



Grid Dynamics

Scalable eCommerce Platform Solutions

Looking for TOMMY HILFINGER



Mikhail Khudnev
Principal Engineer
mkhudnev@griddynamics.com



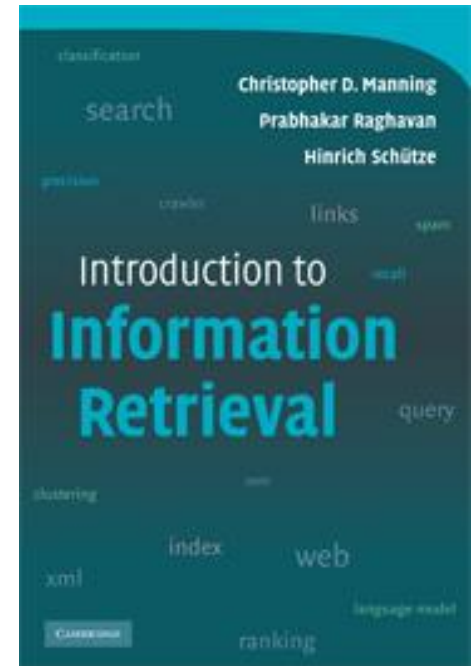
To do this in the search box

Type the important words: `tricolor rat terrier`

Put exact words in quotes: `"rat terrier"`

Type OR between all the words you want: `miniature OR standard`

Put a minus sign just before words you don't want:
`-rodent, -"Jack Russell"`



<http://nlp.stanford.edu/IR-book/>

Does she search this way?



+"michael kors" type:handbag

Search

http://tustinhairsalon.com/wp-content/uploads/2011/02/Blonde_Hair_stylist_s_tylists_blondes_tustin_santa_ana_orange_cou

That's how she searches!

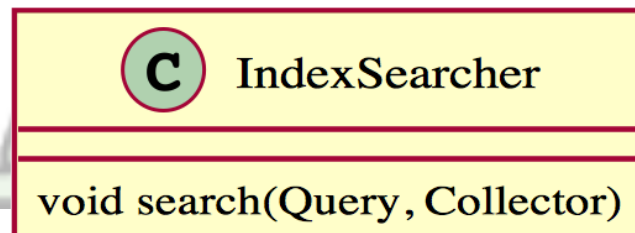


michael kors hand bag

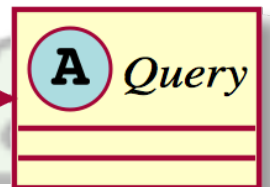
Search

http://tustinhairsalon.com/wp-content/uploads/2011/02/Blonde_Hair_stylist_s_tylists_blondes_tustin_santa_ana_orange_cou

Lucene

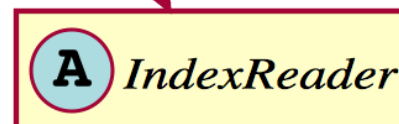
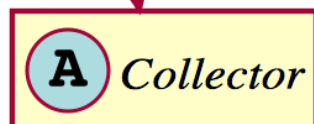


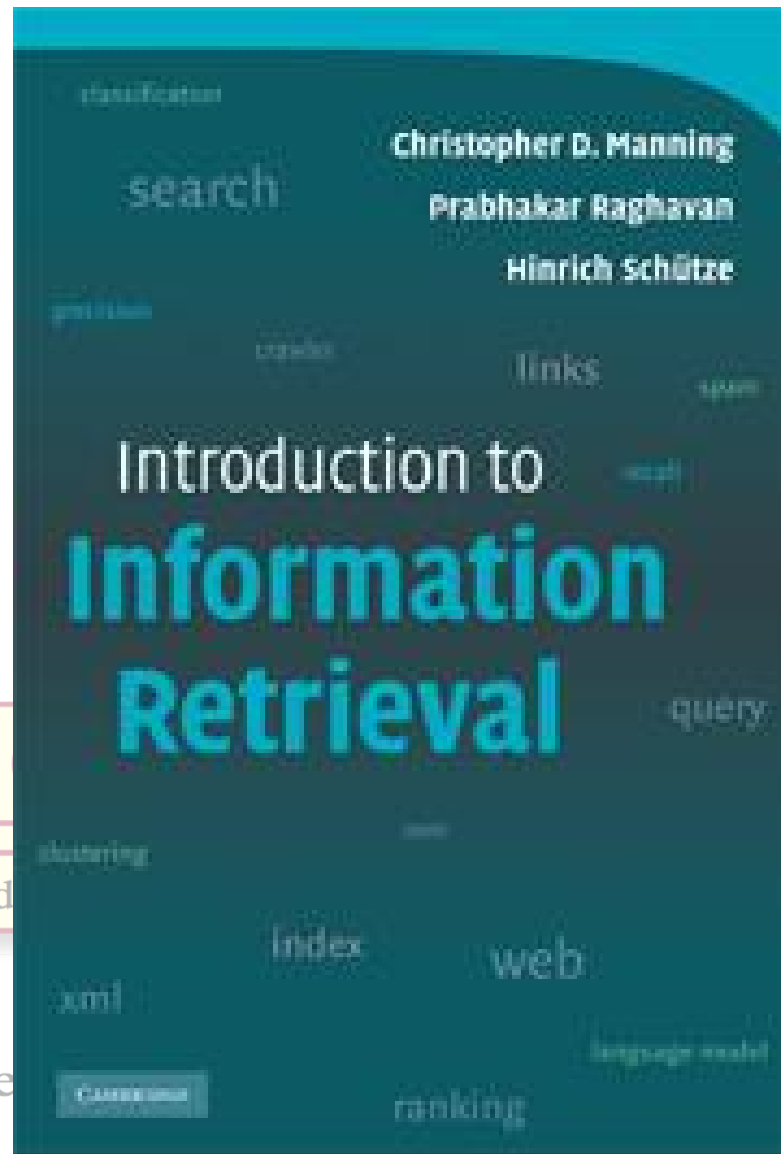
searches for



collects results into

reads index files



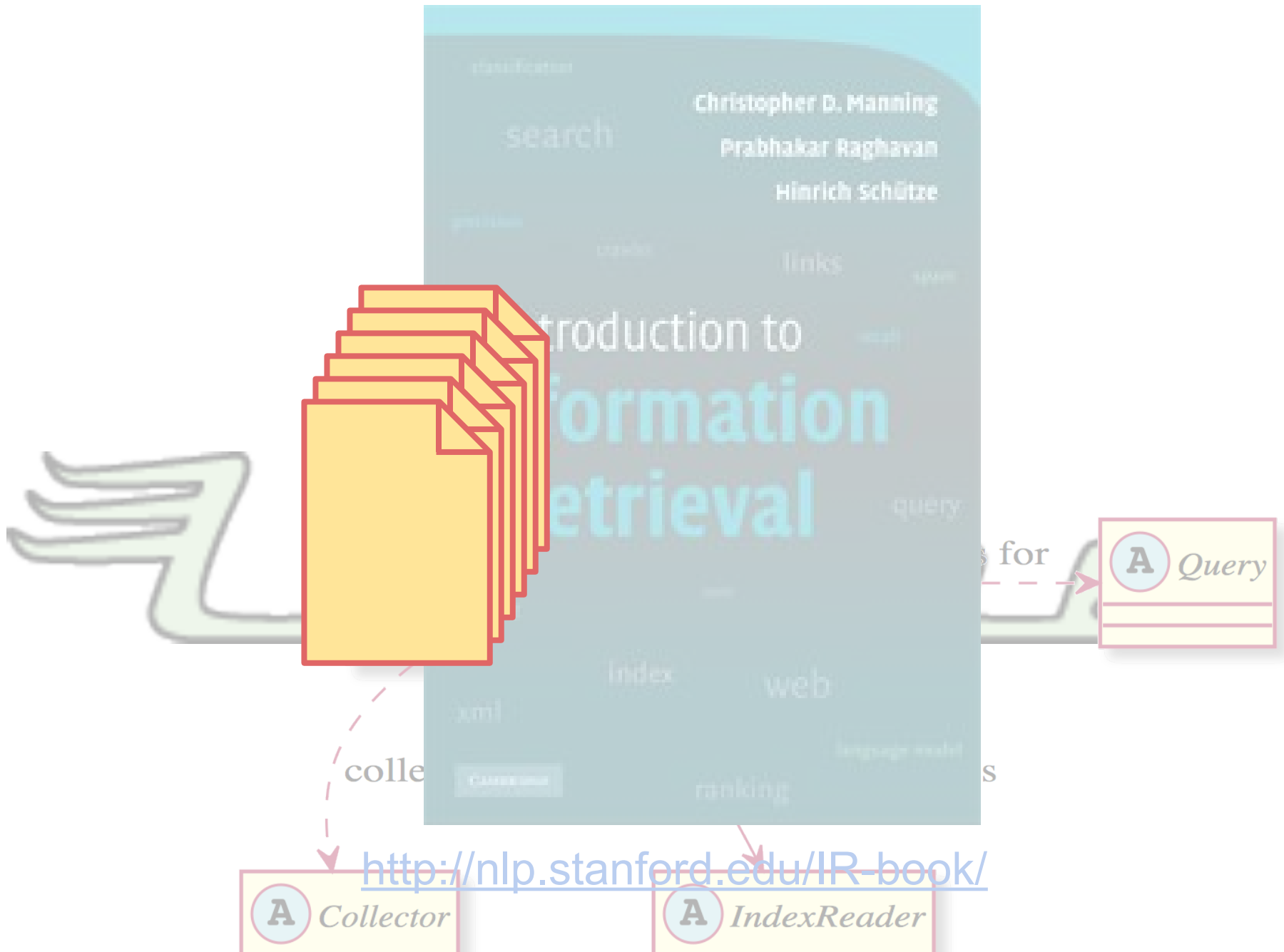


<http://nlp.stanford.edu/IR-book/>

A Collector

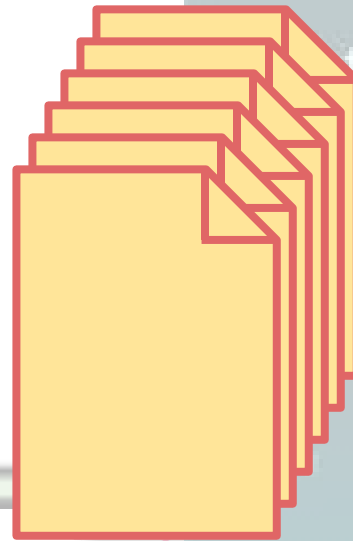
A IndexReader

A Query





"lorem"
"ipsum"
"dolor"
"sit"
"amet"
"consectetur"



<http://nlp.stanford.edu/IR-book/>



$O(n)$

"lorem"
"ipsum"
"dolor"
"sit"
"amet"
"consectetur"

<http://nlp.stanford.edu/IR-book/>

A Collector

A IndexReader

A Query

OOME

Christopher D. Manning
Prabhakar Raghavan
Hinrich Schütze

"lorem"

"ipsum"

"dolor"

"sit"

"amet"

"consectetur"

collector

A Collector

<http://nlp.stanford.edu/IR-book/>

A IndexReader

A Query

WHY SO SERIOUS?

Wu

<http://hdwallpaper.ws/2667/>

RDBMS

ID	BRAND	SIZE	COLOR
----	-------	------	-------

23	Adidas	XL	Blue
----	--------	----	------

45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	----	-----

INSERT INTO .. VALUES

(17, 'Tommy Hilfiger', 'S', 'White')

ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	----	------

45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	----	-----

INSERT INTO .. VALUES

(17, 'Tommy Hilfiger', 'S', 'White')

ID BRAND SIZE COLOR

23 Adidas XL Blue

45 Reebok M Red

61 Nike XL Red



17 Tommy Hilfiger S White

SELECT * FROM ..

WHERE SIZE='XL'

ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	----	------

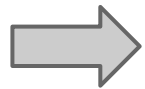
45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	----	-----

17	Tommy Hilfiger	S	White
----	----------------	---	-------

SELECT * FROM ..

WHERE SIZE='XL'



ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	-----------	------

45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	----	-----

17	Tommy Hilfiger	S	White
----	----------------	---	-------

SELECT * FROM ..

WHERE SIZE='XL'

ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	-----------	------

45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	----	-----

17	Tommy Hilfiger	S	White
----	----------------	---	-------

SELECT * FROM ..

WHERE SIZE='XL'

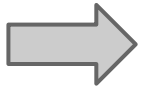
ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	-----------	------

45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	-----------	-----

17	Tommy Hilfiger	S	White
----	----------------	---	-------



SELECT * FROM ..

WHERE SIZE='XL'

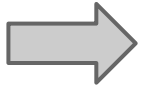
ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	-----------	------

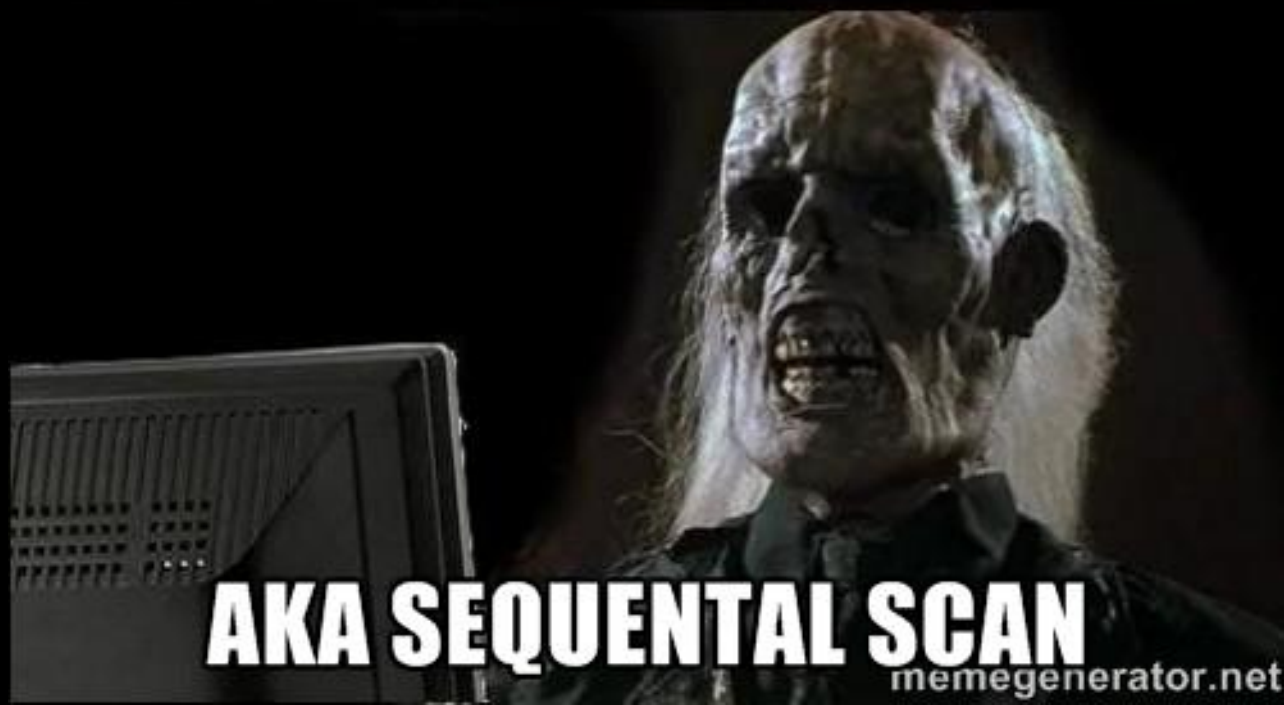
45	Reebok	M	Red
----	--------	---	-----

61	Nike	XL	Red
----	------	-----------	-----

17	Tommy Hilfiger	S	White
----	----------------	---	-------



FULL TABLE SCAN



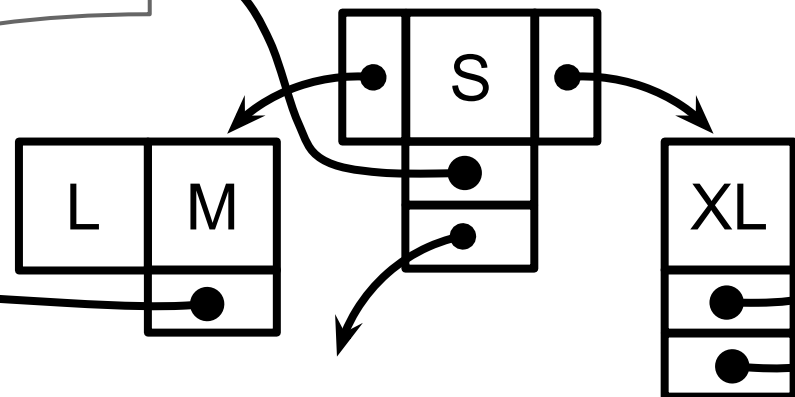
AKA SEQUENTIAL SCAN

memegenerator.net

CREATE INDEX SIZE_FK

ON .. (SIZE)

ID	BRAND	SIZE	COLOR
23	Adidas	XL	Blue
45	Reebok	M	Red
61	Nike	XL	Red
17	Tommy Hilfiger	S	White



SELECT * FROM ..

WHERE SIZE='XL' AND COLOR='Red'

ID	BRAND	SIZE	COLOR
-----------	--------------	-------------	--------------

23	Adidas	XL	Blue
----	--------	-----------	------

45	Reebok	M	Red
----	--------	---	------------

61	Nike	XL	Red
----	------	-----------	------------

17	Tommy Hilfiger	S	White
----	----------------	---	-------

SELECT * FROM ..

WHERE SIZE='XL' AND COLOR='Red'

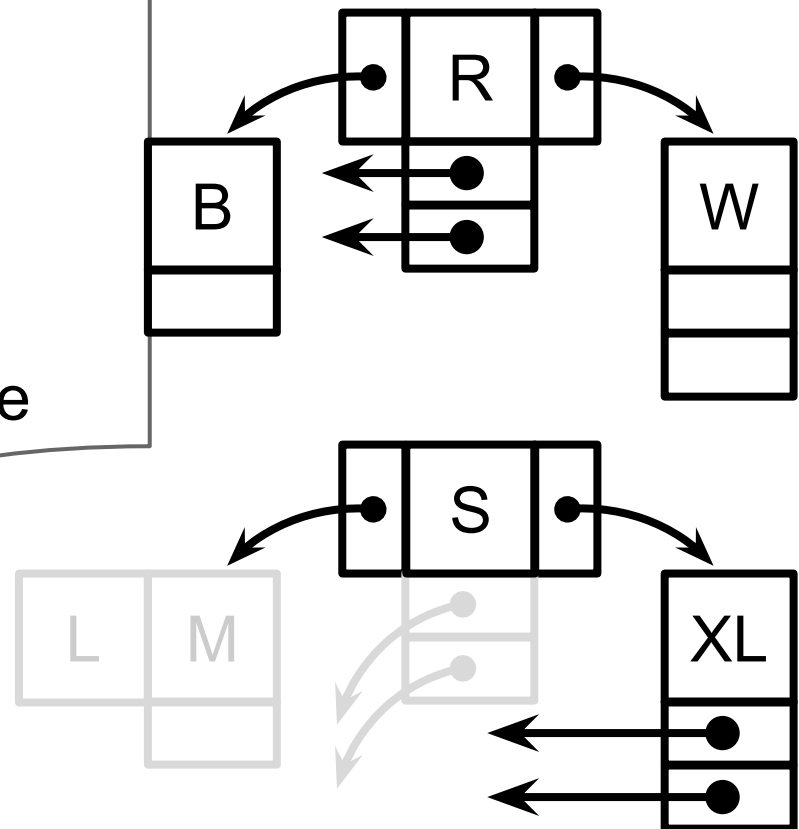
ID	BRAND	SIZE	COLOR
----	-------	------	-------

23	Adidas	XL	Blue
----	--------	-----------	------

45	Reebok	M	Red
----	--------	---	------------

61	Nike	XL	Red
----	------	-----------	------------

17	Tommy Hilfiger	S	White
----	----------------	---	-------

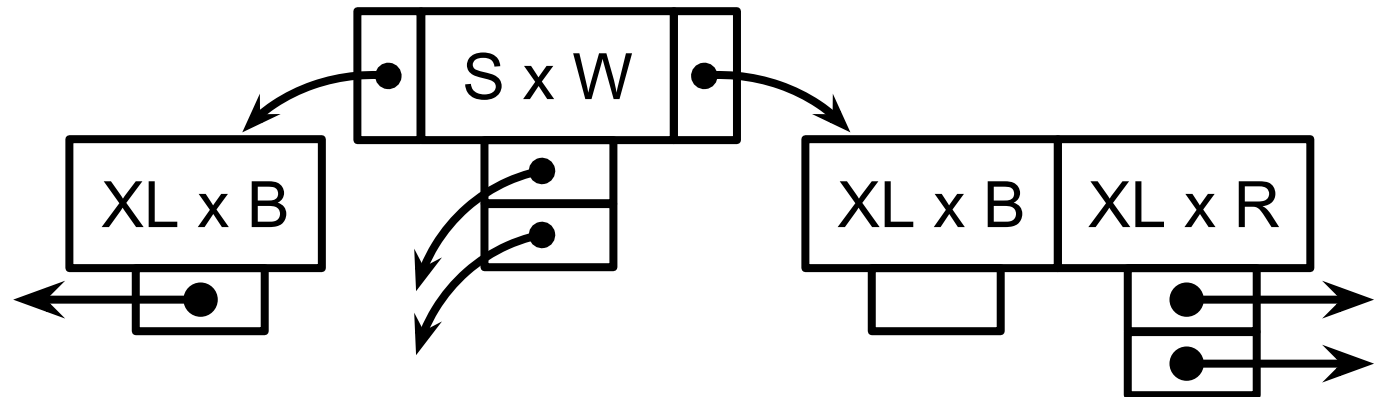


SELECT * FROM ..

WHERE SIZE='XL' AND COLOR='Red'

CREATE INDEX SIZE_COLOR_FK

ON .. (SIZE, COLOR)



Inverted Index

```
T[0] = "it is what it is"  
T[1] = "what is it"  
T[2] = "it is a banana"
```

```
"a":          {2}
"banana":     {2}
"is":         {0, 1, 2}
"it":         {0, 1, 2}
"what":       {0, 1}
```

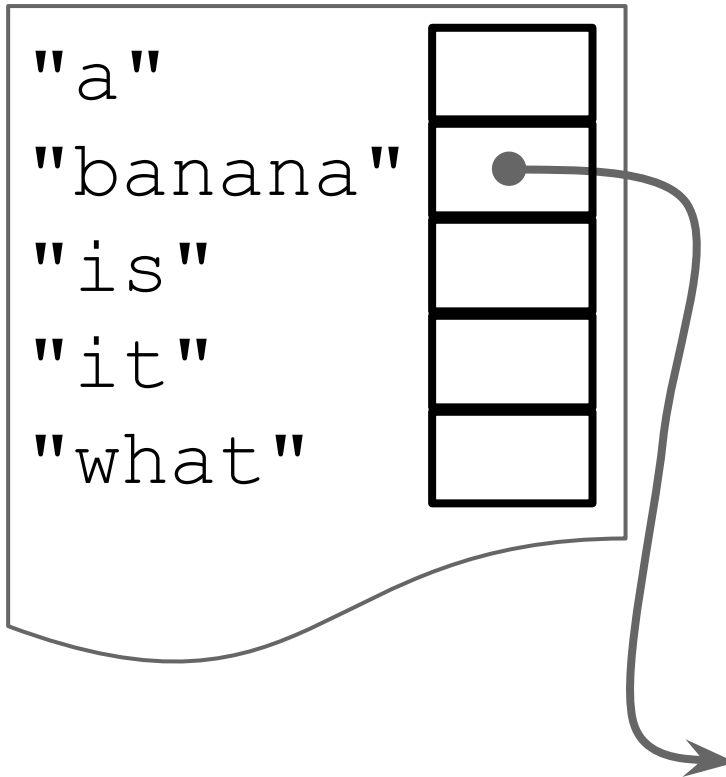
```
T[0] = "it is what it is"
T[1] = "what is it"
T[2] = "it is a banana"
```

"a": {2}
"banana": {2}
"is": {0, 1, 2}
"it": **{0, 1, 2}**
"what": {0, 1}

postings list

term dictionary

index/_1.tis



index/_1.frq

{ 2 }
{ 2 }
{ 0, 1, 2 }
{ 0, 1, 2 }
{ 0, 1 }

Collegii S. Petri & S. Pauli

anno Gandavi 1630.

PA

¶ Tabula brevis & utilis super libello
quodam qui dicitur fasciculus temporum:
& ubi inuenitur punctus ante quem est
in primo latere: vbi vero post in secundo
latere: Incipit feliciter.

Bacū ppha minor	.15
Abbaui martyr	.32.
Abdias ppha	.11 12.
Abdon iudeus	.8.
Abdon & sennes martyres	.32.
Abdo ppha ostra hierosol	.9.
Abel pph ^a martyr innocens	.2.
Abeslan iudeus isrl ^a ton ^a	.8.
Abya rex iuda	.10
Abiabar pontifex gl ^a osus	.9
Abia filius aaron	.7
Abimelech iudeus isrl ^a malus	.8
Abimelech pontifex	.10
Abisue pontifex	.8
Abiab rex fili ^a corobatel	.17
Abesab nascif indoans filia etate	.5.
Abraham moris	.46

A c

¶ Achaz rex iuda	.13
Achitold ^a eph ^a vintoniensi.	.52
Achias monitas ppha	.9
Achymas fili ^a sadoch pontifex	.9
Abimelech sacerdos oili	.9
Achim rex	.19.
Achis fili ^a rex alban ^a	.9.
Achitob pontifex	.9
Achitob sacerdos	.9
Achitob pontifex	.9

African^a iulius

.32

A g

¶ Agmenon rex grecor	.18.
Agar famula sace	.5.
Agapit ^a martyr	.32.
Agapinus martyr	.33
Agapes	.33.
Agapitus papa pph ^a	.39.
Agapit ^a papa secundus	.51.
Agene ^a virdimentis eph ^a	.40
Aggeus pphat	.15
Aggeus increpat populū	.17
Agiale ^a pph ^a rex schibonicoz	.5
Agilis abbas scis	.42.
Agilopit ^a eph ^a colonici.	.45.
Agatho papa	.43.
Agatha virgo & martyr	.32
Agnes virgo & martyr	.33.
Agricola martyr	.33.
Agrippa filius rex italie	.11
Agrippina ciuitas	.28

A b

¶ Abab rex isrl ^a idolatra	.10.
Abialon iudeus	.8.
Abadan ^a eph ^a	.42.
Abmund ^a rex anglie	.49

A l

¶ Alanus doctor vniuersal	.60
Alba filius rex albanie	.9.
Albert ^a magn ^a	.58.
Albertus impator	.60
Albertus impator	.63.
Albert ^a pph ^a hierosolimoz	.57.

Alexander papa secū ^a	.54
Alexander papa tert ^a	.56.
Alexander papa quart ^a	.59
Alexander papa quint ^a	.62.
Alexandria martyr	.36
Almania	.51
Almericus heretic ^a	.57.
Alphonius rex castelle	.59
Altinus discipul ^a sci petri	.27.

A m

¶ Amalech telei p saul	.9
Aman ^a eph ^a aureliacei.	.36.
Amandus vir scis eph ^a	.40
Amarias sacerdos	.9
Amarias pontifex	.12.
Amathus exulaf	.44
Amatonū sine massagetarū re-	
gnum oritur	.4.
Amagones sunt mulieres	.8.
Amagias rex iuda	.12
Ambrosius roman ^a doctor	.35
Amelius comes aluerni.	.47.
Amic ^a & berican ^a martyres	.47.
Amic ^a rex assyrioz	.7.
Aminadab princeps iuda	.7
Aminius peruator indee	.25
Amithus rex assyrioz	.6.
Amnon filius iob	.5.
Amnon monachus	.36
Amnon rex iuda	.14
Amos ppha	.12.
Amos pph ^a 3000. mōachoz	.34.
Amram filius caath	.6.

What is a Scorer?


```
"a":          {2}
"banana":     {2}
"is":         {0, 1, 2}
"it":         {0, 1, 2}
"what":       {0, 1}
```

```
"a":          {2}
"banana":     {2}
"is":         {0, 1, 2}
"it":         {0, 1, 2}
"what":       {0, 1}
```

```
"a":          {2}
"banana":     {2}
"is":         {0, 1, 2}
"it":         {0, 1, 2}
"what":       {0, 1}
```

A *DocIdSetIterator*

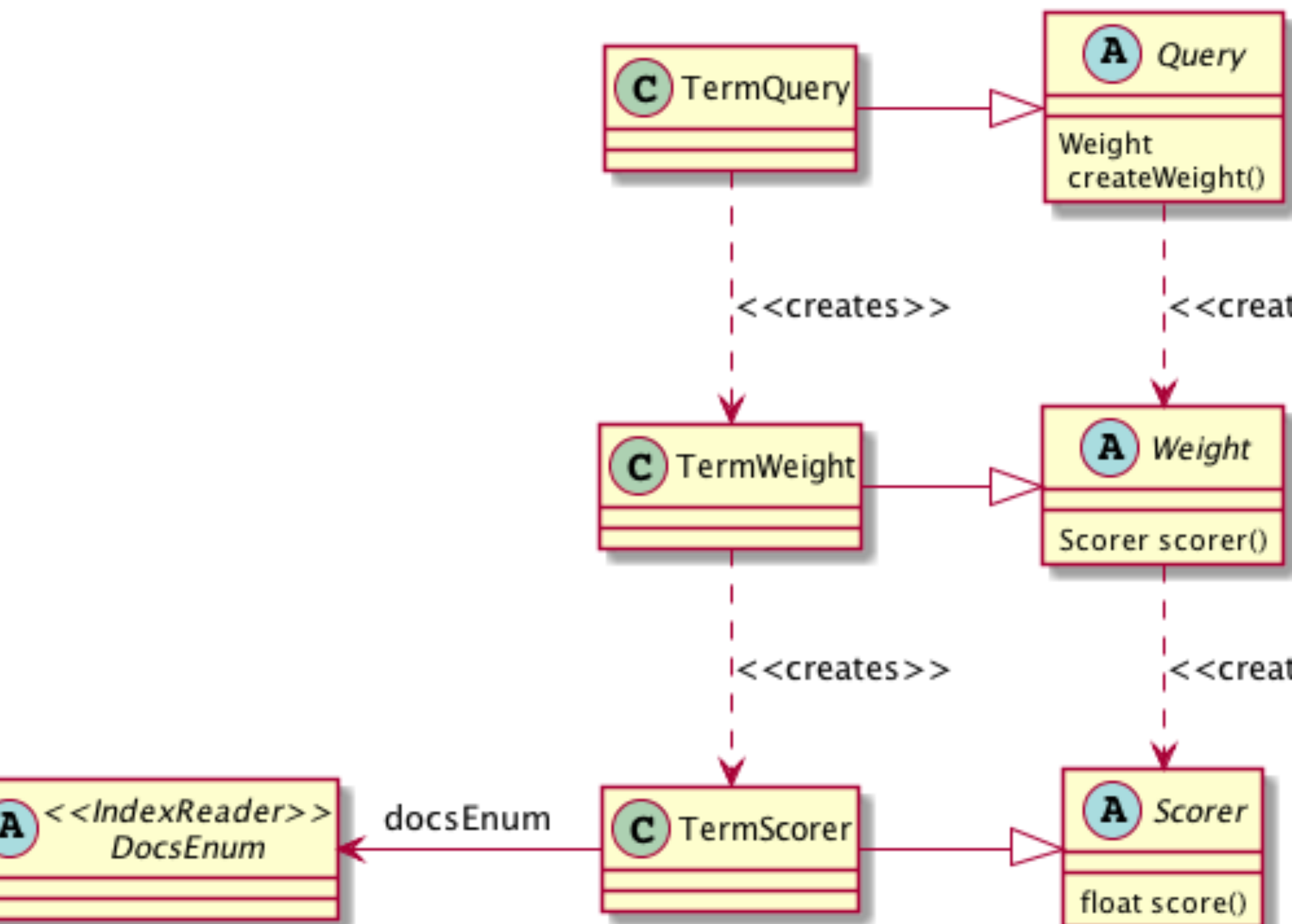
int docID()
int nextDoc()
int advance(int docID)
long cost()



A *Scorer*

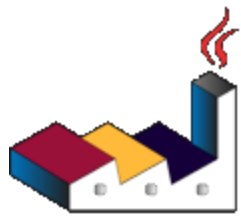
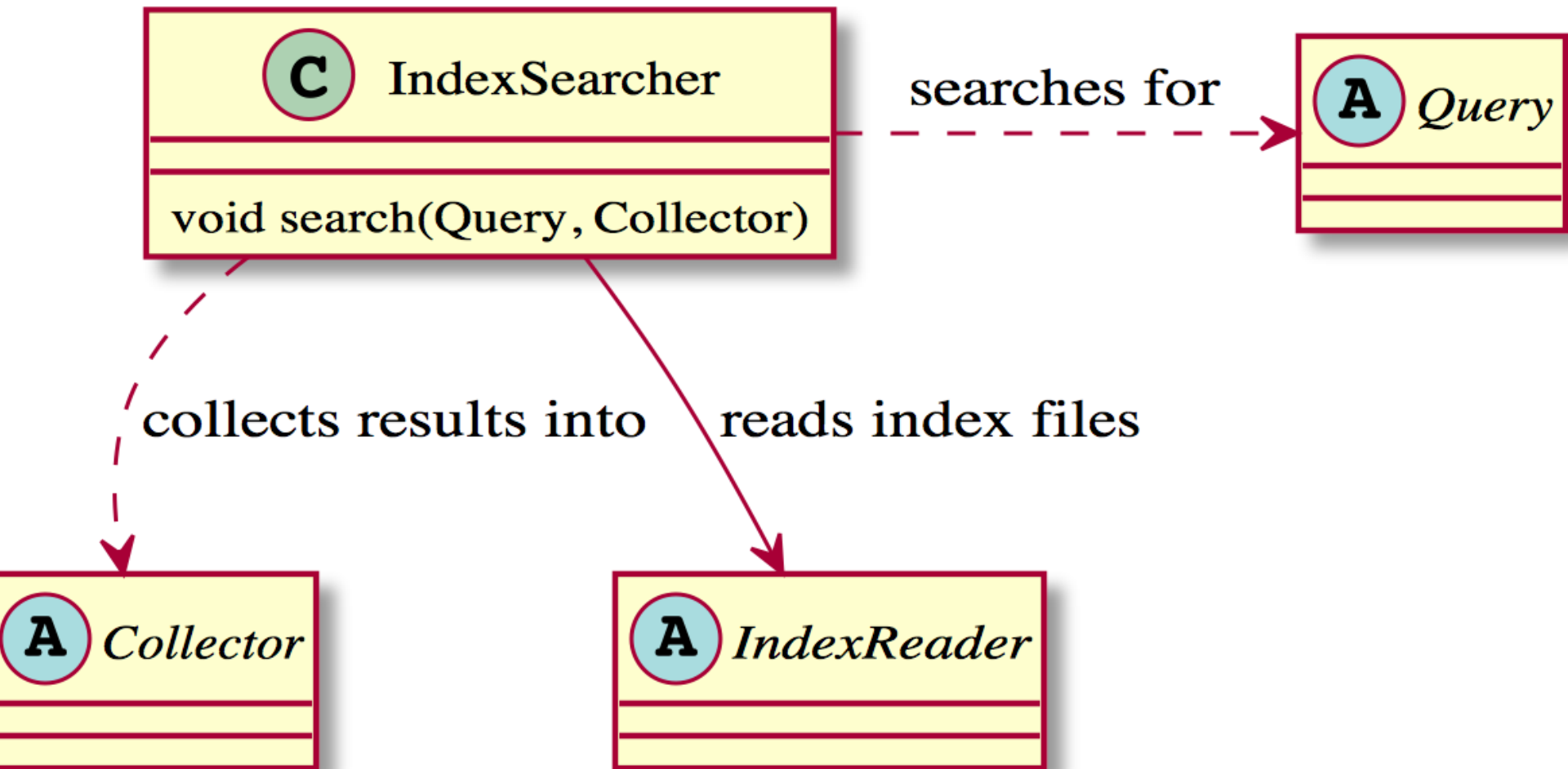
float score()





```
while (
    (doc = nextDoc()) != NO_MORE_DOCS) {

    println("found "+ doc +
        " with score "+score());
}
```



lucene collector



Web

Images

Maps

Shopping

More ▾

Search tools

About 105,000 results (0.19 seconds)

Collector (Lucene 3.0.3 API) - Apache Lucene

lucene.apache.org/core/3_0_3/api/core/org/apache/lucene/collector/Collector.html

org.apache.lucene.search. Class **Collector**. java.lang.Object extended by org.

apache.lucene.search.**Collector**. Direct Known Subclasses: ...

TopScoreDocCollector (Lucene 3.5.0 API) - Apache Lucene

lucene.apache.org/core/3_5_0/.../lucene/.../TopScoreDocCollector.html

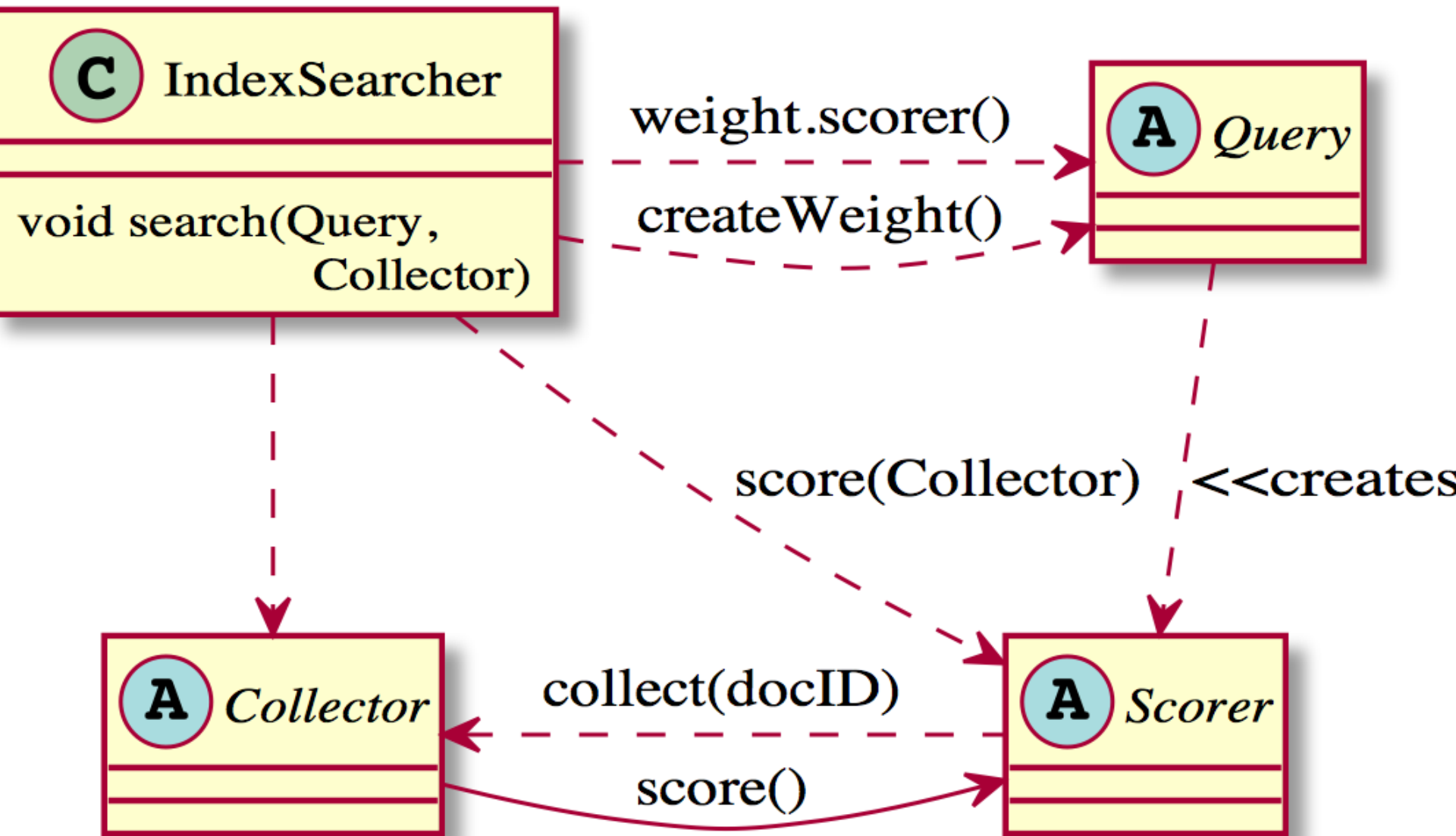
Method Summary. static TopScoreDocCollector · create(int numHits, boolean

docsScoredInOrder) Creates a new TopScoreDocCollector given the number of ...



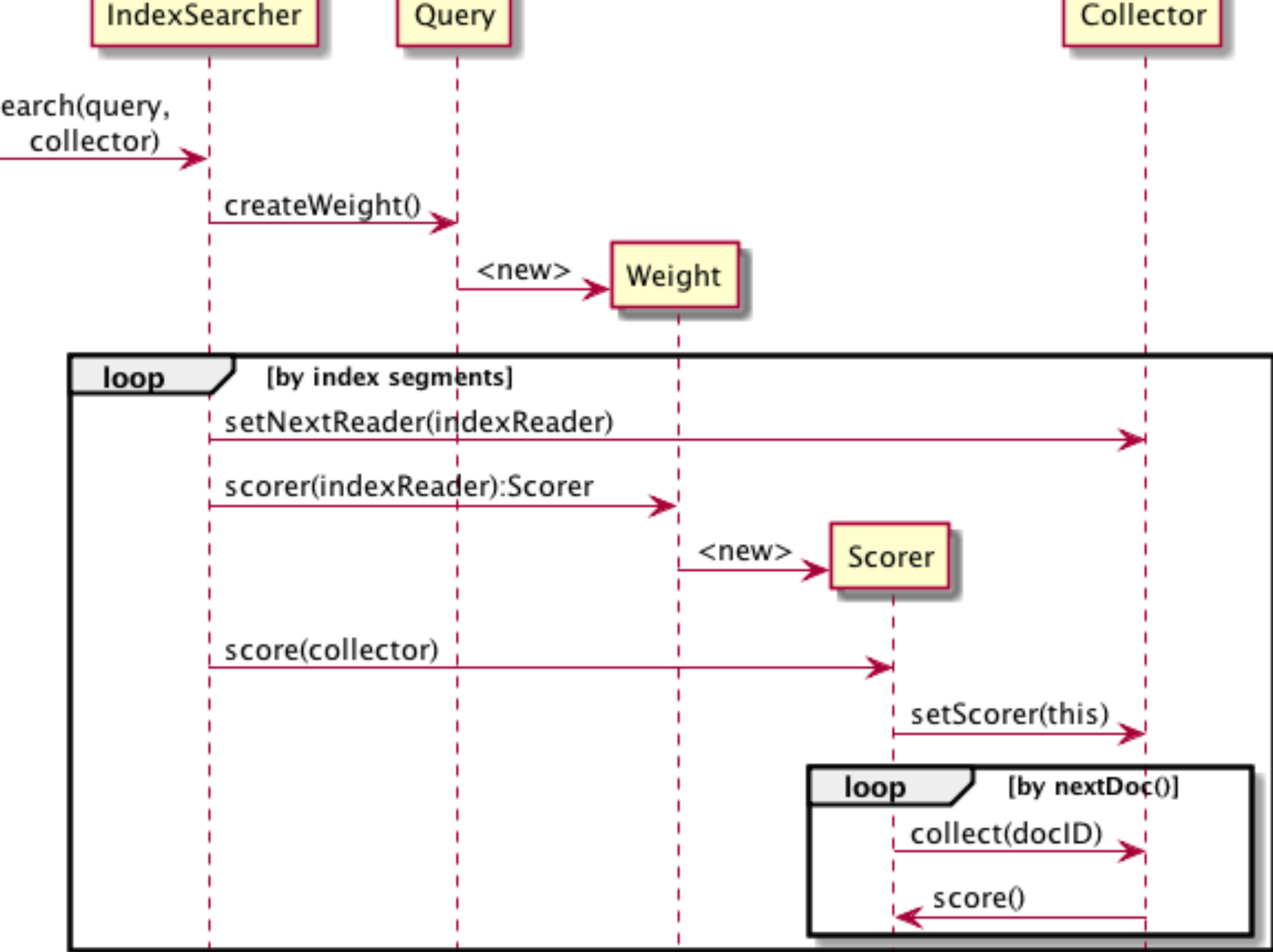
1 2 3 4 5 6 7 8 9 10

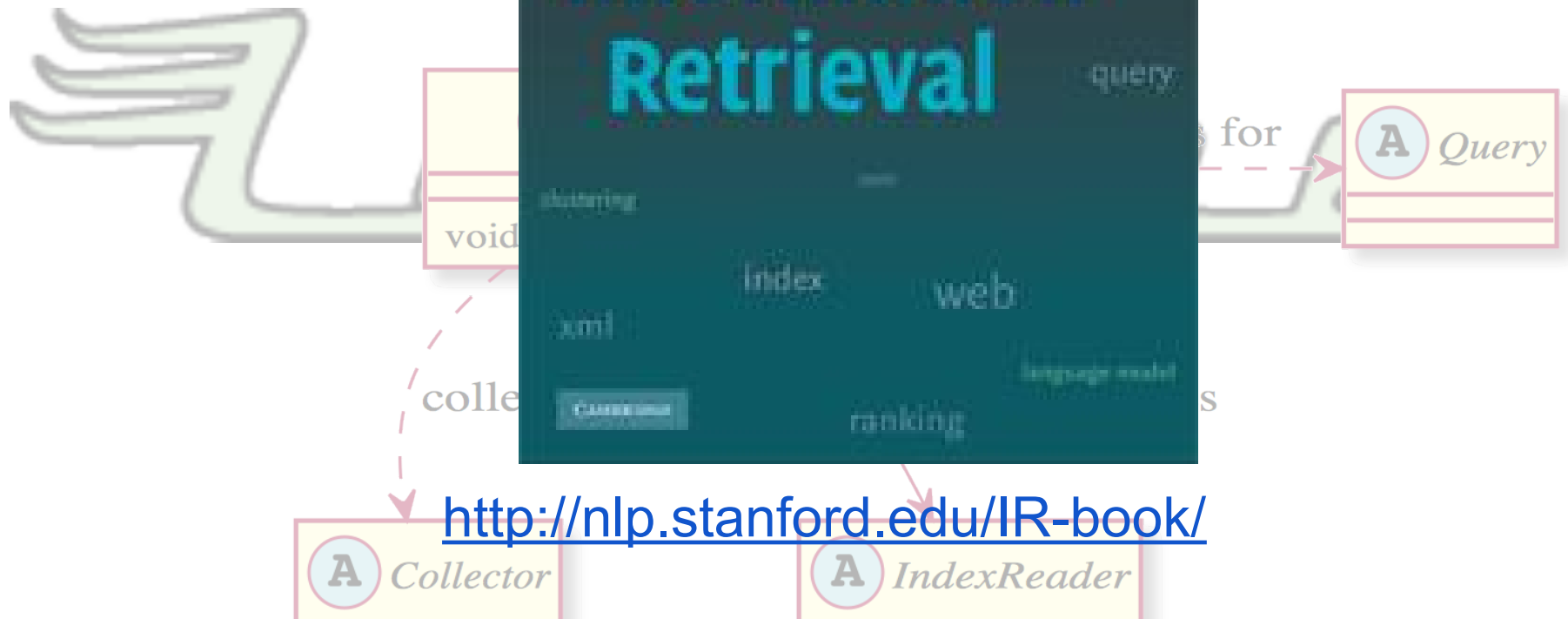
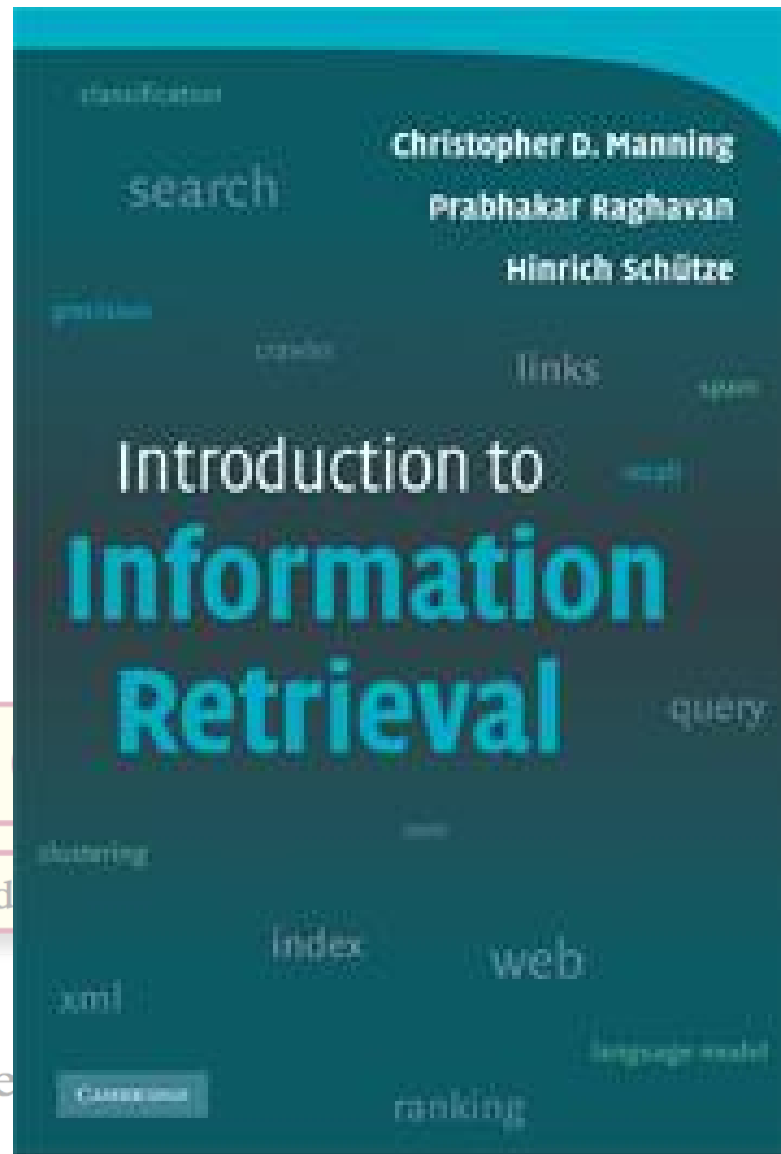
Next



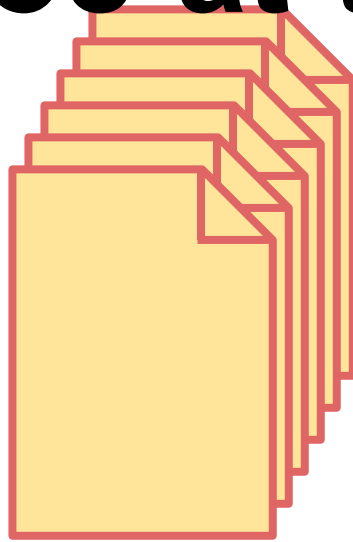
Note: Weight is omitted for sake of compactness







Doc-at-time search



"a": {2}

"banana": {2}

"is": {0, 1, 2}

"it": {0, 1, 2}

"what": {0, 1}



what OR is OR a OR banana

Search

"a" : {2}

"banana" : {2}

"is" : {0, 1, 2}

"it" : {0, 1, 2}

"what" : {0, 1}



what OR is OR a OR banana

Search

"is": {0, 1, 2}

"what": {0, 1}

"a": {2}

"banana": {2}

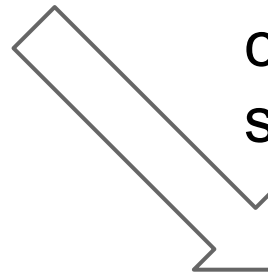
"it": {0, 1, 2}

"is": {0} 1, 2 }

"what": {0} 1 }

"a": {2}

"banana": {2}



collect(0)
score():2

Collector

"is": {0, 1, 2}

"what": {0, 1}

"a": {2}

"banana": {2}

docID×score
0×2

"is": {0, **1**} 2 }

"what": {0, **1**}

"a": {**2**}

"banana": {**2**}

collect(1)
score():**2**

Collector
0x2

"is": {0, 1, 2}

"what": {0, 1}

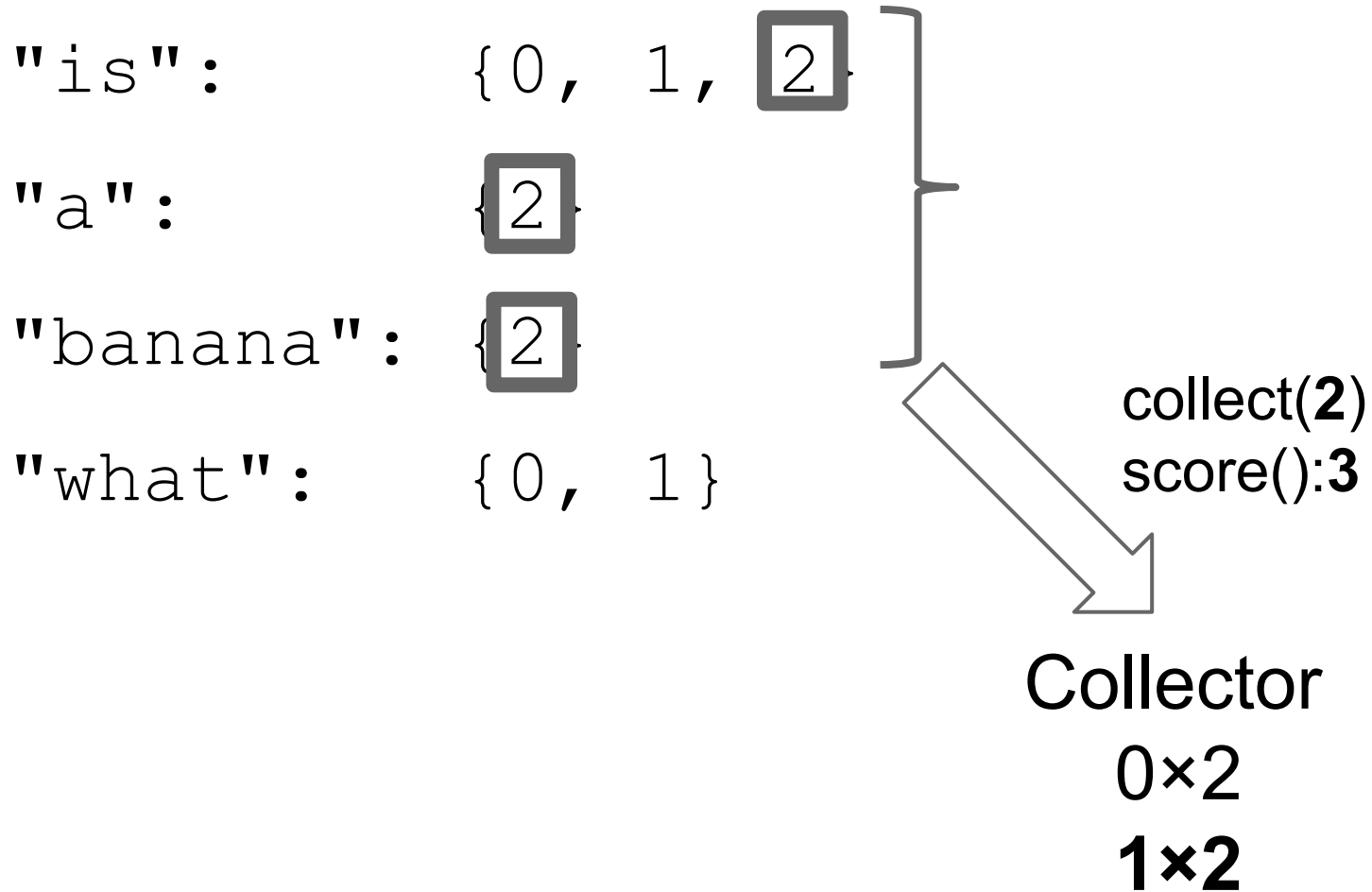
"a": 2}

"banana": 2}

Collector

0×2

1×2



"is": {0, 1, 2}

"a": {2}

"banana": {2}

"what": {0, 1}

Collector

2×3

0×2

1×2

Term-at-time search

"lorem"

"ipsum"

"dolor"

"sit"

"amet"

"consectetur"

"a": {2}

"banana": {2}

"is": {0, 1, 2}

"it": {0, 1, 2}

"what": {0, 1}



what OR is OR a OR banana

Search

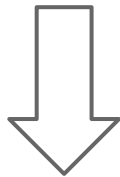
"a": {2}

"banana": {2}

"is": {0, 1, 2}

"it": {0, 1, 2}

"what": **{0, 1}**



Accumulator

... 0×1 ... 1×1 ...

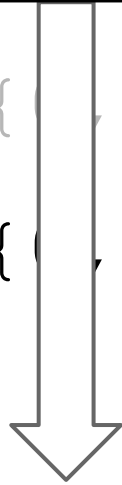
"a": {2}

"banana": {2}

"is": **{0, 1, 2}**

"it": {0, 1, 2}

"what": {0, 1}



Accumulator

... 0×2 ... 1×2 ... **2×1** ...

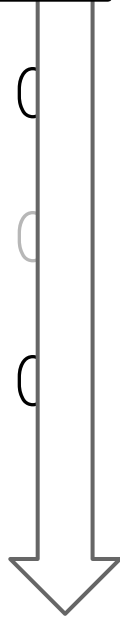
"a": {2}

"banana": **{2}**

"is": {0 1, 2}

"it": {0 1, 2}

"what": {0 1}



Accumulator

... 0×2 ... 1×2 ... **2×2** ...

"a":

{2}

"banana": {1}

"is": {0, 1, 2}

"it": {0, 1, 2}

"what": {0, 1}



Accumulator

... 0x2 ... 1x2 ... **2x3** ...

"a": {2}

"banana": {2}

"is": {0, 1, 2}

"it": {0, 1, 2}

"what": {0, 1}

Accumulator

... **0×2** ... **1×2** ... **2×3** ...



Collector

2×3

0×2

1×2

$O(n)$

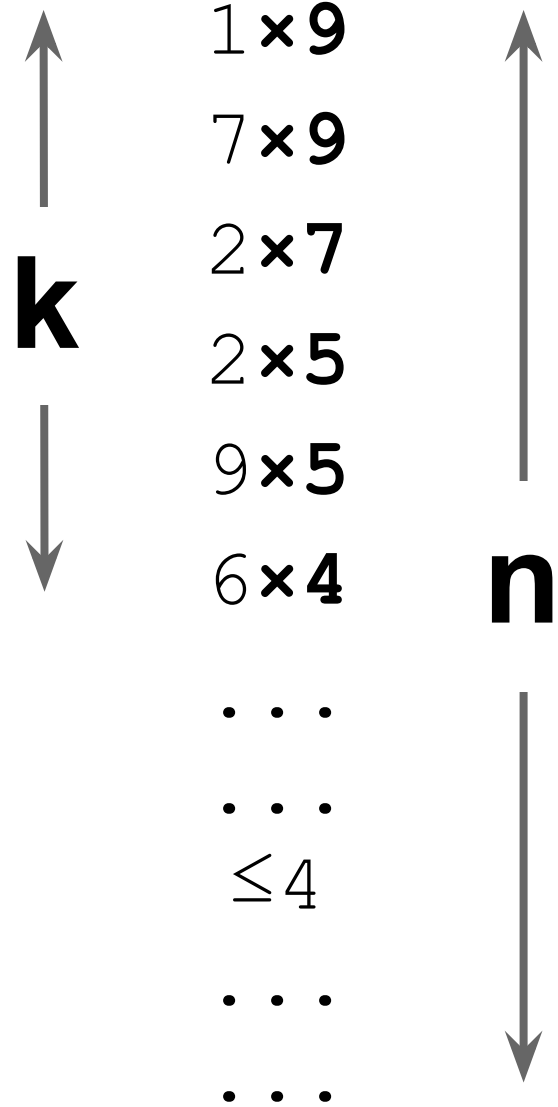
"lorem"
"ipsum"
"dolor"
"sit"
"amet"
"consectetur"

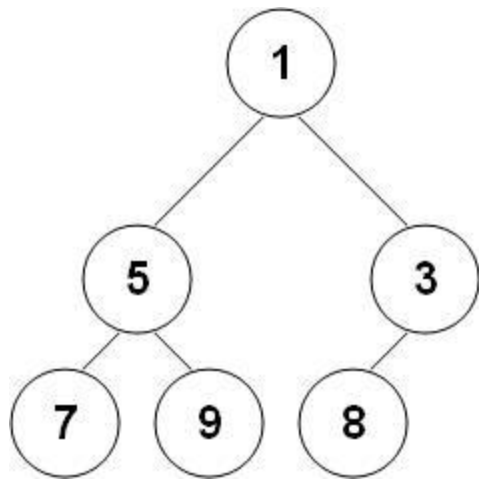
<http://nlp.stanford.edu/IR-book/>

A Collector

A IndexReader

A Query

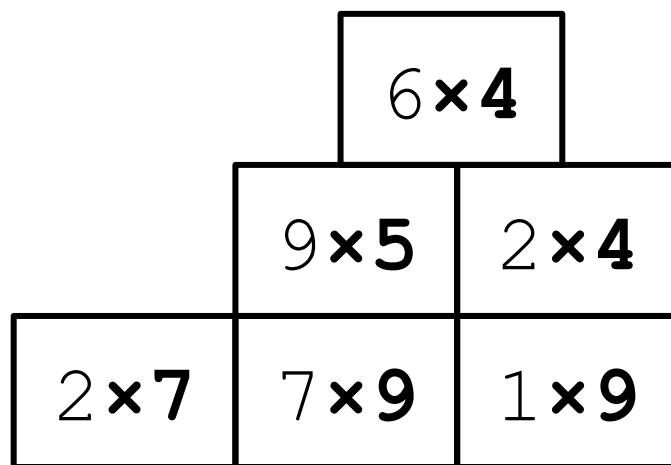




Node	1	5	3	7	9	8
Index	0	1	2	3	4	5

http://en.wikipedia.org/wiki/Binary_heap

$\uparrow$
log k
 $\downarrow$



$\uparrow$
n

...

...

≤ 4

...

...

$\downarrow$

"a":	{2}
"banana":	{2}
"is":	{0, 1, 2}
"it":	{0, 1, 2}
"what":	{0, 1}

p

what OR is OR a OR banana

q

doc at time

term at time

complexity

memory

doc at time

term at time

complexity

$O(p + n \log k)$

memory



q



"a":

{2}

"banana":

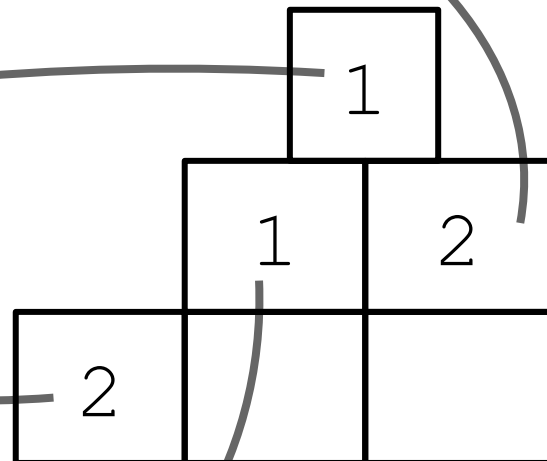
{2}

"is":

{0, 1, 2}

"what":

{0, 1}



doc at time

term at time

complexity

$O(p \log q + n \log k)$

$O(p + n \log k)$

memory

doc at time

term at time

complexity

$O(p \log q + n \log k)$

$O(p + n \log k)$

memory

$q + k$

doc at time

term at time

complexity

$O(p \log q + n \log k)$

$O(p + n \log k)$

memory

$q + k$

n

Space Optimizations for Total Ranking *

Douglass R. Cutting

Excite Inc.

555 Broadway, Redwood City, CA 94063

Jan O. Pedersen

Verity Inc.

894 Ross Dr., Sunnyvale, CA 94089

Abstract

Efficient ranking algorithms for similarity search use an inverted index to avoid scoring documents that have no overlap with the query. Nonetheless, partial scores must be maintained for a significant proportion of the collection. Previous work has focussed on heuristic partial ranking strategies to reduce the memory and time requirements at the cost of no longer computing the true ranks. We present two novel algorithms that efficiently compute the true total ranking with a fixed space requirement independent of the size of the collection.

q=village **OR** operations **OR** years **OR**
disaster **OR** visit

q=village OR operations OR years OR
disaster OR visit OR etc OR map OR
seventieth OR penneplains OR tussock OR sir
OR memory OR character OR campaign OR
author OR public OR wonder OR forker OR
middy OR vocalize OR enable OR race OR
object OR signal OR symptom OR deputy OR
where OR typhous OR rectifiable OR
polygamous OR originally OR look OR
generation OR ultimately OR reasonably OR
ratio OR numb OR apposing OR enroll OR
manhood OR problem OR suddenly OR
definitely OR corp OR event OR material

q=village **AND** operations **AND** years **AND**
disaster **AND** visit

Conjunction

(+, MUST, AND)

"a": {2} 3}

"banana": {2, 3}

"is": {0} 1, 2, 3}

"it": {0} 1, 3}

"what": {0} 1, 3}

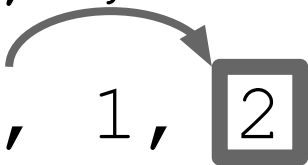
 what AND is AND a AND it

Search

"a": {2, 3}

"banana": {2, 3}

"is": {0, 1, 2, 3}



"it": {0, 1, 3}

"what": {0, 1, 3}

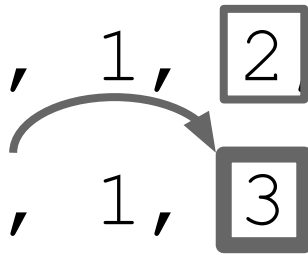
"a": {2, 3}

"banana": {2, 3}

"is": {0, 1, 2, 3}

"it": {0, 1, 3}

"what": {0, 1, 3}



"a": {2, 



"banana": {2, 3}

"is": {0, 1, 2, 



"it": {0, 1, 

"what":  1, 3}

"a": {2, 3}

"banana": {2, 3}

"is": {0, 1, 2, 3}

"it": {0, 1, 3}

"what": {0, 1, 3}



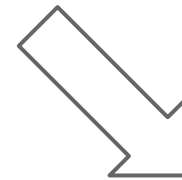
"a": {2, **3**}

"banana": {2, 3}

"is": {0, 1, 2, **3**}

"it": {0, 1, **3**}

"what": {0, 1, **3**}



Collector
3 x 4



$$\Omega(nq + n \log k)$$

Wrap-up

- doc-at-time vs term-at-time
- conjunction & leapfrog



complexity **$O(n)$**

memory **$O(\text{const})$**

finally

One War Story

Search results for
leather handbag

Narrow by

 [Clear All Selections](#)

Special Offers

Sales & Discounts (71)

Gift with Purchase (7)

Clearance/Closeout (42)

Handbag Style  [Show All](#)
Clutch

Brand

Michael Kors (12)

Style&co. (12)

Fossil (10)

Material Girl (8)

INC International Concepts (7)

Sasha Handbags (7)

GUESS? (6)

Kenneth Cole Reaction (6)

The Saks (6)

We found **123** results for *"leather handbag"*

Your Selections Clutch 

Sort by: Featured Items

Show: 40 items



Sold Out
Sold Out
Sold Out
MICHAEL Michael Kors
Handbag, Large Zip Clutch



**NEW! COACH LEGACY SLIM
CLUTCH IN LEATHER**
Will Be \$198.00
Sale \$148.50



**NEW! Frye H
Cameron Stu**
\$498.00

Search results for
leather handbag

Narrow by

☒ [Clear All Selections](#)

Special Offers

Sales & Discounts (71)

Gift with Purchase (7)

Clearance/Closeout (42)

Handbag Style ☒ [Show All](#)

Clutch

Brand

Michael Kors (12)

Style&co. (12)

Fossil (10)

Material Girl (8)

INC International Concepts (7)

Sasha Handbags (7)

GUESS? (6)

Kenneth Cole Reaction (6)

The Sale (6)

We found **123** results for **"leather handbag"**

Your Selections Clutch ☒

Sort by:

Show:



Sold Out
Sold Out
Sold Out
MICHAEL Michael Kors
Handbag, Large Zip Clutch



**NEW! COACH LEGACY SLIM
CLUTCH IN LEATHER**
Will Be \$198.00
Sale \$148.50



**NEW! Frye H
Cameron Stu**
\$498.00

<http://localhost:8983/solr/collection1/select?q=>

(operations OR years OR disaster OR visit OR etc OR
map OR seventieth OR peneplains OR tussock OR sir
OR memory OR character OR campaign OR author OR
public OR wonder OR forker OR middy OR vocalize OR
enable OR race OR object OR signal OR symptom OR
deputy OR where OR typhous OR rectifiable OR
polygamous OR originally OR look OR generation OR
ultimately OR reasonably OR ratio OR numb OR
apposing OR enroll OR manhood OR problem) ...

<http://localhost:8983/solr/collection1/select?q=>

(operations OR years OR disaster OR visit OR etc OR
map OR seventieth OR peneplains OR tussock OR sir
OR memory OR character OR campaign OR author OR
public OR wonder OR forker OR middy OR vocalize OR
enable OR race OR object OR signal OR symptom OR
deputy OR where OR typhous OR rectifiable OR
polygamous OR originally OR look OR generation OR
ultimately OR reasonably OR ratio OR numb OR
apposing OR enroll OR manhood OR problem) ...

AND (id:yes_49912894 OR id:nurse_30134968)

<http://localhost:8983/solr/collection1/select?q=>

(operations OR years OR disaster OR visit OR etc OR
map OR seventieth OR peneplains OR tussock OR sir
OR memory OR character OR campaign OR author OR
public OR wonder OR forker OR middy OR vocalize OR
enable OR race OR object OR signal OR symptom OR
deputy OR where OR typhous OR rectifiable OR
polygamous OR originally OR look OR generation OR
ultimately OR reasonably OR ratio OR numb OR
apposing OR enroll OR manhood OR problem) ...

AND (id:yes_49912894 OR id:nurse_30134968)

&mm=32&...

minShouldMatch=2

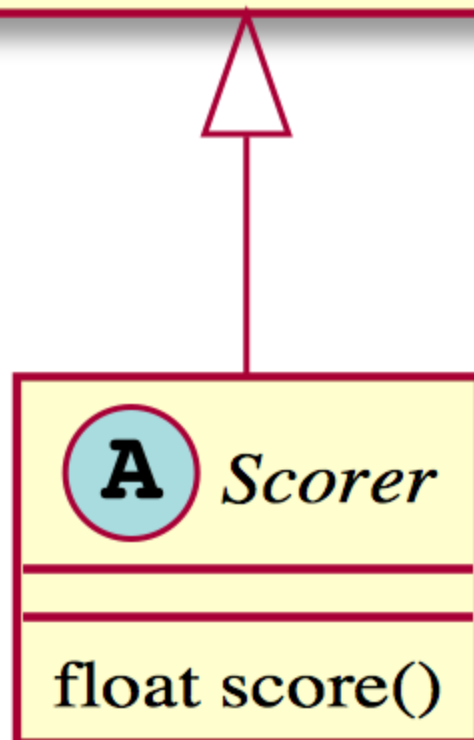
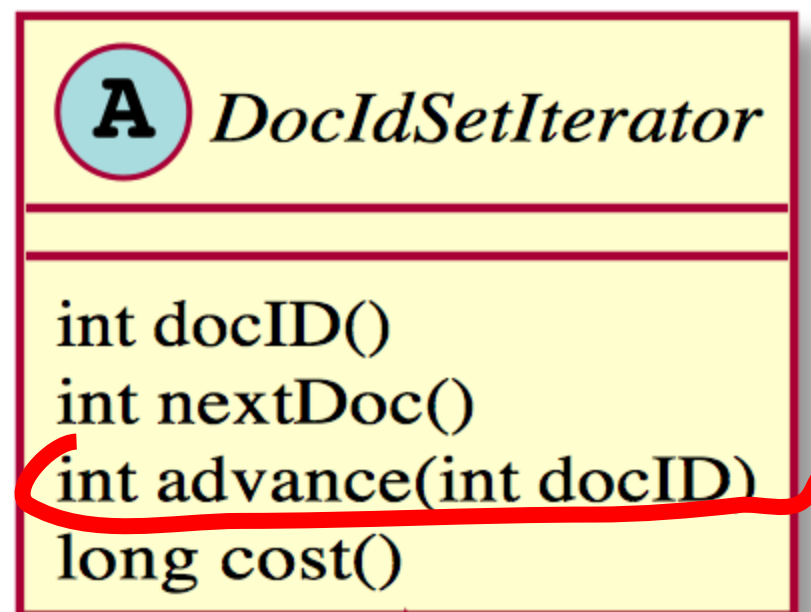
- ✓ straight jeans
- ✓ silver jeans
- ✓ silver jeans straight

✗ ~~jeans~~

✗ ~~silver~~

org.apache.lucene.search.DisjunctionSumScorer

```
int nextDoc() {  
    while(true) {  
        while (subScorers[0].docID() == doc) {  
            if (subScorers[0].nextDoc() != NO_DOCS) {  
                heapAdjust(0);  
            } else {  
                ....  
            }  
        }  
        ...  
        if (nrMatchers >= minimumNrMatchers) {  
            break;  
        }  
    }  
    return doc;  
}
```





mm=3

{1, **3** 7, 10, 27, 30, ...}

{**3** 5, 10, 27, 32, ...}

{ **20**, 27, 31, ...}

{ **30**, 37, ...}



mm=3

{1, 3, **7**, 10, 27, 30, ...}

{3, **5**, 10, 27, 32, ...}

{**20**, 27, 31, ...}

{**30**, 37, ...}



mm=3

{1, 3, **7**, 10, 27, 30, ...}

{3, 5, **10**, 27, 32, ...}

{**20**, 27, 31, ...}

{**30**, 37, ...}



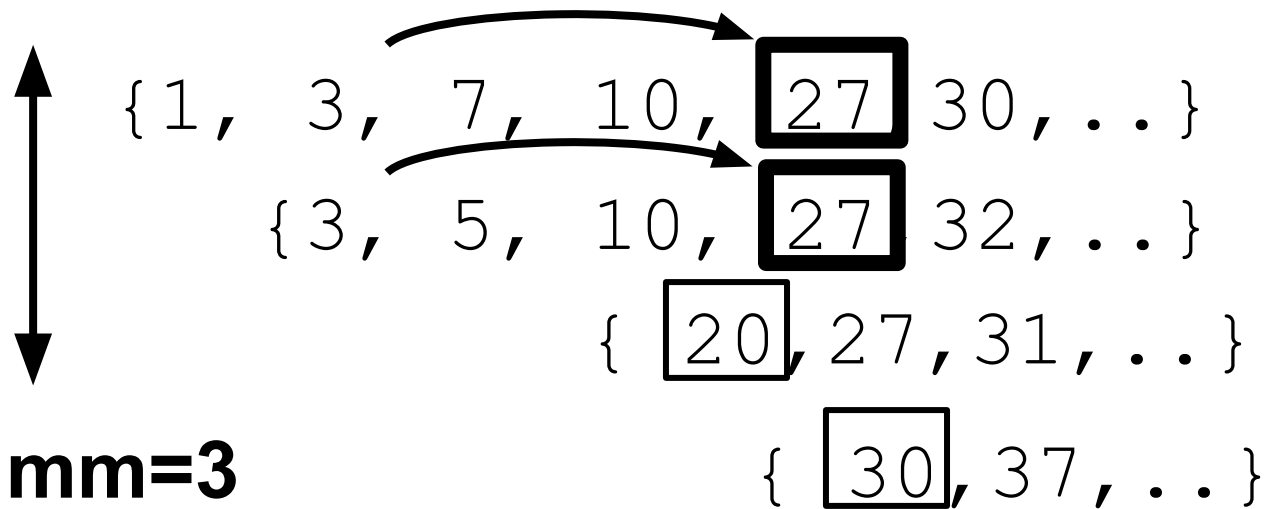
mm=3

{1, 3, 7, **10**, 27, 30, ...}

{3, 5, **10**, 27, 32, ...}

{**20**, 27, 31, ...}

{**30**, 37, ...}





The **Apache Software Foundation**

<http://www.apache.org/>

Dashboards ▾

Projects ▾

Issues ▾

Agile ▾



Lucene - Core / [LUCENE-4571](#)

speedup disjunction with minShouldMatch

Agile Board

More Actions ▾

▼ Details

Type:  Improvement

Priority:  Major

Affects Version/s: 4.1

Component/s: [core/search](#)

Labels: None

Lucene Fields: New

Status:  Resolved

Resolution: Fixed

Fix Version/s: [5.0](#), [4.3](#)

Task	QPS trunk	QPS patch	
PKLookup	251.82	239.58	-4.9%
Low4MinShouldMatch0	8.34	8.62	3.3%
Low1MinShouldMatch0	1.90	1.99	4.8%
Low2MinShouldMatch0	2.41	2.53	4.8%
Low3MinShouldMatch0	3.94	4.13	5.0%
HighMinShouldMatch0	1.61	1.70	5.1%
HighMinShouldMatch2	1.68	1.94	15.6%
Low1MinShouldMatch2	1.99	2.39	19.8%
Low2MinShouldMatch2	2.57	3.27	27.2%
HighMinShouldMatch3	1.71	2.48	45.0%
Low3MinShouldMatch2	4.24	6.21	46.3%
Low1MinShouldMatch3	2.05	3.25	58.9%
HighMinShouldMatch4	1.74	3.52	102.2%
Low2MinShouldMatch3	2.67	5.40	102.3%
Low1MinShouldMatch4	2.08	5.74	175.3%
Low4MinShouldMatch2	10.18	48.51	376.5%
Low3MinShouldMatch3	4.50	40.66	804.3%
Low4MinShouldMatch3	10.22	151.37	1380.9%
Low2MinShouldMatch4	2.71	45.36	1574.3%
Low4MinShouldMatch4	10.22	222.35	2075.6%
Low3MinShouldMatch4	4.50	206.42	4486.1%

Task	QPS trunk	QPS patch	
PKLookup	251.82	239.58	-4.9%
Low4MinShouldMatch0	8.34	8.62	3.3%
Low1MinShouldMatch0	1.90	1.99	4.8%
Low2MinShouldMatch0	2.41	2.53	4.8%
Low3MinShouldMatch0	3.94	4.13	5.0%
HighMinShouldMatch0	1.61	1.70	5.1%
HighMinShouldMatch2	1.68	1.94	15.6%


[Simon Willnauer](#) added a comment - 22/Mar/13 14:13

HOLY SHIT!


[Adrien Grand](#) added a comment - 22/Mar/13 14:16

Agreed, these speedups are awesome!

Low4MinShouldMatch2	10.18	48.51	376.5%
Low3MinShouldMatch3	4.50	40.66	804.3%
Low4MinShouldMatch3	10.22	151.37	1380.9%
Low2MinShouldMatch4	2.71	45.36	1574.3%
Low4MinShouldMatch4	10.22	222.35	2075.6%
Low3MinShouldMatch4	4.50	206.42	4486.1%

Lucene 4.3 Highlights

- minShouldMatch BooleanQuery major performance improvement



Lucene 4.3 Highlights

- minShouldMatch BooleanQuery major performance improvement
- SortingAtomicReader and SortingMergePolicy
- DocIdSetIterator and Scorer now has a cost API
- Analyzing/FuzzySuggester now enable recording an arbitrary byte[] as a payload
- Spatial module: support for query relations Within, Contains, and Disjoint
- Facet module: new method computes facet counts using SortedSetDocValuesField, without a separate taxonomy index.





Grid Dynamics

Scalable eCommerce Platform Solutions



Mikhail Khludnev

Principal Engineer

mkhludnev@griddynamics.com

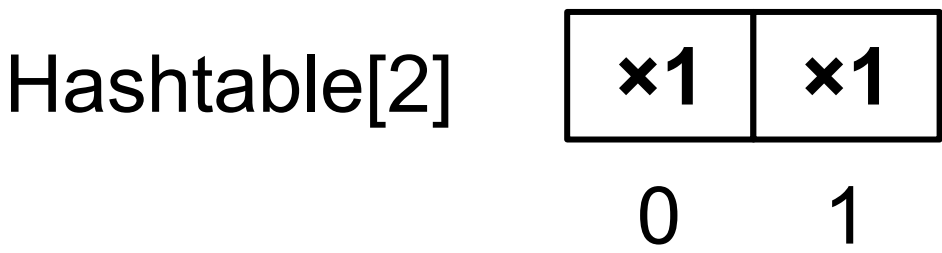
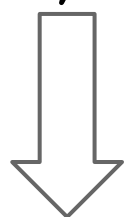
<http://goo.gl/7q8nHm>

BooleanScorer

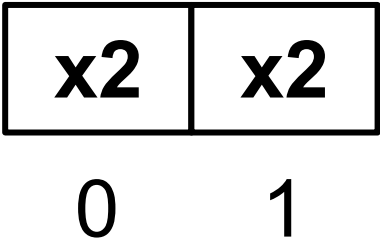
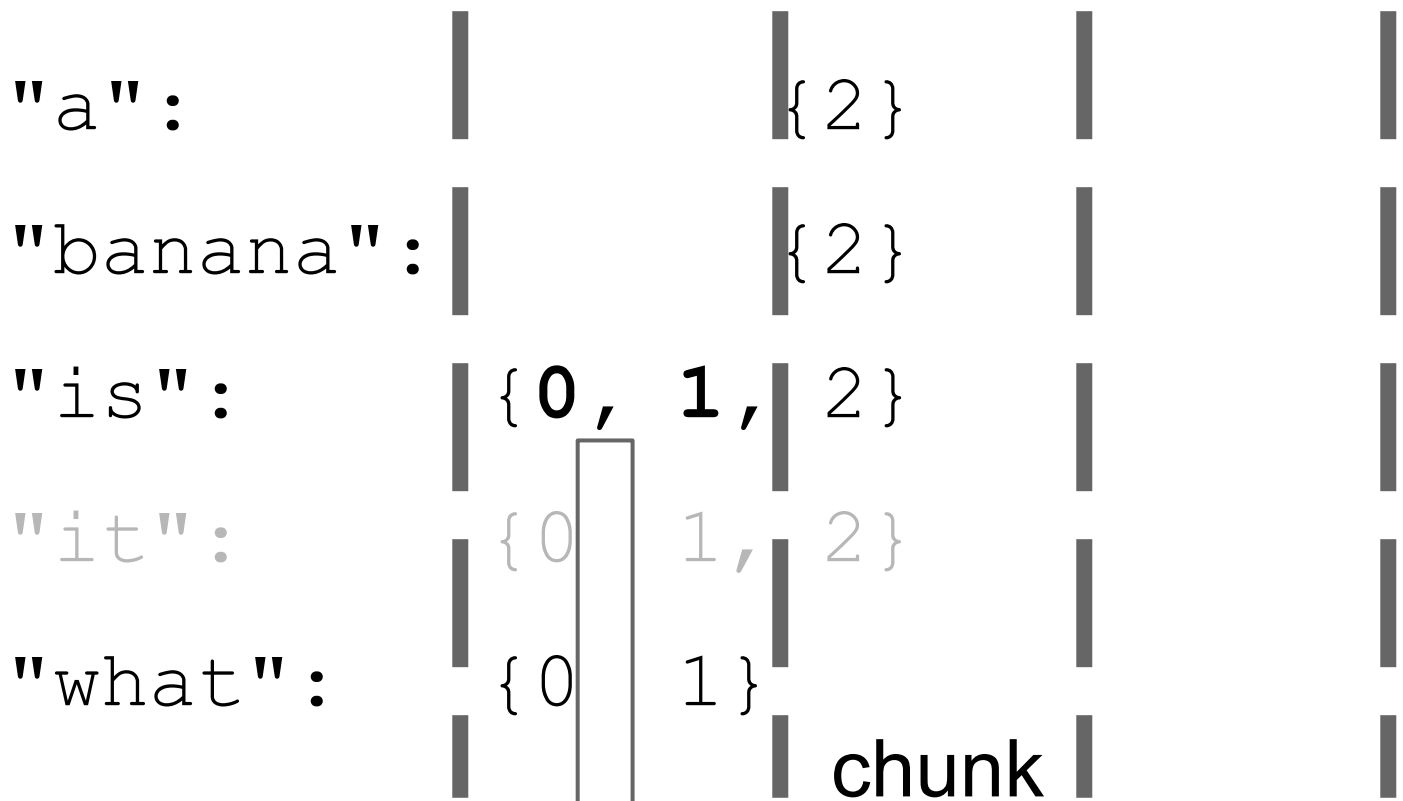
term at time

org.apache.lucene.search.BooleanScorer

"a":				
		{ 2 }		
"banana":				
		{ 2 }		
"is":		{ 0, 1, 2 }		
"it":		{ 0, 1, 2 }		
"what":		{ 0, 1 }		

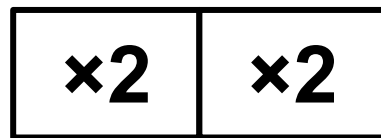


org.apache.lucene.search.BooleanScorer

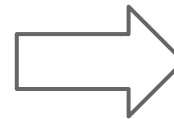


org.apache.lucene.search

"a":				
		{ 2 }		
"banana":				
		{ 2 }		
"is":		{ 0, 1, 2 }		
"it":		{ 0, 1, 2 }		
"what":		{ 0, 1 }		



0 1



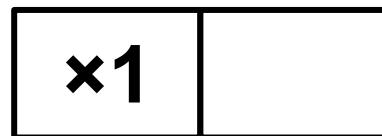
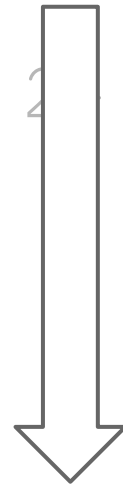
Collector

0x2

1x2

org.apache.lucene.search

"a":			{ 2 }		
"banana":			{ 2 }		
"is":		{ 0, 1,	2 }		
"it":		{ 0, 1,	2 }		
"what":		{ 0, 1 }	}		



0

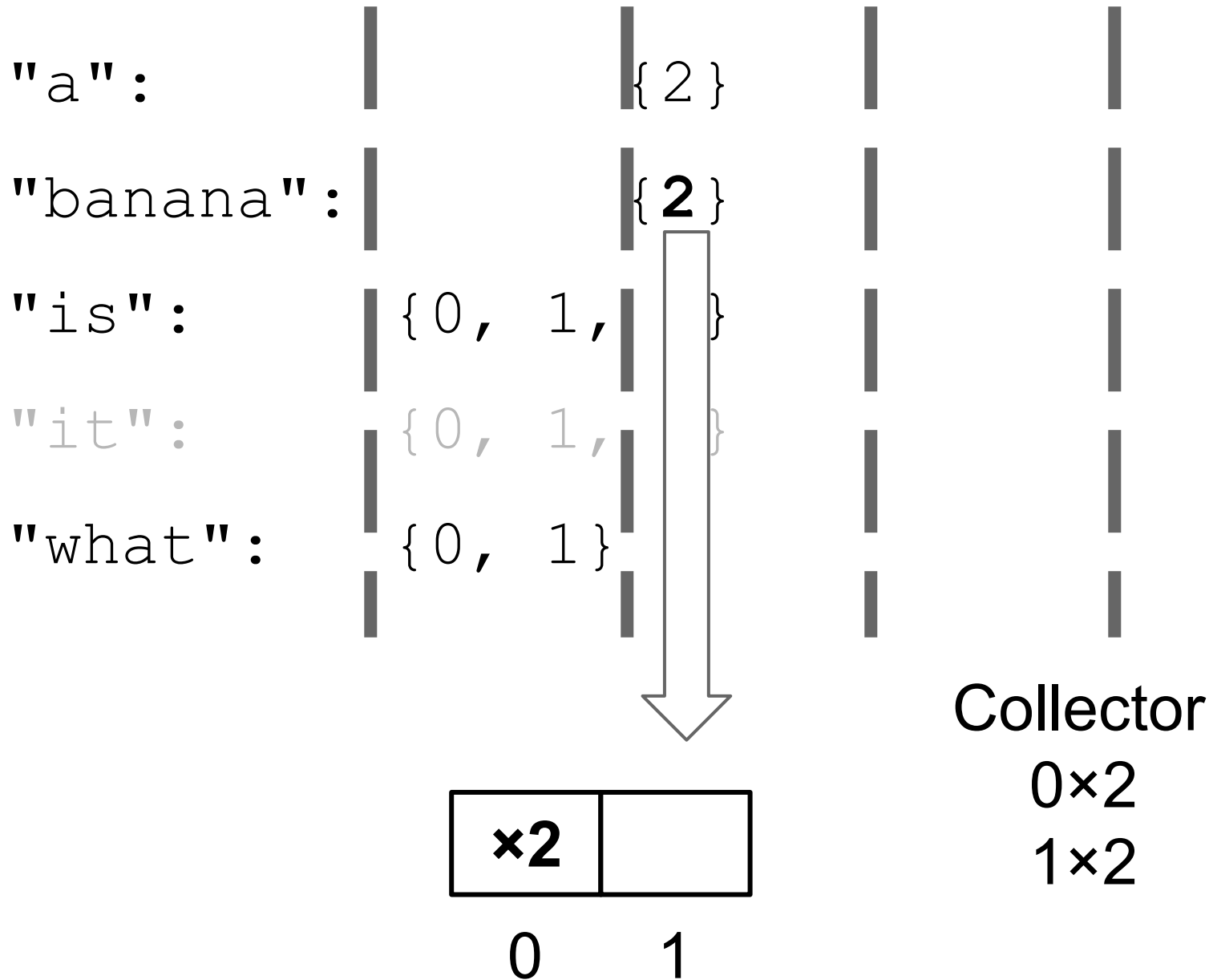
1

Collector

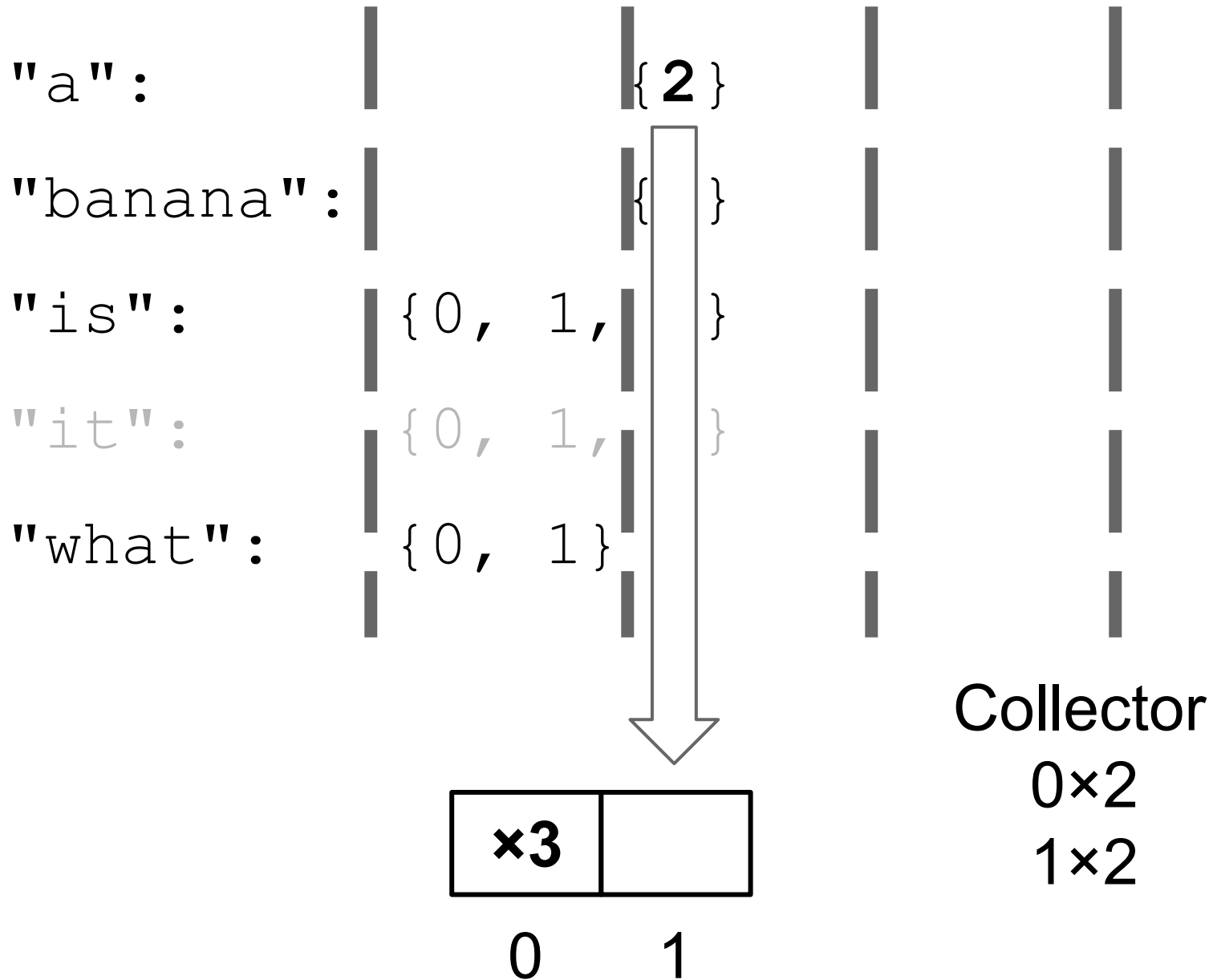
0x2

1x2

org.apache.lucene.search

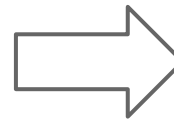
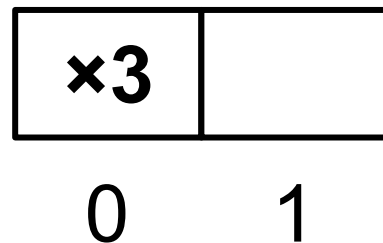


org.apache.lucene.search



org.apache.lucene.search

"a":				
		{ 2 }		
"banana":				
		{ 2 }		
"is":		{ 0, 1, 2 }		
"it":		{ 0, 1, 2 }		
"what":		{ 0, 1 }		



Collector

2x3

0x2

1x2

Linked Open Hash [2K]

