

**Accesso a banche dati
remote ed eterogenee
mediante interfacce di browsing
su thesauri**

Oreste Signore

CNUCE-CNR e Univerità di Lecce

Seminario

Sistema Informativo Nazionale per la Matematica

Lecce

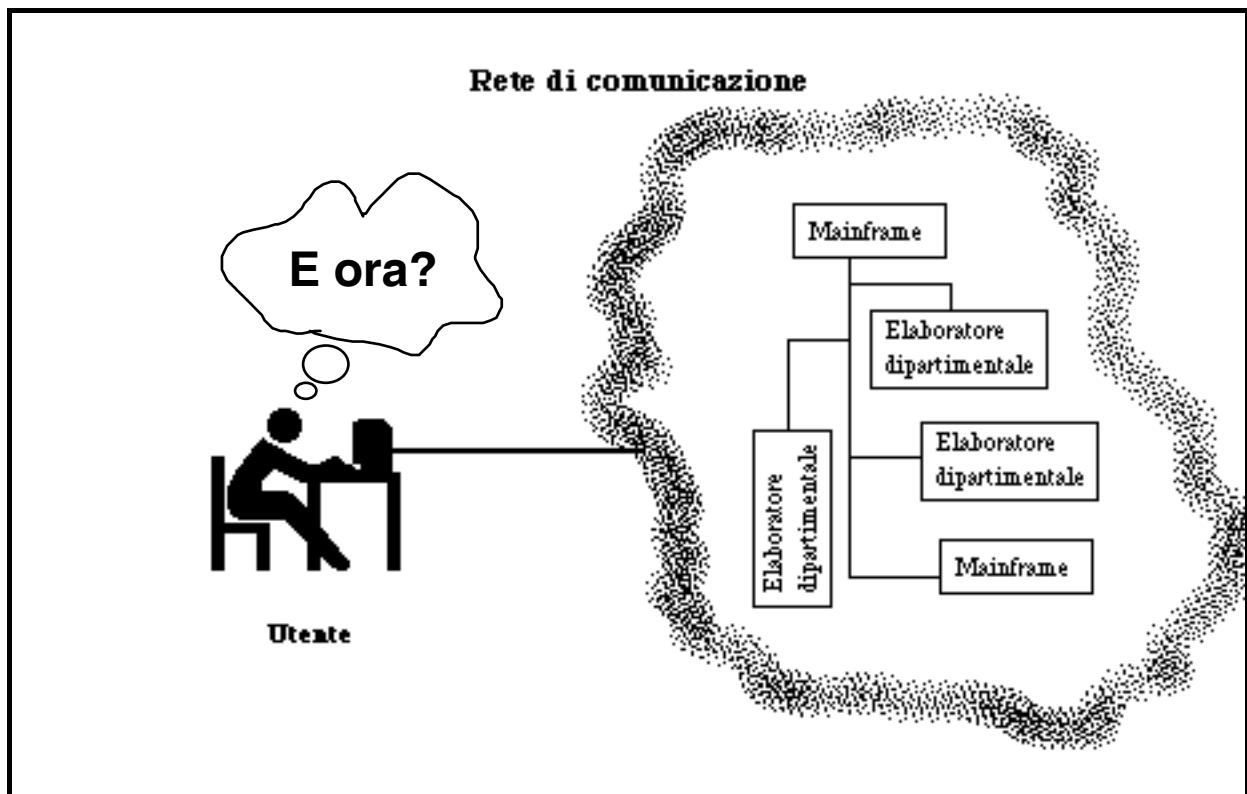
30 settembre - 2 ottobre 1991

Contenuto

- **L' informazione on-line**
- **Le caratteristiche dei sistemi di Information Retrieval**
- **Il problema dell' accesso alle banche dati documentarie**
- **Thesauri e interfaccia utente**
- **Un esempio**
- **Conclusioni**

Il mito dell' informazione on-line

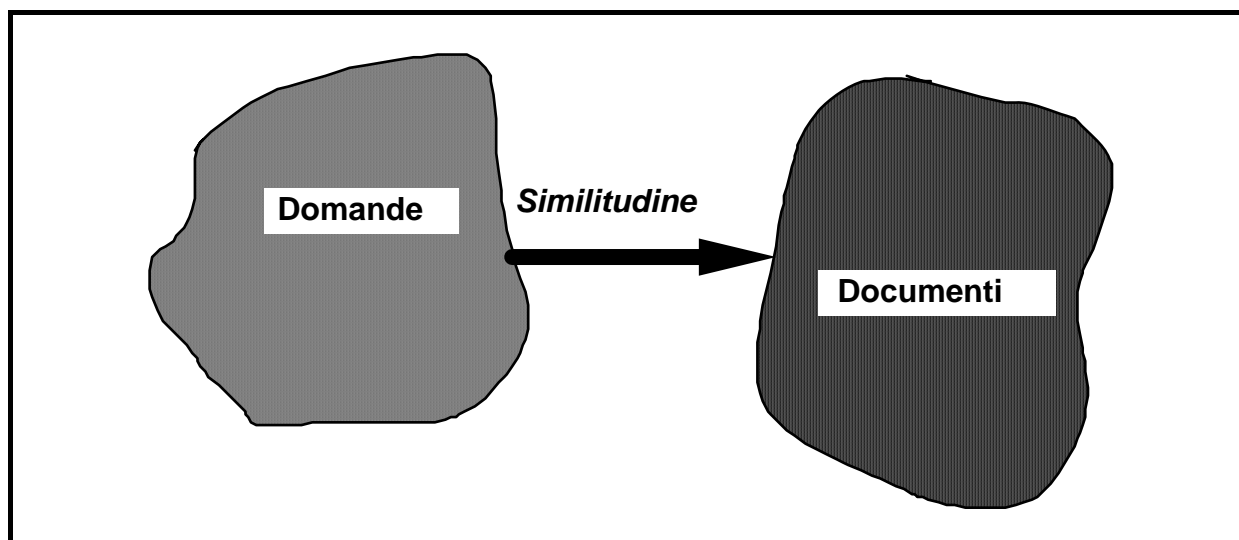
- Sono interessato ad accedere alle banche dati
- Procurati un modem e un identificatore di rete, e poi accedi alla banca dati che ti interessa



- Le banche dati residenti sui nodi (host) sono gestiti da:

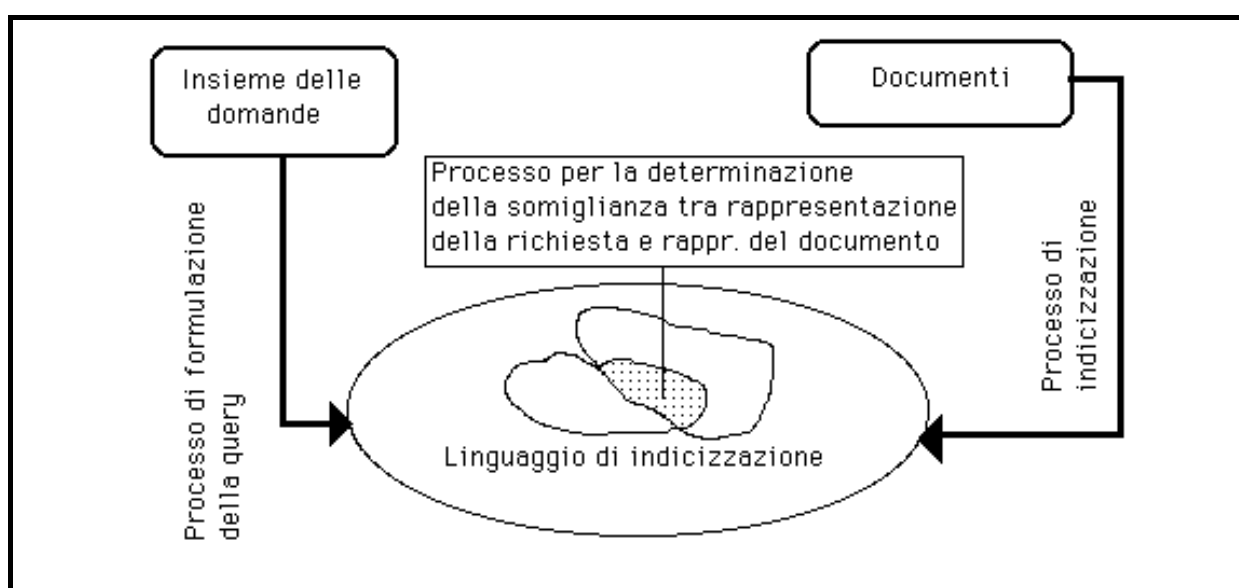
Information Retrieval Systems

Descrizione funzionale di un IRS

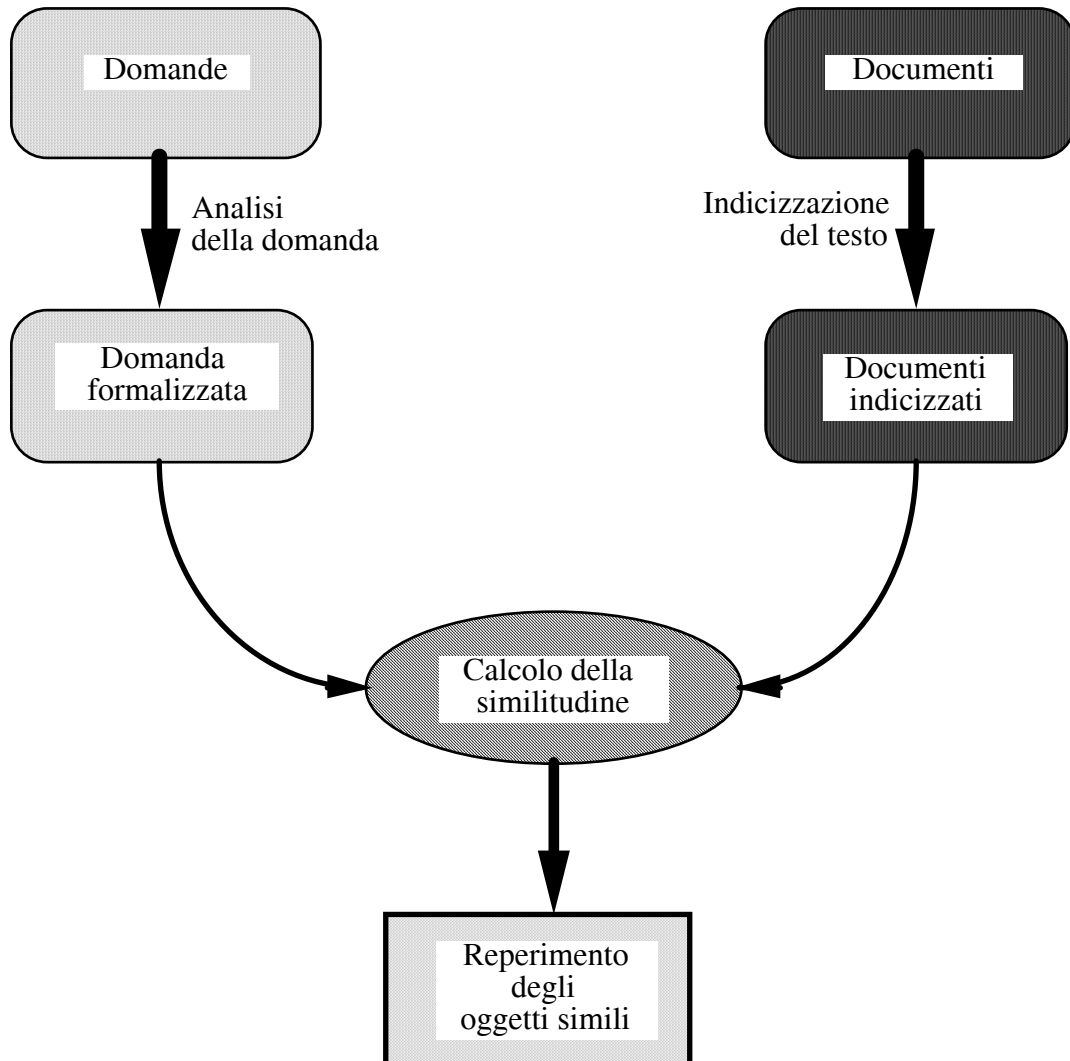


Analisi della corrispondenza tra l' insieme delle domande e l' insieme dei documenti mediante un linguaggio intermedio di rappresentazione: *linguaggio di indicizzazione* (o *linguaggio di classificazione*).

Quindi:



... più in dettaglio ...

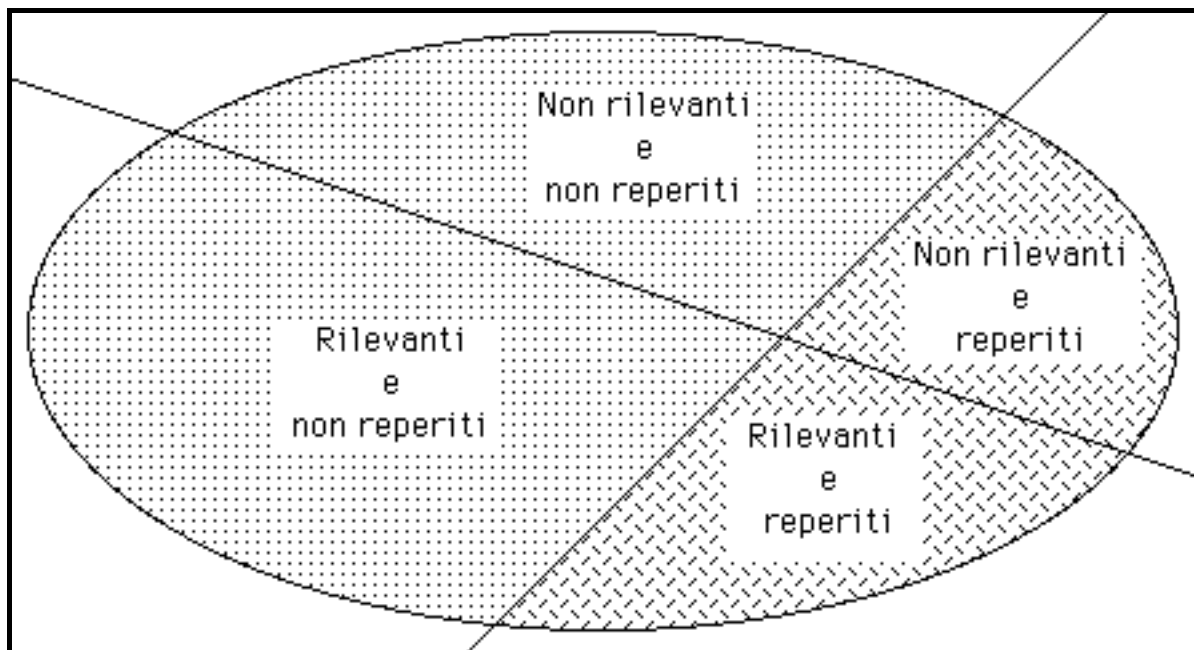


- **I documenti memorizzati vengono rappresentati da un insieme di termini indice.**

... in definitiva ...

Un Information Retrieval System è un sistema in grado di rappresentare documenti e recuperare quelli che sono probabilmente rilevanti per una richiesta dell' utente.

- Dispone di:
 - Metodi per la rappresentazione dei documenti
 - Metodi per il recupero dei documenti



- Efficacia misurata valutando congiuntamente due fattori:

- **Richiamo** =
$$\frac{\text{Numero Docum Rilevanti Recuperati}}{\text{Numero Docum Rilevanti Esistenti}}$$

- **Precisione** =
$$\frac{\text{Numero Docum Rilevanti Recuperati}}{\text{Numero Docum Recuperati}}$$

Il problema dell' accesso alle banche dati documentarie

Crescita del mercato dell' informazione limitata da:

- **problemi tecnologici e organizzativi:**
 - **affidabilità e velocità delle reti di trasmissione**
 - **dialogo per terminali TTY**
 - **vincoli valutari e problemi di addebito**
 - **alti costi di trasmissione e di interrogazione**

- **problemi di interfaccia utente:**
 - **complessità delle procedure di login**
 - **differenze tra i linguaggi di interrogazione**
 - **struttura della banca dati**
 - **contenuto dei campi**

Soluzioni individuate a livello europeo:

- **linguaggio unificato di comandi (CCL, successivamente proposto come standard ISO)**
- **workstation per l' interrogazione**

L' interfaccia utente

- In un approccio Information retrieval è necessario:
 - *identificare i contenuti dei documenti da archiviare*
 - *identificare i contenuti della query dell' utente*
 - *avviare una ricerca per argomento.*
- Gli utenti hanno bisogno di un supporto che gli permetta di operare un raffinamento del concetto ricercato.
- Una interfaccia a menù?

| Immetti i valori richiesti | | |
|----------------------------|---|-------------------------|
| Autore | = | _____ OR _____ OR _____ |
| Anno | = | _____ |
| Argomento | = | _____ OR _____ OR _____ |

No, grazie!

(E' il modo in cui la presunzione e l' ignoranza, travestiti da tecnologia, vogliono soppiantare intelligenza e cultura)

- Solo un approccio basato sulla conoscenza assicura una consistenza nella rappresentazione tra chi indicizza e chi ricerca, ma richiede di avere a disposizione una rete semantica di concetti, accessibile a indicizzatori e utenti.

Rappresentazione della conoscenza sullo specifico dominio

- I problemi tipici:
 - scarsa precisione dei termini utilizzati
 - identificazione errata dell' elemento da ricercare
- Supporto all' utente con la rappresentazione esplicita di:
 - relazioni di sinonimia
 - relazioni di preferenza
 - relazioni gerarchiche

tra i termini, quindi mediante la costituzione di un

thesaurus

- Lista di termini tra cui esistono relazioni semantiche di gerarchia, preferenza ed equivalenza (ISO 2788)
- La disponibilità di un thesaurus consente un aumento considerevole dei fattori di precisione e richiamo.
- I thesauri sono essenziali per poter raffinare il concetto da ricercare.
- L' utente deve essere messo in grado di identificare chiaramente le relazioni intercorrenti tra i vari termini.

Le interfacce di browsing

Le prestazioni dei sistemi di Information Retrieval dipendono dall'efficacia con cui l'utente è in grado di identificare il contenuto dei documenti.

E' necessario disporre di un meccanismo di comunicazione dei concetti chiave, comune anche a chi indicizza o cataloga i documenti.

Una soluzione possibile:
interfaccia di browsing

Una *interfaccia di browsing sui concetti*:

- rende evidente l'organizzazione della banca dati
 - è un meccanismo di comunicazione dei concetti chiave
 - permette il raffinamento dei concetti ricercati
 - consente all'utente di orientarsi
 - permette all'utente di identificare i descrittori
-
- Condizione essenziale per la realizzazione di una interfaccia di browsing è la *disponibilità di un thesaurus di alta qualità*.
-
- Alcuni prototipi:
 - CANSEARCH (Pollitt, 1983)
 - CALIBAN (Frei and Jauslin, 1983)
 - COALSORT (Monarch and Carbonnell, 1987)

Un esempio: Proteus

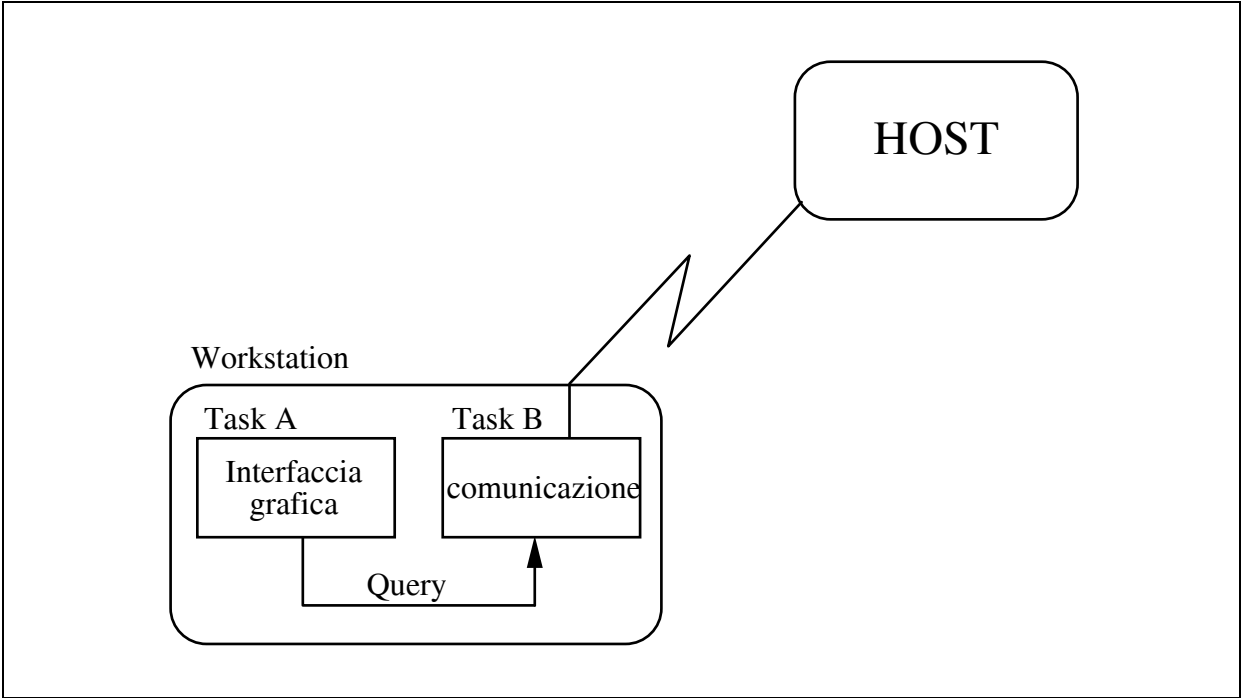
Finalità:

- **Consentire una interazione più agevole con i sistemi tradizionali.**
- **Fornire una rappresentazione del documento indipendente dalla sua effettiva implementazione fisica.**
- **Fornire all' utente strumenti che permettono di identificare i concetti di interesse per formulare una richiesta (query) al sistema nel modo più efficace possibile.**

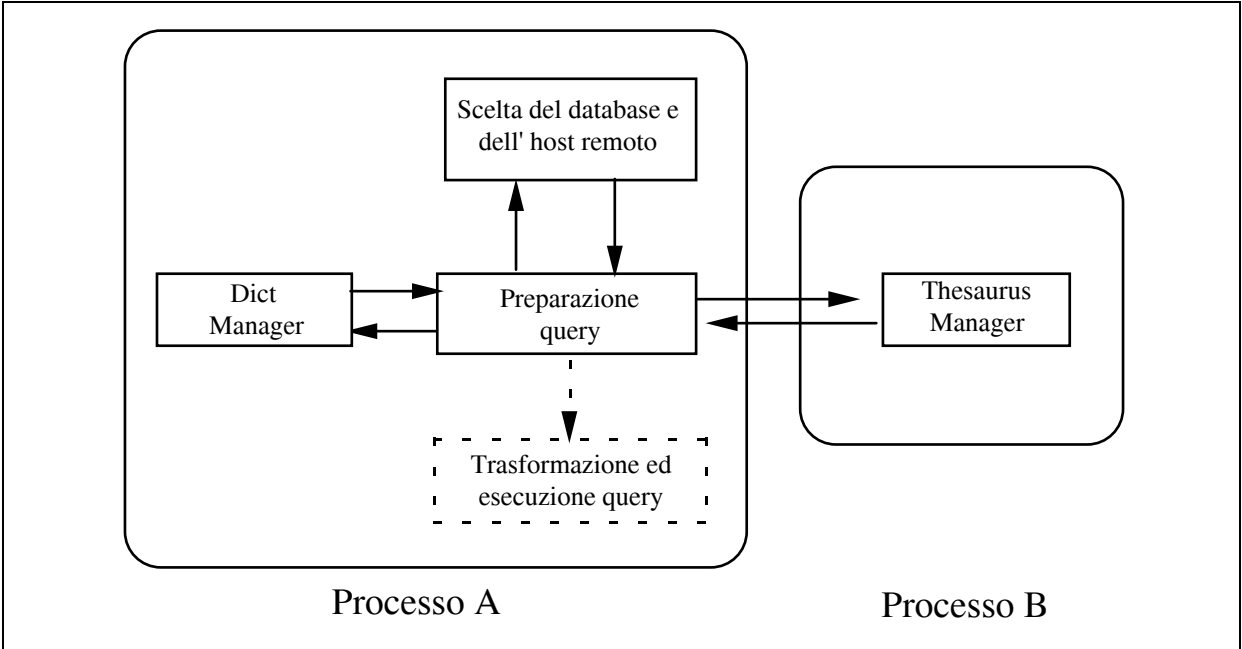
Strumenti:

- **Gestore di dizionari.**
- **Browser grafico su thesauri, grazie a cui l' utente ha una rappresentazione grafica dei concetti di interesse legati da relazioni di gerarchia, equivalenza e preferenza.**
- **Finestre per l' inserimento delle richieste specifiche su un campo.**

Architettura generale



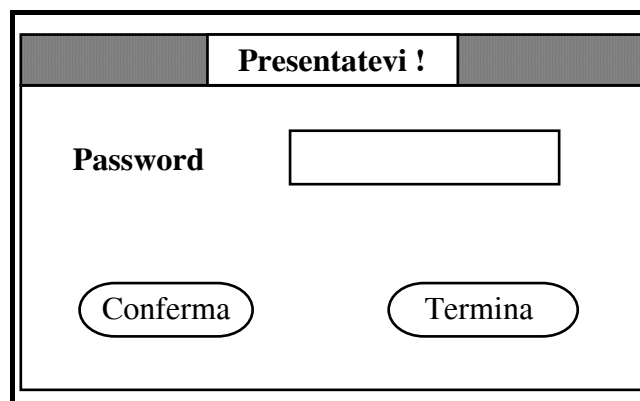
Due task



Due processi

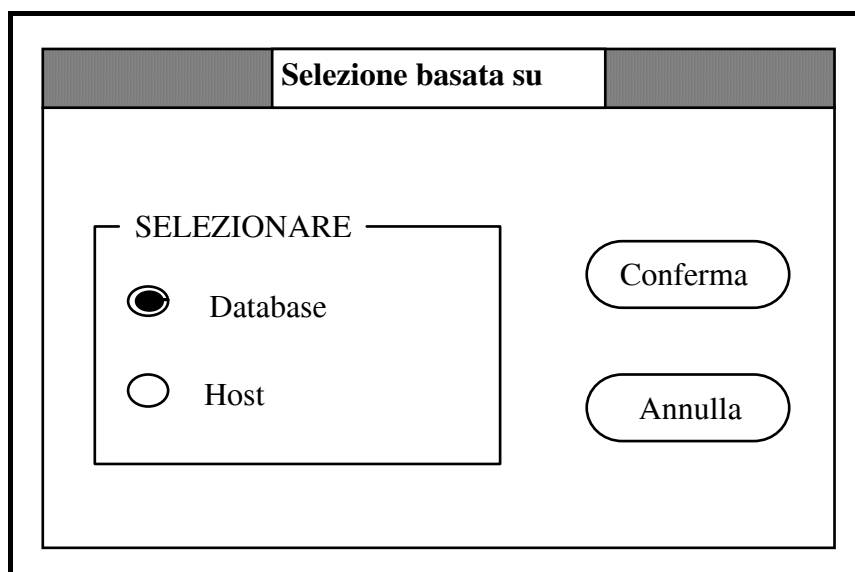
(A può attivare B e proseguono in parallelo)

Sign on



A dialog box titled "Presentatevi !" with a shaded header. The main area contains the label "Password" followed by a rectangular input field. Below the input field are two rounded buttons: "Conferma" on the left and "Termina" on the right.

Scelta della modalità di selezione



A dialog box titled "Selezione basata su" with a shaded header. The main area contains a box labeled "SELEZIONARE" containing two radio button options: "Database" (selected) and "Host". To the right of this box are two rounded buttons: "Conferma" and "Annulla".

Selezione della banca dati

Selezionare i termini per la connessione

| | |
|--|---|
| <p>DB accessibili</p> <div style="border: 1px solid black; height: 100px; width: 100%; position: relative;"><div style="position: absolute; right: -20px; top: 0; bottom: 0; text-align: center;">↑</div><div style="position: absolute; right: -20px; bottom: 0; text-align: center;">↓</div></div> | <p>Host accessibili</p> <div style="border: 1px solid black; height: 100px; width: 100%; position: relative;"><div style="position: absolute; right: -20px; top: 0; bottom: 0; text-align: center;">↑</div><div style="position: absolute; right: -20px; bottom: 0; text-align: center;">↓</div></div> |
| <p>Sistemi di gestione</p> <div style="border: 1px solid black; height: 100px; width: 100%; position: relative;"><div style="position: absolute; right: -20px; top: 0; bottom: 0; text-align: center;">↑</div><div style="position: absolute; right: -20px; bottom: 0; text-align: center;">↓</div></div> | <p style="text-align: center;">Scelta corrente</p> <p>DB <input style="width: 100%;" type="text"/></p> <p>Host <input style="width: 100%;" type="text"/></p> <p>Sistema <input style="width: 100%;" type="text"/></p> |

ProseguìDescrizioneAnnullaTermina

Fig.

Il "Documento virtuale"

Lista dei campi del documento

Nome DB

Tema

Tipo documento

Autore

Sede autore

Titolo originale

Lingua originale

Titolo tradotto

Riferimenti bibliografici

Parole chiave

▲
↑

▼
↓

Termina Selezione

Esegui

Annulla

Oreste Signore

LE-16

Una query "nativa"

The screenshot shows a dialog box titled "Valore per query". It contains a label "Autore" followed by an empty text input field. Below the input field are three buttons: "Conferma", "Salva e continua", and "Annulla".

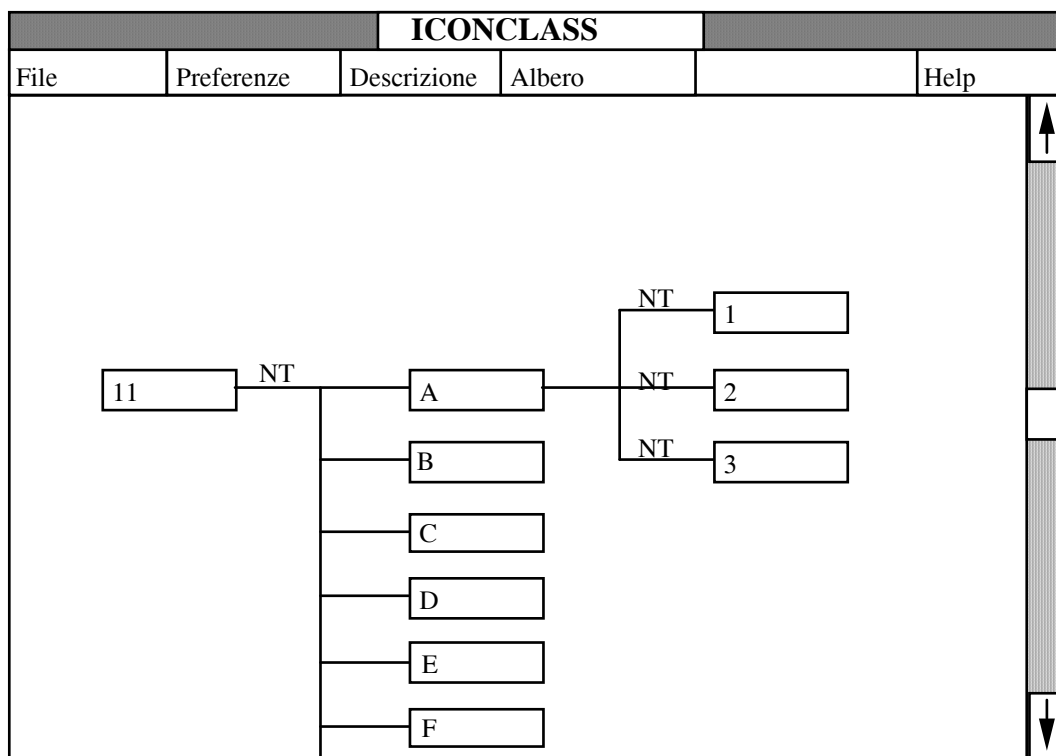
Revisione della query preparata

The screenshot shows a dialog box titled "Query costruita". It features a "Query :" label and a text input field containing "1 AND 2 AND 3". Below this is a scrollable list box with three items:

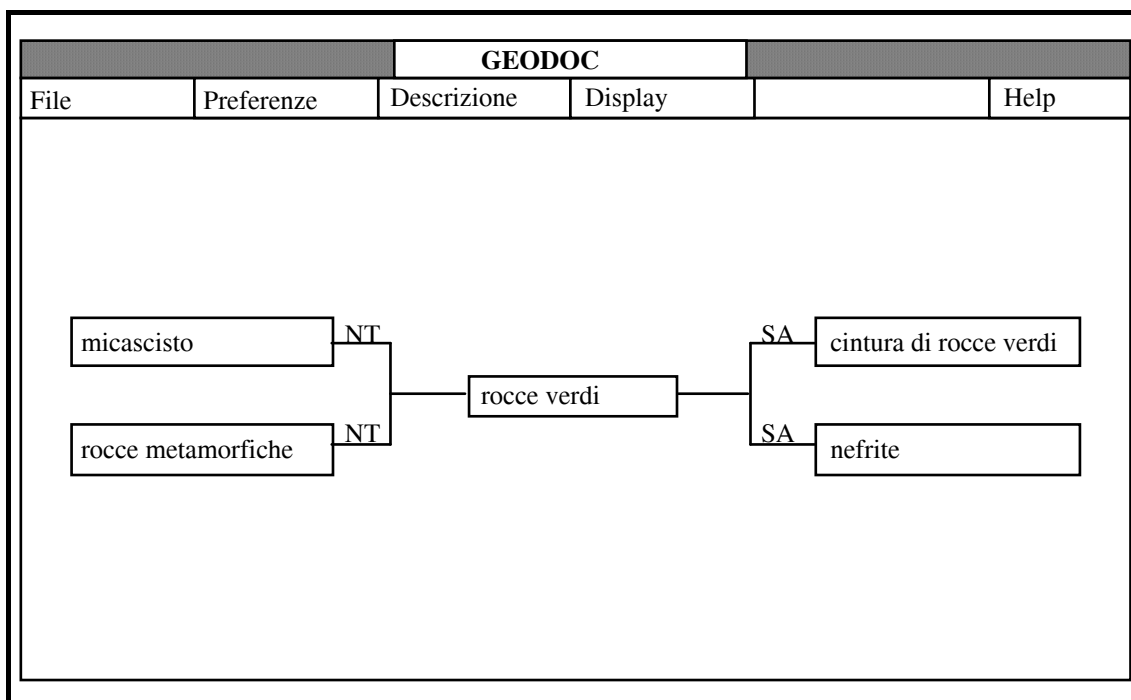
- 1 Autore : Bigazzi OR Bonadonna
- 2 Tema: 227.A.03
- 3 Tipo documento: Congresso

At the bottom of the dialog are three buttons: "Esegui", "Annulla", and "Ritorna alla selezione".

Un display ad albero



Un display a farfalla



Conclusioni

- **La ricerca dei documenti su una banca dati richiede la conoscenza esatta dell' argomento da ricercare**
- **La disponibilità di thesauri permette la condivisione della conoscenza su uno specifico settore tra indicizzatori e utenti**
- **Le interfacce di browsing consentono il raffinamento del concetto ricercato e l' identificazione del sottoinsieme di documenti significativi**
- **L' utilizzo di workstation sofisticate permette di superare molti ostacoli alla diffusione della fruizione dei sistemi in linea.**