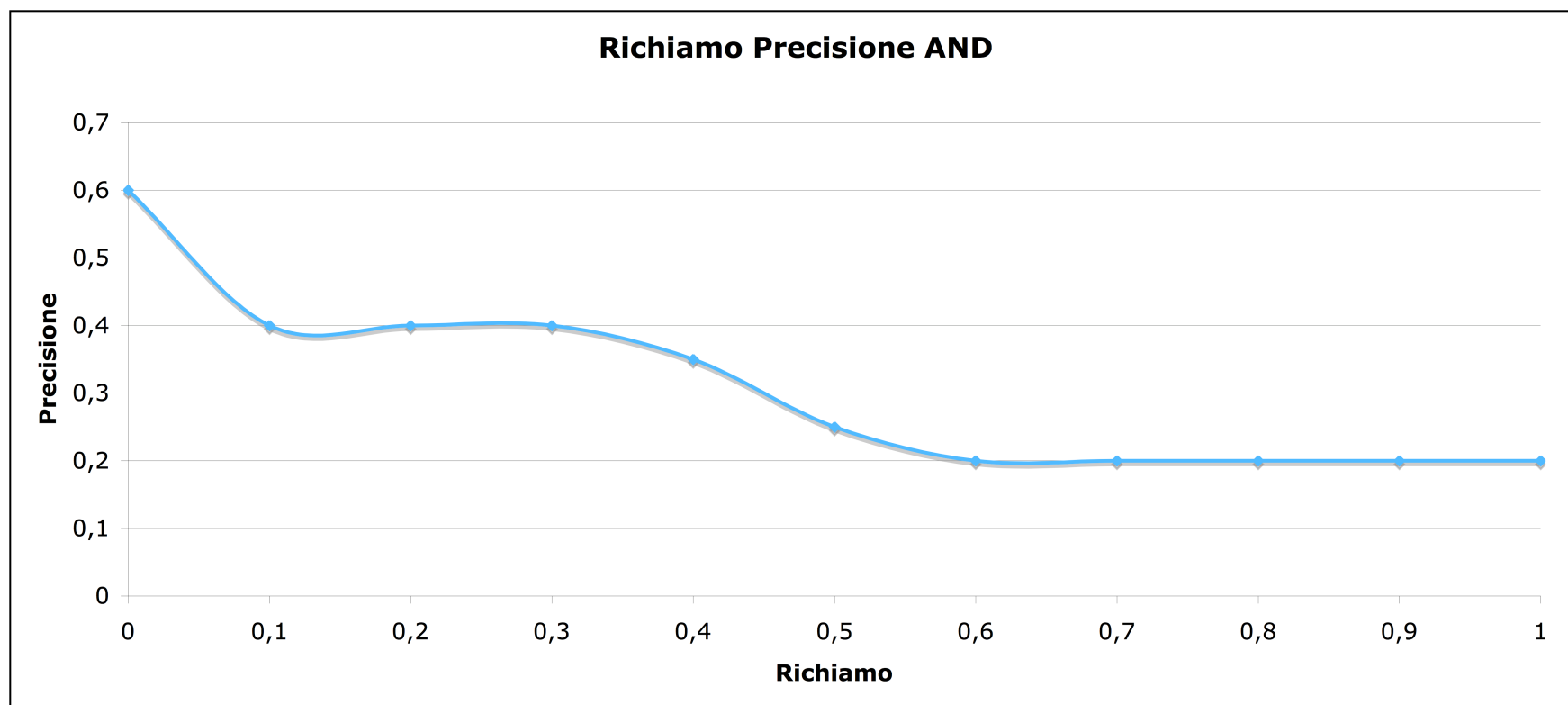
Ricerca tipo file

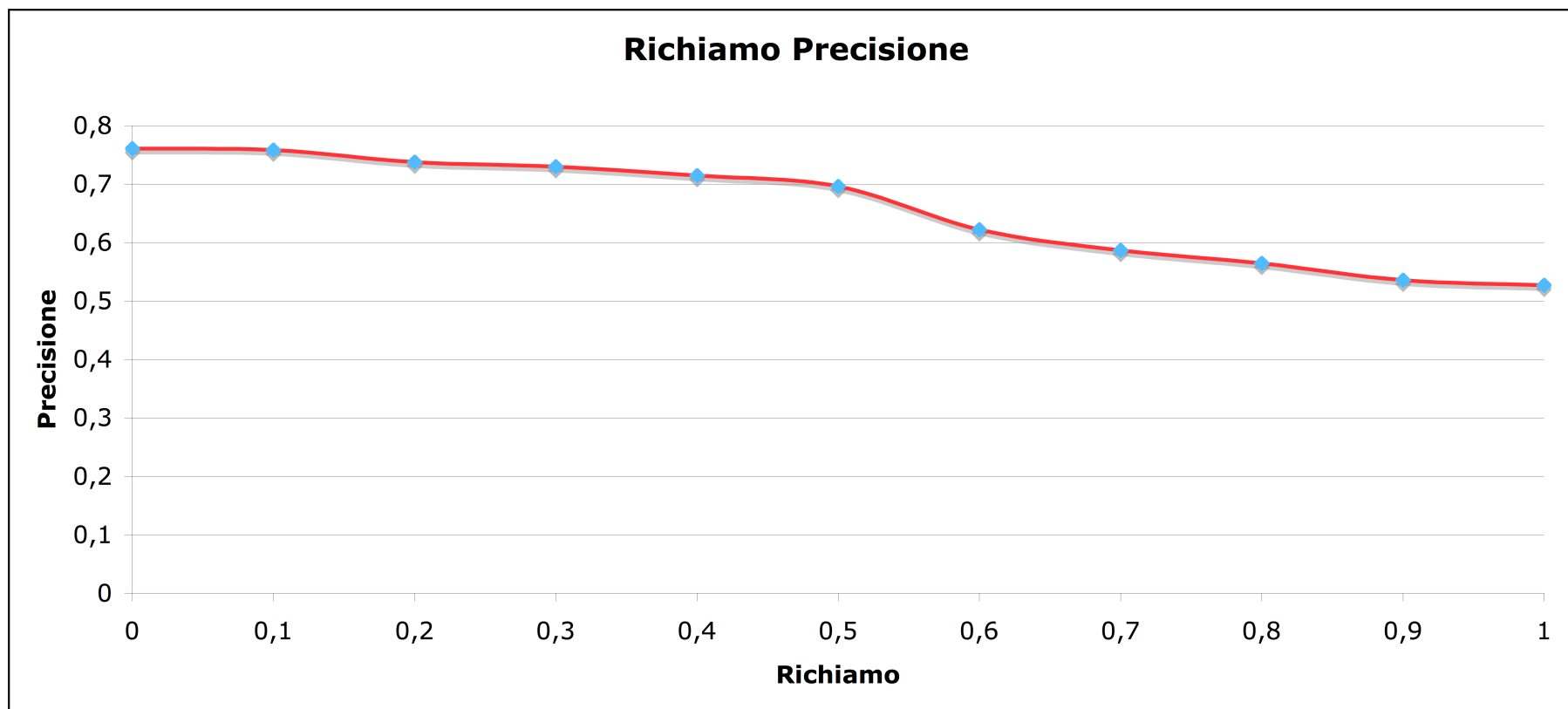
application/msword
application/pdf
application/postscript
application/vnd.ms-excel
application/vnd.ms-powerpoint
application/vnd.sun.xml.calc
application/vnd.sun.xml.calc.template
application/vnd.sun.xml.draw
application/vnd.sun.xml.draw.template

Tutti ---->
Sel ----->
<----- Sel
<----- Tutti

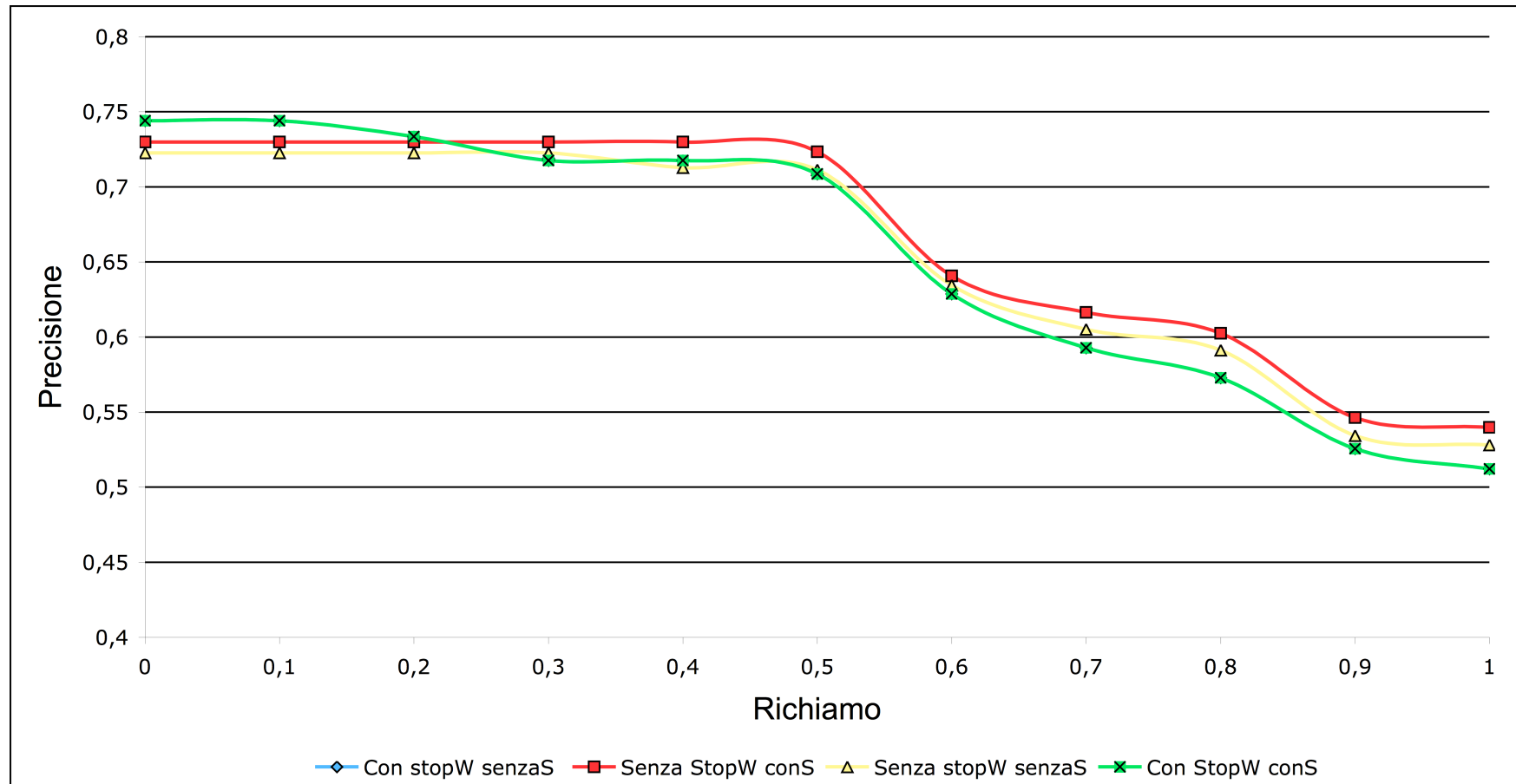
Ignora i files di questo tipo

Limita i risultati alla sotto-cartella:

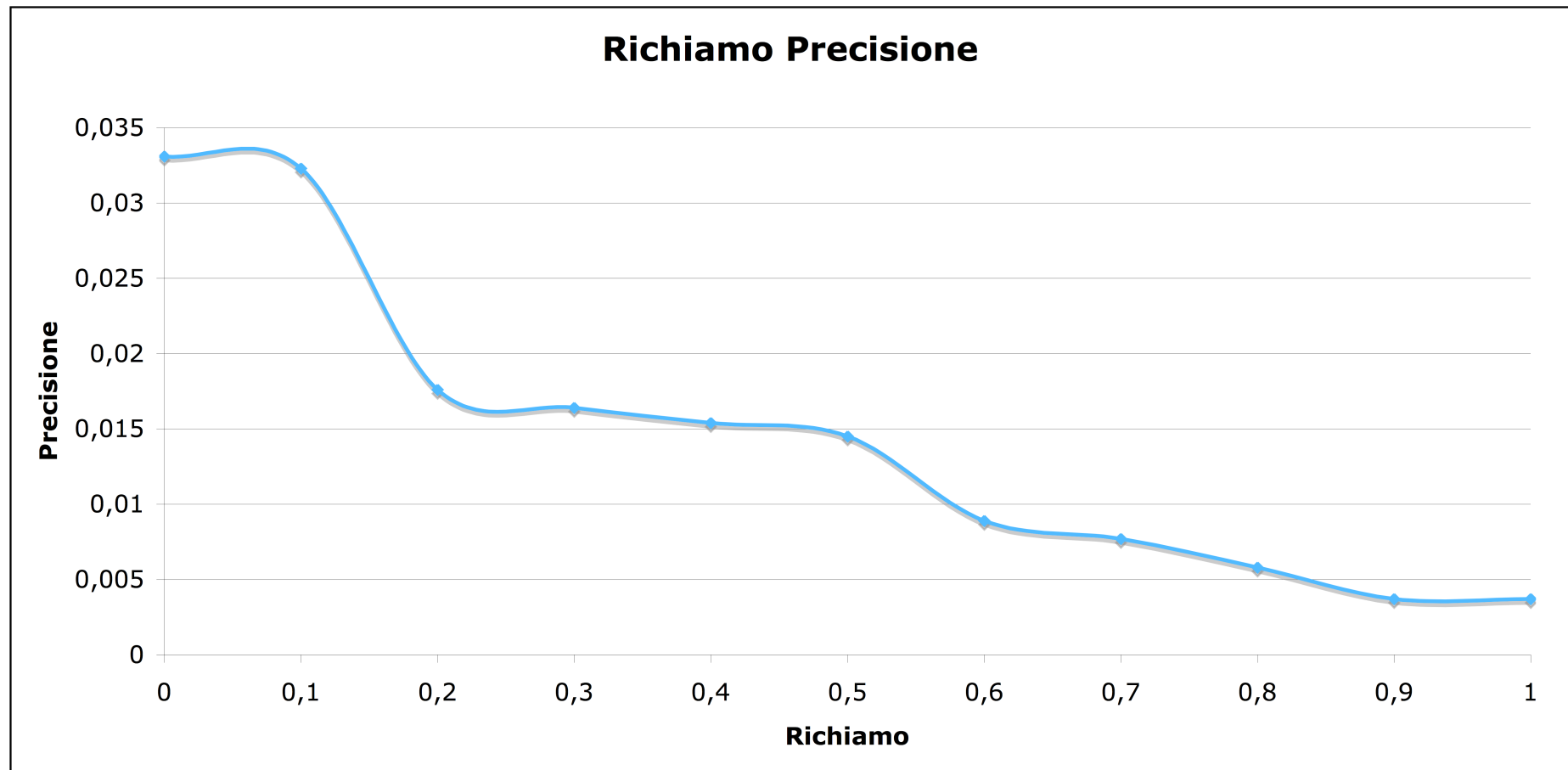




Interrogazione con OR



Confronto tra le varie configurazioni



Con collezione errata

