

Compound Words Query Parser

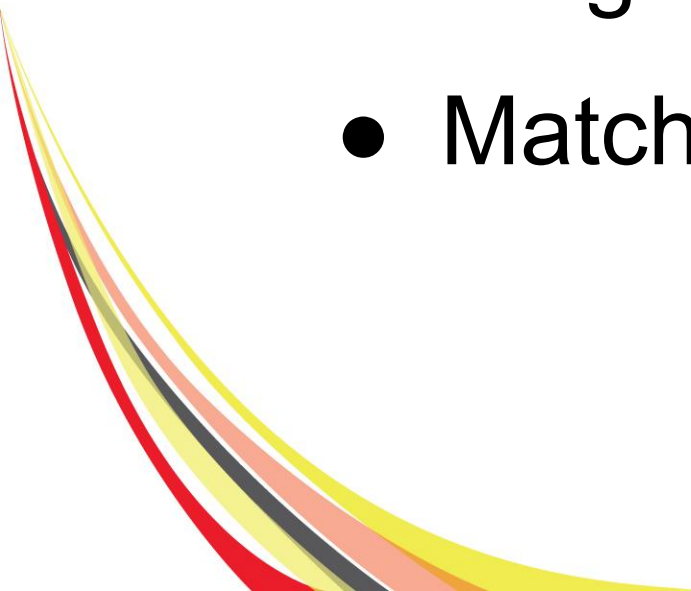
Mikhail Khlodnev



Grid Dynamics

Scalable eCommerce Platform Solutions

Agenda

- eCommerce Search is Special
 - Compound Words Query Parser
 - Staged Search
 - Match Spotting
- 

Part I

eCommerce Search is Special

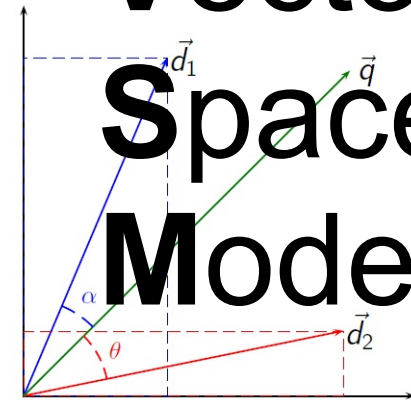


Lucene

is...

Boolean
Retrieval
(+ -)

Vector
Space
Model



Does she search this way?



+"michael kors" type:handbag

Search

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

Plain Text Documents

Else Jeans Skinny Jeans,
Colored Denim Dark Green-
Wash

In a dark green wash perfect
for fall, these Else Jeans skin
jeans hit the colored-denim
trend right on the mark! Lycoc
cotton rayon polyester Lycra
Machine washable Imported
Low rise: approx. 8 inches
Skinny fit Skinny leg Zipper f
with button closure 5-pocket
style Dark green wash, color
denim Waistband with belt
loops Inseam: approx. 30-1/2
inches

Lorem ipsum dolor sit amet,
consectetur adipisicing elit, sed
do eiusmod tempor incididunt
ut labore et dolore magna
aliqua. Ut enim ad minim
veniam, quis nostrud
exercitation ullamco laboris nisi
ut aliquip ex ea commodo
consequat. Duis aute irure
dolor in reprehenderit in
voluptate velit esse cillum
dolore eu fugiat nulla pariatur.
Excepteur sint occaecat
cupidatat non proident, sunt in
culpa qui officia deserunt mollit
anim id est laborum.

Plain Text Documents



approx velit

Search

Else Jeans Skinny Jeans,
Colored Denim Dark Green-
Wash

In a dark green wash perfect
for fall, these Else Jeans skin
jeans hit the colored-denim
trend right on the mark! Lycoc
cotton rayon polyester Lycra
Machine washable Imported
Low rise: approx. 8 inches
Skinny fit Skinny leg Zipper f
with button closure 5-pocket
style Dark green wash, color
denim Waistband with belt
loops Inseam: approx. 30-1/2
inches

Lorem ipsum dolor sit amet,
consectetur adipiscing elit, sed
do eiusmod tempor incididunt
ut labore et dolore magna
aliqua. Ut enim ad minim
veniam, quis nostrud
exercitation ullamco laboris nisi
ut aliquip ex ea commodo
consequat. Duis aute irure
dolor in reprehenderit in
voluptate velit esse cillum
dolore eu fugiat nulla pariatur.
Excepteur sint occaecat
cupidatat non proident, sunt in
culpa qui officia deserunt mollit
anim id est laborum.

Product Catalog Document

| | |
|------------------|------------------|
| BRAND: | Calvin Klein |
| GENDER: | Women's |
| TYPE: | Jeans |
| FIT: | At Waist |
| OCCASION: | Casual |
| LEG: | Classic Straight |
| WEIGHT: | Super Skinny |
| COLOR: | Light Wash |

Polysemy Problem

 clothing & accessories ▾

Polysemy Problem

 clothing & accessories 

COLOR:"**Pink**" TYPE:"Sweater"

BRAND: "Tomas **Pink**" TYPE:"Sweater"

BRAND: "**Pink** Lotus" TYPE:"Sweater"



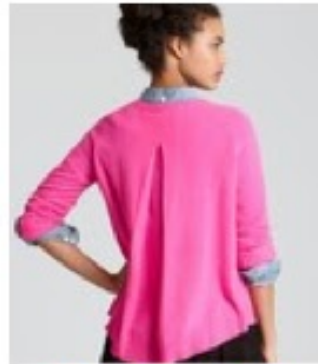
Free Ship \$50+

Aqua Cashmere Sweater
\$168 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater
\$148 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater
\$198 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater
\$178 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater
\$188 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater
\$148 - Bloomingdale's



Free Ship on \$99+

PINK TARTAN Color Block
\$325 - Lord & Taylor



Pink Tartan Print Pocket
~~\$255~~ \$89



Pink Tartan Contrast-Trim
~~\$225~~ \$119



Free Ship, Free Return
Pink Lotus 'Needle Out'
\$78 - Nordstrom



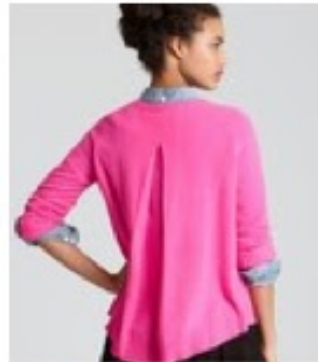
Free Ship \$50+

Aqua Cashmere Sweater \$168 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater \$148 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater \$198 - Bloomingdale's



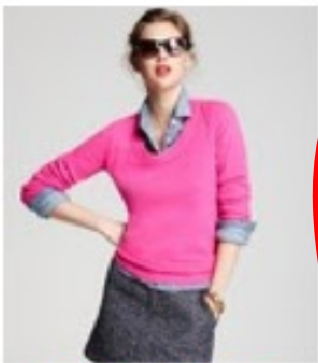
Free Ship \$50+

Aqua Cashmere Sweater \$178 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater \$188 - Bloomingdale's



Free Ship \$50+

Aqua Cashmere Sweater \$148 - Bloomingdale's



Free Ship on \$99+

PINK TARTAN Color Block Sweater \$325 - Lord & Taylor



Pink Tartan Print Pocket Sweater \$255 \$89



Pink Tartan Contrast-Trim Sweater \$225 \$119



Free Ship, Free Return Pink Lotus 'Needle Out' Sweater \$78 - Nordstrom



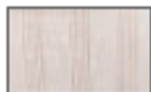
Pink Lotus 'Needle Out Infinity' Wrap Cardigan

Free Ship, Free Return · Ends Sun, Feb 24

Sheer drop-needle stripes texture an airy wrap designed with ultra-long front panels that can be wrapped and tied in a variety of ways to create a new look every day.

- Approx. length from shoulder: front 45"; back 21".
- Cotton; machine wash.
- By Pink Lotus; made in the USA.
- Women's Active & Swim.

Colors:



WHITE

Sizes:

Small, Medium, Large

[Buy at Nordstrom >](#)

[More Pink Lotus Sweaters](#) [Add to My Brands](#)

[Share](#) [Sale Alert](#) [Favorite](#) [Wish List](#)

[Pin it](#)

[Like](#)

QuickLook: [on](#) · [off](#)



Free Ship \$50+

Aqua Cashmere Sweater \$168 - Bloomingdale's



Aqua \$14



Free Ship \$50+

Aqua Cashmere Sweater \$148 - Bloomingdale's

PINK \$3



Free Ship \$50+

Aqua Cashmere Sweater \$188 - Bloomingdale's



Free Ship, Free Return

Pink Lotus 'Needle Out \$78 - Nordstrom



Mink **Pink Women's The Triple Swirl Waterfall Cardigan**

62 \$ - Karmaloop



Pink Tartan Jenna Sweater - Camel - Size X Large

73 \$ - Saks Fifth Avenue



Open Front Womens Salmon Pink Sweater S

39 \$ - Laimarket.com



Pink Cardigan Hollow Sweater

55 \$ - ThatsPoint.com



Liz Claiborne New York Cowl Neck Cable Collar **Sweater -**

46 \$ - QVC.com



Fashion Girls Wonderful Salmon **Pink Sweater XL**

28 \$ - Laimarket.com

pink sweater



Mink Pink Women's The Triple Swirl Waterfall Cardigan

62 \$ - Karmaloop



Pink Tartan Jenna Sweater - Camel - Size X Large

73 \$ - Saks Fifth Avenue



Open Front Womens Salmon Pink Sweater S

39 \$ - Laimarket.com



Pink Cardigan Hollow Sweater

55 \$ - ThatsPoint.com



Liz Claiborne New York Cowl Neck Cable Collar Sweater -

46 \$ - QVC.com



Fashion Girls Wonderful Salmon Pink Sweater XL

28 \$ - Laimarket.com

https://www.google.com/search?hl=ru&tbm=shop&q=pink+sweater&oq=pink+sweater&gs_l=products-cc.3..017j0i5l3.4658.7244.0.7699.12.11.0.1.1.0.298.1153.5j4j1.10.0...0.0...1ac.1.denzbOMINDg#q=pink+sweater&hl=ru&tbm=shop&ei=oDmlUPXgAoqC4gSylYGACg&start=20&sa=N&num=20&bay=on 2 or r. gc r. pw r. of &fn=7898c08770da3440&hnc1=35466521&biw=1280&bih=660

**Brands**

Alex Evenings (1)

August Silk (1)

Bar III (2)

Charter Club (5)

DKNY Jeans (1)

Karen Scott (1)

Nautica (4)

Perry Ellis (1)

Ralph Lauren (6)

Rare Editions (1)

Say What? (1)

Style&co. (1)

Tommy Hilfiger (3)

Karen Scott Sweater,
Long-Sleeve Mock Turtleneck**\$19.98**Ralph Lauren Kids Sweater,
Little Girls Boyfriend
Cardigans**\$49.50**Ralph Lauren Kids Sweater,
Little Girls Peplum Cardigan**\$49.50**Charter Club Plus Size
Cardigan, Three-Quarter-
Sleeve Space-Dyed**\$44.98**Charter Club Plus Size
Sweater, Cashmere
Long-Sleeve Tunic**\$209.00**Charter Club Petite Sweater,
Long-Sleeve Cashmere
Turtleneck**\$119.00**Tommy Hilfiger Knit Cardigan,
Button-Down Shirt & JeggingsDKNY Jeans Pointelle
Sweater & Soho Skinny Jeans**\$59.00 - 79.00**



Brands

Alex Evenings (1)

August Silk (1)

Bar III (2)

Charter Club (5)

DKNY Jeans (1)

Karen Scott (1)

Nautica (4)

Perry Ellis (1)

Ralph Lauren (6)

Rare Editions (1)

Say What? (1)

Style&co. (1)

Tommy Hilfiger (3)



Karen Scott Sweater, Long-Sleeve Mock Turtleneck

\$19.98



Ralph Lauren Kids Sweater, Little Girls Boyfriend Cardigans

\$49.50



Ralph Lauren Kids Sweater, Little Girls Peplum Cardigan

\$49.50



Charter Club Plus Size Cardigan, Three-Quarter-Sleeve Space-Dyed

\$44.98



Charter Club Plus Size Sweater, Cashmere Long-Sleeve Tunic

\$209.00



Charter Club Petite Sweater, Long-Sleeve Cashmere Turtleneck

\$119.00



Tommy Hilfiger Knit Cardigan, Button-Down Shirt & Jeggings



DKNY Jeans Pointelle Sweater & Soho Skinny Jeans

\$59.00 - 79.00

Polysemy Problem No.2

pink rose sweater



BRAND:"**Pink Rose**" TYPE:"Sweater"

COLOR:"**Pink**", "**Rose**" TYPE:"Sweater"



pink rose sweater

Search

Department

Sweaters (26)

Dress (13)

Outfit (11)

Tops (5)

Boots (4)

Hosiery (3)

Jeans (2)

Watches (2)

Baby Sets (1)



[Sweet Heart Rose Girls Dress, Little Girls Dot Dress and Cardigan](#)



[Free People Sweater, Long-Sleeve V-Neck Rose-Print Cardigan](#)



[Pink Rose Juniors Sweater, Long Sleeve Cable-Knit Cocoon](#)



[Pink Rose Juniors Sweater, Long Sleeve Cable-Knit Elbow-Patch](#)

Brand

Pink Rose (48)

DKNY (4)

Alfani (2)

Rewash (2)

Sweet Heart Rose (2)

Vince Camuto (2)

First Impressions (1)

Free People (1)



[Pink Rose Plus Size Dress, Long-Sleeve Cowl-Neck Sweater](#)



[Pink Rose Juniors Sweater, Long Sleeve Cable-Knit](#)



[Pink Rose Juniors Sweater, Long Sleeve Striped Dolman](#)



[Pink Rose Juniors Sweater, Short Sleeve Tulip Tunic](#)



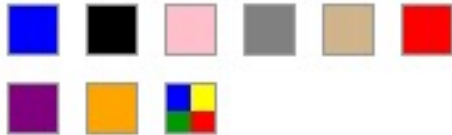
pink rose sweater

Search

Brand

Pink Rose (26)

Color



FOB

Juniors (25)

Plus Sizes (1)

Gender

Female (26)

PRODUCT_TYPE

SWEATER (26)



Pink Rose Juniors Sweater, Three-Quarter Sleeve Striped High-Low



Pink Rose Juniors Sweater, Short Sleeve Cable-Knit Turtleneck



Pink Rose Juniors Sweater, Long Sleeve Striped Ribbed



Pink Rose Plus Size Sweater, Short-Sleeve Belted Tunic



Pink Rose Juniors Sweater, Three-Quarter Sleeve Pointelle-Knit Cardigan



Pink Rose Juniors Dress, Long Sleeve Belted Cable-Knit



Pink Rose Juniors Sweater, Three-Quarter Sleeve Ribbed Sheer



Pink Rose Juniors Dress, Short Sleeve Cable-Knit Sweater

Ranking Can't Help

Gender

Womens (427)

Girls (106)

Mens (86)

Juniors (78)

Boys (1)

Department

Dress (367)

Outfit (151)

Pants (123)

Suit (43)

Blazer (28)

Loafers (20)

Oxfords (18)

Handbags (15)

Boots (9)

Pumps (9)

more...

Brand

Alfani (93)

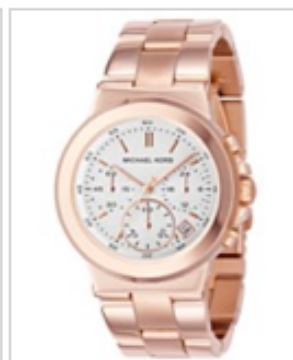
red dress



MORE COLORS AVAILABLE
[Calvin Klein Handbag, Evening Clutch](#)



[Sequin Hearts Girls Dress, Girls Red Ruffle Dress](#)



[Michael Kors Watch, Women's Chronograph Dylan Rose Gold Tone Stainless Steel Bracelet 32mm MKS223](#)



[Calvin Klein Dress, Sleeveless Ruffled Cap Sleeve Red Sheath](#)



[Calvin Klein Petite Dress, Sleeveless Ruffled Cap Sleeve Red Sheath](#)



NEW! [Alfani RED Pants, Grey Plaid with Blue Deco Slim Fit](#)



NEW! [Alfani RED Suit Separates, Grey Plaid with Blue Deco Slim Fit](#)



NEW! [Alfani RED Pants, Slim-Fit Black Ministripe](#)

Ranking Can't Help

Gender

Womens (427)

Girls (106)

Mens (86)

Juniors (78)

Boys (1)

Department

Dress (367)

Outfit (151)

Pants (123)

Suit (43)

Blazer (28)

Loafers (20)

Oxfords (18)

Handbags (15)

Boots (9)

Pumps (9)

more...

red dress



MORE COLORS AVAILABLE
[Calvin Klein Handbag, Evening Clutch](#)



[Sequin Hearts Girls Dress, Girls Red Ruffle Dress](#)



[Michael Kors Watch, Women's Chronograph Dylan Rose Gold Tone Stainless Steel Bracelet 32mm MK5223](#)



[Calvin Klein Dress, Sleeveless Ruffled Cap Sleeve Red Sheath](#)



[Calvin Klein Petite Dress, Sleeveless Ruffled Cap Sleeve Red Sheath](#)



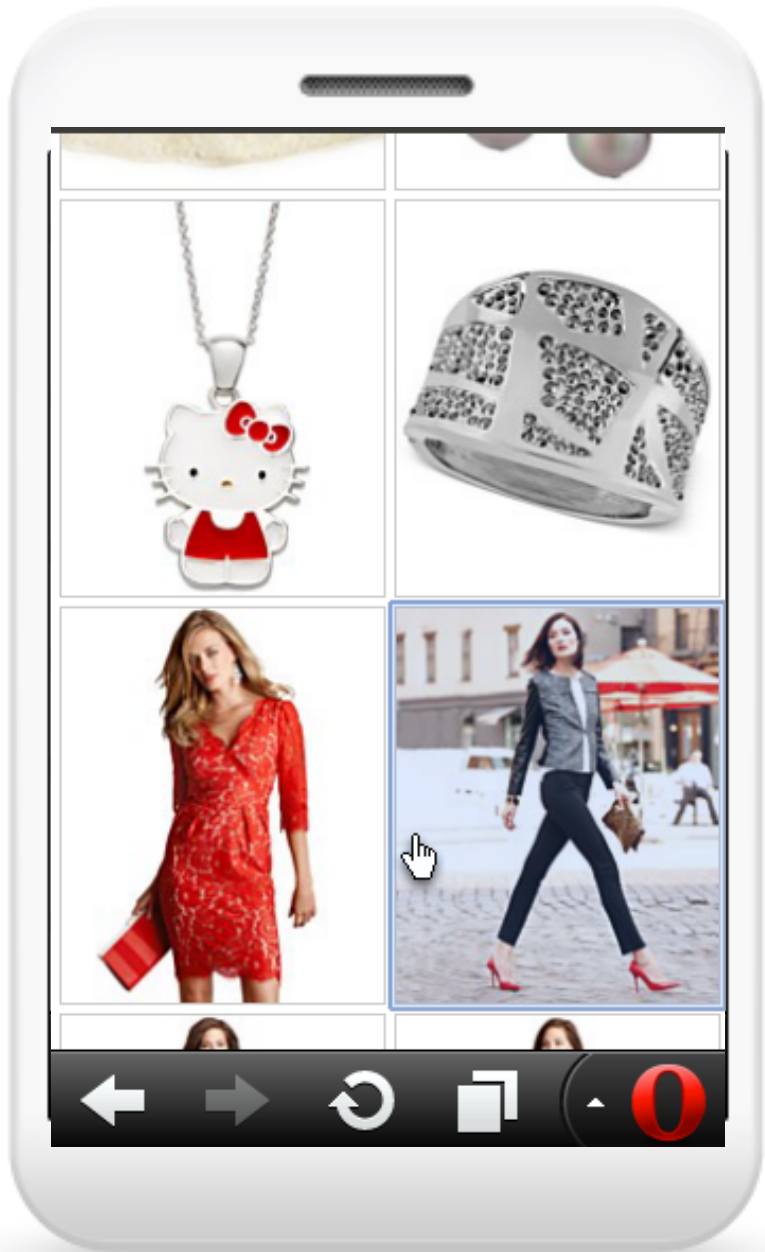
NEW!
[Alfani RED Pants, Grey Plaid with Blue Deco Slim Fit](#)



NEW!
[Alfani RED Suit Separates, Grey Plaid with Blue Deco Slim Fit](#)



NEW!
[Alfani RED Pants, Slim-Fit Black Minisripe](#)



**Precision
is a MUST!**

Think About the User

bieber black boot bride co cole
collection concepts
cookware cuisinart dinnerware dresses
ensemble fairfield flannel food furniture gift
girlfriend handbags hotel impulse inc
international italian jackets
jeans justin
kenneth kloth kut lancome
lauren martha
mikasa mother nicole number occasion
pants piece processor purchase
ralph reaction richie simpson
square stewart swimwear



silver jeans



Silver Jeans



Clothing

Women

Tops & Tees

Jeans

Shorts

Maternity

Men

Jeans

Sale & Clearance in Clothing

[View All Silver Jeans](#)



Wrap-Up

- Product Descriptions Differs to Plain Text
- No Query Syntax
- Polysemy Problem
- Ranking Can't Help When Precision is Low

eCommerce Search is Special!



Part II

Compound Words Query Parser



Don't tokenize at index time

```
<fieldType name="istring" class="solr.TextField" >  
  <analyzer type="index">  
    <tokenizer class="solr.KeywordTokenizerFactory"/>  
    <filter class="solr.LowerCaseFilterFactory"/>  
  </analyzer>  
..
```

```
</fieldType>
```

lauren ralph lauren

calvin klein

ralph lauren

inc international

concepts

tommy hilfiger

charter club

effy collection

nourison

anne klein

betsey johnson

bar iii

giani bernini

jones new york



black and decker toaster oven

Search

"black"

"black and"

"black and decker"

... "and"

"and decker"

"and decker toaster"

... "decker"

"decker toaster"

"decker toaster oven"

"toaster oven"



black and decker toaster oven

Search

"black"

"black and"

"black and decker"

... "and"

"and decker"

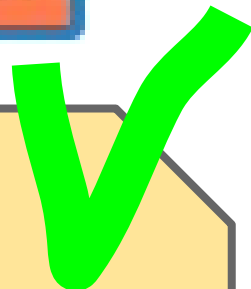
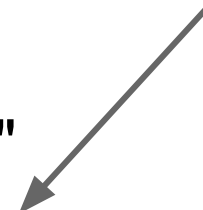
"and decker toaster"

... "decker"

"decker toaster"

"decker toaster oven"

... "toaster oven"



ID: 123
BRAND: **"black and decker"**
TYPE: **"toaster oven"**
COLOR: white



black and decker toaster oven

Search

"black"

"black and"

"black and decker" ←

... "and"

"and decker"

"and decker toaster"

... "decker"

"toaster"

"decker toaster oven" →

... **"oven"**



| | |
|--------|---------------------------|
| ID: | 456 |
| BRAND: | "black and decker" |
| TYPE: | "oven" |
| COLOR: | white |



black and decker toaster oven

Search

"black"

"black and"

"black and decker"

... "and"

"and decker"

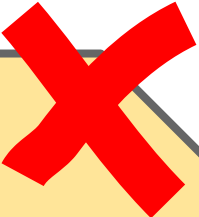
"and decker toaster"

... "decker"

"decker toaster"

"decker toaster oven"

... "toaster oven"



| | |
|--------|---------------------|
| ID: | 789 |
| BRAND: | "decker toaster" |
| TYPE: | "toaster oven" |
| COLOR: | black |

Brand

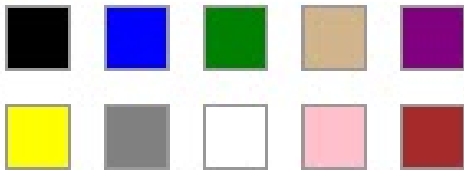
Free People (24)



fee people sweater

Search

Color



FOB

Women's (24)

Men's (1)

Gender

Female (24)

PRODUCT_TYPE

SWEATER (24)

PANTS (4)

TOP (2)

JEANS (2)

SHORTS (1)



Free People Sweater, Long-Sleeve Crew-Neck Wool-Blend Textured Cardigan



Free People Sweater, Long-Sleeve Scoop-Neck Wool-Blend Colorblocked Top



Free People Sweater, Long-Sleeve Scoop-Neck Cable-Knit Top



Free People Sweater, Long-Sleeve Scoop-Neck Patterned Open-Stitch Wool-Blend Top



Free People Long-Sleeve Boatneck Leopard-Print Top & Faux-Leather Colored Shorts



Free People Sweater, Open-Front Long-Sleeve Floral Cardigan



Free People Sweater, Long-Sleeve Striped Cardigan



Free People Long-Sleeve V-Neck Sweater & Printed Leggings

Concept Search: Semantics Enabled Syntactic Search

Fausto Giunchiglia, Uladzimir Kharkevich, and Ilya Zaihrayeu

Department of Information Engineering and Computer Science
University of Trento, Italy

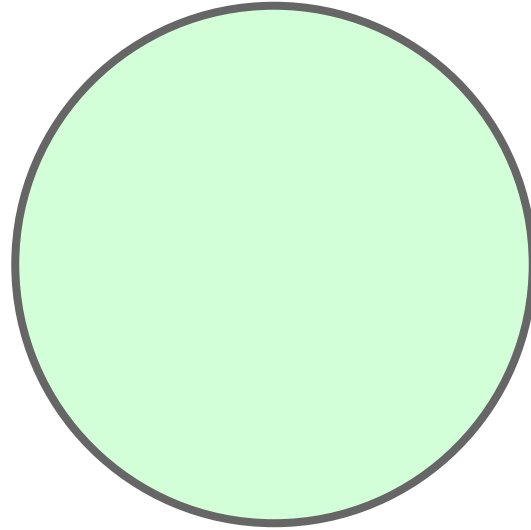
{fausto,kharkevi,ilya}@disi.unitn.it

Abstract. Historically, information retrieval (IR) has followed two principally different paths that we call syntactic IR and semantic IR. In syntactic IR, terms are represented as arbitrary sequences of characters and IR is performed through the computation of string similarity. In semantic IR, instead, terms are represented as concepts and IR is performed

Part III

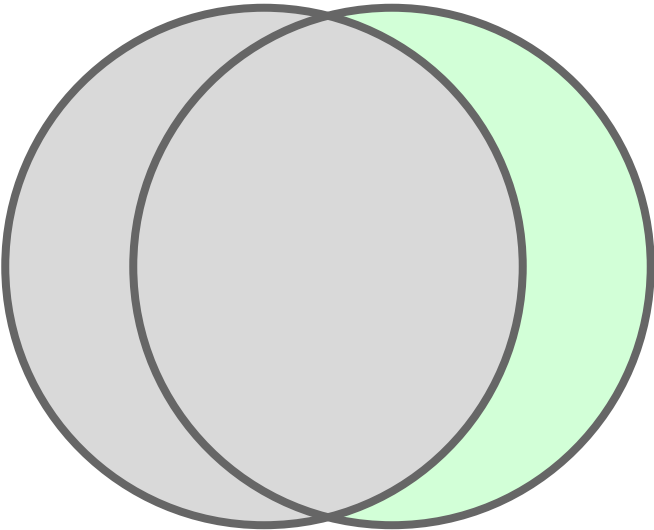
Staged Search





Expected

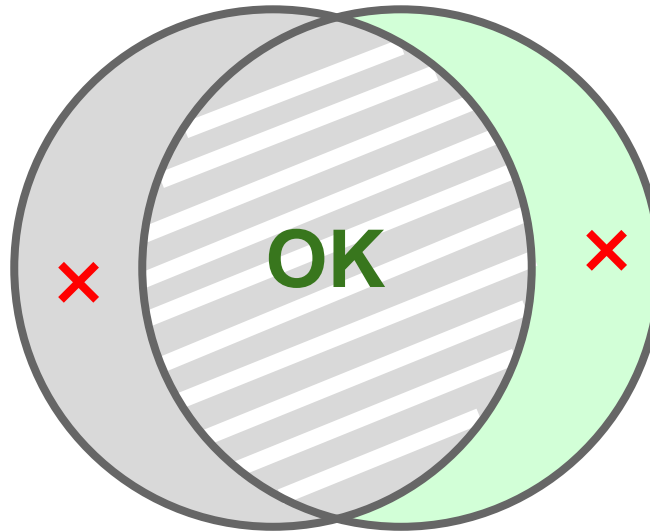
Actual



Expected

Actual

Expected

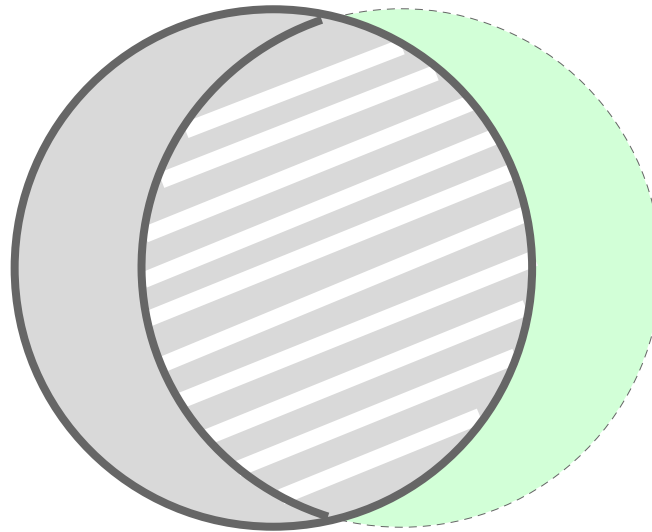


False
Positive

True
Positive

False
Negative

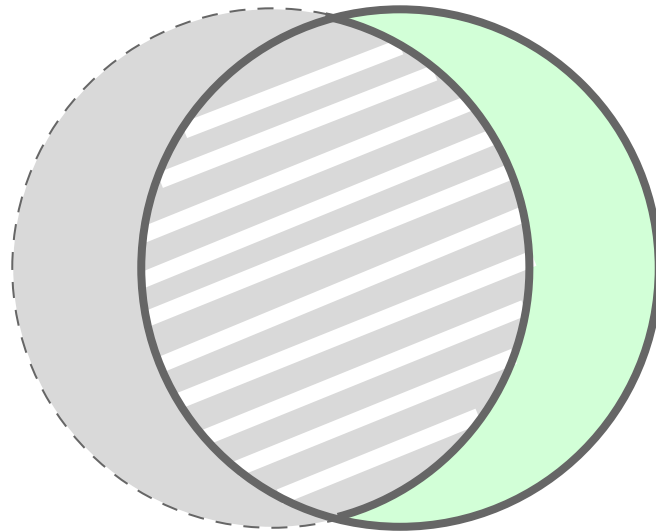
Actual



Expected

$$\text{Precision} = \frac{\text{Actual} \cap \text{Expected}}{\text{Actual}}$$

Actual



Expected

$$\text{Recall} = \frac{\text{Actual} \cap \text{Expected}}{\text{Expected}}$$

100% Recall

Actual

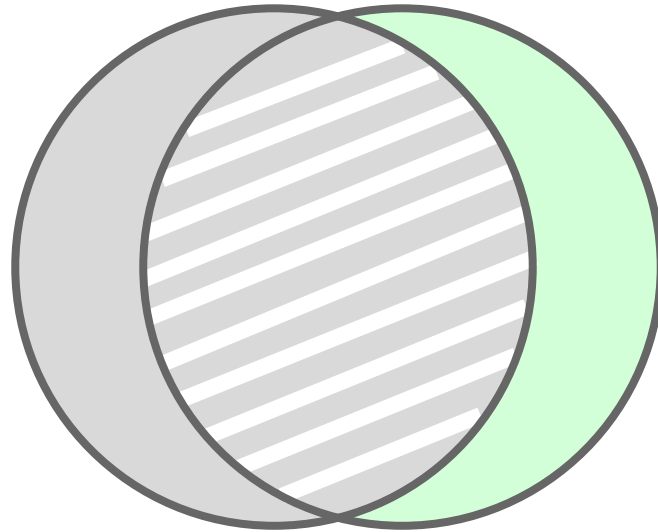


Expected

Tuning Trade-Off

Actual

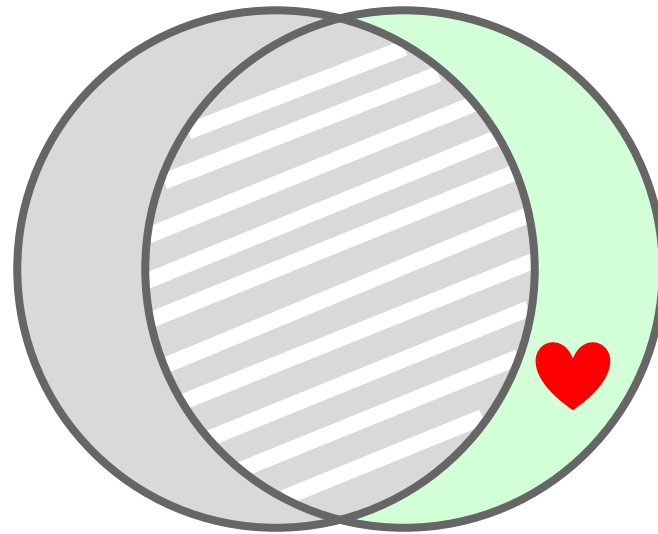
Expected



Tuning Trade-Off

Actual

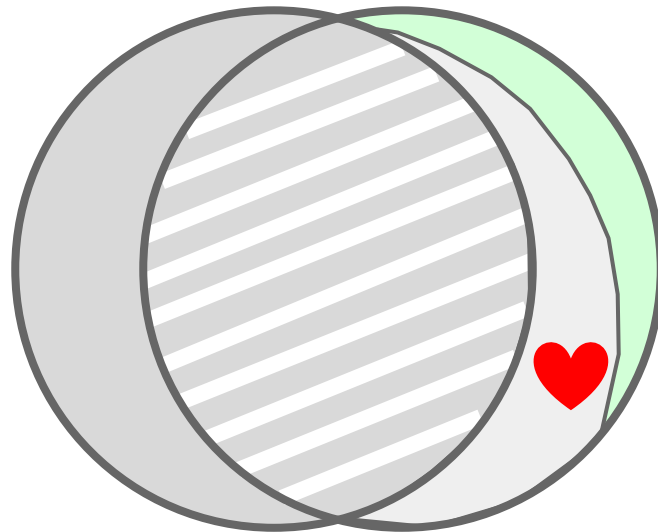
Expected



Tuning Trade-Off

Actual

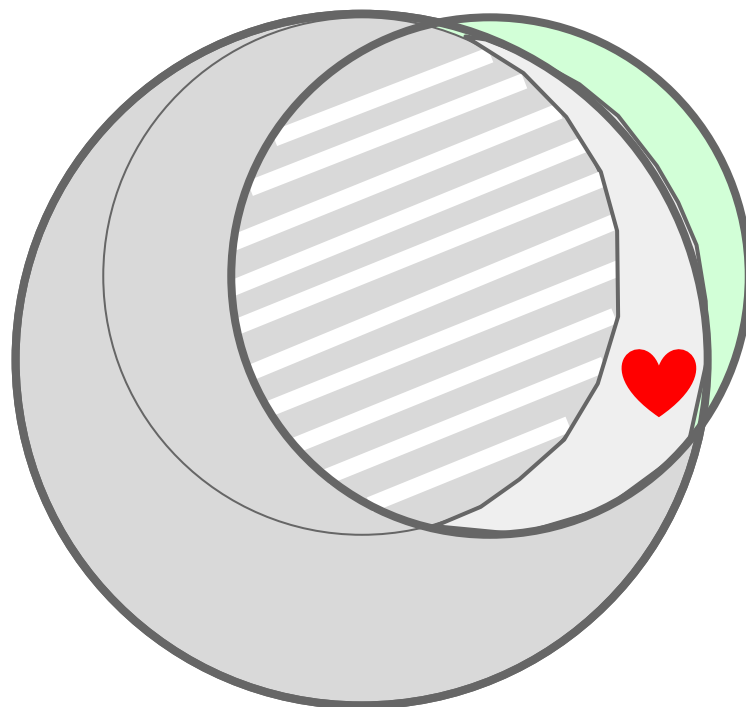
Expected



Tuning Trade-Off

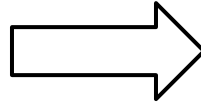
Actual

Expected



defaultOperator

AND



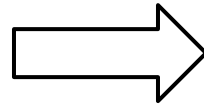
OR

numFound=100

numFound=1000

defaultSearchField / queryFields

BRAND



BRAND

TYPE

STYLE

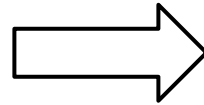
text

numFound=100

numFound=1000

Stemming & Synonyms

text



text_stemmed

numFound=100

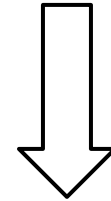
numFound=1000

ORDER by DESCending Precision

BRAND:(Silver AND Jeans)

numFound=100

BRAND:(Silver OR Jeans)



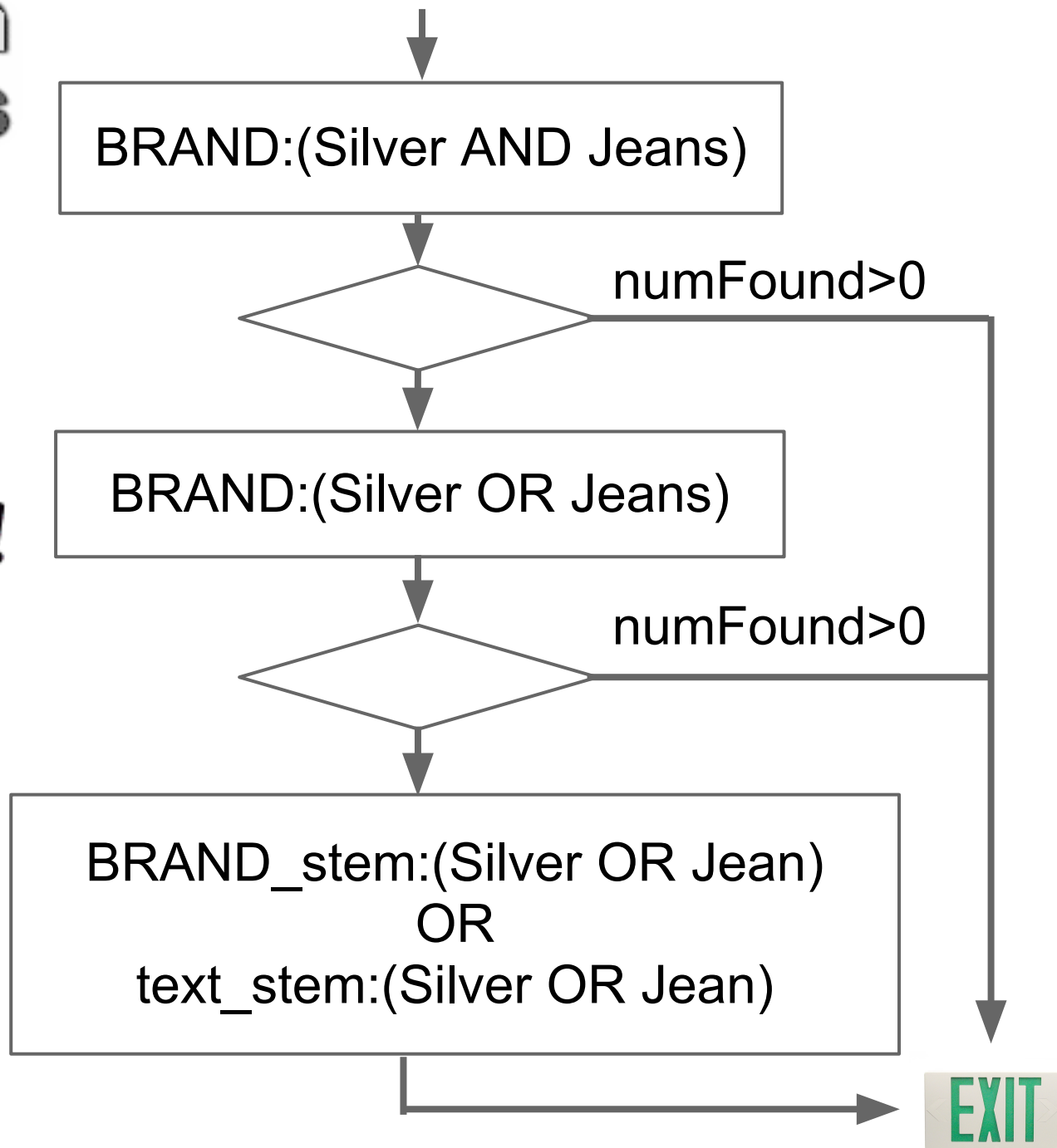
BRAND_stem:(Silver OR Jean)
OR
text_stem:(Silver OR Jean)

numFound=1000

Captain Obvious



to the rescue!



Searching Complexity

$$O(n \log p)$$

num found

page size

It Depends™

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 score 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 novel algorithms that efficiently compute the true total ranking with a fixed space requirement independent of the size of the collection.

Stages Ratios

Compounds - Exact Match

50%

Compounds - Omit Words

30%

Text Match

20%

Think About The User

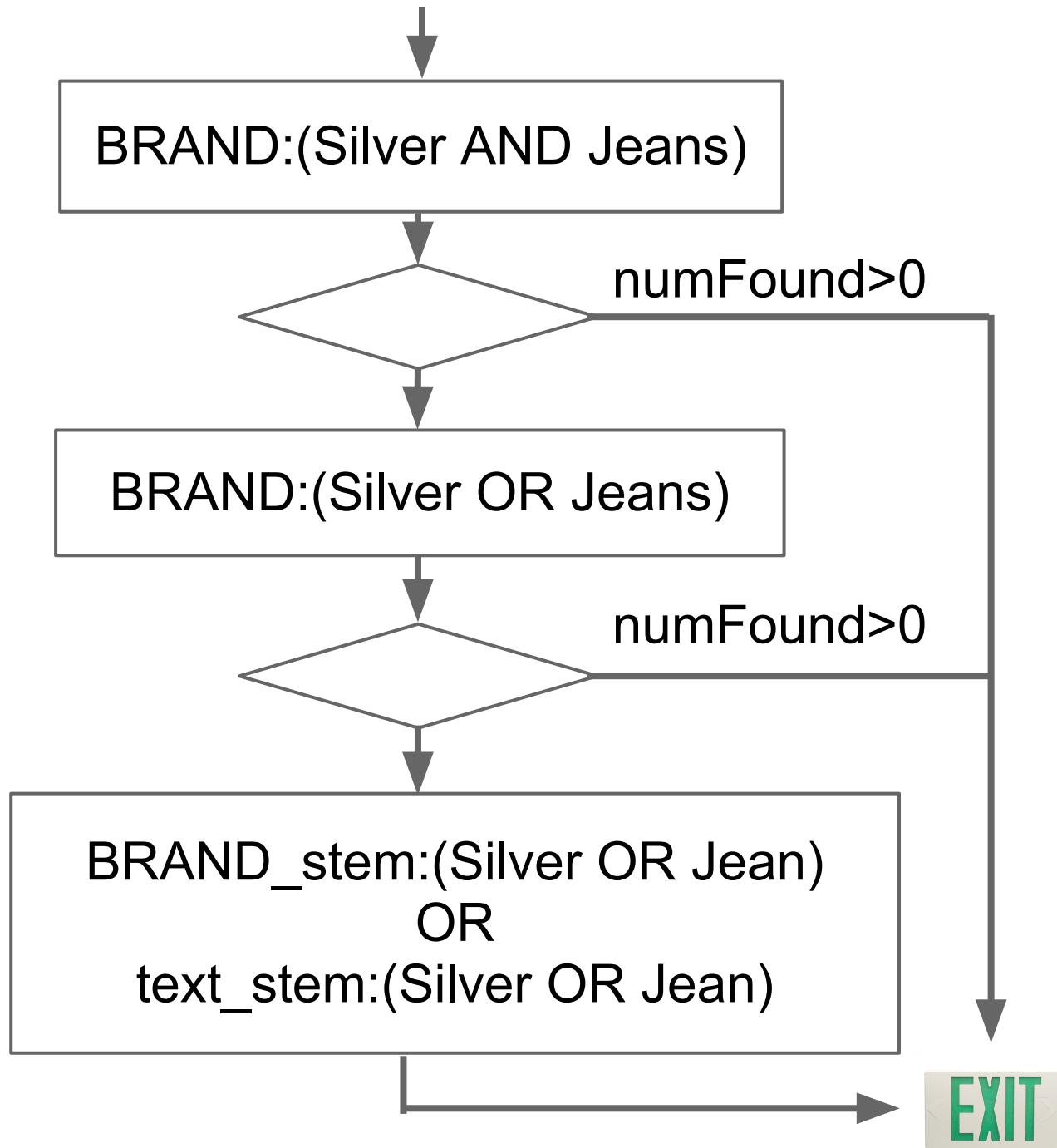
bieber black boot bride co cole
collection concepts
cookware cuisinart dinnerware dresses
ensemble fairfield flannel food furniture gift
girlfriend handbags hotel impulse inc
international italian jackets
jeans justin
kenneth kloth kut lancome
lauren martha
mikasa mother nicole number occasion
pants piece processor purchase
ralph reaction richie simpson
square stewart swimwear



Part IV

Match Spotting



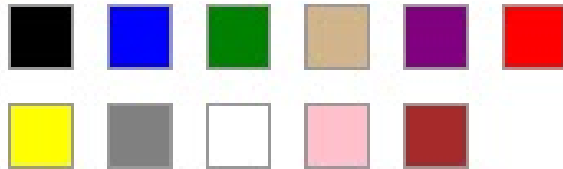


Facet Counts

Brand

Free People (24)

Color



FOB

Women's (24)

Men's (1)

Gender

Female (24)

PRODUCT_TYPE

SWEATER (24)

PANTS (4)

TOP (2)

JEANS (2)

Explain Info

15.110098 = (MATCH) sum of:

11.197777 = (MATCH) weight(**BRAND:Free People in 28807**), product of:

0.81810623 = queryWeight(BRAND:Free People), product of:

9.678479 = idf(docFreq=232, maxDocs=1368899)

0.08452839 = queryNorm

13.687436 = (MATCH) fieldWeight(BRAND:Free People in 28807), product of:

1.4142135 = tf(termFreq(BRAND:Free People)=2)

9.678479 = idf(docFreq=232, maxDocs=1368899)

1.0 = fieldNorm(field=BRAND, doc=28807)

3.912321 = (MATCH) weight(**PRODUCT_TYPE:SWEATER in 28807**), product of:

0.5750671 = queryWeight(PRODUCT_TYPE:SWEATER), product of:

6.8032427 = idf(docFreq=4130, maxDocs=1368899)

0.08452839 = queryNorm

6.8032427 = (MATCH) fieldWeight(PRODUCT_TYPE:SWEATER in 28807), product of:

1.0 = tf(termFreq(PRODUCT_TYPE:SWEATER)=1)

6.8032427 = idf(docFreq=4130, maxDocs=1368899)

1.0 = fieldNorm(field=PRODUCT_TYPE, doc=28807)

Explain
Info **+** **Facet**
Counts **=** **Match**
Spotting

silver jeans



found **190**

BRAND: "Silver Jeans" - **(156)**

COLOR: "Silver" **TYPE:** "Jeans" - **(34)**

Silver Jeans



Clothing

Women

Tops & Tees

Jeans

Shorts

Maternity

Men

Jeans

Sale & Clearance in Clothing

[View All Silver Jeans](#)



Wrap-Up

- eCommerce Search is Special
- Compound Words Query Parser
- Staged Search
- Match Spotting



<http://goo.gl/2OHk2>