

Progetto Pedigree

Installazione, studio e valutazione di “Beagle”

A.A. 2005/2006

Corso di Reperimento dell'Informazione

Prof.sa Maristella Agosti

Team:

Argenton Matteo

Buzzi Lorenzo

Gatto Giorgio

Molinaro Matteo

Zorzan Emmanuele



- Beagle: dipendenze e funzionamento
- L'esigenza informativa
- Caratteristiche tecniche
- Prestazioni



Cos'è Beagle?

- Beagle è un esempio di “Desktop Search Utility” (DSU)
- I DSU sono strumenti per la ricerca di documenti all'interno di un personal computer.
- Un DSU si colloca tra le applicazioni di Information Retrieval perché ne realizza lo schema di Architettura Funzionale.
- Un DSU si distingue dai sistemi di IR “canonici” (es. OPAC, cataloghi on-line, etc.) per il fatto che cambia l'esigenza informativa dell'utente, come verrà illustrato nelle slide successive.



Dipendenze e Funzionamento

- **Dipendenze:**

- Mono



- Gnome



- **Funzionamento:**

- Demone **beagled** in background

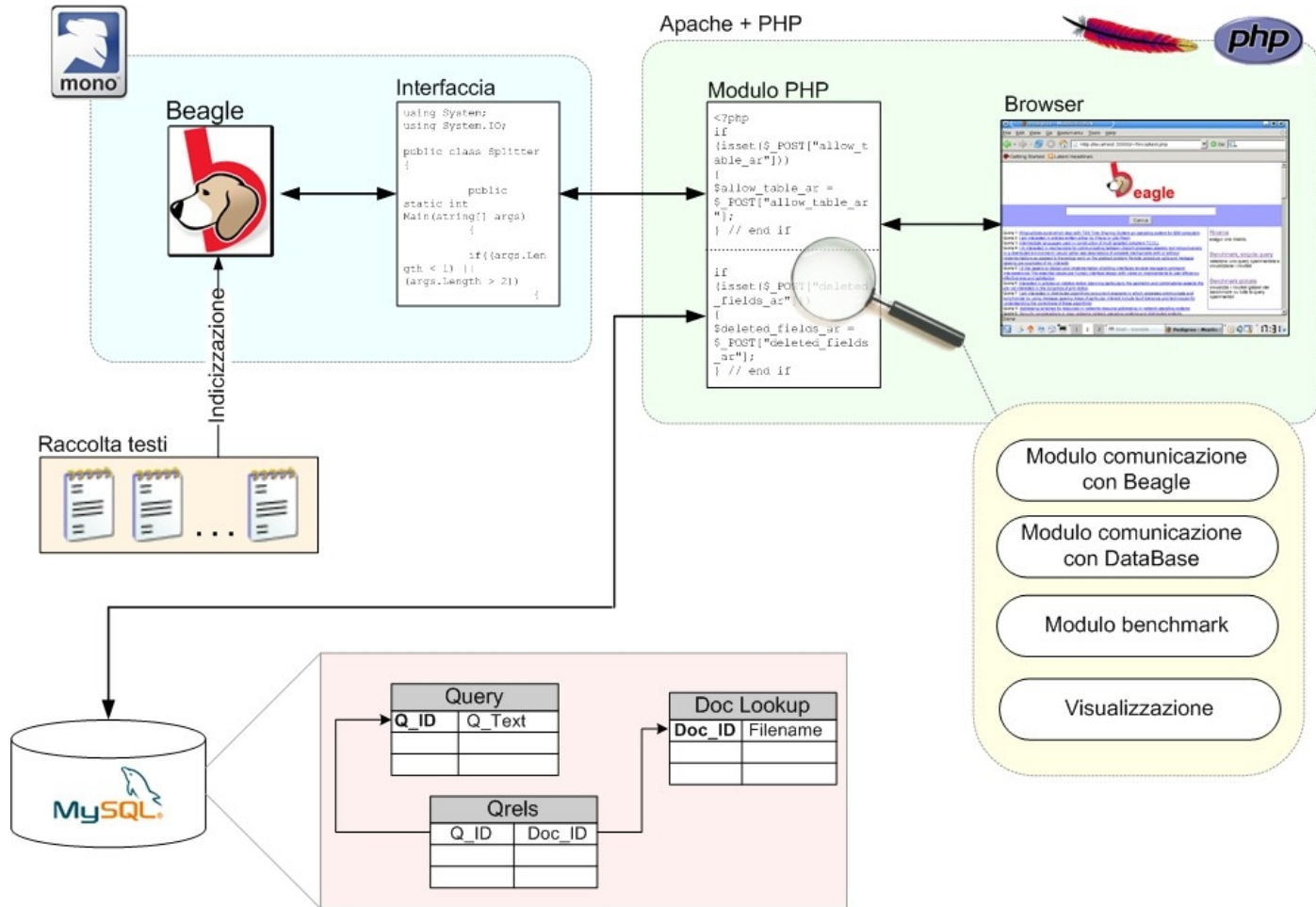
- Indicizzazione

- Funzione di Reperimento

- Front-end grafico semplice



Schema del Sistema





Quale Esigenza Informativa?

- **Collezione:**
 - “Limitata”
 - “Nota” a priori
 - Aggiornata all'ultima modifica (stato corrente del File System)
- **Esigenza Informativa:**
 - L'utente **non** desidera trovare **nuovi** documenti, riguardanti un argomento specifico, all'interno di una vasta collezione (es. web)
 - La ricerca è mirata verso un preciso documento
 - L'esistenza del documento è certa (a meno di amnesie 😊)



AND-filling VS OR-filling

Esigenza Informativa

- Utente di una DSU: trovare un **documento preciso**, che contenga tutti i descrittori della query.

soddisfatta con AND

- Utente che desidera trovare i **documenti** che contengono il **maggior numero** di descrittori presenti nell'interrogazione, sperando che tra essi vi sia la risposta alla propria necessità

soddisfatta con OR

➔ Beagle esegue l'AND-filling automatico ←




Caratteristiche Tecniche 1/3

- Tipo di **indicizzazione**: incrementale
- Caratteristiche della **Collezione**:
 - Monolingua
 - Multimediale
 - Documenti di testo in vari formati
 - Immagini
 - Audio
 - e-mail
 - **Codifica**: UTF-8



Caratteristiche Tecniche 2/3

- **Modello:** Booleano con Ranking.
- **Ranking:** basato su term-frequency.
- **Posting File:** realizzato con una struttura di tipo B-tree.
- **Presentazione:** Interfaccia di tipo Desktop
 - Altre interfacce (es int. web) possono essere implementate.
 - Pedigree: interfaccia web in php; gira su 
- **Livello:** Almeno per i documenti testuali viene presentato un piccolo estratto.



Caratteristiche Tecniche 3/3

- **Utenza** (multiutenza Linux-like)
- **Collezione** (multicollezione)
- **Log**
 - current
 - exceptions
 - index-helper
- **Sessione:** NO
- **Personalizzazione:** directory e backend



Verifiche Sperimentali

Data la **scarsità** della **documentazione**, alcune caratteristiche di Beagle sono state ricavate empiricamente via **reverse-engineering**:

- **Stemming**
- **Stop List**
- Sistema di **Ranking**
- Dizionario: **Keyword** / Termini

11



18



Parametro	AND	#queries	OR	#queries
P_{AVG}	0.16	9	0.01	64
R_{AVG}	0.02	52	0.11	52
F_{AVG}	< 0.01	64	0.03	64



Screenshot 1/5

Desktop Search: computer develop

Search Sort Help

Find: com [] Find Now

Date Modified
Name
• Relevance

Documents 1-32 of 61

1771.txt Feb 20	2197.txt Feb 20	2919.txt Feb 20	2311.txt Feb 20	2181.txt Feb 20	1769.txt Feb 20	1745.txt Feb 20	1099.txt Feb 20
1572.txt Feb 20	2985.txt Feb 20	2899.txt Feb 20	2812.txt Feb 20	2322.txt Feb 20	2999.txt Feb 20	1805.txt Feb 20	2931.txt Feb 20
1003.txt Feb 20	2977.txt Feb 20	1008.txt Feb 20	2930.txt Feb 20	2829.txt Feb 20	2163.txt Feb 20	2319.txt Feb 20	2304.txt Feb 20
1071.txt Feb 20	2357.txt Feb 20	2924.txt Feb 20	2188.txt Feb 20	2153.txt Feb 20	1697.txt Feb 20	2303.txt Feb 20	2378.txt Feb 20

Title: 2919.txt
Last Edited: February 20
Full Path: /home/flinco/public_html/cacm_collection/2919.txt
s Workbench A Machine for Software **Development** On almost all software **development** projects t...

Schermata principale di Beagle dopo una ricerca



Screenshot 2/5

The screenshot displays a Mozilla Firefox browser window with the title "Pedigree - Mozilla Firefox". The address bar contains the URL "http://localhost/~flinco/". The main content area features the Pedigree logo (a dog's head in a red 'b') and the word "eagle" in red. Below this is a search bar with a "Cerca" button. A sidebar on the right lists search options: "Ricerca" (esegui una ricerca), "Benchmark, singola query" (seleziona una query sperimentale e visualizzane i risultati), and "Benchmark globale" (visualizza i risultati globali del benchmark su tutte le query sperimentali). The system tray at the bottom shows the time as 17:31 and includes icons for network, volume, and other system functions.

Schermata principale dell'interfaccia web realizzata




Screenshot 3/5

Pedigree - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost/~flinco/

Getting Started Latest Headlines

 eagle

Cerca

Query 1: [What articles exist which deal with TSS Time Sharing System an operating system for IBM computers](#)

Query 2: [I am interested in articles written either by Prievé or Udo Pooch](#)

Query 3: [Intermediate languages used in construction of multi targeted compilers TCOLL](#)

Query 4: [I am interested in mechanisms for communicating between disjoint processes possibly but not exclusively in a distributed environment I would rather see descriptions of complete mechanisms with or without implementations as opposed to theoretical work on the abstract problem Remote procedure calls and message passing are examples of my interests](#)

Query 5: [I'd like papers on design and implementation of editing interfaces window managers command interpreters etc The essential issues are human interface design with views on improvements to user efficiency effectiveness and satisfaction](#)

Query 6: [Interested in articles on robotics motion planning particularly the geometric and combinatorial aspects We are not interested in the dynamics of arm motion](#)

Query 7: [I am interested in distributed algorithms concurrent programs in which processes communicate and synchronize by using message passing Areas of particular interest include fault tolerance and techniques for understanding the correctness of these algorithms](#)

Query 8: [Addressing schemes for resources in networks resource addressing in network operating systems](#)

Query 9: [Security considerations in local networks network operating systems and distributed systems](#)

Ricerca
esegui una ricerca

[Benchmark, singola query](#)
seleziona una query sperimentale e visualizzane i risultati

[Benchmark globale](#)
visualizza i risultati globali del benchmark su tutte le query sperimentali

Done

Shell - Konsole Pedigree - Mozilla 17:31

Interfaccia web: interrogazione scegliendo una query dall'elenco

15

U
I
I
18



Screenshot 4/5

Pedigree - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost/~flinco/test.php?qid=19

Getting Started Latest Headlines

Query 19: Parallel algorithms

Documenti rilevanti : 11
Documenti rilevanti reperiti: 4
Risultati trovati: 11

Precision: 0.363636363636
Recall: 0.363636363636
Fallout: 0.00219229564673

[Documento 2973](#)
[Documento 2838](#)
[Documento 2289](#)
[Documento 2266](#)
[Documento 2401](#)
[Documento 2433](#)
[Documento 2195](#)
[Documento 2114](#)
[Documento 1601](#)
[Documento 1613](#)
[Documento 950](#)

[Ricerca](#)
esegui una ricerca

[Benchmark, singola query](#)
seleziona una query sperimentale e visualizzane i risultati

[Benchmark globale](#)
visualizza i risultati globali del benchmark su tutte le query sperimentali

Done

Shell - Konsole Pedigree - Mozilla Firefox 17:32

Risultati dell'interrogazione




Screenshot 5/5

Pedigree - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost/~flinco/index.php

Getting Started Latest Headlines

 eagle

computer develop

Cerca

Query: computer develop

Trovati 81 risultati:

[Documento 3024](#)
[Documento 3027](#)
[Documento 2999](#)
[Documento 2977](#)
[Documento 2985](#)
[Documento 3028](#)
[Documento 3039](#)
[Documento 3059](#)

[Ricerca](#)
esegui una ricerca

[Benchmark, singola query](#)
seleziona una query sperimentale e visualizzane i risultati

[Benchmark globale](#)
visualizza i risultati globali del benchmark su tutte le query sperimentali

Done

Shell - Konsole Pedigree - Mozilla Firefox 17:34

Risultati della ricerca inserendo una query qualsiasi



Grazie per la cortese attenzione

il Pedigree Team