Lucidworks

Who are we?

Based in San Francisco

Offices in Cambridge, Bangalore, Bangkok, New York City, Raleigh, Munich

Over 300 customers across the Fortune 1000

Fusion, a Solr-powered platform for search-driven apps

Consulting and support for organizations using Solr

Problem Statement

User searched for "red case for macbook.pro"

See this:

	Lowepro - CompuDaypack Laptop and Camera Bag - Slate Gray
	$79.99
	Designed for use with most laptops and digital cameras; DryFlo shoulder harness; padded camera and

	OtterBox - Universal Defender Series Case for Apple® iPhone® 4 - Black
	$49.99
	Compatible with black Apple® iPhone® 4; polycarbonate shell with silicone midlayer; ratcheting belt clip holster

	Case Logic - Case for Most Compact Digital Cameras - Black
	$19.99
	Fits most compact digital cameras, MP3 players and other personal electronic devices; neoprene material; wrist strap

	Speck - Hardshell Laptop Case - Red
	$49.99
	Compatible with select Apple MacBook Pro and unibody MacBook laptops with 13" displays; hard polycarbonate plastic construction ports, battery check button and CD/DVD drive

	Vater - Safe 'N Sound Ear Plugs
	$24.99
	Ear protection for all musicians; 2 filter styles; included case

Should see this:

	Hard Candy Cases - Bubble Shell Laptop Case for 13" Apple® MacBook® Pro - Red
	$19.95
	Fits 13" Apple MacBook Pro (aluminum body, black keyboard); polycarbonate shell; form-fitting, snap-on design; bubble texture

	Speck - Hardshell Laptop Case - Red
	$31.87
	Fits most 15" Apple MacBook Pro models; hard polycarbonate plastic material; vented case; built-in

	Speck - Hardshell Laptop Case - Red
	$49.99
	Compatible with select Apple MacBook Pro and unibody MacBook laptops with 13" displays; hard polycarbonate plastic construction ports, battery check button and CD/DVD drive

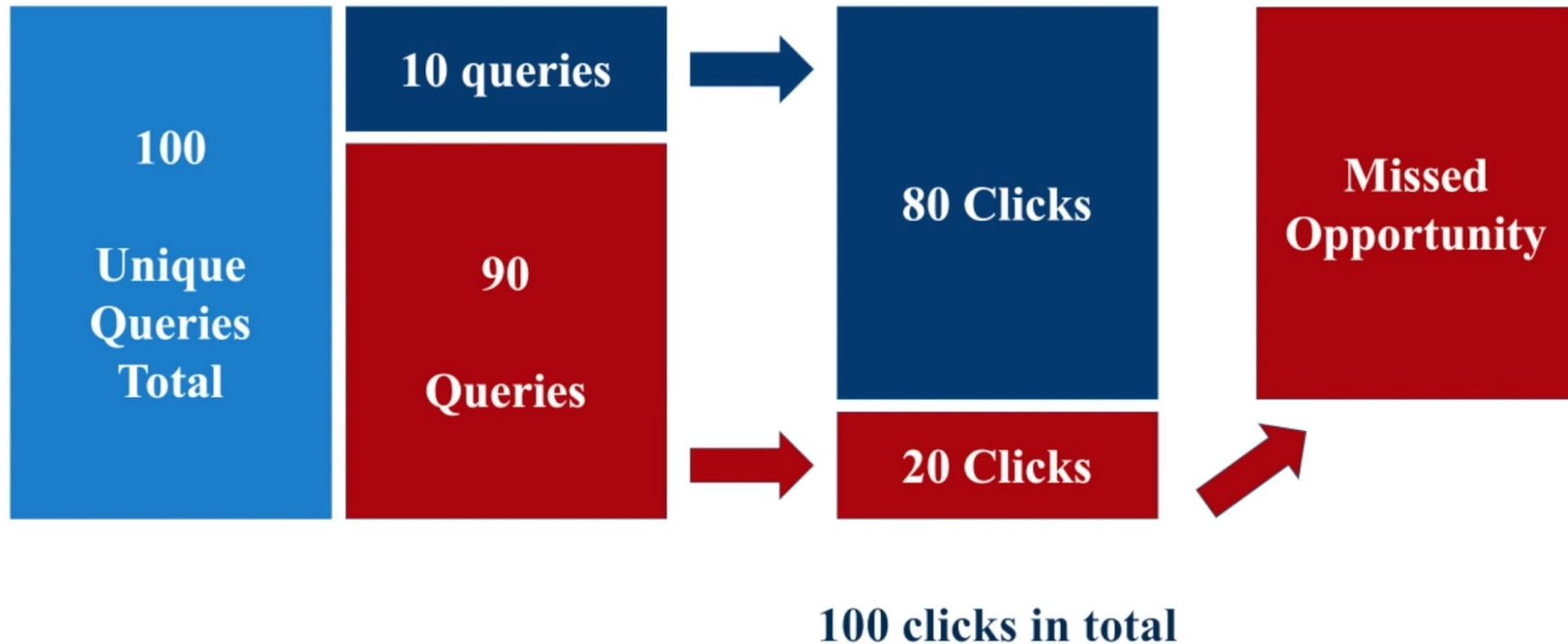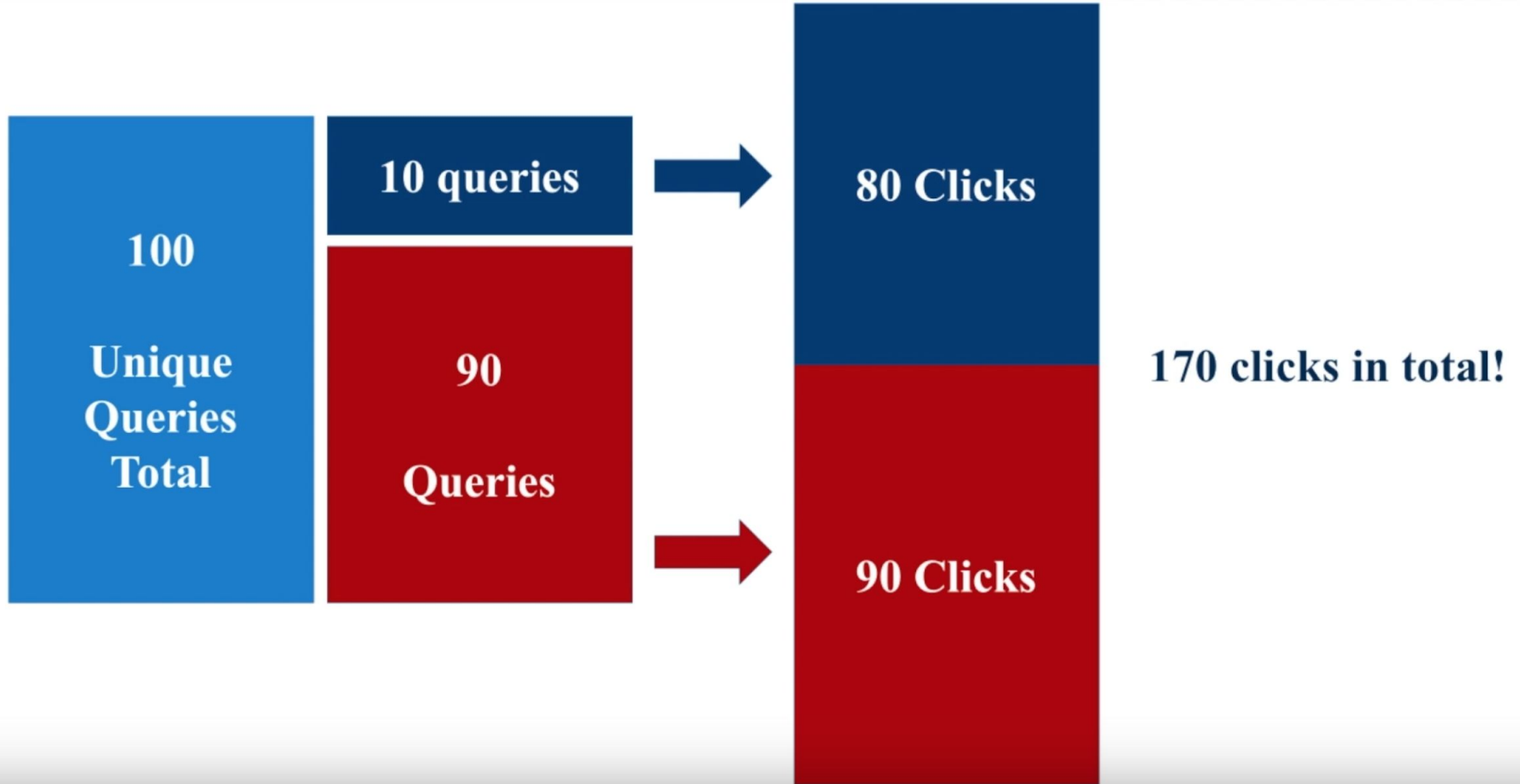|	Speck - Hard Shell Case for Select Apple® MacBook® Pro Laptops - Red
	$49.99
	Compatible with Apple® Macbook® Pro up to 15"; polycarbonate plastic material; rubber feet for added support

|	Cocoon - Case for 15" Apple® MacBook® and MacBook® Pro - Red
	$59.99
	Accommodates up to 15" MacBook or MacBook Pro; expandable gusset; adjustable shoulder strap; padded handles

Based on **Pareto effect**:

- Pareto principle: **80/20 rule**, roughly 80% of the effects come from 20% of the causes.
- Top causes are considered as Heads
- Remaining elements as Torso/Tails aggregated based on criteria
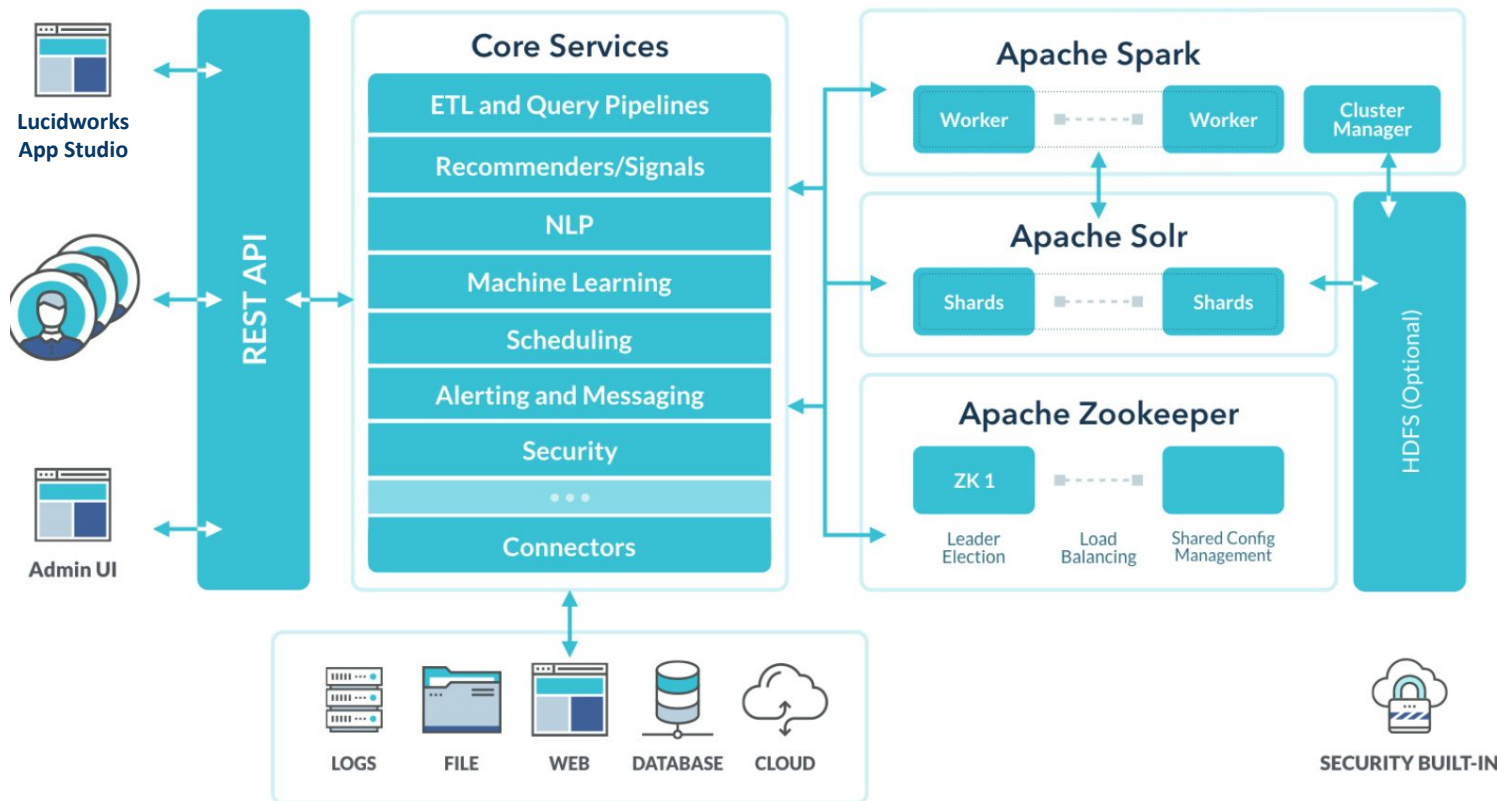
Includes three elements:

- Ranking
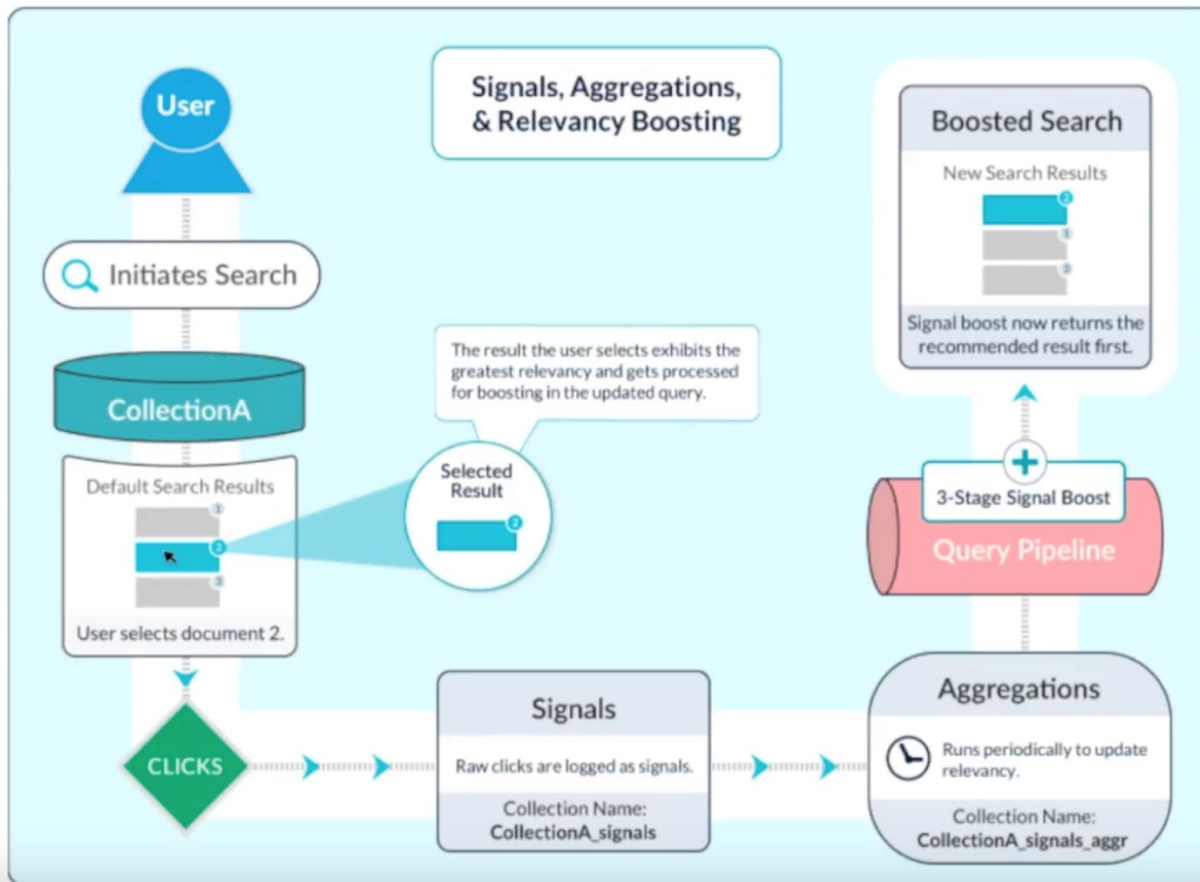- Pareto analysis
- Pie Chart

Problem Statement

It's not uncommon to see the Pareto effect in query-related signals, i.e., approximately 50% to 75% of your traffic might be coming from less than 1% of the queries.

Head Tail Plot
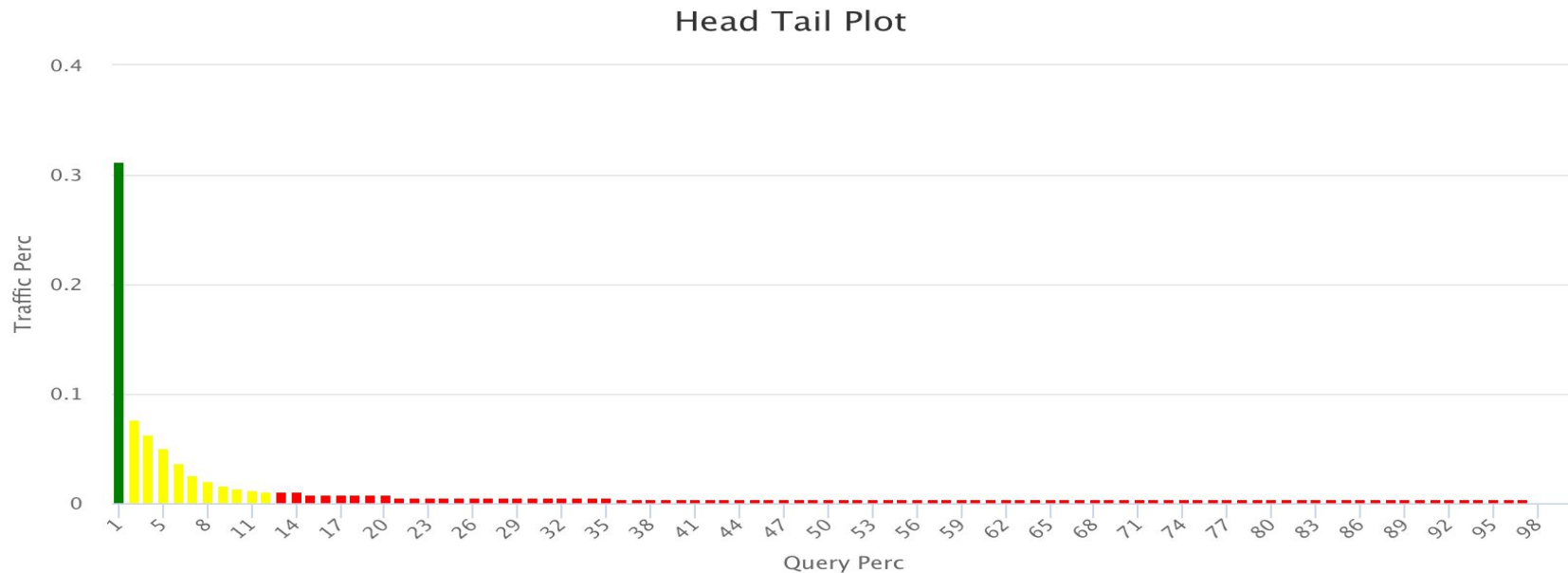
Fusion Overview

Lucidworks Fusion Signals & Aggregations

Fusion signals and aggregation jobs capture and summarize interesting user events, such as queries, result clicks, purchases, and their associated metadata and metrics.

Head – The queries that generate most of your traffic and conversions

Query	Searches	Clicks
lcd tv	35,000	30,000

Tail – The queries that generate very few or no clicks

Query	Searches	Clicks
Outdoor rock speaker	200	5

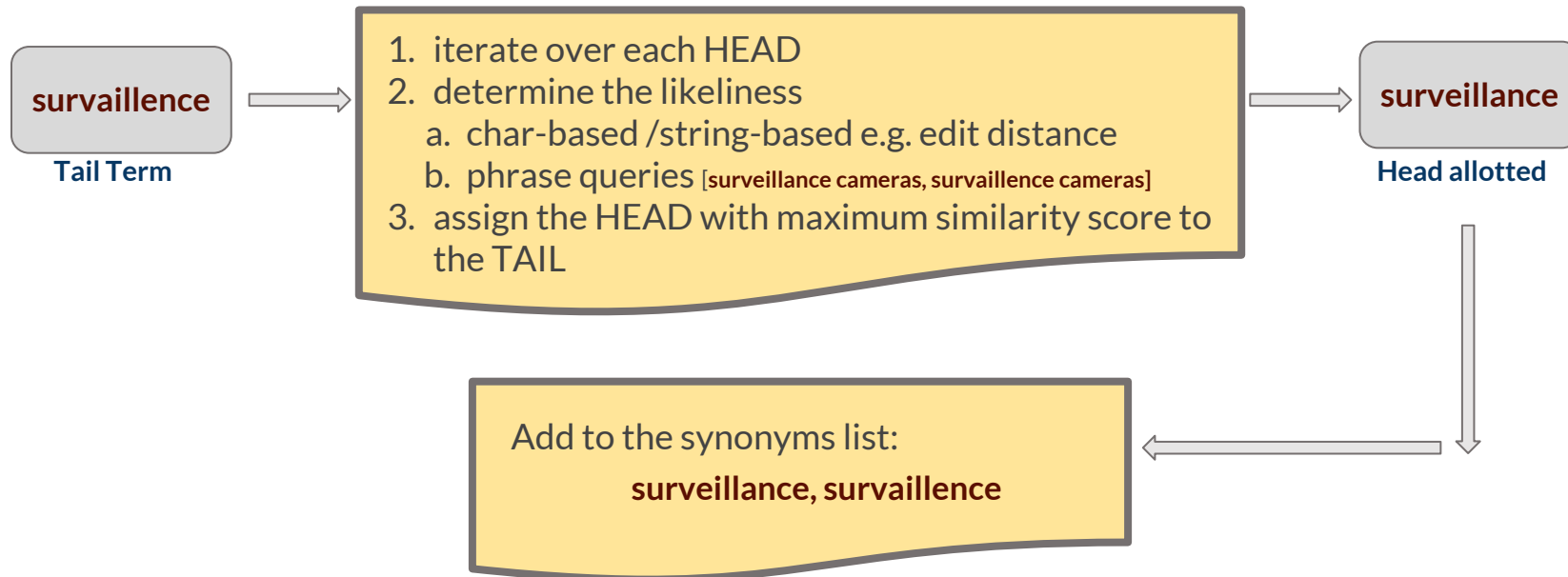|	
doc_id	1854151
doc_id_s	1854151
query_s	lcd tv
query_t	["lcd tv"]
score	50.176533
timestamp_tdt	2018-02-13T00:36:41.158Z
tx_timestamp_txt	["Mon 2018-02-12 19:36:41.158 -0500"]
type	click
type_s	click
tz_timestamp_txt	["Sat 2011-10-08 08:37:04.205 UTC"]
user_id	c0cb0ed57fbfcdc8bd5fefdaaa59a88999296948
user_id_s	c0cb0ed57fbfcdc8bd5fefdaaa59a88999296948

Common tail query categories:

- Misspellings
- Numbers/Dimensions search: 2 x 2, 2 by 2
- Specific query: query contains a head query and descriptive words
  e.g *foldable* water bottle
- Hard to find certain attributes such as color and brand
- Stopwords e.g. case *for* ipad
- Rare term e.g securefit wireless bluetooth *headphone*
- Rewording e.g. dryer and washer

Based on clicks counts aggregation:

HEAD	iphone x	**surveillance**	game chair	.... *more*
TAIL	chair game	iphonex	**survaillence**	.... *more*

survaillence

Tail Term

1. iterate over each HEAD
2. determine the likeliness
   a. char-based /string-based e.g. edit distance
   b. phrase queries [surveillance cameras, survaillence cameras]
3. assign the HEAD with maximum similarity score to the TAIL

surveillance

Head allotted

Add to the synonyms list:
   surveillance, survaillence
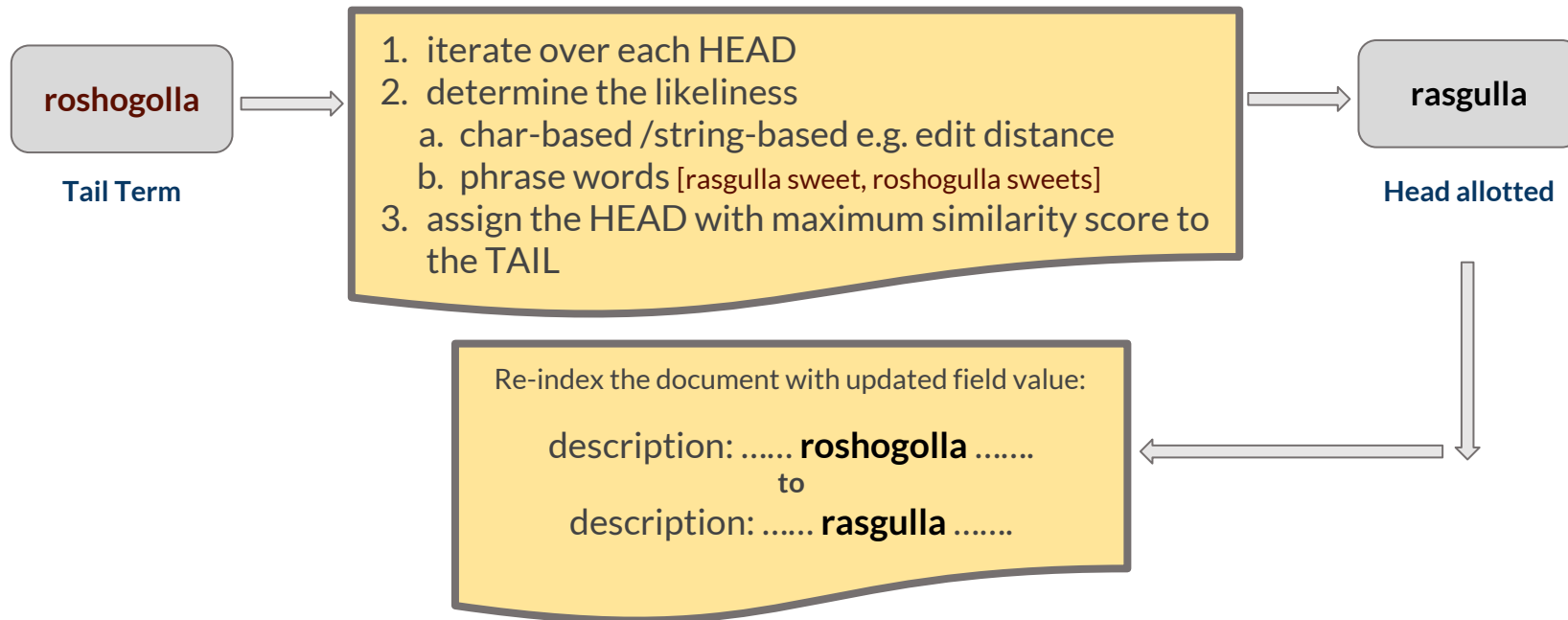
Fusion Token & Phrase Spell Correction Job

Run the spell checker job on signal data or non-signal data.

The configuration must specify:

- **Input Collection**: contains the signals
- **Query Field Name**: field in the collection contains the query
- **Count Field**: field contains the count of the event

Allows you to analyze query performance based on two different events:

- main event e.g. click count
- filtering/secondary event. e.g. query count

Token and phrase spell correction
Provide token and phrase level spell correction which can be put into synonym list

Advanced

* **Spark Job ID**
spell_check

* **Input Collection**
ecommerce_signals

* **Query Field Name**
query_s

Signals data filter query
type_s:click OR type_s:query

Stopwords blob
stopword.txt

Dictionary Collection
ecommerce

Dictionary Field
description

* **Count Field**
count_i

* **Main Event Type**
click

Fusion Head & Tail Analysis Job

User specify signal information in the configuration:

- **Input Collection:** collection contains the signals
- **Query Field Name:** field contains the query string
- **Event Count Field Name:** field contains the count of the event
- **Event Type:** type of events used for analysis

Click Run>Start

Head-n-Tail Analysis

Perform head tail analysis of queries from raw or aggregated signals collection.

Advanced

* Spark Job ID

ecommerce-head-tail

* Input Collection

ecommerce_signals

* Query Field Name

query_s

Signals data filter query

type_s:click OR type_s:query

* Event Count Field Name

count_i

* Main Event Type

click

Filtering Event Type

query

Minimum Main Event Count

1

Minimum Filtering Event Count

20

Minimum Query Length

2

Keywords blob name

ecommerce_keyword.csv

Fusion Head & Tail Analysis: Pareto effect

Summary Stats

Statistics	Value
total number of unique queries	22231
top 100 queries leads to ?% total events	21.24%
top 1.0% queries leads to ?% total events	31.06%
number of queries that constitute the top 75.0% of all events	3060
number of queries that constitute the top 50.0% of all events	680
number of queries that constitute the top 25.0% of all events	140
last 1.0% of total events spreaded in how many queries	1613
how many queries leads to zero events.	628
how many queries have less than 5 events	19734
events threshold for tail.	5
events threshold for head.	39

Head Tail Plot

Lucidworks
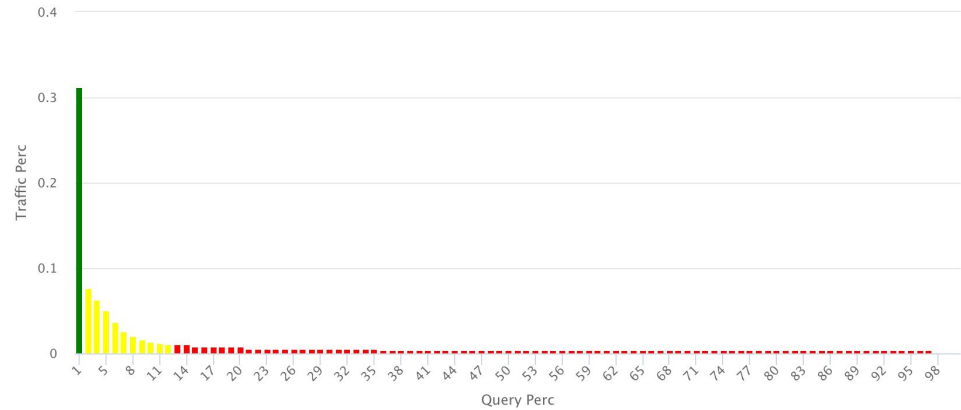
Top Tokens

Tokens	Count
tv	675
the	613
samsung	609
sony	543
case	514
laptop	476
iphone	442
dvd	396
ipod	395
hp	376
car	359
wireless	356

Query Length

Avg String Length	One Word Perc	Two Words Perc	Three Words Perc	Four Words Perc	Five Words Perc	Six+ Words Perc
13	17.89	40.06	27.01	10.07	3.36	1.61

Tail Reasons

Count	Tail Reasons	Proportions
1249	number	6.61%
1129	other-specific	5.98%
930	brand	4.92%
565	spelling	2.99%
205	other-extra	1.08%
156	color	0.82%
83	stopwords	0.43%
75	re-wording	0.39%
68	rare-term	0.36%
42	others	0.22%
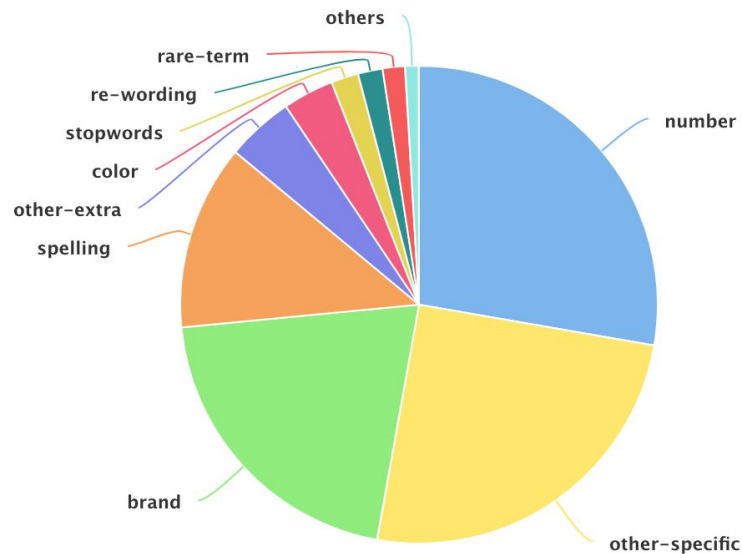
Fusion Automatic Tail Re-writing Suggestions

Extensive feature of Fusion Head & Tail Analysis Job:

Tail Query	Suggested Rewriting	Matched Head	Reason Code	Matched Head (No Num)	Number	Other Specific	Number Unit	Tail Traffic
car amplifier	amplifier (car)^1.5	amplifier	other-specific	amplifier		car		5
headphone bluetooth	bluetooth headphone^2	bluetooth headphone	re-wording	bluetooth headphone				5
subwofer	subwoofer	subwoofer	spelling	subwoofer				5
lord of the ring blu ray	(lord of the ring)^2 blu ray	lord of the ring	other-extra	lord of the ring		blu ray		5
wireless n router	wireless router	wireless router	stopwords	wireless router				5
mw2	mw 2 mw2	mw	number	mw	2		mw2	5
hosa cable	cable hosa	cable	rare-term	cable		hosa		1
the lord of the ring blu ray	(lord of the ring)^2 the blu ray	lord of the ring	other-extra	lord of the ring		the blu ray		1

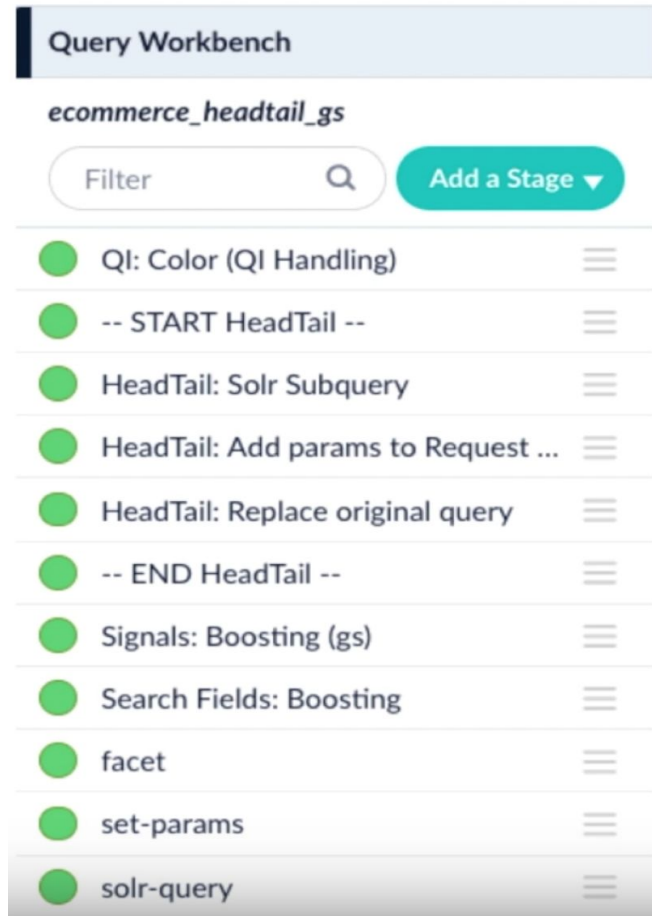# Fusion Tail Rewriting Query Pipeline

Query Parsing and Rewriting in
Fusion Query Workbench:

Query: "red case for macbook.pro"

- query parser match color: "**red**"

- drop stopwords: "**for**"

- HEAD for tail query "case macbook.pro":

  "**macbook pro case**"

Query rewrites to:
"macbook pro case"~10^2 color:red

Query Workbench

ecommerce_headtail_gs

Filter 🔍 | Add a Stage ▼

- 🟢 QI: Color (QI Handling) ≡
- 🟢 -- START HeadTail -- ≡
- 🟢 HeadTail: Solr Subquery ≡
- 🟢 HeadTail: Add params to Request ... ≡
- 🟢 HeadTail: Replace original query ≡
- 🟢 -- END HeadTail -- ≡
- 🟢 Signals: Boosting (gs) ≡
- 🟢 Search Fields: Boosting ≡
- 🟢 facet ≡
- 🟢 set-params ≡
- 🟢 solr-query ≡

The Solution

User searches for "red case for macbook.pro"

Sees this:

Lowepro - CompuDaypack Laptop and Camera Bag - Slate Gray
$79.99

Designed for use with most laptops and digital cameras; DryFlo shoulder harness; padded camera and

OtterBox - Universal Defender Series Case for Apple® iPhone® 4 - Black
$49.99

Compatible with black Apple® iPhone® 4; polycarbonate shell with silicone midlayer; ratcheting belt clip holster

Case Logic - Case for Most Compact Digital Cameras - Black
$19.99

Fits most compact digital cameras, MP3 players and other personal electronic devices; neoprene material; wrist strap

Speck - Hardshell Laptop Case - Red
$49.99

Compatible with select Apple MacBook Pro and unibody MacBook laptops with 13" displays; hard polycarbonate plastic construction
ports, battery check button and CD/DVD drive

Vater - Safe 'N Sound Ear Plugs
$24.99

Ear protection for all musicians; 2 filter styles; included case

After apply query rewriting: "macbook pro case"~10^2 color:red

Hard Candy Cases - Bubble Shell Laptop Case for 13" Apple® MacBook® Pro - Red
$19.95

Fits 13" Apple MacBook Pro (aluminum body, black keyboard); polycarbonate shell; form-fitting, snap-on design; bubble texture

Speck - Hardshell Laptop Case - Red
$31.87

Fits most 15" Apple MacBook Pro models; hard polycarbonate plastic material; vented case; built-in

Speck - Hardshell Laptop Case - Red
$49.99

Compatible with select Apple MacBook Pro and unibody MacBook laptops with 13" displays; hard polycarbonate plastic construction
ports, battery check button and CD/DVD drive

Speck - Hard Shell Case for Select Apple® MacBook® Pro Laptops - Red
$49.99

Compatible with Apple® Macbook® Pro up to 15"; polycarbonate plastic material; rubber feet for added support

Cocoon - Case for 15" Apple® MacBook® and MacBook® Pro - Red
$59.99

Accommodates up to 15" MacBook or MacBook Pro; expandable gusset; adjustable shoulder strap; padded handles

Discovering reasons for low traffic from torso/tail queries, and rewriting tail queries can help improve:

- **Overall conversions**
- **Query models**: color fields needs to be boosted more
- **Product Catalogs:** gap between customer's query and product description
- **SEO/SEM strategies:** learn more about customer's vocabulary or new demand

Never settle, experiment!
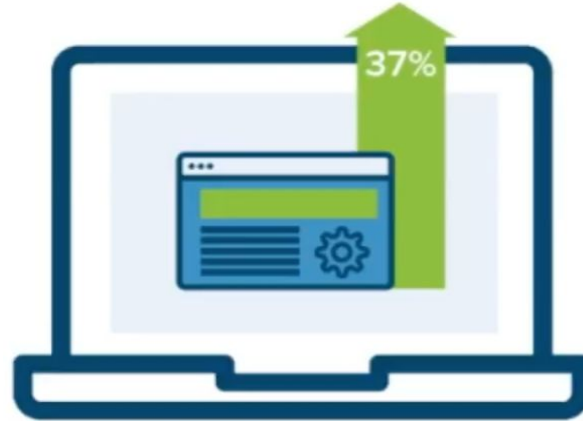
Variant A   **Variant B**

A/B Tests

Based on: *__Head-n-Tail Analysis in Fusion 4__*

Resources

- OfficeMix : Head to Tail Vector Analysis

- Head vs. Long Tail Keywords Analyzed: Impressions, Clicks | SEW

- Head/Tail Analysis Jobs - Lucidworks Documentation

- Fusion 4 Overview | Lucidworks

- Signals - Lucidworks Documentation

- Experiments API (experimental) - Lucidworks Documentation