



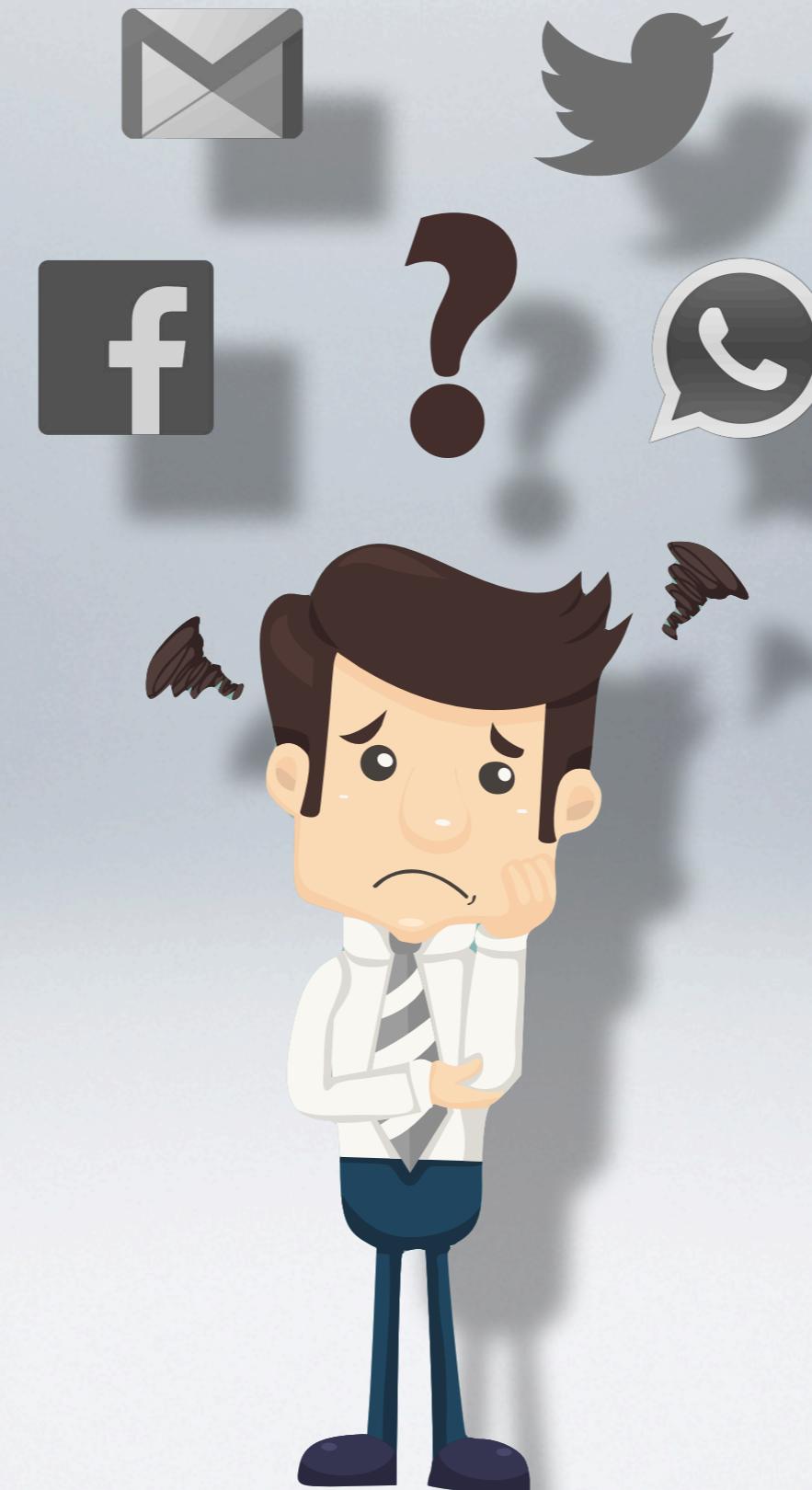
UNIVERSITÀ DEGLI STUDI DI BERGAMO

DIPARTIMENTO DI INGEGNERIA

Moogle: un Web crawler e applicazione
di ricerca full-text per dati privati
distribuiti in rete

Paolo Coffetti
Relatore: Prof. Stefano Paraboschi







World Wide Web

Cloud pubblico



Cloud privato





World Wide Web

Cloud pubblico



Cloud privato





Moogle
my own google

“ Come Google ma
per i miei dati privati ”



Dashboard

Moogle

← → ⌂ <http://moogle.io/dashboard> ⌂

Moogle paolo ?

You are currently using these providers

Add new providers!



Ricerca

Moogle

← → ⌂ http://moogle.io ≡

Moogle paolo ?

Search Submit

A screenshot of the Moogle search interface. At the top, there's a navigation bar with icons for back, forward, refresh, and a menu. Below it is a header with the 'Moogle' logo and a user profile for 'paolo'. The main area features five large social media sharing icons: Gmail, Google Drive, Facebook, Twitter, and Dropbox. At the bottom is a search bar with the placeholder 'Search' and a 'Submit' button.



Risultati

Moogle

↔ C http://moogle.io?q=work

Moogle paolo ?

Email

✉ motivation letter 18 Apr
to Sarah Zhang
New version of motivation letter: To whom it may concern, Hereby I apply for the SPRINT program, held in the...
...highly formative because I was **working** side by side with...

✉ [2] Lettera di motivazione 18 Apr
to stefania calabrese
Ho appena notato " Advance is a solid company with a proved ..." che correggo in " Advanse International Inc...."
...highly formative because I was **working** side by side with...

Get more ***

Social network

f Marcello Coffetti > Paulo Van Der Kofje 12 Mar 12
Unfortunately the printer is not **working** anymore...

f Tommaso Amicabile > Paulo Van Der Kofje 9 Mar 12
... and how long have you been **working** there?

Get more ***

Cloud storage

Cloud storage Description 25 Apr
Shared by puntonim@gmail.com M O O GLE MY OWN GOOGLE DRAFT Table of content...
...laptop, my home and **working** computer, my mobile...

Cloud storage SVN comandi 16 Apr
Shared by puntonim@gmail.com Tue 13 Dec 2011 03:43:10 PM CET SVN COMANDI 1. ...
...mio computer cd ~/workspace/print2book - Lancio...

Get more ***

Mobile

c Sarah 18 Apr
I'm going to **work** after 10 today because I have to stop by ...

w Sarah 18 Apr
And also you have to consider the fact that in the last few months the situation here was pretty fair and this might have ...
... have you been **working** with him lately? If so, how do you ...

Get more ***



Progetto

- Venture capitals ad Amsterdam e Washington DC
- Durata: 1 anno
- Lingua inglese

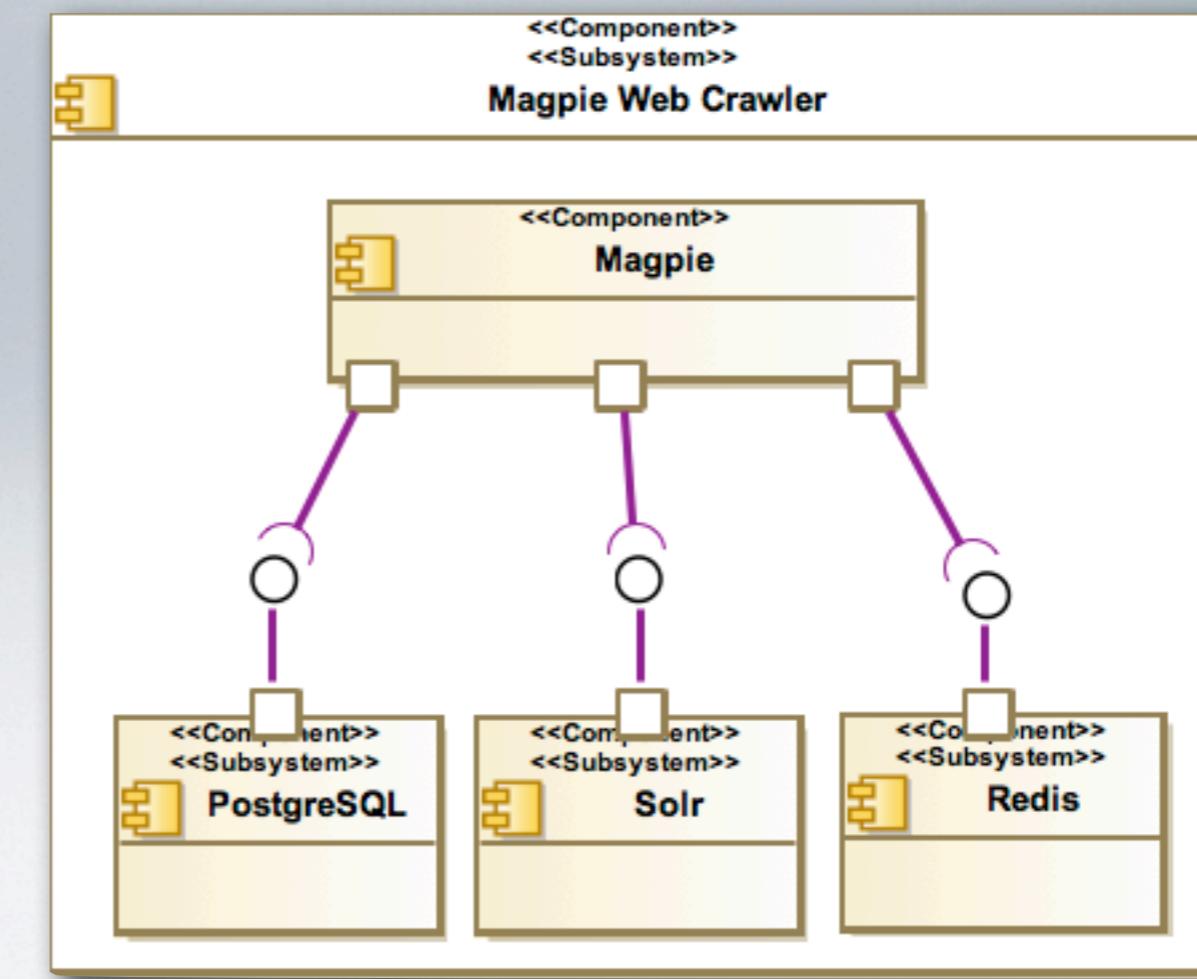
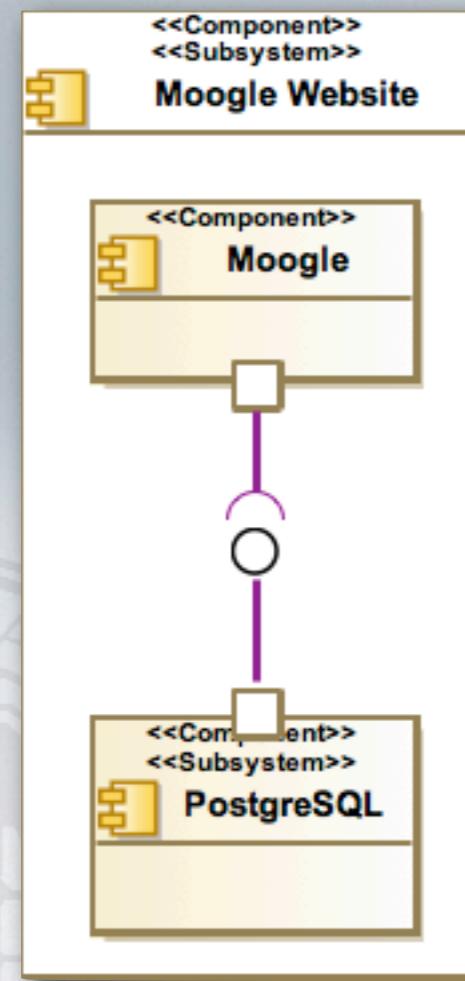


Specifiche tecniche

- **Provider supportati:** Gmail, Drive, Dropbox, Facebook, Twitter
- **Requisito: estendibilità**
- **Tecnologie comuni:** OAuth, RESTful webservice
Ma API diverse...
- **Architettura: astrazione e ereditarietà**

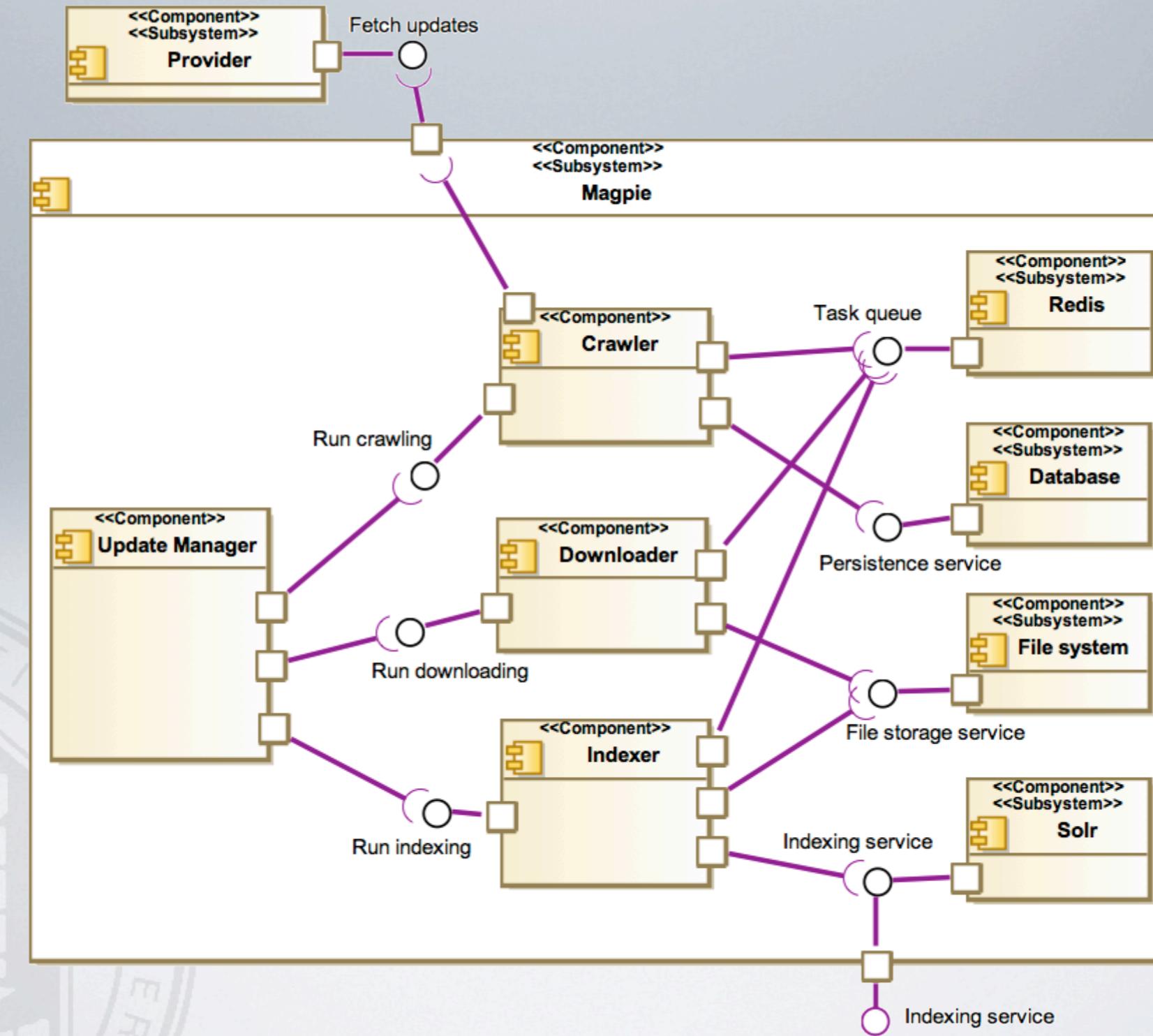


Architettura





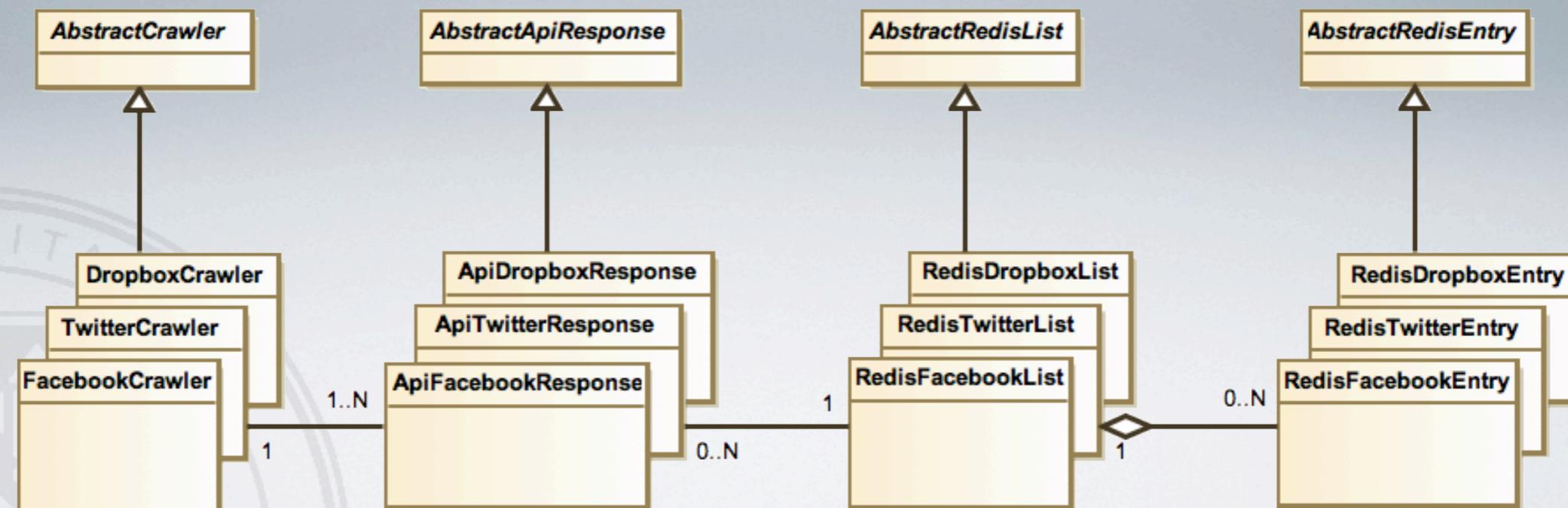
Componente: Magpie





Astrazione e Ereditarietà

Template Pattern





Tecnologie e aspetti più interessanti

- **Redis** per gestire la coda di task
- **Apache Solr** come full-text search engine
- **Docker** per il deployment
- Processo di sviluppo software **iterativo e incrementale**
(RUP di Booch, Kruchten)



Risultati

- Architettura
- Prestazioni
 - 2x Amazon EC2 micro
 - Indice: ~10k documenti
 - Trisp < 2 sec
- Usabilità
 - ✓ UX
 - ✗ Privacy



Sviluppi futuri

- Nuovi **provider** di dati
- Analisi testuale con **identificazione** della lingua
- **Privacy**: nuove piattaforme come router, smart TV, hard drive esterni, smart phone, ...