

By:
Cody Hansen

ColumbuScout

Towards Building Local Search Engines over Large Databases

Professor:
Feifei Li

Small Site Search

Please fill in any information for what you want to search for.
Please select what you want to search for: Publication

Author(s):

Title:

Publisher:

Abbreviation:

Keyword(s):

Status:

Publication Type:

Why have a form based search?

- Slow
- No Spell Checking
- No Suggestions
- Search in Specific Boxes

Large Site Search

Cody Hans

cody hansen
cody hansen hinder
cody hansen facebook
cody hansen married
cody hansen twitter
cody hansen flickr
cody hansen photography
cody hansen facebook
cody hansen biography
cody hansen hockey

[Cody Hans | Facebook](#)
www.facebook.com/codyhans

[Cody Hans is on Facebook](#). Join Facebook to connect with Cody Hans and others you may know. Facebook gives people the power to share and makes it...

[Cody Hans Profiles - Facebook](#)
www.facebook.com/public/CodyHans

View the profiles of people named Cody Hans on Facebook. Join Facebook to connect with Cody Hans and others you may know. Facebook gives people the...

When we are all used to this.

- Fast
- Spell Checking
- Suggestions
- Search Any

Combination of Words

A Search Engine With the Functionality of a Large Search Engine that is Deployable on your own Database

Uses:

- Ranking Based on Rarity and Accuracy
- Keyword Search
- Approximate String Search
- Section Based Parsing and Merging
- Prefix Search
- Associative Keyword Search
- Map-Reduce Parsing

- Efficient
- Effective

Query Times

Keywords	Full Craigslist		Full LinkedIn	
	1	3	1	3
Top 200	0.0286	0.0669	0.0157	0.0408
Top D	0.0353	0.0889	0.0506	0.1359

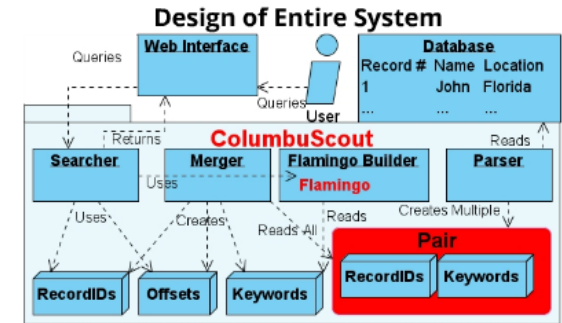
File Triplet

Keywords	Rids File	Offsets
w ₁	n ₁ rid ₁ , 1, . . . , rid ₁ , n ₁	offset ₁
w ₂	n ₂ rid ₂ , 1, . . . , rid ₂ , n ₂	offset ₂
...
w _m	n _m rid _m , 1, . . . , rid _m , n _m	Offset _m

With these three files a quick keyword search can be done to return all the records a given keyword is in

Search 1.7 trillion records in fractions of a second
Parse 1.7 trillion records in less than an hour

How It Works



This figure describes the system as a whole which is largely about the construction of the system

A Typical Search

LinkedIn Search

Show Page: 1 2 3 4 5 6 7 8 9 10 12 Million Records Number of Records Matched: 275320

Enter Exact Keywords Here

The keyword(s) cody cody florida matched 1 record(s)

The keyword(s) cody florida matched 107 record(s)

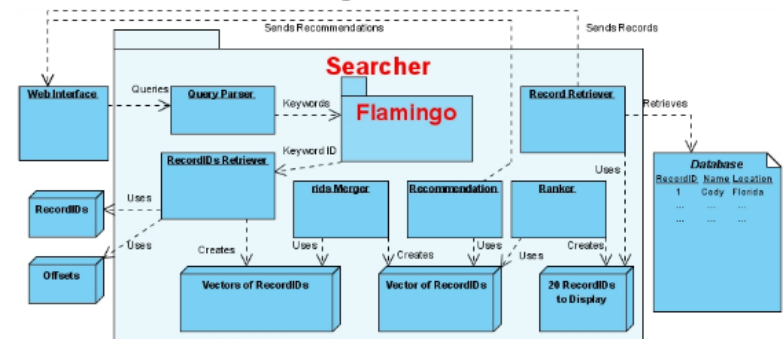
The keyword(s) cody florida matched 1 record(s)

The keyword(s) codd florida matched 1 record(s)

The keyword(s) loria cory matched 1 record(s)

ID	Name	Title	Location	Education	Ranking Value
2548790 Original Post	cody moan	Student at University of Florida	Jacksonville, Florida Area	Education	1
2216446 Original Post	Cody Pierce	Student at University of Central Florida	Sarasota, Florida Area	Education	1

Design of Searcher



This figure describes how a search works