Balancing the *Dimensions* of *User Intent*

Trey Grainger
Chief Algorithms Officer
Lucidworks

HAYSTACK EU

October 28, 2019

About Me

Trey Grainger

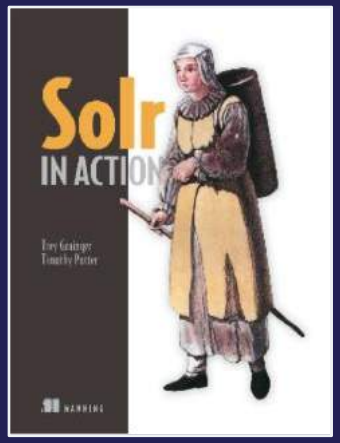
Chief Algorithms Officer

Lucidworks

- **Previously**: SVP of Engineering @ Lucidworks; Director of Engineering @ **CareerBuilder**
- **Georgia Tech – MBA, Management of Technology**
- **Furman University – BA, Computer Science, Business, & Philosophy**
- **Stanford University – Information Retrieval & Web Search**

Other fun projects:

- Co-author of *Solr in Action*, plus numerous research publications
- Advisor to Presearch, the decentralized search engine
- Lucene / Solr contributor



Agenda

- About Lucidworks
- What is AI-powered Search?
- The Dimensions of User Intent
- ***Content Understanding***:
  - *Keyword Search*
- **User Understanding**:
  - *Collaborative Recommendations*
- ***Content Understanding*** + ***User Understanding***:
  - *Personalized Search*
- ***Domain Understanding***:
  - *Knowledge Graphs*
- ***Domain Understanding*** + ***User Understanding***:
  - *Domain-aware Matching*
- ***Content Understanding*** + **Domain *Understanding***:
  - *Semantic Search*
  - Balancing Approaches:
    - Keyword vs. Vector vs. Knowledge Graph Search
    - Vector Search
    - Knowledge Graph Search
- Combining it all together

Lucidworks

Who are we?

300+

EMPLOYEES

400+

CUSTOMERS ACROSS THE FORTUNE 1000

OFFICES IN

San Francisco, CA (HQ)
Raleigh-Durham, NC
Cambridge, UK
Bangalore, India
Hong Kong

COMPANY BEHIND

ACTIVATE

The Search & AI Conference

Solr

DEVELOPMENT, HOSTING, & SUPPORT

FUSION Server

FUSION AI

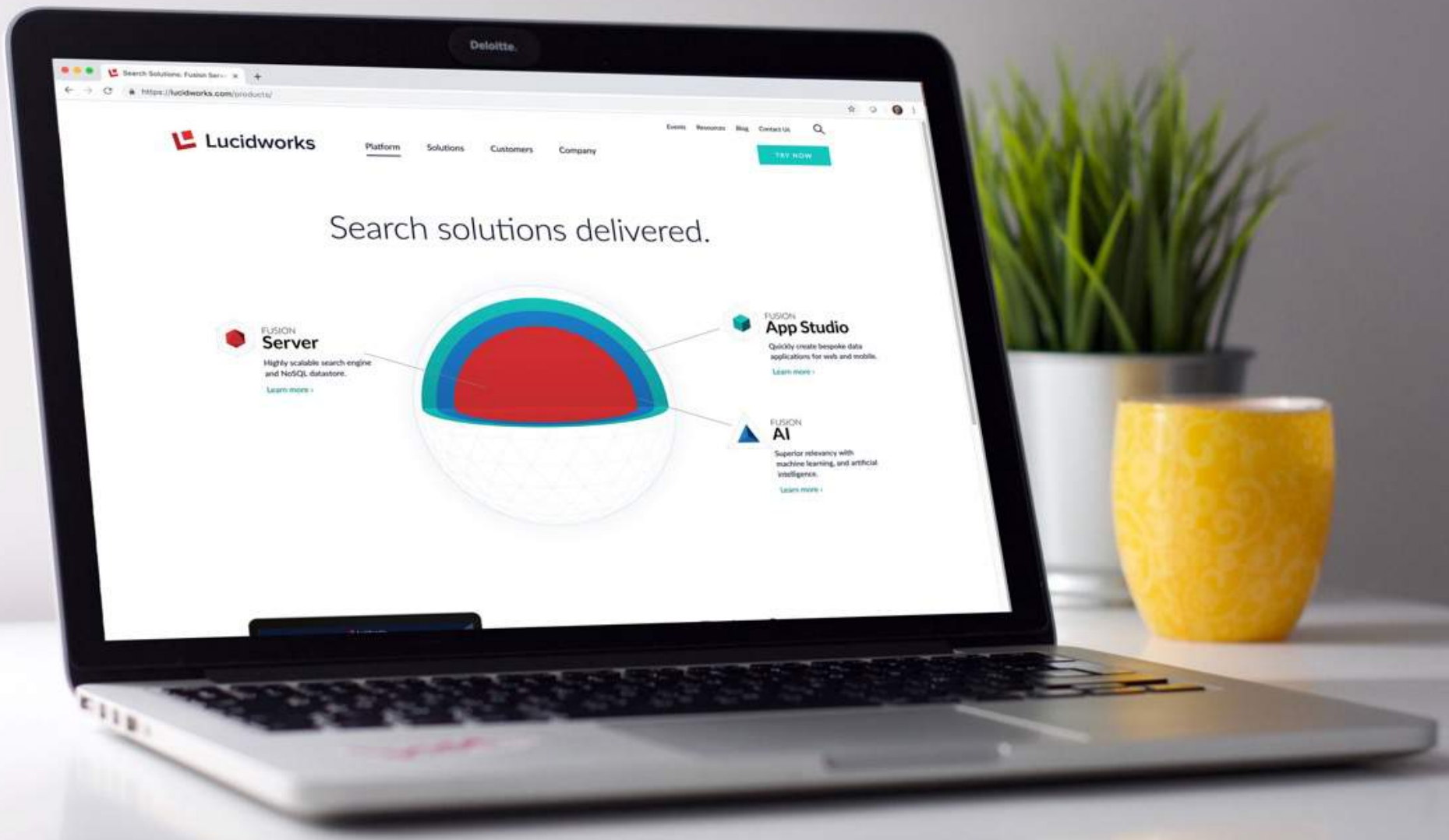
FUSION App Studio

Lucidworks

Proudly built with open-source tech at its core: Apache Solr & Apache Spark

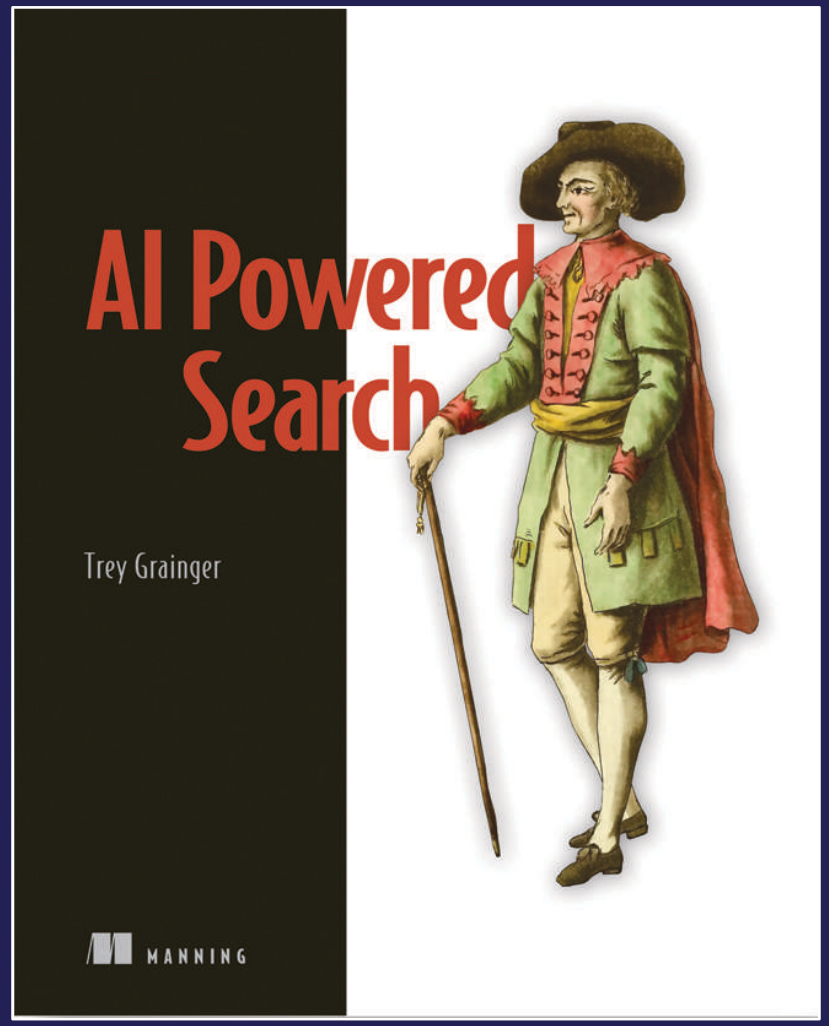
Personalizes search with applied machine learning

Proven on the world's biggest information systems



What is

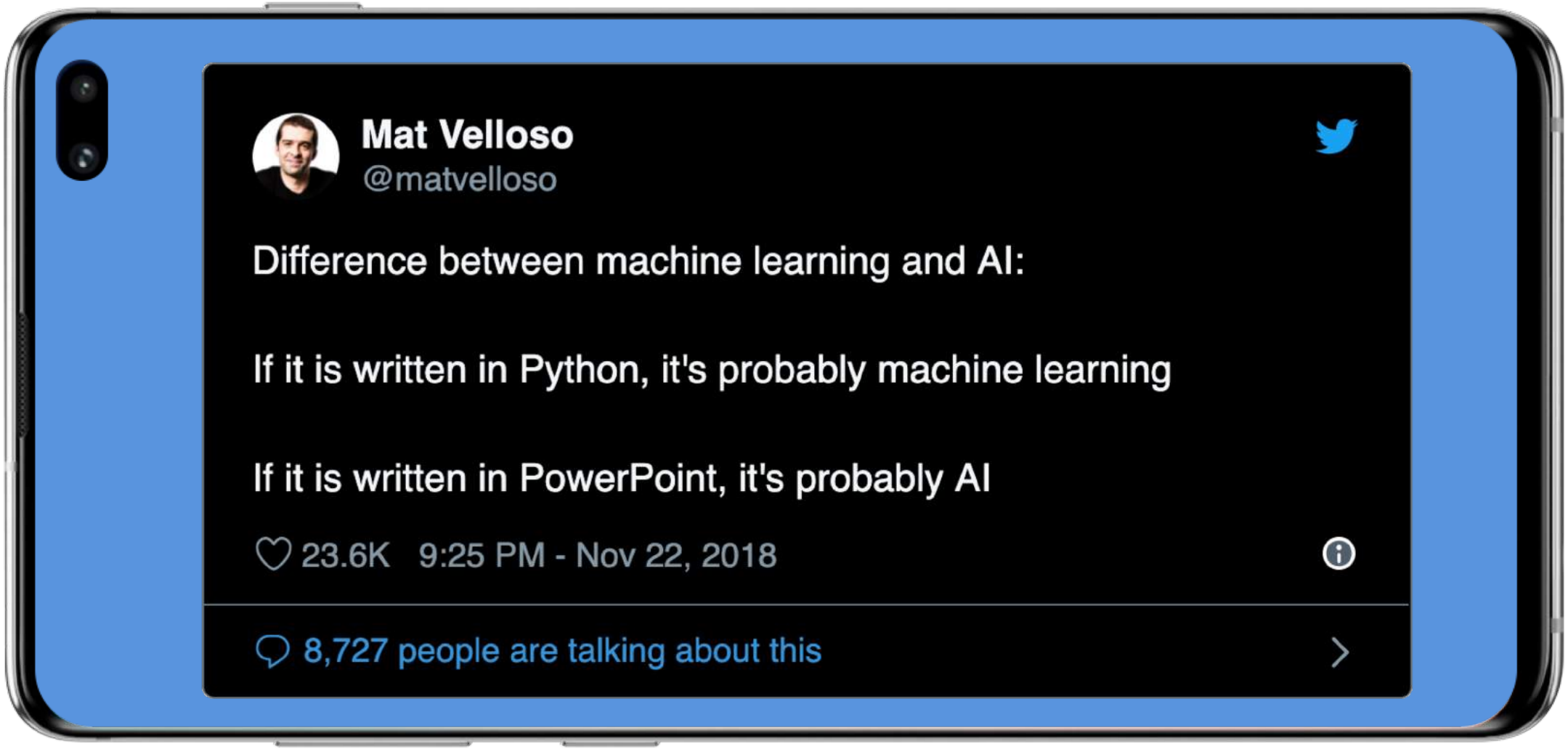
AI-Powered Search ?

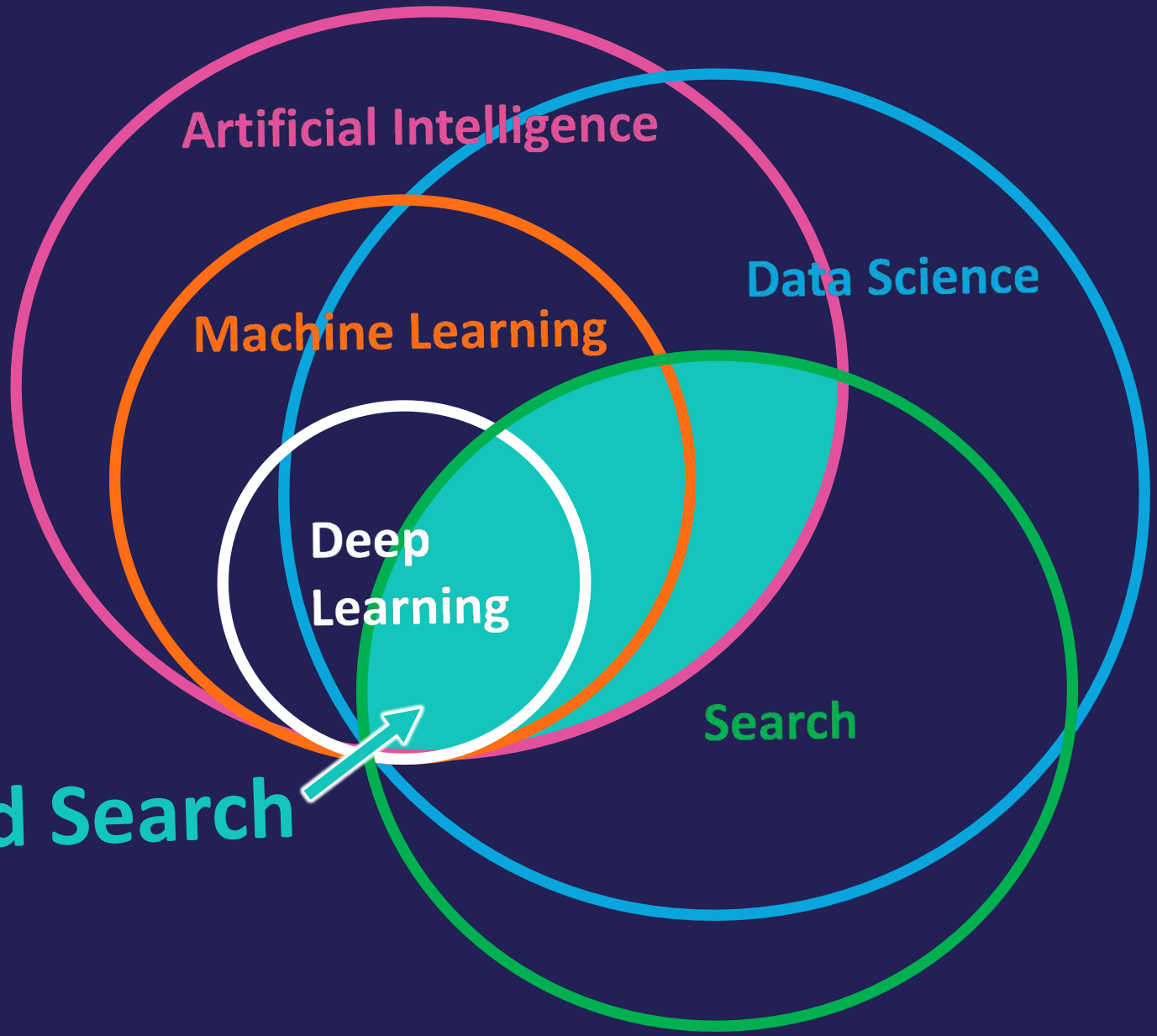


(Haystack discount code: **ctwhay19**)

… is my new **book!**

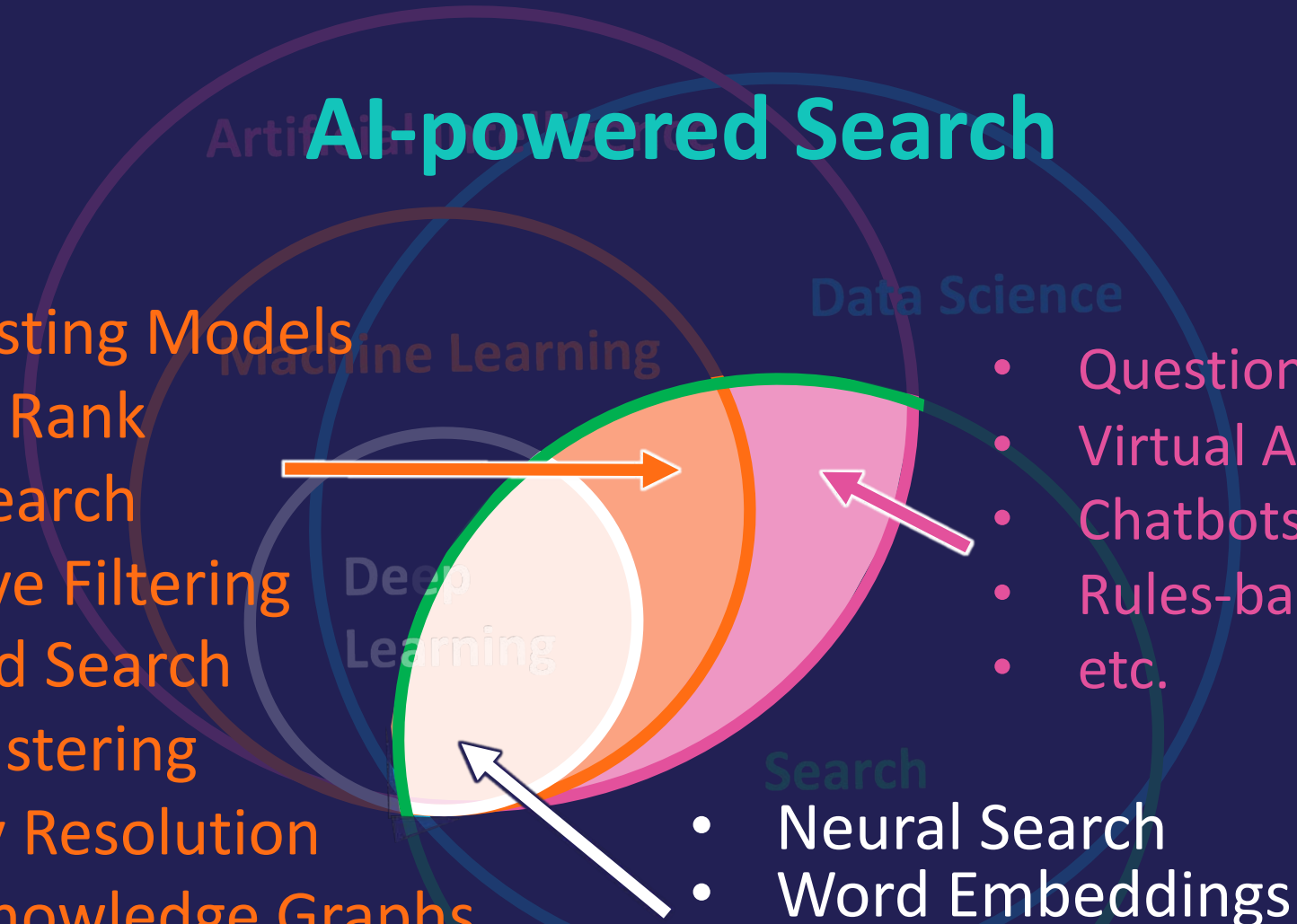
*http://***aiPoweredSearch***.com*


Mat Velloso
@matvelloso
Difference between machine learning and AI:
If it is written in Python, it's probably machine learning
If it is written in PowerPoint, it's probably AI
23.6K 9:25 PM - Nov 22, 2018
8,727 people are talking about this



AI-powered Search

AI-powered Search



- Signals Boosting Models
- Learning to Rank
- Semantic Search
- Collaborative Filtering
- Personalized Search
- Content Clustering
- NLP / Entity Resolution
- Semantic Knowledge Graphs
- Document Classification
- etc.

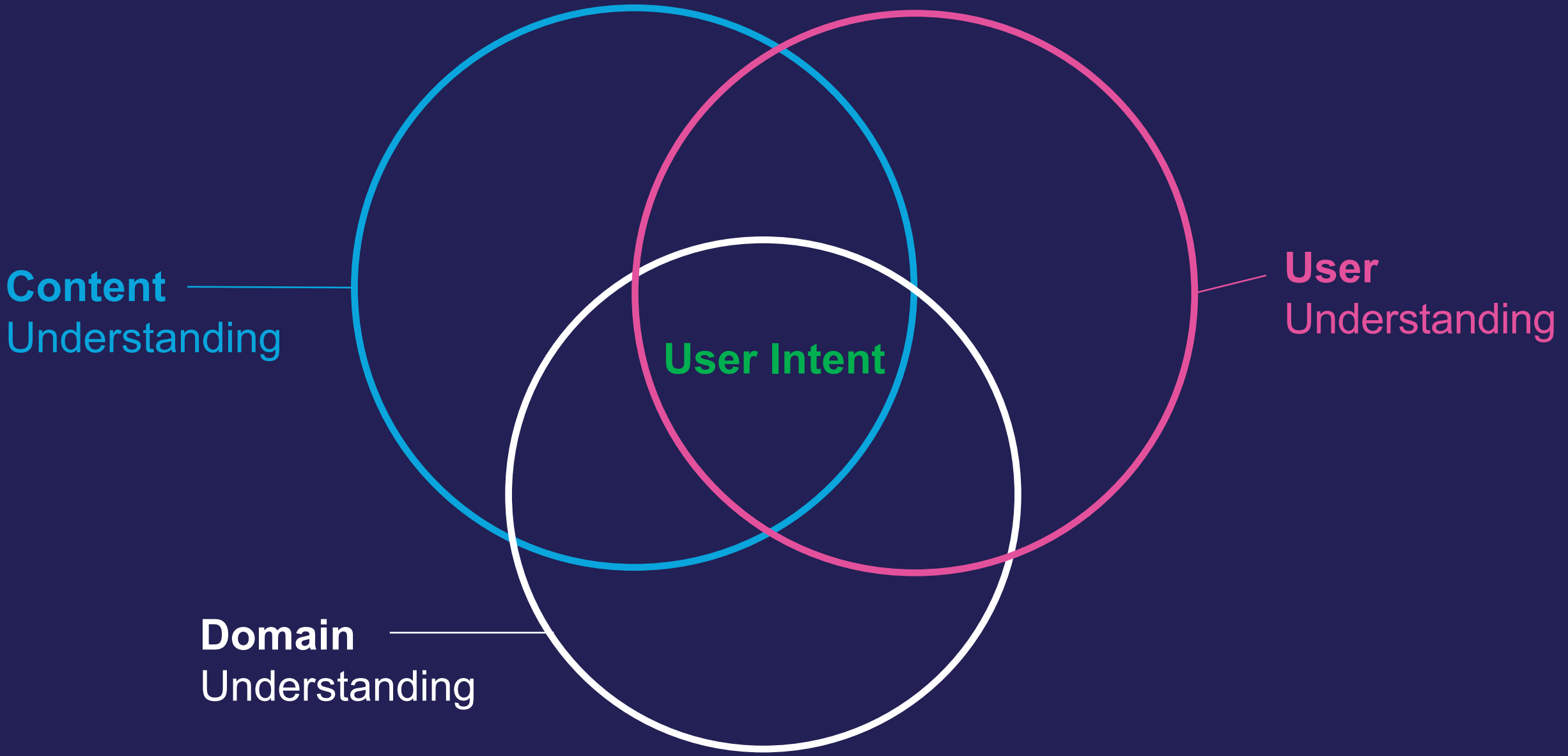
- Question / Answer Systems
- Virtual Assistants
- Chatbots
- Rules-based Relevancy
- etc.

- Neural Search
- Word Embeddings
- Vector Search
- Image / Voice Search
- etc.

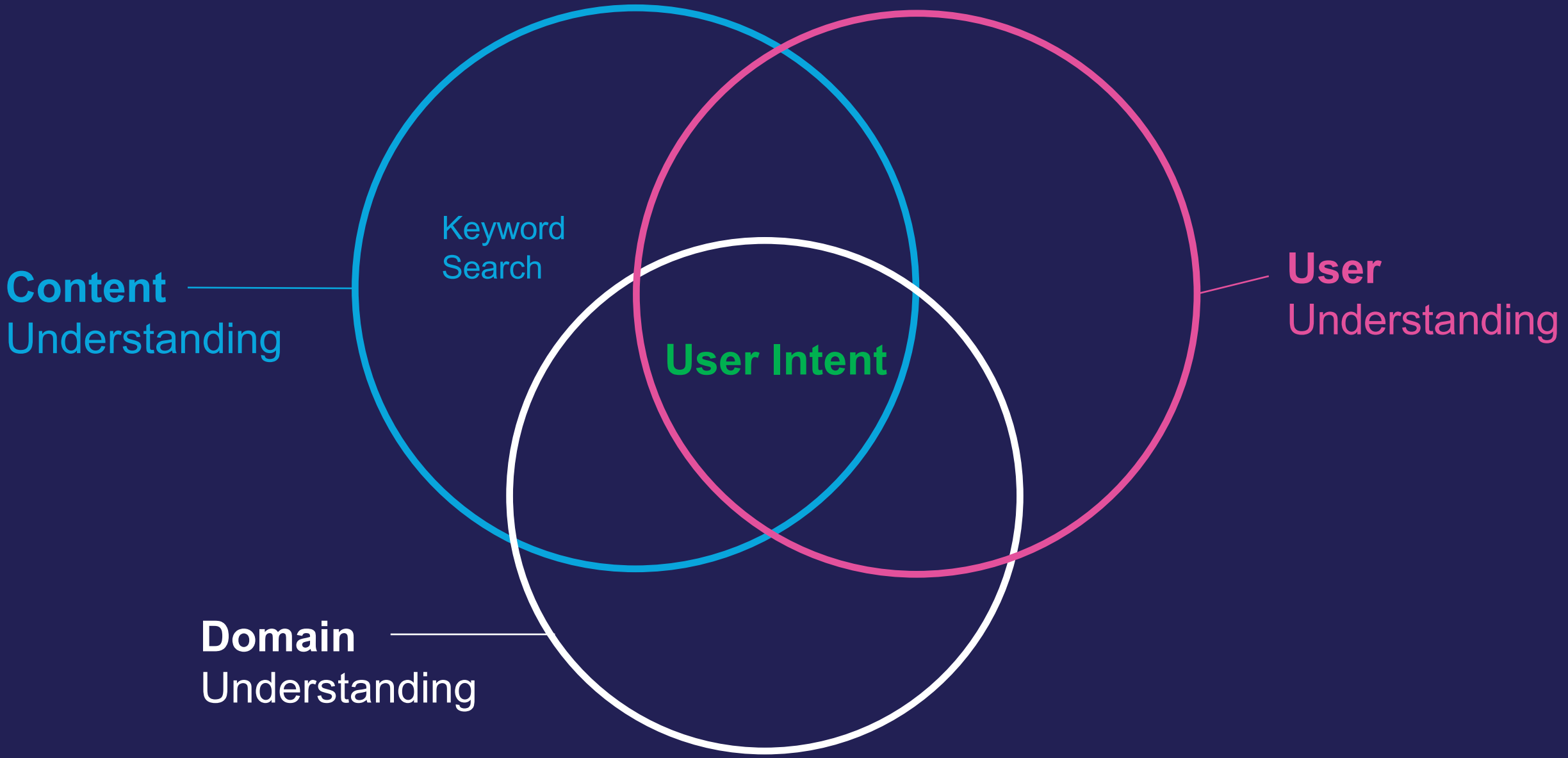
We have a big toolbox - great!

But how do we properly apply those tools?

Dimensions of User Intent



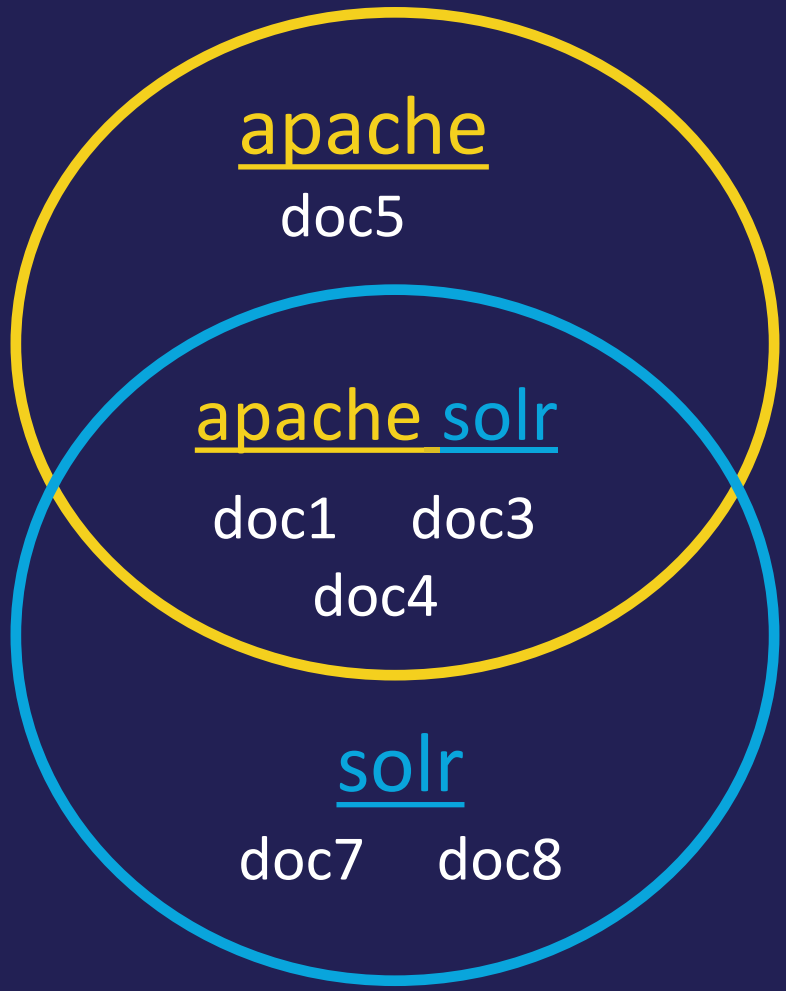
Dimensions of User Intent



Matching queries to documents

*/solr/collection/select/?q=*apache solr

Term	Documents
…	…
apache	doc1, doc3, doc4, doc5
…	
lucene	doc2, doc4, doc6
…	…
solr	doc1, doc3, doc4, doc7, doc8
…	…



BM25 (Relevance Scoring between Query and Documents)

$$\textbf{Score}(q, d) = \sum_{t \text{ in } q} \textbf{idf}(t) \cdot \textbf{(}\ \textbf{tf}(t \text{ in } d) \cdot (k + 1)\ \textbf{)} \,/\, (\ \textbf{tf}(t \text{ in } d) + k \cdot (1 - \textbf{b} + \textbf{b} \cdot |d| \,/\, \textbf{avgdl}\ )$$

Where:

t = term; **d** = document; **q** = query; **i** = index

$$\textbf{tf}(t \text{ in } d) = \text{numTermOccurrencesInDocument}\ ½$$

$$\textbf{idf}(t) = 1 + \log\,(\text{numDocs} \,/\, (\text{docFreq} + 1))$$

$$|d| = \sum_{t \text{ in } d} 1$$

$$\textbf{avgdl} = = (\ \sum_{d \text{ in } i} |d|\ )\,/\,(\ \sum_{d \text{ in } i} 1\ )\ )$$

k = Free parameter. Usually ~1.2 to 2.0. Increases term frequency saturation point.

b = Free parameter. Usually ~0.75. Increases impact of document normalization.

DIJIHUB

ipad

Home | Deal of the Day

Showing 1 - 12 of 2,690

Type

HardGood	2,558
Software	102
Bundle	30

Department

COMPUTERS	1,145
PHOTO/COMMODITIES	928
AUDIO	165
MAGNOLIA HOME THEATR	99
COMPUTER SOFTWARE	98
DTS	75
DIGITAL COMMUNICATIO	63
ACCESSORIES	48

Show more

Reg. Price

0.0 - 50.0 500.0

Format

DVD-ROM	83
CD-ROM	6

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Navy

SKU 5470056

$29.99

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Graphite

SKU 5470038

$29.99

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Lush Flower

SKU 5470047

$29.99

Computers & Tablets

Rocketfish™ - Bluetooth Speaker for Apple® iPad®, iPad 2 and iPad (3rd Generation)

SKU 3640066

Computers & Tablets

Harley-Davidson - Envelope Case for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Black

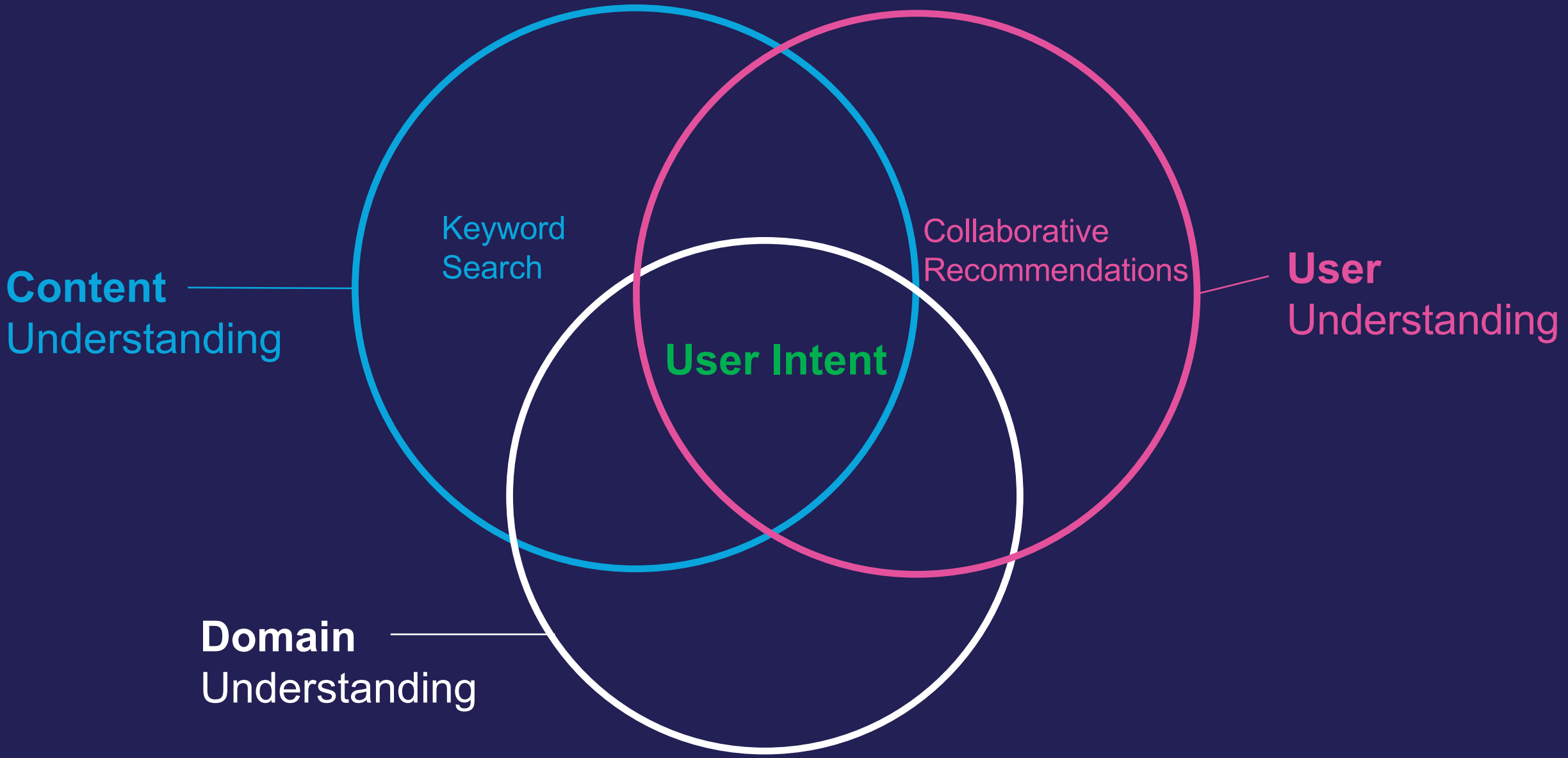
SKU 5422949

Computers & Tablets

Forward Industries - G-Form Extreme Sleeve for Apple® iPad®, iPad 2 and the new iPad

SKU 4979513

Dimensions of User Intent



DIJIHUB

Search...

Home | Deal of the Day

Computers & Tablets

Apple® - iPad® 2 with Wi-Fi - 64GB - Black

9.7" widescreen display; 802.11a/b/g/n Wi-Fi; Bluetooth; iBooks support; measures just 0.34" thin and weighs only 1.35 lbs.

Shop Now

Trey, recommended for you!

TV & Home Theater

LG - 47" Class / LED / 1080p / 120Hz / 3D / HDTV 3D Blu-ray Player Bundle

SKU 2970264

$1099.99

TV & Home Theater

LG - 47" Class - LCD - 1080p - 120Hz - HDTV

SKU 2138282

$999.99

TV & Home Theater

LG - 55" Class / LED / 1080p / 120Hz / 3D / HDTV 3D Blu-ray Player Bundle

SKU 2970291

$1599.99

TV & Home Theater

LG - 55" Class / 1080p / 120Hz / LCD HDTV

SKU 2138291

$1499.99

Collaborative Filtering (Recommendations)

Recommendations for Alonzo:
- **doc22**: "iPad Pro"
- **doc12**: "Kindle Fire"
- …

ipad

User Searches

User Sees Results

User takes an action

Users' actions inform system improvements

User	Query
Alonzo	**ipad**
Elena	**printer**
Ming	**ipad**
…	…

User	Item	Weight
Alonzo	**doc22**	**1.0**
Alonzo	**doc12**	**0.4**
…	…	…
Ming	doc12	0.9
Ming	doc22	0.6
…	…	…

Matrix Factorization

User	Action	Document
Alonzo	**click**	doc22
Elena	**click**	doc17
Ming	**click**	doc12
Alonzo	**purchase**	doc22
Ming	**click**	doc22
Ming	**purchase**	doc12
Elena	**click**	doc2
…	…	…

Recommendations (User-Item, Item-Item, Query-Item)

Recommendations for Alonzo:
- **doc22**: "iPad Pro"
- **doc12**: "Kindle Fire"

…

User	Item	Weight
Alonzo	doc22	1.0
Alonzo	doc12	0.4
…	…	…
Ming	doc12	0.9
Ming	doc22	0.6
…	…	…

Recommendations for Doc22:
- **doc22**: "iPad Pro"
- **doc12**: "Kindle Fire"

…

Item	Item	Weight
doc22	doc22	1.0
doc22	doc12	0.85
…	…	…
doc12	doc12	1.0
doc12	doc22	0.83
…	…	…

Recommendations for "ipad":
- **doc22**: "iPad Pro"
- **doc12**: "Kindle Fire"

…

Query	Item	Weight
ipad	doc22	0.98
ipad	doc12	0.6
…	…	…
kindle	doc12	0.96
apple	doc22	0.90
…	…	…

Matrix Factorization

DIJIHUB

ipad

Home | Deal of the Day

Showing 1 - 12 of 2,690

Type

HardGood	2,558
Software	102
Bundle	30

Department

COMPUTERS	1,145
PHOTO/COMMODITIES	928
AUDIO	165
MAGNOLIA HOME THEATR	99
COMPUTER SOFTWARE	98
DTS	75
DIGITAL COMMUNICATIO	63
ACCESSORIES	48

Show more

Reg. Price

0.0 - 50.0 500.0

Format

DVD-ROM	83
CD-ROM	6

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Navy

SKU 5470056

$29.99

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Graphite

SKU 5470038

$29.99

Computers & Tablets

Built NY - Slim Sleeve for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Lush Flower

SKU 5470047

$29.99

Computers & Tablets

Rocketfish™ - Bluetooth Speaker for Apple® iPad®, iPad 2 and iPad (3rd Generation)

SKU 3640066

Computers & Tablets

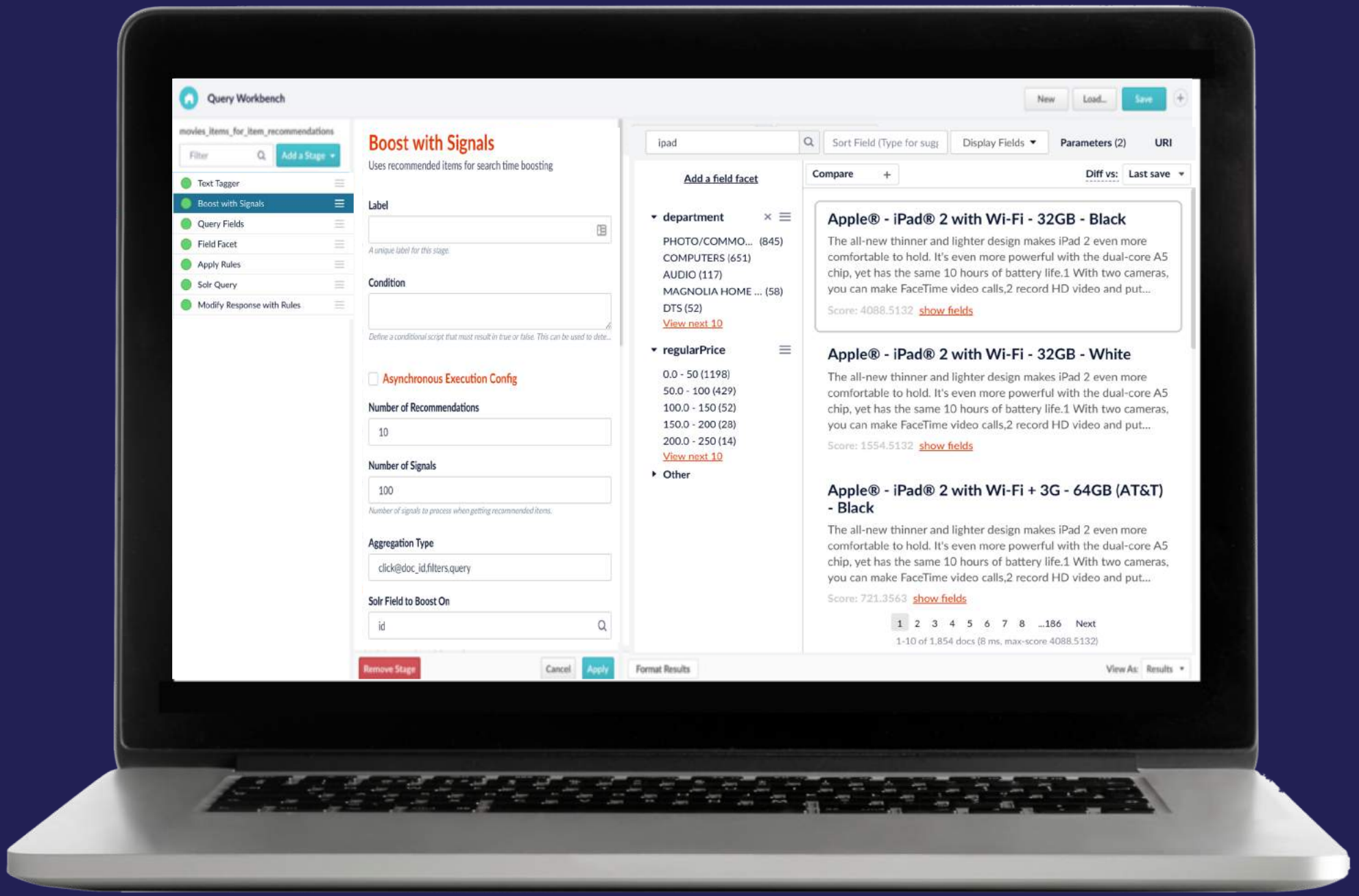
Harley-Davidson - Envelope Case for Apple® iPad®, iPad 2 and iPad (3rd Generation) - Black

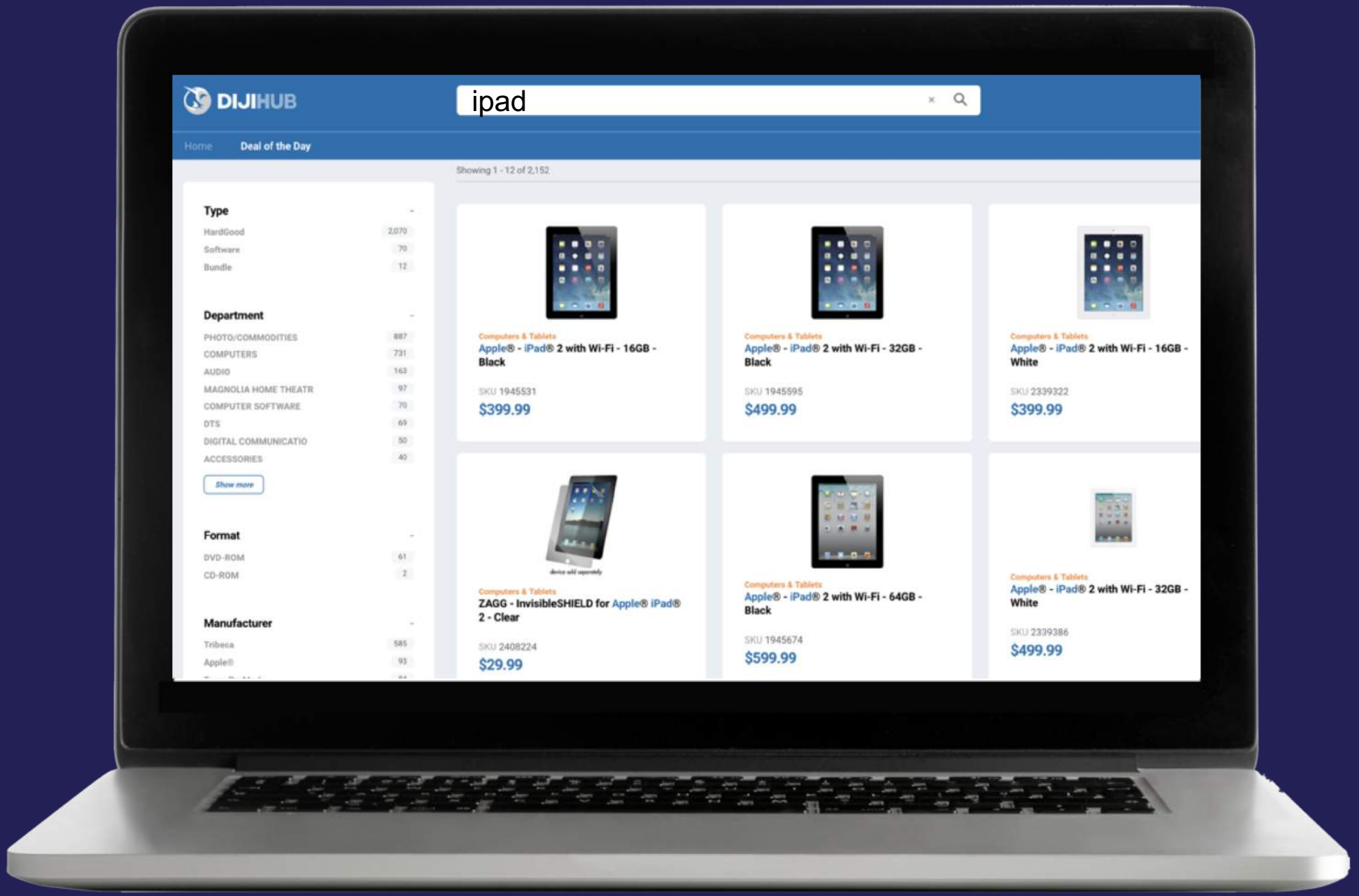
SKU 5422949

Computers & Tablets

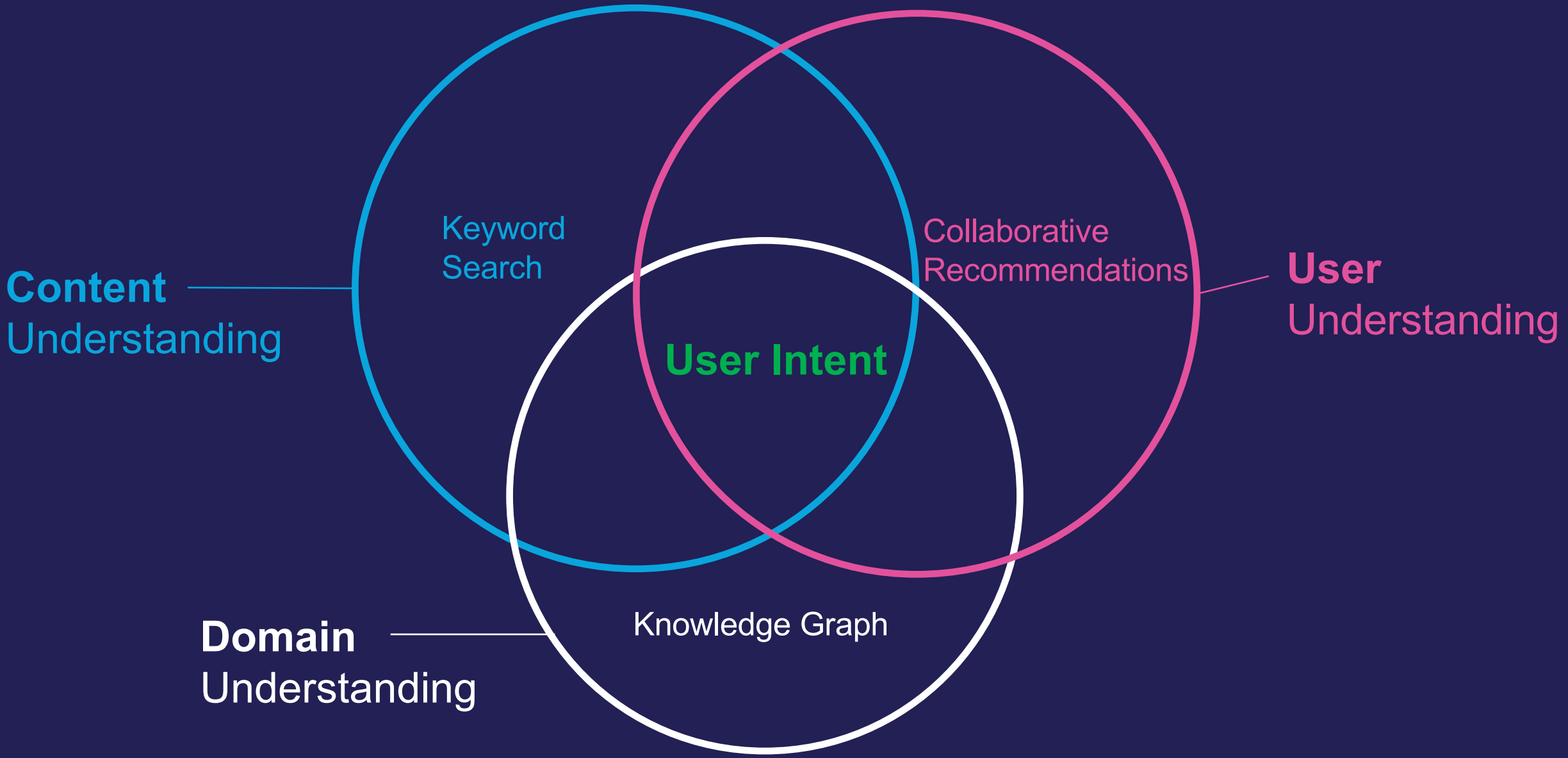
Forward Industries - G-Form Extreme Sleeve for Apple® iPad®, iPad 2 and the new iPad

SKU 4979513


Query Workbench
New
Load...
Save
movies_items_for_item_recommendations
Filter
Add a Stage
Text Tagger
Boost with Signals
Query Fields
Field Facet
Apply Rules
Solr Query
Modify Response with Rules
Boost with Signals
Uses recommended items for search time boosting
Label
A unique label for this stage.
Condition
Define a conditional script that must result in true or false. This can be used to dete...
Asynchronous Execution Config
Number of Recommendations
10
Number of Signals
100
Number of signals to process when getting recommended items.
Aggregation Type
click@doc_id,filters,query
Solr Field to Boost On
id
Remove Stage
Cancel
Apply
Format Results
ipad
Sort Field (Type for sug
Display Fields
Parameters (2)
URI
Add a field facet
Compare
Diff vs:
Last save
department
PHOTO/COMMO... (845)
COMPUTERS (651)
AUDIO (117)
MAGNOLIA HOME ... (58)
DTS (52)
View next 10
regularPrice
0.0 - 50 (1198)
50.0 - 100 (429)
100.0 - 150 (52)
150.0 - 200 (28)
200.0 - 250 (14)
View next 10
Other
Apple® - iPad® 2 with Wi-Fi - 32GB - Black
The all-new thinner and lighter design makes iPad 2 even more comfortable to hold. It's even more powerful with the dual-core A5 chip, yet has the same 10 hours of battery life.1 With two cameras, you can make FaceTime video calls,2 record HD video and put...
Score: 4088.5132 show fields
Apple® - iPad® 2 with Wi-Fi - 32GB - White
The all-new thinner and lighter design makes iPad 2 even more comfortable to hold. It's even more powerful with the dual-core A5 chip, yet has the same 10 hours of battery life.1 With two cameras, you can make FaceTime video calls,2 record HD video and put...
Score: 1554.5132 show fields
Apple® - iPad® 2 with Wi-Fi + 3G - 64GB (AT&T) - Black
The all-new thinner and lighter design makes iPad 2 even more comfortable to hold. It's even more powerful with the dual-core A5 chip, yet has the same 10 hours of battery life.1 With two cameras, you can make FaceTime video calls,2 record HD video and put...
Score: 721.3563 show fields
1 2 3 4 5 6 7 8 ...186 Next
1-10 of 1,854 docs (8 ms, max-score 4088.5132)
View As: Results


DIJIHUB
ipad
Home
Deal of the Day
Showing 1 - 12 of 2,152
Type
HardGood 2,070
Software 70
Bundle 12
Department
PHOTO/COMMODITIES 887
COMPUTERS 731
AUDIO 163
MAGNOLIA HOME THEATR 97
COMPUTER SOFTWARE 70
DTS 69
DIGITAL COMMUNICATIO 50
ACCESSORIES 40
Show more
Format
DVD-ROM 61
CD-ROM 2
Manufacturer
Tribeca 585
Apple® 93
Computers & Tablets
Apple® - iPad® 2 with Wi-Fi - 16GB - Black
SKU 1945531
$399.99
Computers & Tablets
Apple® - iPad® 2 with Wi-Fi - 32GB - Black
SKU 1945595
$499.99
Computers & Tablets
Apple® - iPad® 2 with Wi-Fi - 16GB - White
SKU 2339322
$399.99
Computers & Tablets
ZAGG - InvisibleSHIELD for Apple® iPad® 2 - Clear
SKU 2408224
$29.99
Computers & Tablets
Apple® - iPad® 2 with Wi-Fi - 64GB - Black
SKU 1945674
$599.99
Computers & Tablets
Apple® - iPad® 2 with Wi-Fi - 32GB - White
SKU 2339386
$499.99

Dimensions of User Intent



What is a Knowledge Graph?

(vs. Ontology vs. Taxonomy vs. Synonyms, etc.)

Overly Simplistic Definitions

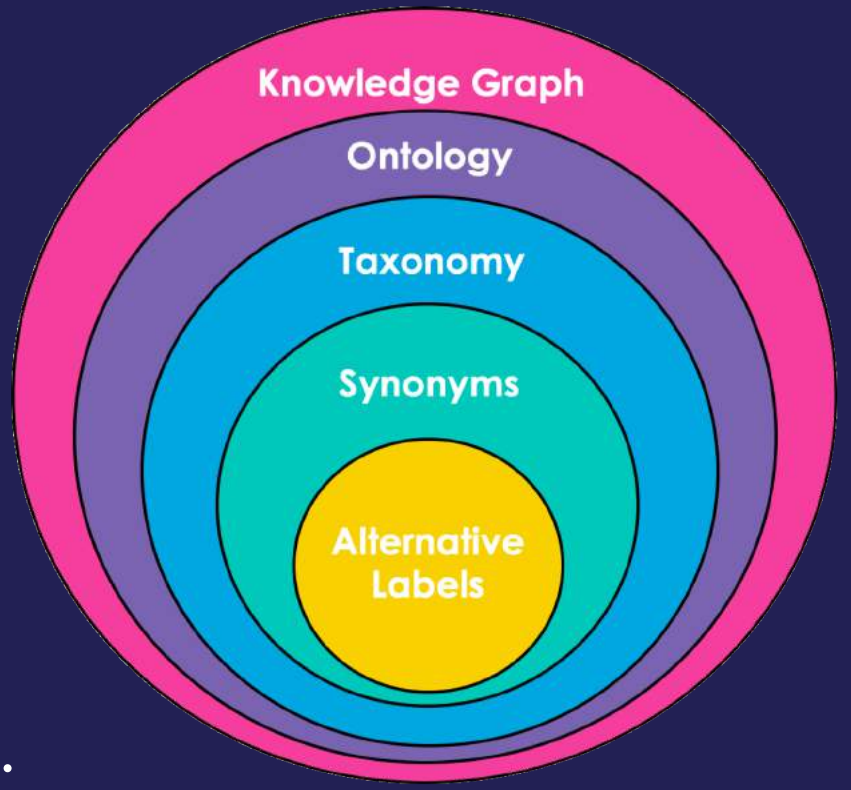
Alternative Labels: Substitute words with identical meanings
[CTO => Chief Technology Officer; specialise => specialize]

Synonyms List: Provides substitute words that can be used to represent the same or very similar things
[human => homo sapien, mankind; food => sustenance, meal]

Taxonomy: Classifies things into Categories
[john is Human; Human is Mammal; Mammal is Animal]

Ontology: Defines relationships between types of things
[animal eats food; human is animal]

Knowledge Graph: Instantiation of an Ontology (contains the things that are related)
[john is human; john eats food]



A Knowledge Graph subsumes the other types.

Enterprise Search

Theo

localhost:8080/#/search?q=gartner

Lucidworks

Find documents, people, departments...

Search

DATE MODIFIED

2012 2018

DEPARTMENT

marketing 1

EXPERTISE

Search...

- case studies 2
- fusion 2
- gartner 2
- it 2
- use cases 2
- analysts 1
- cloud 1
- collateral 1
- communications 1
- documentation 1

Show more

FILE FORMAT

Google Docs 19

All 84 | GoogleDrive 82 | Q&A 0 | Documentation 0 | Website 0 | People 2

Sort by relevance

SHOWING 1 - 10 OF 84 gartner ×

Andy Wibbels Offline

Dir of Marketing

Department marketing

Expertise cloud, documentation, fusion, it, use cases, case studies

Gartner Slide.key

Author will.hayes@lucidworks.com, Modified 12 Oct 2016, 18:51

Source GoogleDrive

Gartner Critical Capcabilities

GARTNER CRITICAL CAPABILITIES Operational Database Management Systems (ODBMS) · High

Author alea.abed@lucidworks.com, Modified 13 Nov 2017, 21:11

Source GoogleDrive

Gartner Resource Planning

Author john.underwood@lucidworks.com, Modified 10 Jan 2018, 21:56

Source GoogleDrive

Gartner Go To Market Notes

Call with Hank Barnes, Gartner GTM Analyst - 11/15/17 Hank's notes: The core thing that

Author alea.abed@lucidworks.com, Modified 14 Dec 2017, 23:37

Source GoogleDrive

Looking for marketing content?

We've brought all marketing documents, conversations and experts together to help you find your answer...

Gartner

Gartner, Inc. is an American research and advisory firm providing information technology related insight for IT and other business leaders located across the world. Its headquarters are in Stamford, Connecticut, United States. It was known as Gartner Group, Inc until 2000, when it was then changed ...

Industry Research

Type company

From Wikipedia, the free encyclopedia

Slack Conversations

Laurel Gaddie view in slack

Enterprise Search

Theo

localhost:8080/#/department/marketing

Lucidworks

Find documents, people, departments...

Search

Marketing –

30 Members

4000 Files

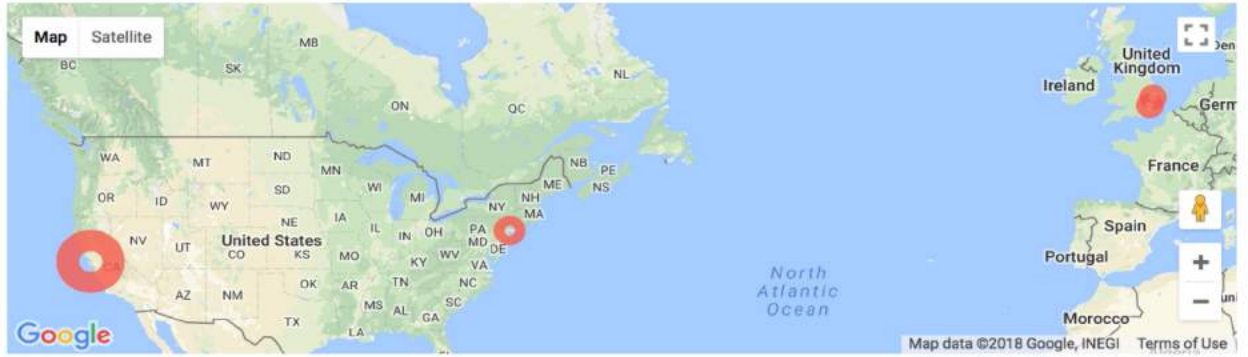
7 Channels

Team skills

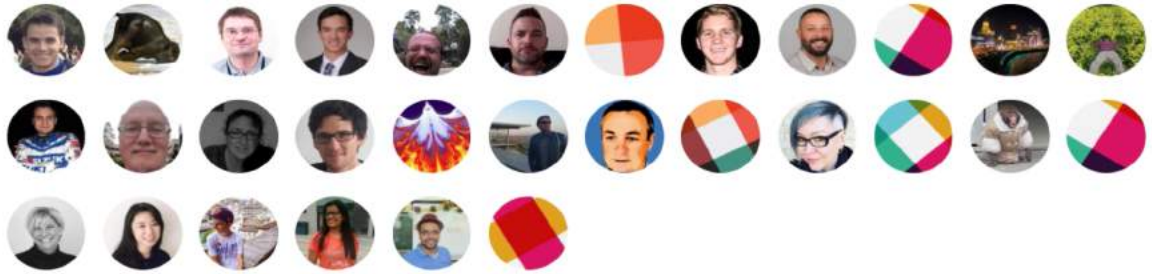
Search...

- it 16
- documentation 13
- fusion 13
- das 8
- null 7
- build 6
- security 6
- cloud 5
- testing 5
- connectors 4
- insights 4
- machine learning 4
- scala 4
- spark 4
- sales 3

Show more



Team members



Latest marketing documents

see all ›

Case Study: Structured Data
updated 2 days ago

Handling archive deletions in Connectors SDK
updated 6 days ago

PowerPoint Presentation
updated 6 days ago

Gartner MQ 2018 Process [INTERNAL ONLY]

Leadership

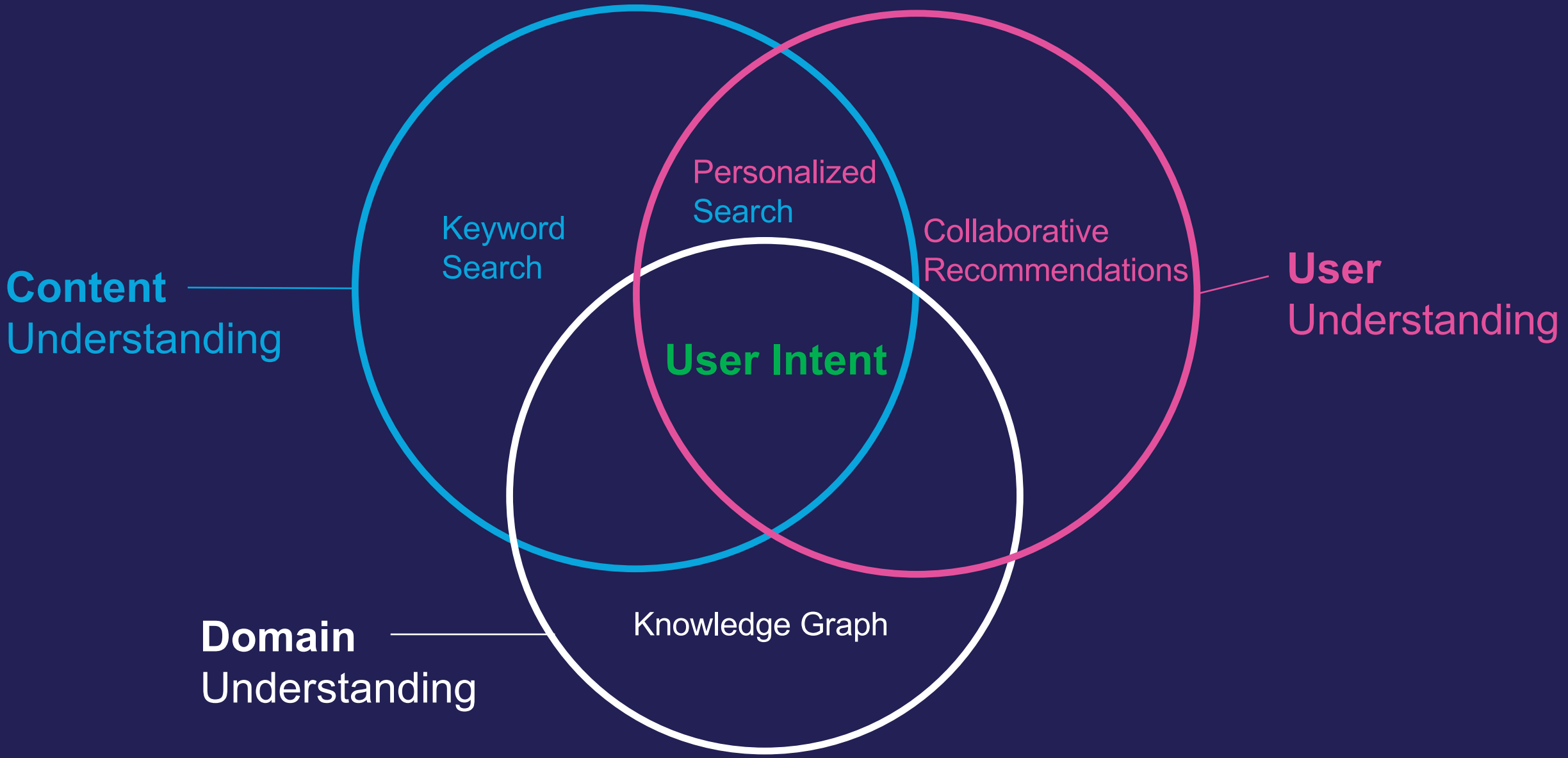
Curt Janka
Creative Director

John Western
Sales Director, EMEA

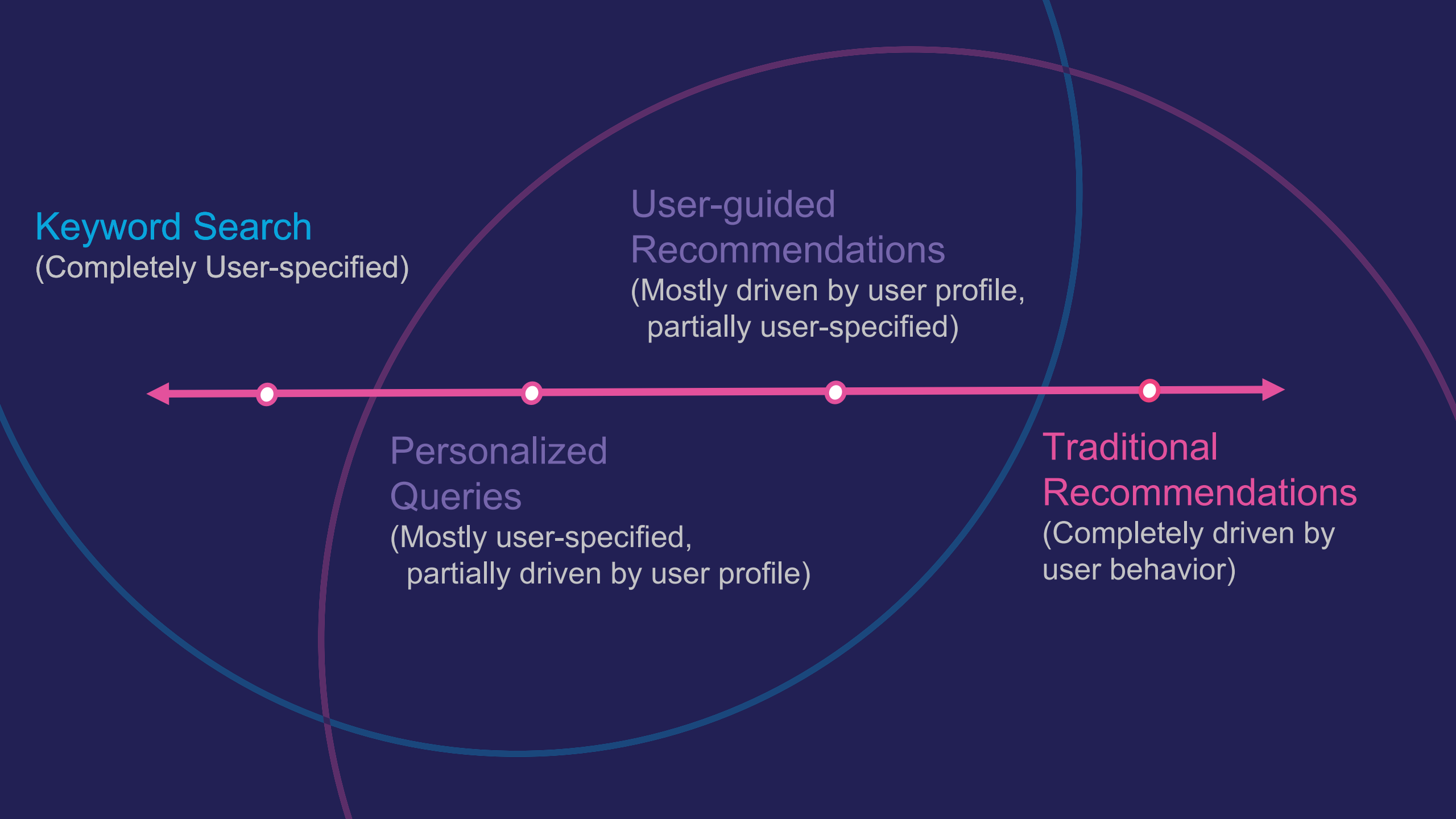
Slack Channels

- **fusion-help** 6233 messages
- **fusion-dev** 3534 messages
- **documentation** 2743 messages
- **fusion-build** 1858 messages
- **marketing-discussion** 1354 messages
- **general** 962 messages
- **training_development** 948 messages

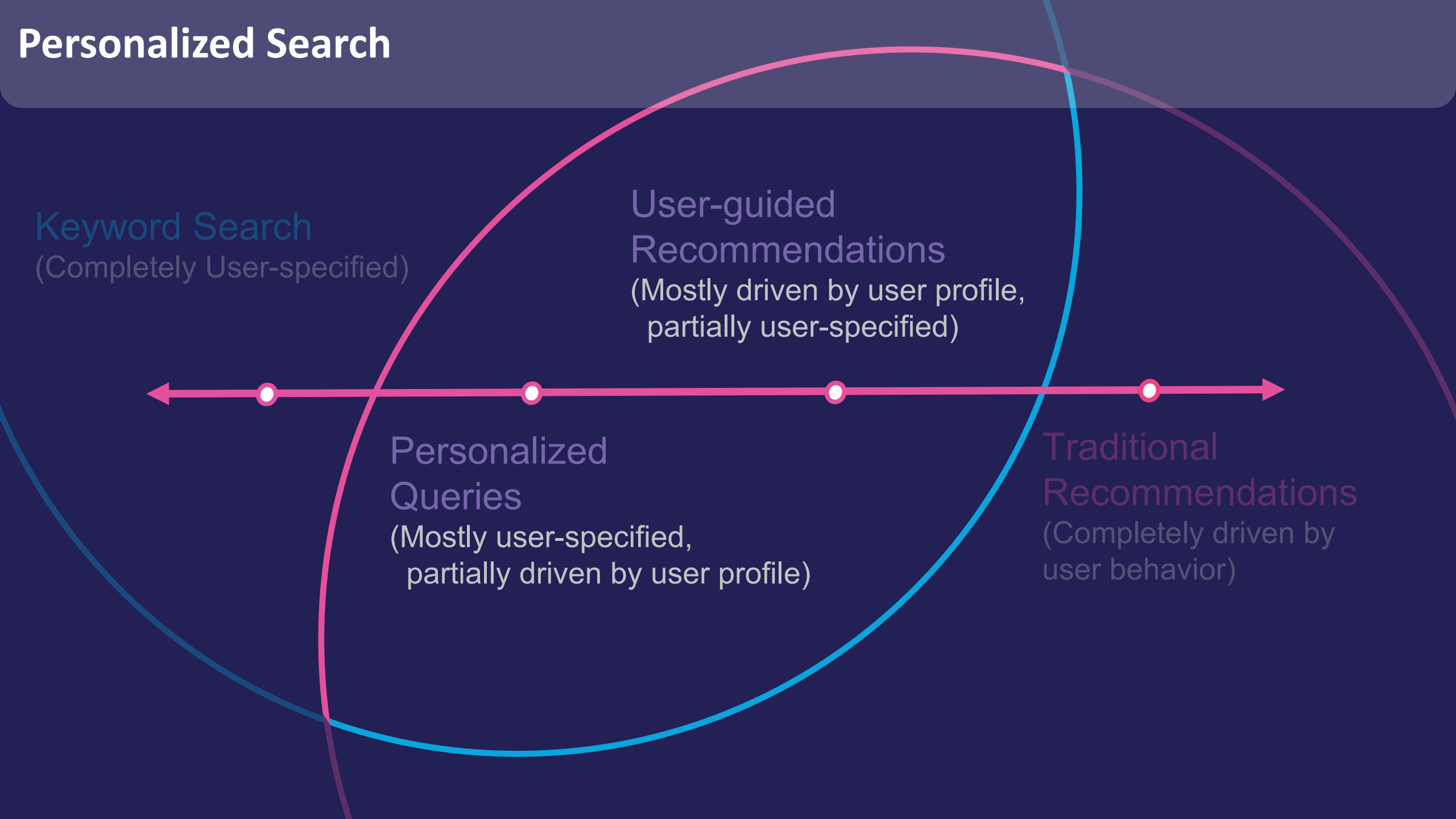
Dimensions of User Intent




Keyword Search
(Completely User-specified)
Traditional Recommendations
(Completely driven by user behavior)


Keyword Search
(Completely User-specified)
User-guided Recommendations
(Mostly driven by user profile, partially user-specified)
Personalized Queries
(Mostly user-specified, partially driven by user profile)
Traditional Recommendations
(Completely driven by user behavior)

Personalized Search



DIJIHUB

Search... Search Clear

Trends | Search | Search with Query Rewrite | Search with Signals Aggr | Compare | Personalization Aggregation

Personalization

User Details for 008f7b53f33aaeb279f13d9c222edca9610c0cb6

Launch Personalized Search Page

User Click History | Items for User | Personalization Aggregation

Image	Timestamp	Query	Product	Doc Id
	2018-10-23T08:53:06.400Z	samsung refrigerator	Samsung - 25.6 Cu. Ft. Side-by-Side Refrigerator with Thru-the-Door Ice and Water - Stainless-Steel	9747183
	2018-10-23T08:50:28.420Z	samsung refrigerator	Samsung - 25.5 Cu. Ft. Side-by-Side Refrigerator with Thru-the-Door Ice and Water - Stainless-Steel	9791769
	2018-10-23T08:45:12.490Z	samsung range	Samsung - 30" Self-Cleaning Freestanding Electric Convection Range - Stainless-Steel/Black	9212504
	2018-10-23T08:44:09.910Z	samsung range	Samsung - 30" Self-Cleaning Freestanding Electric Convection Range - Stainless-Steel	9212498
	2018-10-06T16:08:07.830Z	samsung dishwasher	Samsung - 24" Tall Tub Built-In Dishwasher - Stainless-Steel	9980455
	2018-10-04T15:57:40.500Z	lg range electric	LG - 30" Self-Cleaning Freestanding Electric Range - Stainless-Steel	2429379
	2018-10-04T15:56:10.880Z	lg range electric	LG - 30" Self-Cleaning Freestanding Double Oven Electric Range - Stainless-Steel	2428282

User Details for 008f7b53f33aaeb279f13d9c222edca9610c0cb6

Launch Personalized Search Page

User Click History | Items for User | Personalization Aggregation

User Id	Department	Brand	Color	Count
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance		Stainless-Steel	22
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	Samsung		9
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	KitchenAid		6
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	LG		5
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance		Monochromatic Stainless-Steel	3
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	Bosch		2
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	Frigidaire		2
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	Whirlpool		1
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Magnolia Home Theatr	Russound		1
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	GE		1
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance	Electrolux		1
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance		Stainless-Steel/Black	1
008f7b53f33aaeb279f13d9c222edca9610c0cb6	Appliance		Black/Stainless-Steel	1

All results loaded

DIJIHUB

microwave

Search

Clear

Trends Search Search with Que

Search for "microwave"

PRODUCTS


GE - 0.7 Cu. Ft. Compact Microwave - Black GE


Hello Kitty - 0.7 Cu. Ft. Compact Microwave Hello Kitty


GE - 1.1 Cu. Ft. Mid-Size Microwave - White GE

SUGGESTED SPELLINGS

panasonic microwave Phrase => Phrase

frigidaire microwave Phrase => Phrase

ge microwave Phrase => Phrase

SEARCH HISTORY

microwave

SUGGESTED WORDS/PHRASES

microwave

microwave stainless

microwave black

microwave stainless steel

microwave white

microwave package

microwave ovens ge over the range microwave

Sort by relevance

Brand -

Search...

GE 293
Whirlpool 149
Sharp 101
Frigidaire 88
KitchenAid 78
LG 63
Panasonic 55
Maytag 38
Samsung 33
Electrolux 22
Goldstar 16
Bodum 13
Show more

Department -

Search...

Appliance 1,030
Computers 9
Interactive Software 3
Mobile Audio 3
Accessories 2
Photo/Commodities 2

Ft. Compact Microwave

ts in this microwave that
mable recipes for customized
sign for optimal placement.
ties Brand Hello Kitty

ID 9419729
$59

Score 1034.6353

DIJIHUB

microwave × Search Clear

Trends | Search | Search with Query Rewrite | **Search with Signals Aggr** | Compare | Personalization Aggregation

Brand -

Search...

GE 293
Whirlpool 149
Sharp 101
Frigidaire 88
KitchenAid 78
LG 63
Panasonic 55
Maytag 38
Samsung 33
Electrolux 22
Goldstar 16
Bodum 13
Show more

Department -

Search...

Appliance 1,030
Computers 9
Interactive Software 3
Mobile Audio 3
Accessories 2
Photo/Commodities 2

Regular Search Results:

GE - 0.7 Cu. Ft. Compact Microwave - Black

Cook or reheat your meals in minutes with this compact 700-watt microwave oven that features instant-on LED controls which heat food with the simple touch of a button. It also provides peace of mind with a control lockout feature that locks the microwave when it's not in use.

Dept **Appliance** Brand **GE** Color **Black**

ID **9419729**
$59
Score 1373.8041

Hello Kitty - 0.7 Cu. Ft. Compact Microwave

Cook your favorite treats in this microwave that features 6 preprogrammable recipes for customized use and a compact design for optimal placement.

Dept **Photo/Commodities** Brand **Hello Kitty**

ID **2817088**
$89
Score 1034.6353

GE - 1.1 Cu. Ft. Mid-Size Microwave - White

Prepare hassle-free meals and snacks with this

Personalized Search Results:

User: **008f7b53f33aaeb279f13d9c222edca9610c0cb6**

Samsung - Mid-Size Microwave - Stainless-steel

This mid-size microwave is large enough for almost any meal, but compact enough to fit into almost any space. Take an interactive product tour of this Samsung Microwave! (479KB Flash demo)

Dept **Appliance** Brand **Samsung** Color **Stainless-Steel**

ID **4564643**
$58
Score 1689.5975

Samsung - 1.7 Cu. Ft. Over-the-Range Microwave - Stainless-Steel

The capabilities of a toaster oven, convection oven, range hood and microwave combine in this microwave that features a 1.7 cu. ft. capacity for simultaneously cooking multiple dishes and sensor cook options to take the guesswork out of quick meal prep.

Dept **Appliance** Brand **Samsung** Color **Stainless-**

Nice - personalization is awesome!

Let’s roll it out everywhere!

DIJIHUB

iphone × Clear

Trends | Search | Search with Query Rewrite | Search with Signals Aggr | Compare | Personalization Aggregation

Brand

Search...

- Tribeca 180
- OtterBox 97
- Incase 86
- Apple 85
- Belkin 59
- Griffin Technology 57
- IFrogz 52
- IHome 45
- Pangea Brands 44
- Sony 37
- Mophie 32
- Rocketfish Mobile 31

Show more

Department

Search...

- Digital Communication 690
- Photo/Commodities 441
- Audio 338
- Mobile Audio 101
- Computers 76
- DTS 55
- Computer Software 45
- Magnolia Home Theatr 31
- Accessories 28
- Interactive Software 27
- Video 3

Color

Suggested Searches

iphone unlock | laguna beach | unlocked iphone | i phone | iphone4 unlocked | iphone verizon | bodyguardz | pantech breeze | iphone refurbished | refurbished iphone

iphone ×

SHOWING 1 - 10 OF 1,863 • Debug

Sort by relevance

1 2 3 4 5 next

Ugh...

Samsung - Audio Dock for Apple® iPod® and iPhone®

Listen to tunes on your Apple iPod or iPhone or compatible Samsung Galaxy mobile phone with this Samsung DAE570ZA audio dock that delivers 10W total system power for clear sound. Bluetooth technology makes it easy to wirelessly connect compatible devices.

Dept Audio Brand Samsung

ID 5563122

$179

Score 5.7202854

Samsung - 840W 7.2-Ch. A/V Home Theater Receiver

This powerful 7.2-channel home theater receiver features 12 DSP soundfields for a lush soundscape. An Apple® iPod® and iPhone® dock lets you listen to your favorite music.

Dept Audio Brand Samsung

ID 1151846

$499

Score 5.5088468

Samsung - Audio Dock for Apple® iPod®, iPhone® and iPad®

Rock out to your favorite tunes with this Samsung DAE750ZA dock that features woven glass fiber speakers and a down-firing subwoofer for powerful audio. The dual docking system allows you to listen to tunes on your Apple iPod, iPhone, iPad and more.

Dept Audio Brand Samsung

ID 5563159

$699

Score 5.3621306

Samsung - Audio Dock for Apple® iPod®, iPhone® and iPad®

Enjoy your favorite songs with this Samsung DAE670ZA dock that features a dual docking system for listening to tunes on your Apple iPod, iPhone, iPad or compatible Samsung Galaxy device. The woven glass fiber speakers deliver stunning audio.

Dept Audio Brand Samsung

ID 5563104

$359

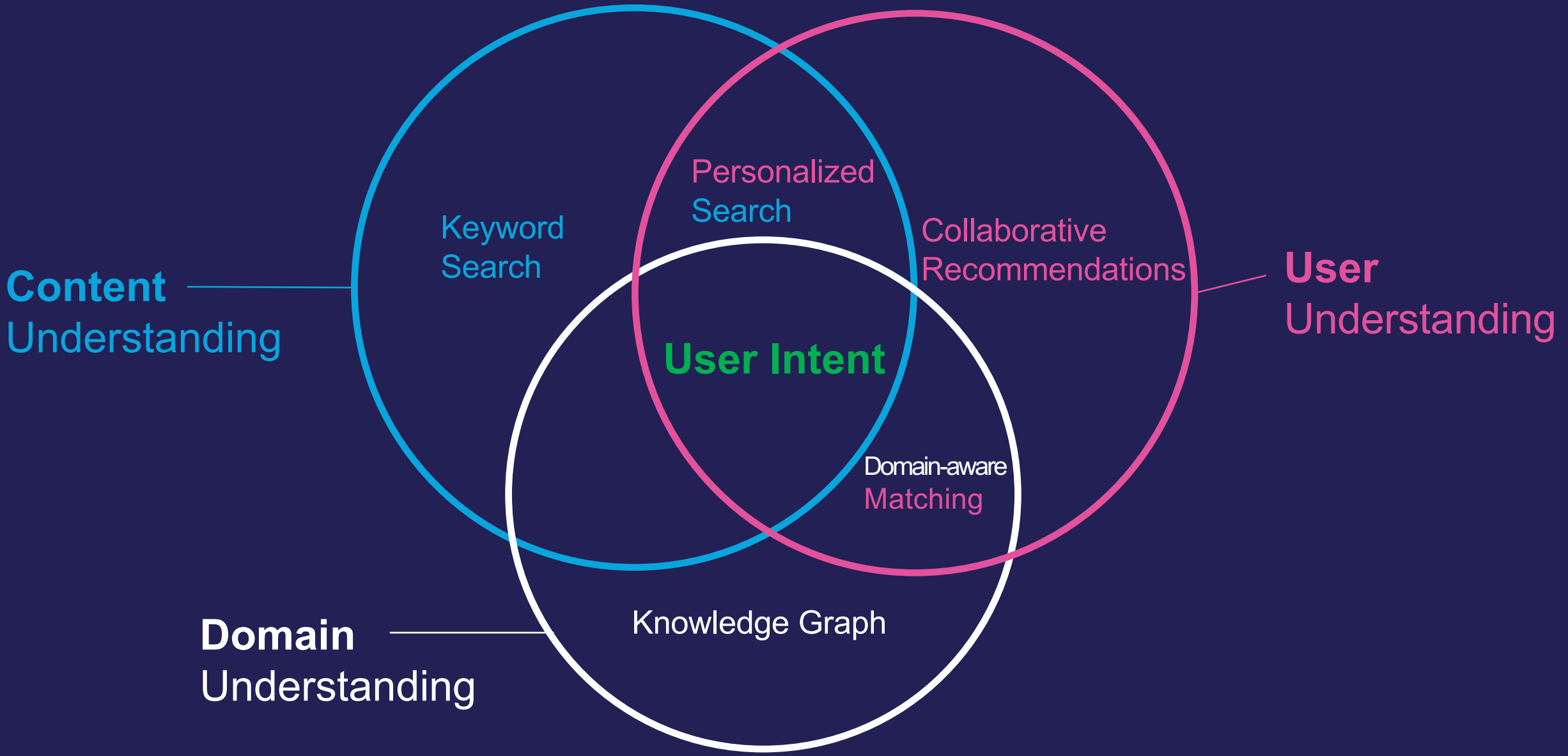
Score 5.3621306

Samsung - Wireless Cradle for Apple® iPod® and iPhone® - Black

Samsung - 250W Compact Shelf System

This compact shelf system features an FM tuner and a built-in dock

Dimensions of User Intent



Collaborative Recommendations
(Completely driven by user behavior)

Knowledge Graph
(Understanding conceptual and logical relationships between domain-specific entities)

Knowledge Graph

(Understanding conceptual and logical relationships between domain-specific entities)

Personas / User Profiles

(User attributes and preferences in knowledge graph)

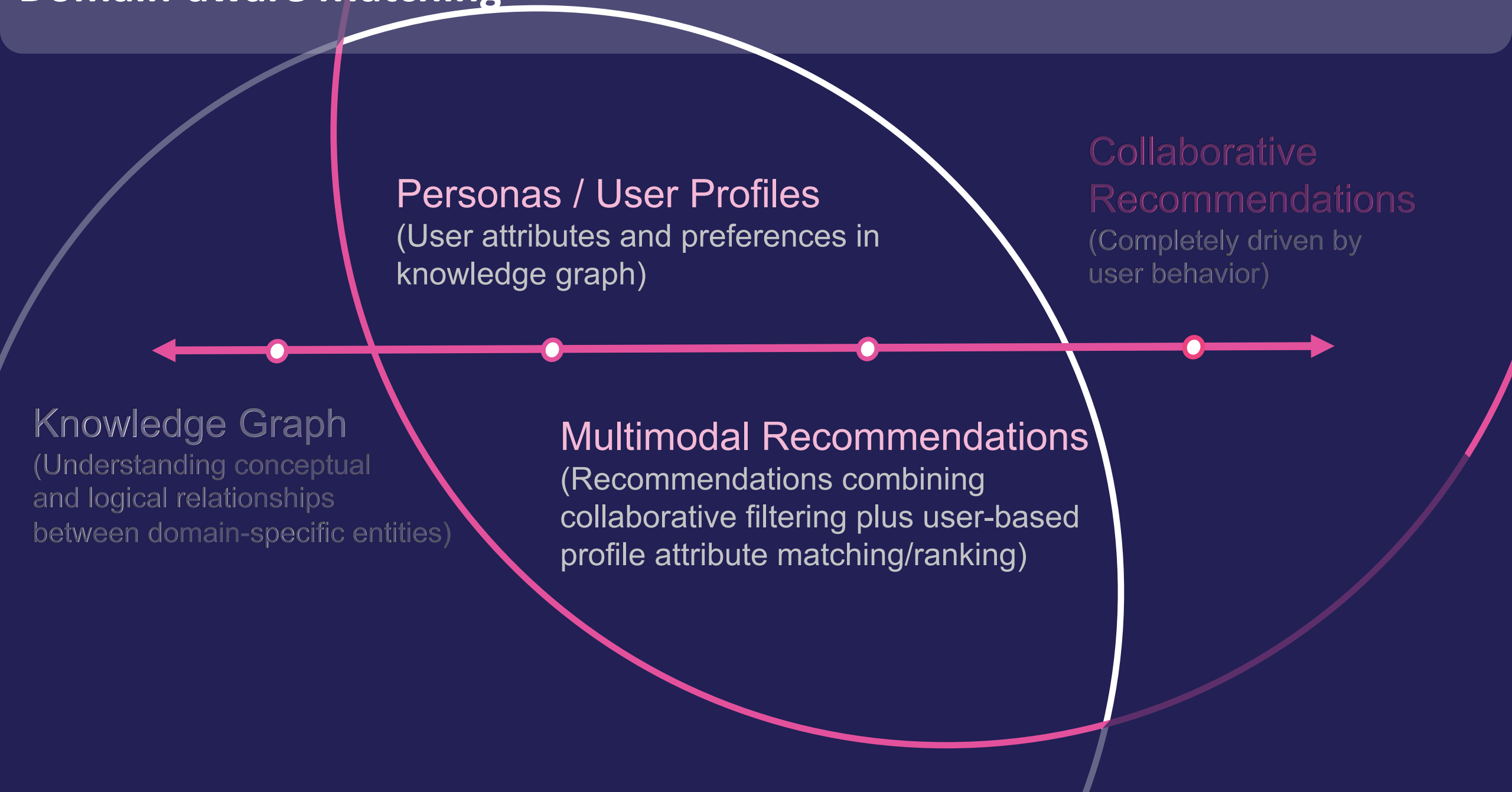
Multimodal Recommendations

(Recommendations combining collaborative filtering plus user-based profile attribute matching/ranking)

Collaborative Recommendations

(Completely driven by user behavior)

Domain-aware Matching



Lucidworks

Find documents, people, departments... Search

Alea Abed

Field Marketing

Marketing
alea Offline
alea.abed@lucidworks.com
-

What connects you and Alea Abed?

You're both skilled in fusion

Most popular documents authored by Alea Abed

see 189 more ›

pdf **Search firm Lucidworks looks to stand-alone applications after Twigkit buy**
Location GoogleDrive

pdf **Microsoft Word - Revolution 2015 Sponsorship Contract.docx**
Location GoogleDrive

xls **Rev 2015 Attendee List.xlsx**
Location GoogleDrive

doc **Revolution 2015 Sponsorship Contract.docx**
Location GoogleDrive

pdf **Reviolution 2015 Sponsorship Prospectus.key**
Location GoogleDrive

doc **Revolution 2015 Blog Posts & Emails.docx**
Location GoogleDrive

doc **Revolution 2015 testimonial tweets.docx**
Location GoogleDrive

Revolution 2016 Signage.pdf
Location GoogleDrive

Alea Abed's skills

fusion search it collateral analysts communications

Colleagues with similar skills


Bjarki Holm
VP of Solutions
management

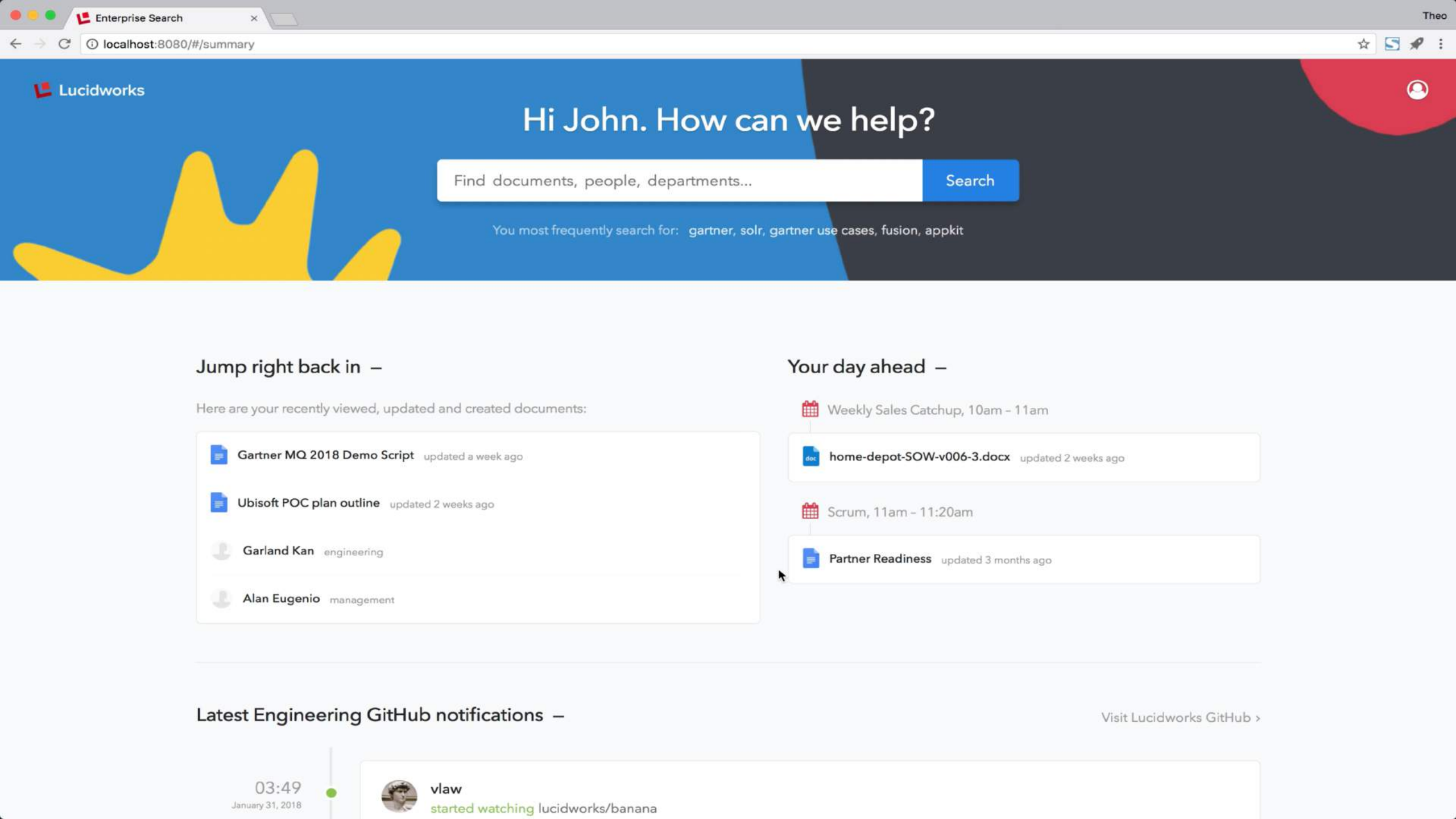

Al Cole
Director of Consulting Services
development


Paul Stark
Sales Director, Northeast
development

Channel contributions

marketing-discussion
327 messages

meetup-planning


Enterprise Search
Theo
localhost:8080/#/summary
Lucidworks
Hi John. How can we help?
Find documents, people, departments...
Search
You most frequently search for: gartner, solr, gartner use cases, fusion, appkit
Jump right back in –
Here are your recently viewed, updated and created documents:
Gartner MQ 2018 Demo Script updated a week ago
Ubisoft POC plan outline updated 2 weeks ago
Garland Kan engineering
Alan Eugenio management
Your day ahead –
Weekly Sales Catchup, 10am – 11am
home-depot-SOW-v006-3.docx updated 2 weeks ago
Scrum, 11am – 11:20am
Partner Readiness updated 3 months ago
Latest Engineering GitHub notifications –
Visit Lucidworks GitHub ›
03:49
January 31, 2018
vlaw
started watching lucidworks/banana

Garland Kan engineering

Alan Eugenio management

Partner Readiness updated 3 months ago

Latest Engineering GitHub notifications –

Visit Lucidworks GitHub ›

03:49
January 31, 2018

vlaw
started watching lucidworks/banana

00:25
January 31, 2018

vthacker
pushed to lucidworks/lucene-solr
Commit messages
SOLR-11880: MDCAwareThreadPoolExecutor creating new exceptions for every request made (temp fix)

22:38
January 30, 2018

vthacker
pushed to lucidworks/lucene-solr
Commit messages
SOLR-11901: Remove Class name capture from logging

Explore departments –

more ›

Marketing
30 people 3999 files 7 channels

Engineering
132 people 132 files 24 channels

HR
22 people 32 files 15 channels

Detect Sentences Index Stage

Steve Tsuchiyama support

Latest Marketing department tweets –

lucidworks ›

22:25
January 29, 2018

Blog Post: #Search is critical for #Omnichannel #Retail customers. Here's why. #Retailtech #MachineLearning #AI
https://t.co/eoXWFj4plM

21:40
January 29, 2018

Are you an #NLP expert? Lucidworks is searching for an NLP Research Engineer to join our #CTO's team. #Hiring…
https://t.co/FupVludArY

21:47
January 26, 2018

Check out @gsingers's newest article, How to Develop Tools for Great Customer Experience - DZone Agile
https://t.co/1bIZqkl8V4

Explore departments –

more ›

Marketing
30 people. 3999 files. 7 channels.

Engineering
132 people. 132 files. 24 channels.

HR
22 people. 32 files. 15 channels.

Multimodal Recommendations

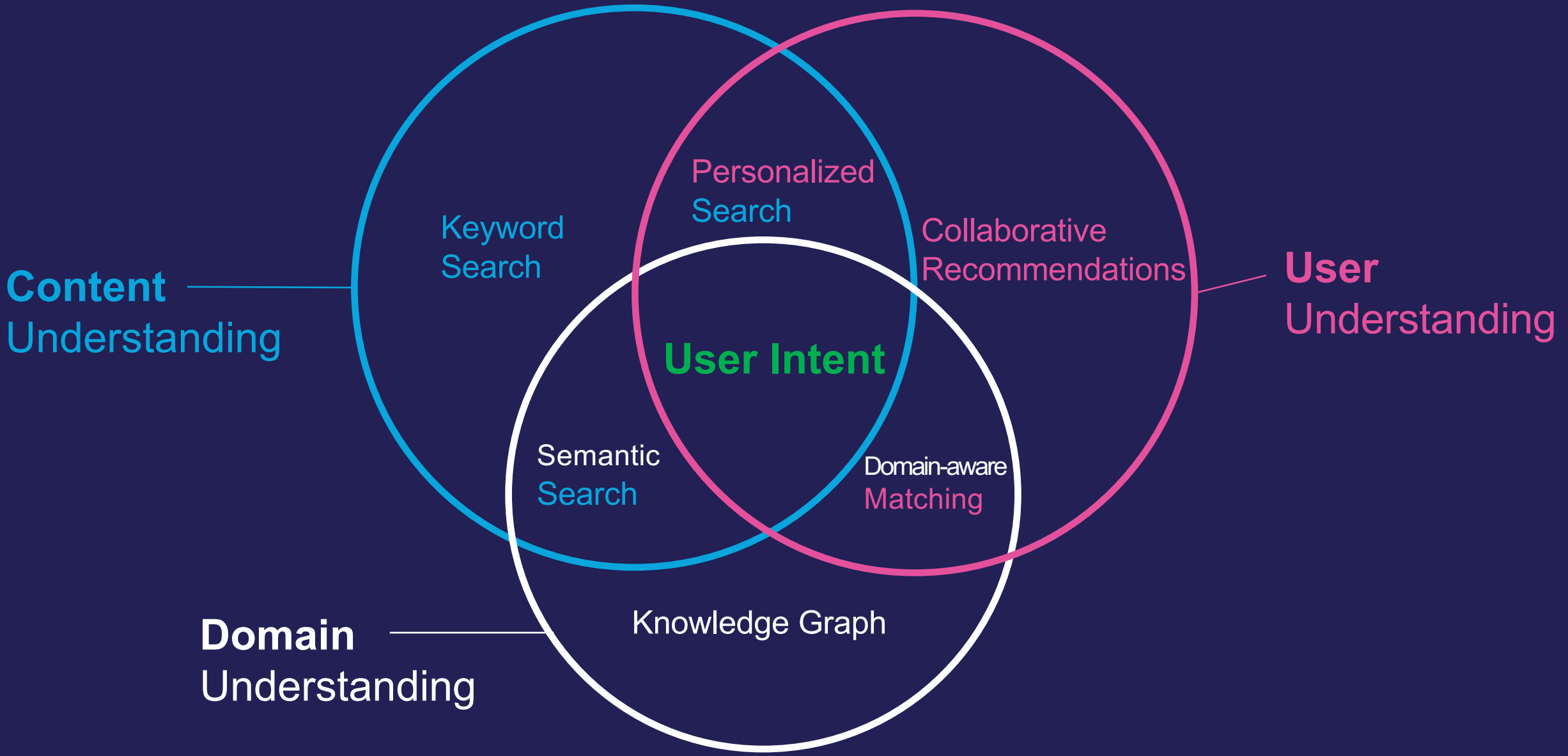
Jane is a nurse educator in Boston seeking between $40K and $60K

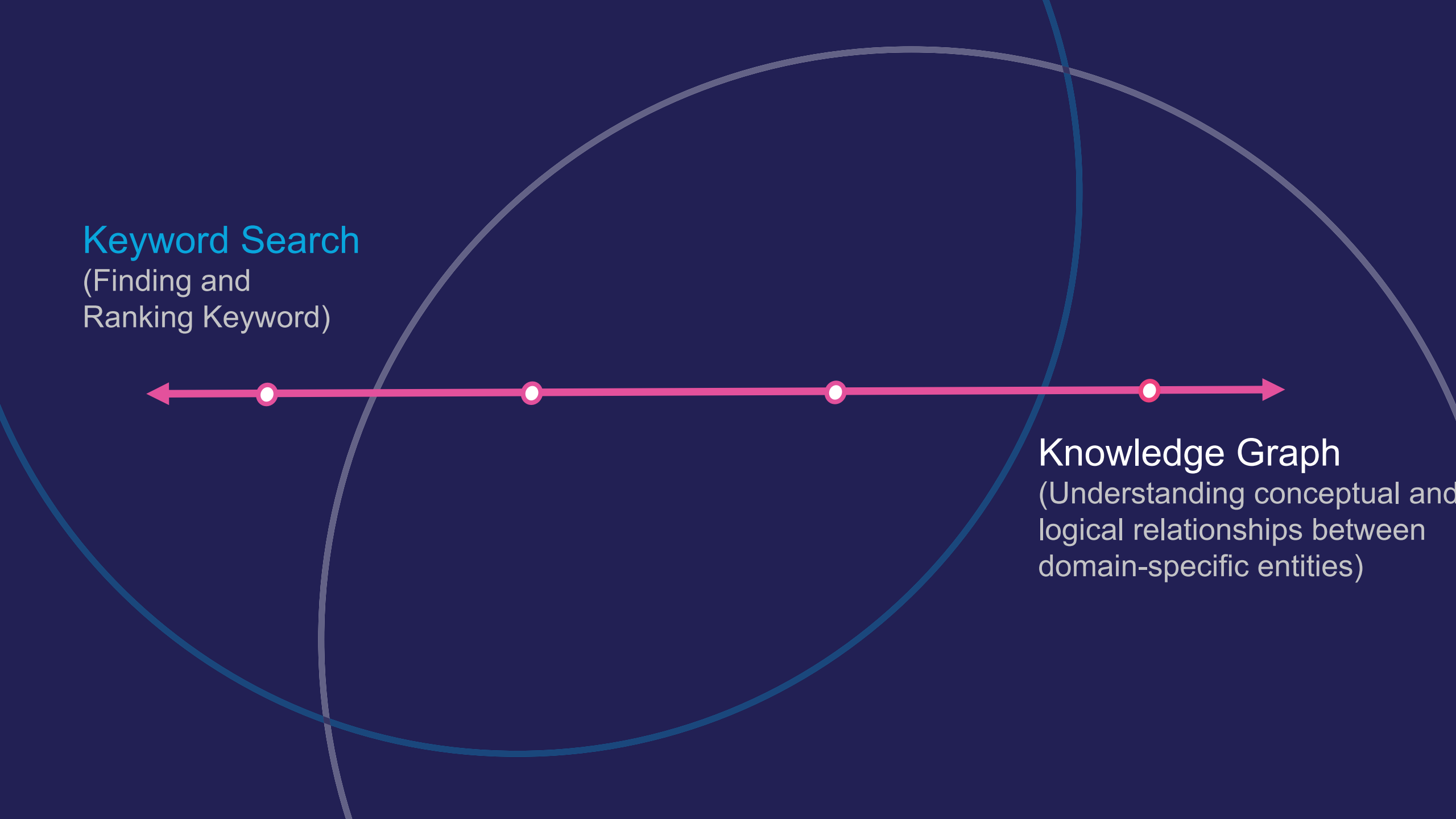
She has interacted with the same content as the following users:
u99,u1,u50,u2311,u253,u70,u99

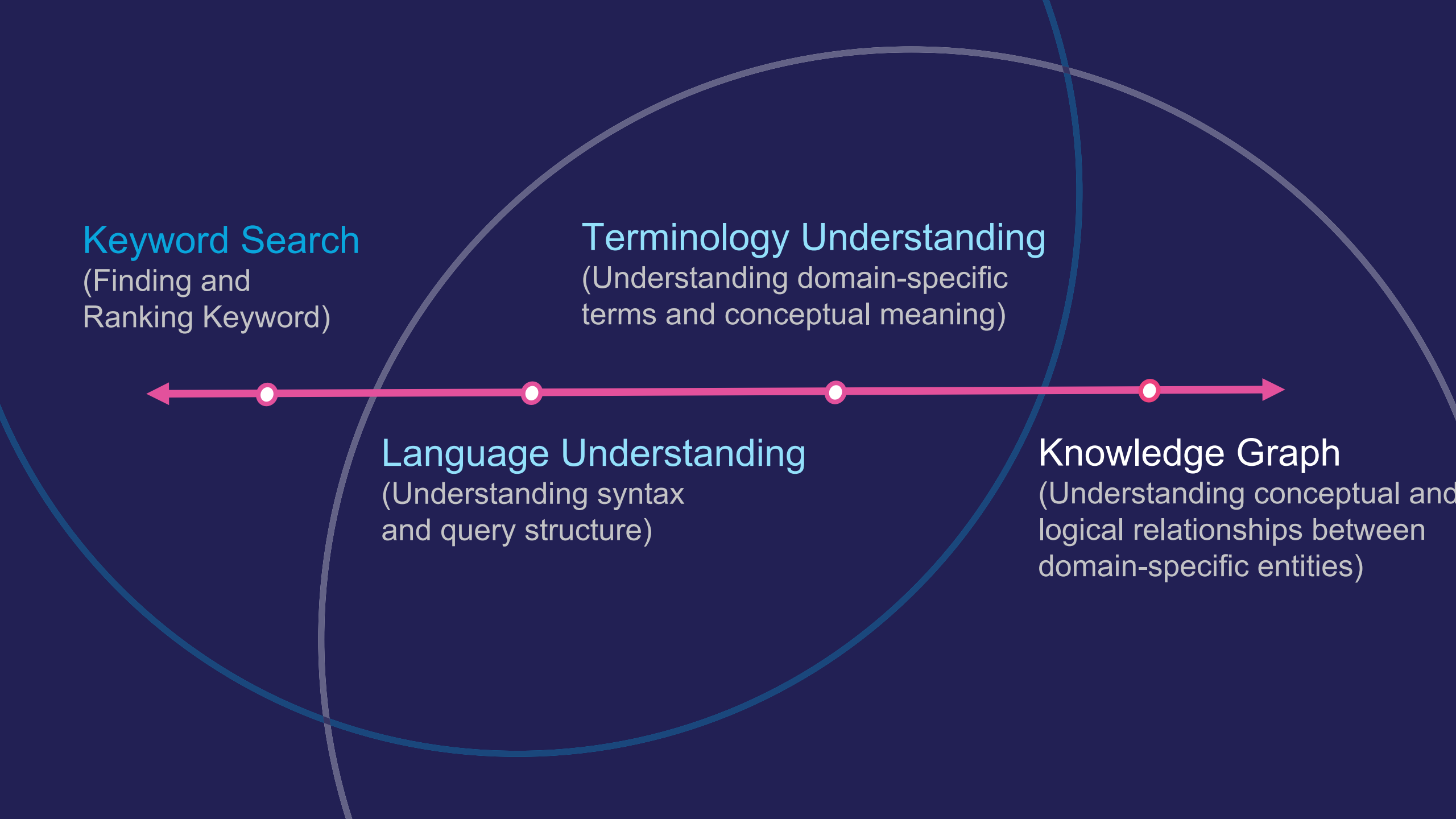
```
http://localhost:8983/solr/jobs/select/?
   fl=jobtitle,city,state,salary&
  q=(
      jobtitle:"nurse educator"^25 OR jobtitle:(nurse educator)^10
    )
    AND (
          (city:"Boston" AND state:"MA")^15
           OR state:"MA")
    AND (_val_:"map(salary, 40000, 60000,10, 0)"
    OR {!edismax mm=2}similar_users:(u99 u1 u50 u2311 u253 u70 u99))
```

*Example derived from chapter 16 of *Solr in Action*

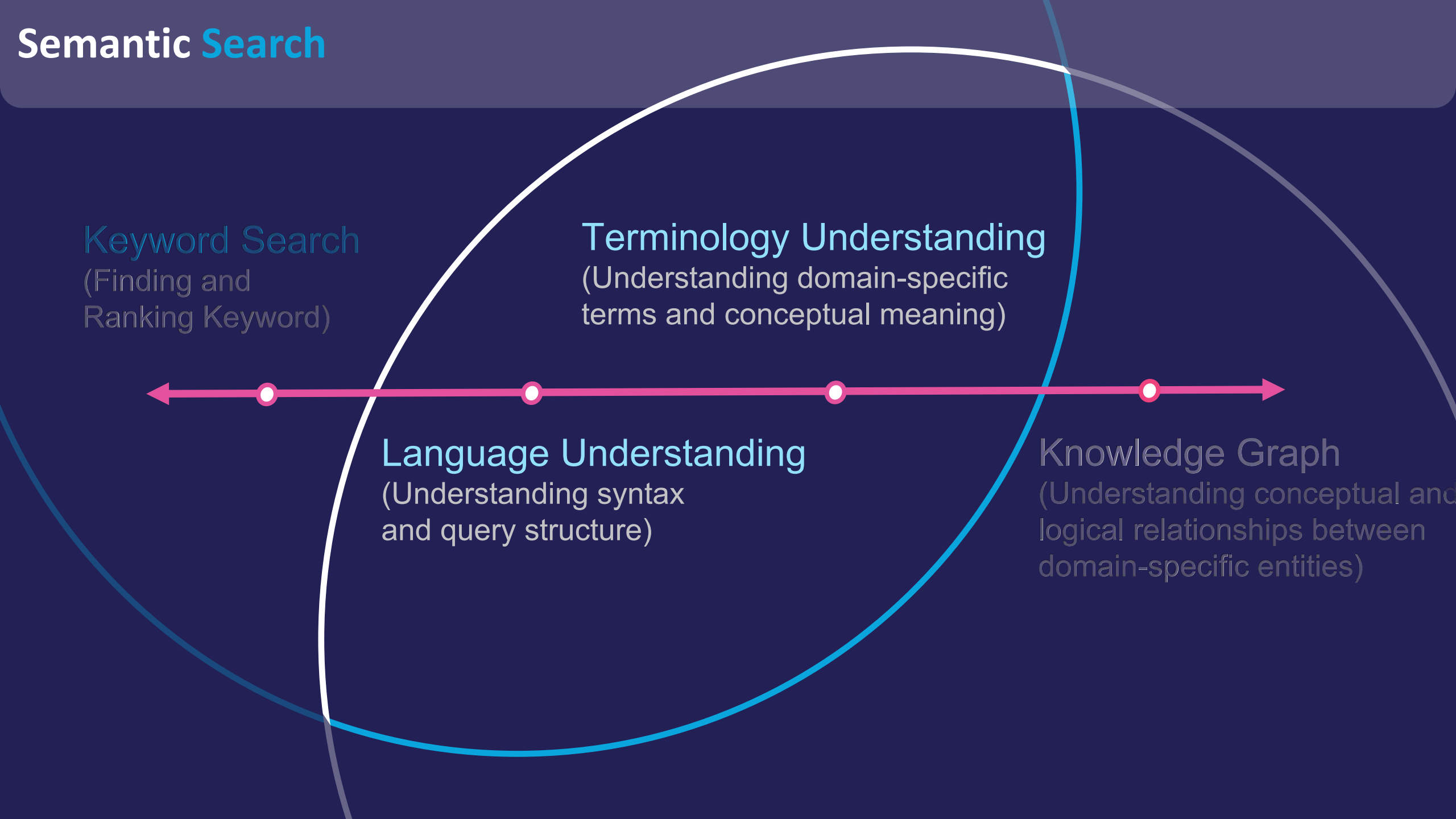
Dimensions of User Intent

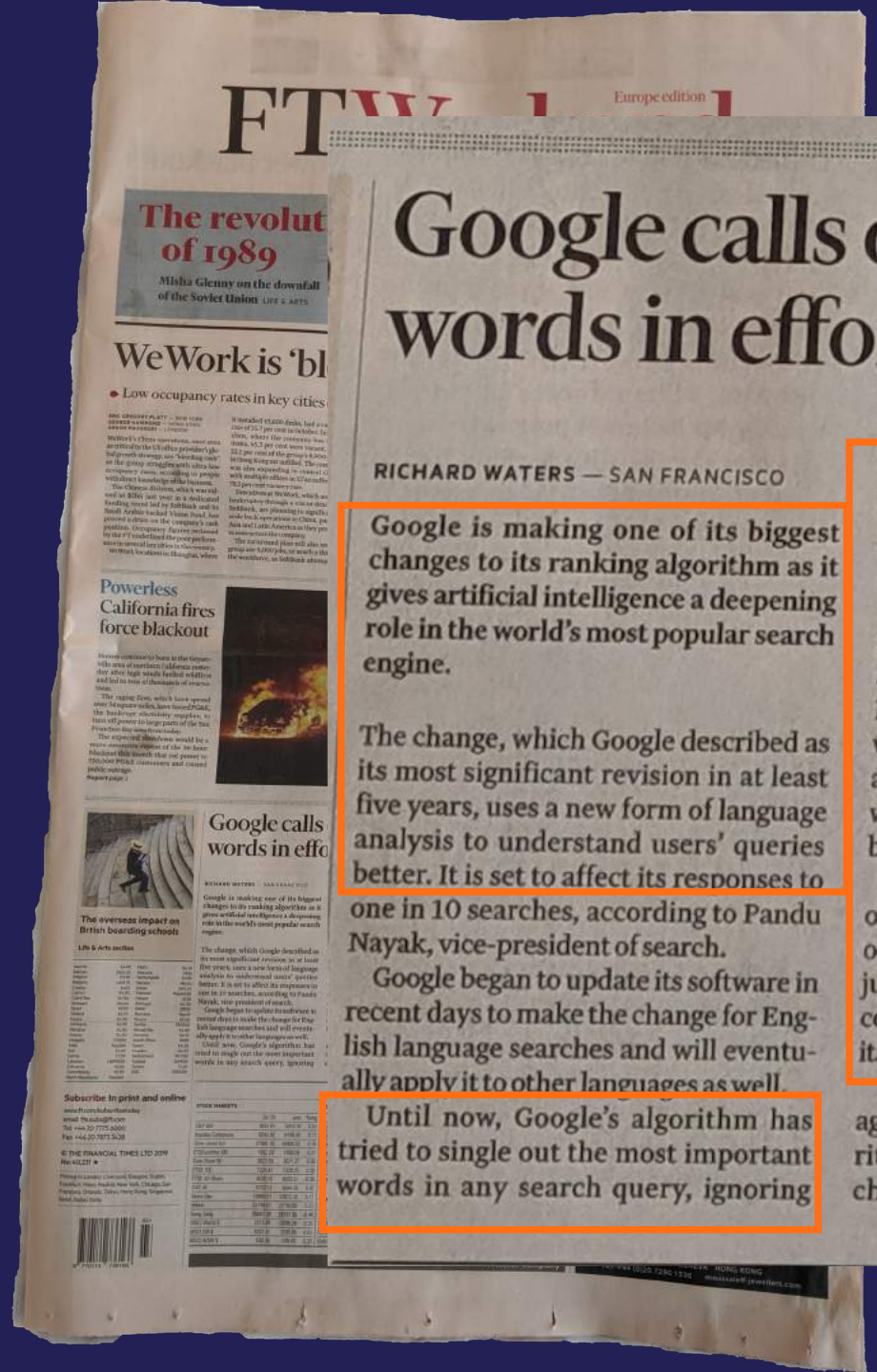



Keyword Search
(Finding and
Ranking Keyword)
Knowledge Graph
(Understanding conceptual and
logical relationships between
domain-specific entities)


Keyword Search
(Finding and
Ranking Keyword)
Terminology Understanding
(Understanding domain-specific
terms and conceptual meaning)
Language Understanding
(Understanding syntax
and query structure)
Knowledge Graph
(Understanding conceptual and
logical relationships between
domain-specific entities)

Semantic Search




FT

Europe edition

The revolut of 1989

Misha Glenny on the downfall of the Soviet Union LIFE & ARTS

WeWork is 'bl

Low occupancy rates in key cities

Powerless

California fires force blackout

Google calls words in effo

The overseas impact on British boarding schools

Life & Arts section

Subscribe in print and online

Google calls on Bert to look at the little words in effort to improve its answers

RICHARD WATERS — SAN FRANCISCO

Google is making one of its biggest changes to its ranking algorithm as it gives artificial intelligence a deepening role in the world's most popular search engine.

The change, which Google described as its most significant revision in at least five years, uses a new form of language analysis to understand users' queries better. It is set to affect its responses to one in 10 searches, according to Pandu Nayak, vice-president of search.

Google began to update its software in recent days to make the change for English language searches and will eventually apply it to other languages as well.

Until now, Google's algorithm has tried to single out the most important words in any search query, ignoring smaller or common words that seem less significant. This enables it to zero in on the main subject, but often results in it misunderstanding a precise request.

The new technique, known as Bert, relies on a language model built up from analysis of vast amounts of text online. Rather than read a string of search words sequentially it analyses them all at the same time — including smaller words that would have been ignored before.

One example Google gave of the types of question it could now handle — "How old was Taylor Swift when Kanye jumped on the stage?" — points to more complex queries that have been beyond its reach before.

Since Google was founded 21 years ago periodic updates to its ranking algorithm have sometimes resulted in big changes in the amount of traffic it directs to outside websites, angering companies whose businesses depend on the search engine. But Google predicted the latest revision would not have a large impact on online traffic.

The update marks the first application of a piece of research on natural language processing from last year that has drawn considerable attention in AI circles. Understanding language is one of the most difficult problems in AI given the fluidity of language depending on the context and the person using it.

The search company said the Bert technique would return more useful responses to many queries. But it also said that in some instances the new algorithm had produced worse results than before. Jeff Dean, Google's head of AI, said: "I don't think we're anywhere close to solving language, but this is a good step."

Lucidworks

Can my wife drive on my insurance?

Search

- Does My Husband Have To Be On My Car Insurance? +
- Can Husband And Wife Have Separate Auto Insurance? +
- Does My Spouse Have To Be On My Auto Insurance? +
- Can My Girlfriend Be On My Car Insurance? +
- Does My Spouse Have My Auto Insurance? +
- Can I Keep My Ex Wife On My Car Insurance? +
- Is My Spouse Covered On My Auto Insurance? +
- Can I Add My Girlfriend To My Auto Insurance Policy? +
- Can I Get Car Insurance Without My Husband? +
- Does My Husband Have To Be On My Auto Insurance If He Doesn't Drive The Car? +

Score: 0.85

The answer is yes, unless your husband has coverage on a separate auto insurance policy. In Wisconsin at least every person in the household must be listed on an auto insurance policy or they must be listed on anouther auto insurance policy. Typically having more than one person listed on a policy will lessen the cost of coverage as opposed to having seperate policies. Contact your local agent to find out the premium difference between a combined policy and individual policies.

Score: 0.81

The answer to this question could be yes or no, depending on a variety of factors. If your husband has an individual policy, not related to his employer, than you can certainly remain on his policy. If your husband's coverage is employer sponsored, then the rules set by the employer's plan would be the guiding factor in this answer.

Score: 0.81

In California all licensed drivers residing in the household must be included on the car policy. That information is more readily available and your insurance company and agent probably know about your wife and her driving record. She can, however, be excluded. In that case the insurance company is not obligated to honor any claim when your wife is operating the car. Of course, that is a very scary proposition.

Score: 0.8

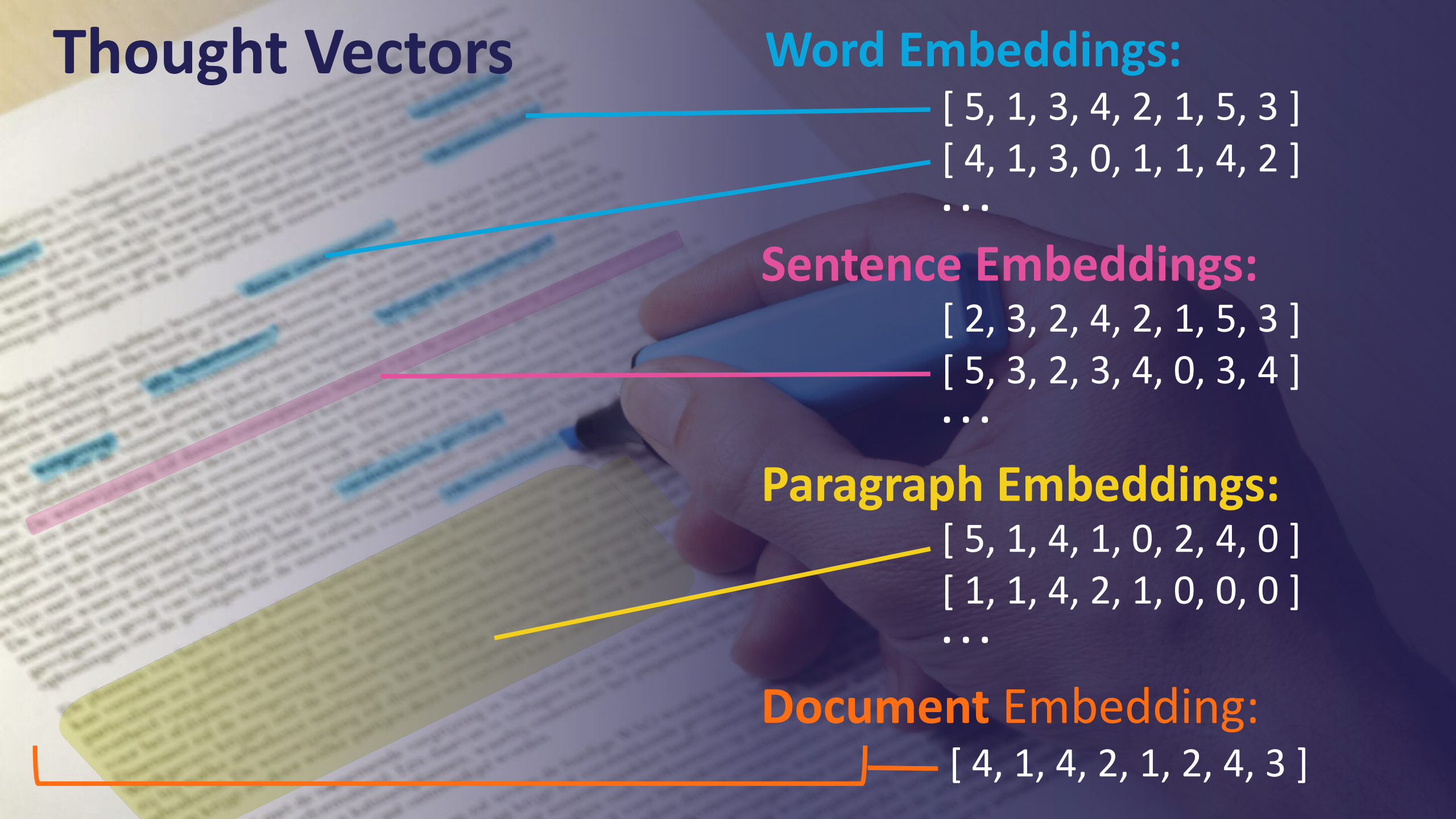
Yes, it is possible; but is typically frowned upon. Typically you are able to secure better rates by having both husband and wife on the same policy. If your husband secures auto insurance on his own then it would be possible for you to do the same. Not all carriers will allow this though. Contact a local agent to discuss your coverage needs and pricing in your area.

Score: 0.79

Based on my education and experience in the industry up to this point, yes it is possible for a husband and wife to have separate auto insurance. This practice is frowned upon though by the carriers and will typically make the cost of the auto insurance more costly for both the husband and the wife.

Score: 0.77

Thought Vectors



Word Embeddings:

[5, 1, 3, 4, 2, 1, 5, 3]

[4, 1, 3, 0, 1, 1, 4, 2]

. . .

Sentence Embeddings:

[2, 3, 2, 4, 2, 1, 5, 3]

[5, 3, 2, 3, 4, 0, 3, 4]

. . .

Paragraph Embeddings:

[5, 1, 4, 1, 0, 2, 4, 0]

[1, 1, 4, 2, 1, 0, 0, 0]

. . .

Document Embedding:

[4, 1, 4, 2, 1, 2, 4, 3]

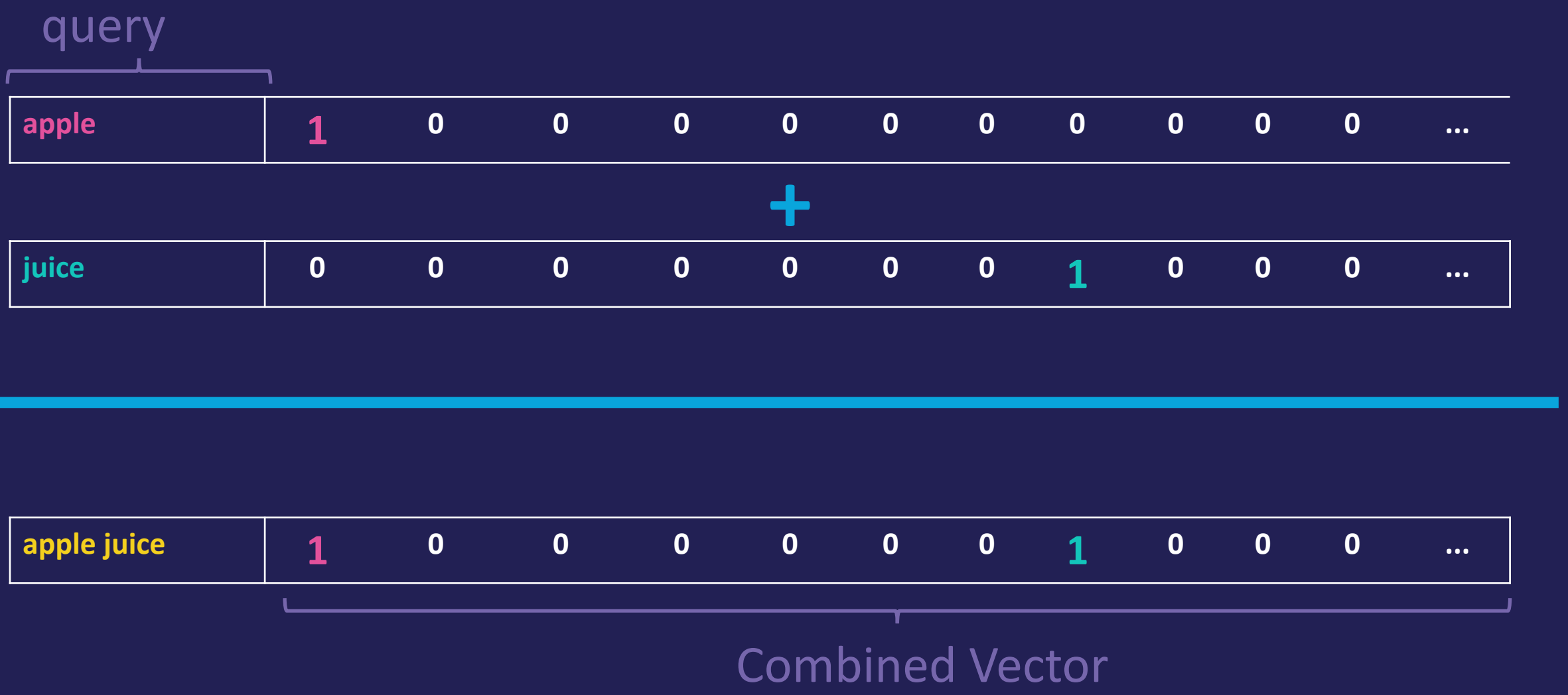
Single Term Searches (as a Vector)

query

exact term lookup in inverted index

	apple	caffeine	cheese	coffee	drink	donut	food	juice	pizza	tea	water	… *term N*
cappuccino	0	0	0	0	0	0	0	0	0	0	0	…
apple	1	0	0	0	0	0	0	0	0	0	0	…
juice	0	0	0	0	0	0	0	1	0	0	0	…
cheese	0	0	1	0	0	0	0	0	0	0	0	…
pizza	0	0	0	0	0	0	0	0	1	0	0	…
donut	0	0	0	0	0	1	0	0	0	0	0	…
green	0	0	0	0	0	0	0	0	0	0	0	…
tea	0	0	0	0	0	0	0	0	0	1	0	…
bread	0	0	1	0	0	0	0	0	0	0	0	…
sticks	0	0	0	0	0	0	0	0	0	0	0	…

Multi-term Query Vectors



Multi-term Searches

query

exact term lookup in inverted index

	apple	caffeine	cheese	coffee	drink	donut	food	juice	pizza	tea	water	… *term N*
latte	0	0	0	0	0	0	0	0	0	0	0	…
cappuccino	0	0	0	0	0	0	0	0	0	0	0	…
apple juice	1	0	0	0	0	0	0	1	0	0	0	…
cheese pizza	0	0	1	0	0	0	0	0	1	0	0	…
donut	0	0	0	0	0	1	0	0	0	0	0	…
soda	0	0	0	0	0	0	0	0	0	0	0	…
green tea	0	0	0	0	0	0	0	0	0	1	0	…
water	0	0	0	0	0	0	0	0	0	0	1	…
cheese bread sticks	0	0	1	0	0	0	0	0	0	0	0	…
cinnamon sticks	0	0	0	0	0	0	0	0	0	0	0	…

Dimensionality Reduction

	food	drink	dairy	bread	caffeine	sweet	calories	healthy
apple juice	0	5	0	0	0	4	4	3
cappuccino	0	5	3	0	4	1	2	3
cheese bread sticks	5	0	4	5	0	1	4	2
cheese pizza	5	0	4	4	0	1	5	2
cinnamon bread sticks	5	0	1	5	0	3	4	2
donut	5	0	1	5	0	4	5	1
green tea	0	5	0	0	2	1	1	5
latte	0	5	4	0	4	1	3	3
soda	0	5	0	0	3	5	5	0
water	0	5	0	0	0	0	0	5

Vector Similarity Scoring

Phrase:	Vector:
apple juice:	[0, 5, 0, 0, 0, 4, 4, 3]
cappuccino:	[0, 5, 3, 0, 4, 1, 2, 3]
cheese bread sticks:	[5, 0, 4, 5, 0, 1, 4, 2]
cheese pizza:	[5, 0, 4, 4, 0, 1, 5, 2]
cinnamon bread sticks:	[5, 0, 4, 5, 0, 1, 4, 2]
donut:	[5, 0, 1, 5, 0, 4, 5, 1]
green tea:	[0, 5, 0, 0, 2, 1, 1, 5]
latte:	[0, 5, 4, 0, 4, 1, 3, 3]
soda:	[0, 5, 0, 0, 3, 5, 5, 0]
water:	[0, 5, 0, 0, 0, 0, 0, 5]

Vector Similarity (a, b):

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \times |\mathbf{b}|}$$

Ranked Results: **Cheese Pizza**	
0.99	cheese bread sticks
0.91	cinnamon bread sticks
0.89	donut
0.47	latte
0.46	apple juice
…	…
0.19	water

Ranked Results: **Green Tea**	
0.94	water
0.85	cappuccino
0.80	latte
0.78	apple juice
0.60	soda
…	…
0.19	donut

Performance Considerations

Problem: Vector Scoring is Slow

- Unlike keyword search, which looks up pre-indexed answers to queries, Vector Search must instead calculate similarities between the query vector and every document's vectors to determine best matches, which is slow at scale.

Solution: Quantized Vectors

- "Quantization" is the process for mapping vectors features to discrete values.
- Creating "tokens" which map to a similar vector space, enables matching on those tokens to perform an ANN (Approximate Nearest Neighbor) search
- This enables converting vector scoring into a search problem (term lookup and scoring), which is fast again, at the expense of some recall and scoring accuracy

Recommended Approach: Quantized Vector Search + Vector Similarity Reranking

- Combine the best of both worlds by running an initial ANN search on a quantized vector representation, and then re-rank the top-N results using full Vector similarity scoring.

Solr Implementation Options

Option 1: Streaming Expressions

Streaming Expressions

Send Documents to Solr:

```
curl -X POST -H "Content-Type: application/json" \
 http://localhost:8983/solr/food/update?commit=true \
 --data-binary ' [
 {"id": "1", "name_s":"donut", "vector_fs":[5.0,0.0,1.0,5.0,0.0,4.0,5.0,1.0]},
 {"id": "2", "name_s":"apple juice",
             "vector_fs":[1.0,5.0,0.0,0.0,0.0,4.0,4.0,3.0]},
 {"id": "3", "name_s":"cappuccino",
             "vector_fs":[0.0,5.0,3.0,0.0,4.0,1.0,2.0,3.0]},
 {"id": "4", "name_s":"cheese pizza",
             "vector_fs":[5.0,0.0,4.0,4.0,0.0,1.0,5.0,2.0]},
 {"id": "5", "name_s":"green tea",
             "vector_fs":[0.0,5.0,0.0,0.0,2.0,1.0,1.0,5.0]},
 {"id": "6", "name_s":"latte", "vector_fs":[0.0,5.0,4.0,0.0,4.0,1.0,3.0,3.0]},
 {"id": "7", "name_s":"soda", "vector_fs":[0.0,5.0,0.0,0.0,3.0,5.0,5.0,0.0]},
 {"id": "8", "name_s":"cheese bread sticks",
             "vector_fs":[5.0,0.0,4.0,5.0,0.0,1.0,4.0,2.0]},
 {"id": "9", "name_s":"water", "vector_fs":[0.0,5.0,0.0,0.0,0.0,0.0,0.0,5.0]},
 {"id": "10", "name_s":"cinnamon bread sticks",
              "vector_fs":[5.0,0.0,1.0,5.0,0.0,3.0,4.0,2.0]}
]'
```

Solr

- Dashboard
- Logging
- Cloud
- Collections
- Java Properties
- Thread Dump
- Suggestions

food

- Overview
- Analysis
- Dataimport
- Documents
- Files
- Query
- Stream
- Schema

Streaming Expression (expr)

```
top(
  select(
    search(food, q="*:*", fl="id,vector_fs,name_s", sort="id asc"),
    cosineSimilarity(vector_fs, array(5.2,0.0,1.0,5.0,0.0,4.0,5.0,1.0)) as cos, id, name_s),
    n=5,
    sort="cos desc"
)
```

Execute with explanation

http://localhost:8983/solr/food/stream?expr=top(select(search(food%2C%20q%3D%22*%3A*%22%2C%20fl%3D%22id%2Cvector_fs%2Cname_s%22%2C%2


```
{
  "result-set": {
    "docs": [
      {
        "name_s": "donut",
        "cos": 0.99984612,
        "id": "1"
      },
      {
        "name_s": "cinnamon bread sticks",
        "cos": 0.98628004,
        "id": "10"
      },
      {
        "name_s": "cheese pizza",
        "cos": 0.89078353,
        "id": "4"
      },
      {
```

Option 2:
Streaming Expressions Query Parser

Streaming Expressions Query Parser

Request:

```
http://localhost:8983/solr/food/select?q=*:*&fl=id,name_s&
  fq={!streaming_expression}top(
   select(
    search(food, q="*:*", fl="id,vector_fs", sort="id asc"),
    cosineSimilarity(vector_fs, array(5.1,0.0,1.0,5.0,0.0,4.0,5.0,1.0)) as cos, id),
    n=5,  sort="cos desc”
  )
```

Response:

```
{  "responseHeader":{
    …  },
  "response":{"numFound":5,"start":0,"docs":[
     {      "name_s":"donut",   "id":"1"},
     {      "name_s":"apple juice",   "id":"2"},
     {      "name_s":"cheese pizza", "id":"4"},
     {      "name_s":"cheese bread sticks", "id":"8"},
     {      "name_s":"cinnamon bread sticks",  "id":"10"}]
  }}
```

Option 3:
Solr Vector Scoring Plugin

Solr Vector Scoring Plugin

Send Documents to Solr:

```
curl -X POST -H "Content-Type: application/json"
http://localhost:8983/solr/{your-collection-name}/update?commit=true --
data-binary '
[
 {"name":"example 0", "vector":"0|1.55 1|3.53 2|2.3 3|0.7 4|3.44 5|2.33"},
 {"name":"example 1", "vector":"0|3.54 1|0.4 2|4.16 3|4.88 4|4.28 5|4.25"},
 {"name":"example 2", "vector":"0|1.11 1|0.6 2|1.47 3|1.99 4|2.91 5|1.01"},
 {"name":"example 3", "vector":"0|0.06 1|4.73 2|0.29 3|1.27 4|0.69 5|3.9"},
 {"name":"example 4", "vector":"0|4.01 1|3.69 2|2 3|4.36 4|1.09 5|0.1"},
 {"name":"example 5", "vector":"0|0.64 1|3.95 2|1.03 3|1.65 4|0.99 5|0.09"}
]'
```

Solr Vector Scoring Plugin

Request:

```
http://localhost:8983/solr/{your-collection-name}/query?fl=name,score,vector&q={!vp f=vector
  vector="0.1,4.75,0.3,1.2,0.7,4.0”
}
```

Response:


```
{ "responseHeader":{ "status":0, "QTime":1}},
  "response":{ "numFound":6,"start":0,"maxScore":0.99984086,
    "docs":[
      { "name":["example 3"], "vector":["0|0.06 1|4.73 2|0.29 3|1.27 4|0.69 5|3.9 "],
"score":0.99984086},
{ "name":["example 0"], "vector":["0|1.55 1|3.53 2|2.3 3|0.7 4|3.44 5|2.33 "], "score":0.7693964},
{ "name":["example 5"], "vector":["0|0.64 1|3.95 2|1.03 3|1.65 4|0.99 5|0.09 "], "score":0.76322395},
{ "name":["example 4"], "vector":["0|4.01 1|3.69 2|2 3|4.36 4|1.09 5|0.1 "], "score":0.5328145},
{ "name":["example 1"], "vector":["0|3.54 1|0.4 2|4.16 3|4.88 4|4.28 5|4.25 "], "score":0.48513117},
{ "name":["example 2"], "vector":["0|1.11 1|0.6 2|1.47 3|1.99 4|2.91 5|1.01 "], "score":0.44909418}]
}}
```

Option 4:
Solr Vector Scoring + LSH Plugin

Solr Vector Scoring + LSH Plugin

Send Documents to Solr:

```
curl -X POST -H "Content-Type: application/json" http://localhost:8983/solr/{your-collection-name}/update?update.chain=LSH&commit=true --data-binary '
[
  {"id":"1", "vector":"1.55,3.53,2.3,0.7,3.44,2.33"},
  {"id":"2", "vector":"3.54,0.4,4.16,4.88,4.28,4.25"}
]'
```

Request:

```
http://localhost:8983/solr/{your-collection-name}/query?fl=name,score,vector&q={!vp f=vector vector="1.55,3.53,2.3,0.7,3.44,2.33" lsh="true" reRankDocs="5"}&fl=name,score,vector,_vector_,_lsh_hash_
```

Solr Vector Scoring + LSH Plugin

Request:

```
http://localhost:8983/solr/{your-collection-name}/query?fl=name,score,vector&q={!vp f=vector
vector="1.55,3.53,2.3,0.7,3.44,2.33" lsh="true"
reRankDocs="5"}&fl=name,score,vector,_vector_,_lsh_hash_
```

Response:


```
{
 "responseHeader":{ "status":0, "QTime":8, "response":{"numFound":1,"start":0,"maxScore":36.65736,
 "docs":[
  { "id": "1", "vector":"1.55,3.53,2.3,0.7,3.44,2.33",
   "_vector_":"/z/GZmZAYeuFQBMzMz8zMzNAXCj2QBUeuA==",
   "_lsh_hash_":["0_8", "1_35", "2_7", "3_10", "4_2", "5_35", "6_16", "7_30", "8_27", "9_12", "10_7",
                 "11_32", "12_48", "13_36", "14_10", "15_7", "16_42", "17_5", "18_3", "19_2", "20_1",
                 "21_0", "22_24", "23_18", "24_42", "25_31", "26_35", "27_8", "28_1", "29_24", "30_47",
                 "31_14", "32_22", "33_39", "34_0", "35_34", "36_34", "37_39", "38_27", "39_27",
                 "40_45", "41_10", "42_21", "43_34", "44_41", "45_9", "46_31", "47_0", "48_4", "49_43"],
   "score":36.65736}
  ] } }
```

Option 5 (Work in Progress):
First-class Vector Fields in Lucene/Solr

issues.apache.org/jira/browse/SOLR-12890

Dashboards Projects Issues

Solr / SOLR-12890

Vector Search in Solr (Umbrella Issue)

Details

Type:	New Feature	Status:	OPEN
Priority:	Major	Resolution:	Unresolved
Affects Version/s:	None	Fix Version/s:	None
Component/s:	None		
Labels:	None		

Description

We have recently come across a need to index documents containing vectors using solr, and have even worked on a small POC. We used an URP to calculate the LSH(we chose to use the superbit algorithm, but the code is designed in a way that algorithm picked can be easily chagned), and store the vector in other sparse dense forms in a binary field.
Perhaps an addition of an LSH URP in conjunction with a query parser that uses the same properties to calculate LSH(or maybe ktree, or some other algorithm all together) should be considered as a Solr feature?

Issue Links

is duplicated by

~~SOLR-13500~~ Vector Search with Solr — RESOLVED

relates to

~~SOLR-12879~~ Query Parser for MinHash/LSH — CLOSED

links to

Now In Progress

issues.apache.org/jira/browse/LUCENE-9004

Dashboards Projects Issues Search

Lucene - Core / LUCENE-9004

Approximate nearest vector search

Details

Type:	New Feature	Status:	OPEN
Priority:	Major	Resolution:	Unresolved
Affects Version/s:	None	Fix Version/s:	None
Component/s:	None		
Labels:	None		
Lucene Fields:	New		

Description

"Semantic" search based on machine-learned vector "embedding" representing terms, queries and documents is becoming a must-have feature for a modern search engine. SOLR-12890 is exploring various approaches to this, including providing vector-based scoring functions. This is a spinoff issue from that.

The idea here is to explore approximate nearest-neighbor search. Researchers have found an approach based on navigating a graph that partially encodes the nearest neighbor relation at multiple scales can provide accuracy > 95% (as compared to exact nearest neighbor calculations) at a reasonable cost. This issue will explore implementing HNSW (hierarchical navigable small-world) graphs for the purpose of approximate nearest vector search (often referred to as KNN or k-nearest-neighbor search).

At a high level the way this algorithm works is this. First assume you have a graph that has a partial encoding of the nearest neighbor relation, with some short and some long-distance links. If this graph is built in the right way (has the hierarchical navigable small world property), then you can efficiently traverse it to find nearest neighbors (approximately) in log N time where N is the number of nodes in the graph. I believe this idea was pioneered in [1]. The great insight in that paper is that if you use the

People

Assignee: Unassigned

Reporter: Michael S

Votes: 3 Vote for this

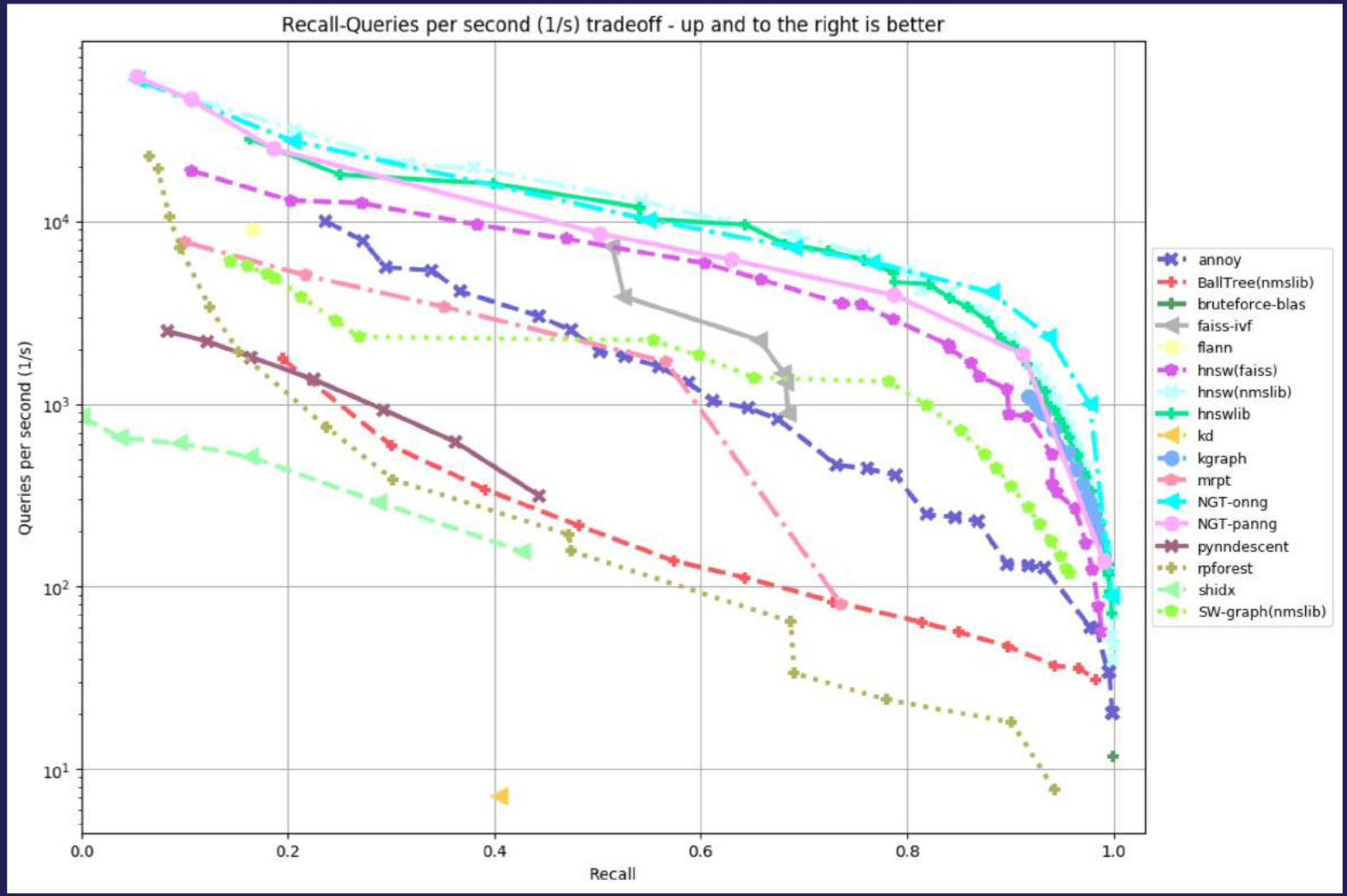
Watchers: 11 Start watching

Dates

Created: 4 days ago

Updated: Yesterday

ANN Benchmarks
(Approximate Nearest Neighbor)



https://github.com/erikbern/ann-benchmarks

Vector Encoders

Vector Encoders

- Take queries, documents, sentences, paragraphs, etc. and transform them into vectors.

- Usually leverage deep learning, which can discover rich language usage rules and map them to combinations of features in the vector

- **Popular Libraries:**
  - Bert
  - Elmo
  - Universal Sentence Encoder
  - Word2Vec
  - Sentence2Vec
  - Glove
  - fastText
  - *many more …*

FT Europe edition

The revolut of 1989

Misha Glenny on the downfall of the Soviet Union LIFE & ARTS

WeWork is 'bl

Low occupancy rates in key cities

Powerless

California fires force blackout

The overseas impact on British boarding schools

Life & Arts section

Google calls words in effo

Subscribe in print and online

Google calls on Bert to look at the little words in effort to improve its answers

RICHARD WATERS — SAN FRANCISCO

Google is making one of its biggest changes to its ranking algorithm as it gives artificial intelligence a deepening role in the world's most popular search engine.

The change, which Google described as its most significant revision in at least five years, uses a new form of language analysis to understand users' queries better. It is set to affect its responses to one in 10 searches, according to Pandu Nayak, vice-president of search.

Google began to update its software in recent days to make the change for English language searches and will eventually apply it to other languages as well.

Until now, Google's algorithm has tried to single out the most important words in any search query, ignoring smaller or common words that seem less significant. This enables it to zero in on the main subject, but often results in it misunderstanding a precise request.

The new technique, known as Bert, relies on a language model built up from analysis of vast amounts of text online. Rather than read a string of search words sequentially it analyses them all at the same time — including smaller words that would have been ignored before.

One example Google gave of the types of question it could now handle — "How old was Taylor Swift when Kanye jumped on the stage?" — points to more complex queries that have been beyond its reach before.

Since Google was founded 21 years ago periodic updates to its ranking algorithm have sometimes resulted in big changes in the amount of traffic it directs to outside websites, angering companies whose businesses depend on the search engine. But Google predicted the latest revision would not have a large impact on online traffic.

The update marks the first application of a piece of research on natural language processing from last year that has drawn considerable attention in AI circles. Understanding language is one of the most difficult problems in AI given the fluidity of language depending on the context and the person using it.

The search company said the Bert technique would return more useful responses to many queries. But it also said that in some instances the new algorithm had produced worse results than before. Jeff Dean, Google's head of AI, said: "I don't think we're anywhere close to solving language, but this is a good step."

Keyword Search vs. Vector Search

Query Type	Likely Outcome
Obscure keyword combinations *Q. (software OR hardware) AND enginee**	• Keyword search succeeds • Vector Search fails
Natural Language Queries *Q. Can my wife drive on my insurance?*	• Keyword search might get lucky, but probably fails • Vector Search succeeds
Fuzzy Language Queries *Q. famous french tower*	• Keyword search mismatch yields poor results • Vector Search succeeds
Structured Relationship Queries *Q. popular bbq near Activate*	• Keyword search fails • Vector search fails • **Need a Knowledge Graph!**

Giant Graph of Relationships...

Trey Grainger works for Lucidworks.

He spoke at the Activate 2019 conference.

#Activate19 (Activate) wqs held in Washington, DC September 9-12, 2019.

Trey got his masters degree from Georgia Tech.

Trey's Voicemail

Semantic Knowledge Graph

Documents

id: 1
job_title: Software Engineer
desc: software engineer at a great company
skills: .Net, C#, java

id: 2
job_title: Registered Nurse
desc: a registered nurse at hospital doing hard work
skills: oncology, phlebotemy

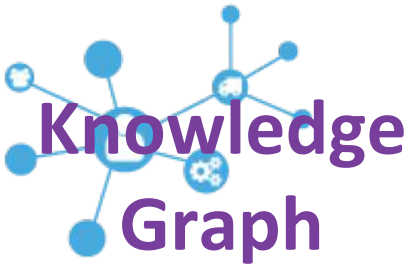
id: 3
job_title: Java Developer
desc: a software engineer or a java engineer doing work
skills: java, scala, hibernate

Docs-Terms Forward Index

field	doc	term
desc	1	a
		at
		company
		engineer
		great
		software
	2	a
		at
		doing
		hard
		hospital
		nurse
		registered
		work
	3	a
		doing
		engineer
		java
		or
		software
		work
job_title	1	Software Engineer
...	...	...

Terms-Docs Inverted Index

field	term	postings list	
		doc	pos
desc	a	1	4
		2	1
		3	1, 5
	at	1	3
		2	4
	company	1	6
	doing	2	6
		3	8
	engineer	1	2
		3	3, 7
	great	1	5
	hard	2	7
	hospital	2	5
	java	3	6
	nurse	2	3
	or	3	4
	registered	2	2
	software	1	1
		3	2
	work	2	10
		3	9
job_title	java developer	3	1
...	...	...	...



***Source**: Trey Grainger, Khalifeh AlJadda, Mohammed Korayem, Andries Smith."The Semantic Knowledge Graph: A compact, auto-generated model for real-time traversal and ranking of any relationship within a domain". DSAA 2016.*

Related term vector (for query concept expansion)

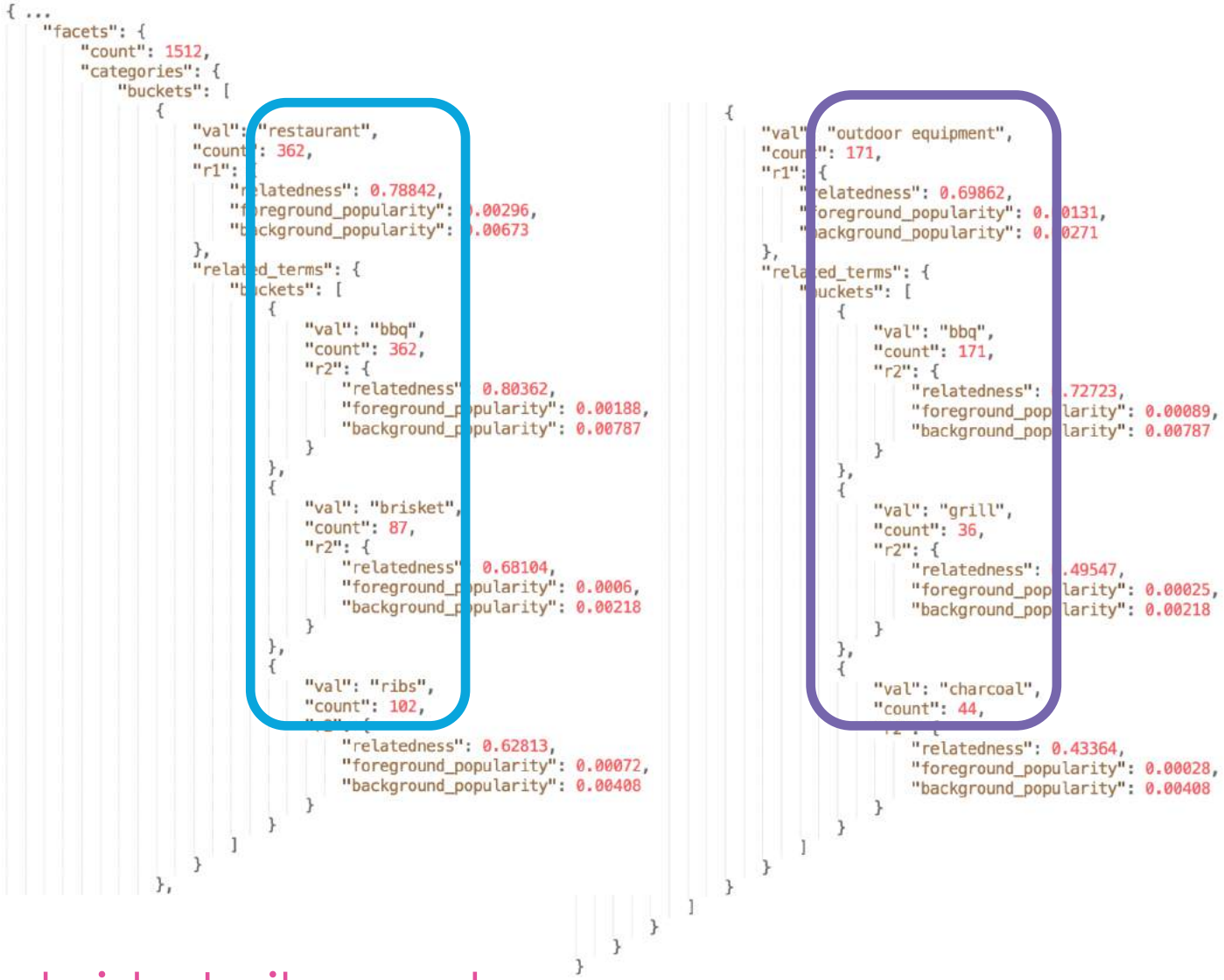
`http://localhost:8983/solr/stack-exchange-health/skg`

```
{
  "queries": [
    "advil"
  ],
  "compare": [
    {
      "type": "Body_t",
      "limit": 20,
      "sort": "relatedness"
    }
  ]
}
```


```
{
    "data": [
        {
            "type": "Body_t",
            "values": [
                {
                    "name": "advil",
                    "relatedness": 0.62076,
                    "popularity": 909,
                    "foreground_popularity": 909,
                    "background_popularity": 909
                },
                {
                    "name": "motrin",
                    "relatedness": 0.51264,
                    "popularity": 649,
                    "foreground_popularity": 649,
                    "background_popularity": 779
                },
                {
                    "name": "aleve",
                    "relatedness": 0.4601,
                    "popularity": 390,
                    "foreground_popularity": 390,
                    "background_popularity": 390
                },
                {
                    "name": "naproxen",
                    "relatedness": 0.26633,
                    "popularity": 520,
                    "foreground_popularity": 520,
                    "background_popularity": 2598
                },
                {
                    "name": "ibuprofen",
                    "relatedness": 0.2508,
```

Disambiguation by Category Example

```
{
  params: {
    fore: "bbq",
    back: "*:*",
    qf: "text_t",
    defType: "edismax"
  },
  query: "bbq",
  limit: 0,
  facet: {
    categories : {
      type : type,
      field : "type_ss",
      limit: 2,
      sort: { "r1": "desc" },
      facet : {
        r1 : "relatedness($fore,$back)",
        related_terms : {
        type : terms,
        field : "text_t",
        limit: 3,
        sort: { "r2": "desc" },
        facet : {
          r2 : "relatedness($fore,$back)"
          }
        }
      }
    }
  }
}
```



Meaning 1: Restaurant => bbq, brisket, ribs, pork, …
Meaning 2: Outdoor Equipment => bbq, grill, charcoal, propane, …

Search

STATUS
Auto (30962)

PUBLISHED
Yes (30870)
No (92)

CONFIDENCE
0 1

Misspelling Detection

SIMULATOR >>

92 unpublished changes

PUBLISH

Fusion replaces misspelled query terms with their spelling corrections for better search results. Schedule the Token and Phrase Spell Correction job to get AI-generated spelling corrections based on your signals collection.

Misspelling	Suggested Correction	Confidence	Edited On ↓	Status	Published
laptopo	laptop	0.5	04/09/19 08:39	Auto	Yes
joybird	jaybird	1	04/09/19 08:39	Approved / Denied	Yes
e case	encase	0.5	04/09/19 08:39		Yes
shredde	shredder	0	04/09/19 08:39	Auto	Yes
flatsceen	flatscreen	1	04/09/19 08:39	Auto	Yes
i live	ilive	1	04/09/19 08:39	Auto	Yes
antholigy	anthology	1	04/09/19 08:39	Auto	Yes
dazzale	dazzle	1	04/09/19 08:39	Auto	Yes
rep blic	republic	1	04/09/19 08:39	Auto	Yes
reveng	revenge	0.5	04/09/19 08:39	Auto	Yes
acvessory	accessory	0.5	04/09/19 08:39	Auto	Yes
storagw	storage	0.5	04/09/19 08:39	Auto	Yes
correl	corel	1	04/09/19 08:39	Auto	Yes
beat bry	beat by	0.5	04/09/19 08:39	Auto	Yes
a droid	android	0.5	04/09/19 08:39	Auto	Yes
flyerhtc	flyer htc	1	04/09/19 08:39	Auto	Yes

Displaying 1 - 20 of 30962 results

< Previous 1 2 3 Next >

DIJIHUB

samsung black laptopo × Search Clear

Trends | Search | Search with Query Rewrite | Search with Signals Aggr | Compare | Personalization Aggregation

Brand -

Search...

Samsung 2,280
HP 763
Sony 745
Tribeca 736
GE 727
Lenmar 502
Whirlpool 461
Toshiba 441
Frigidaire 397
Dell 371
KitchenAid 300
Metra 294
Show more

Department -

Search...

Photo/Commodities 6,098
Video/Compact Disc 5,571
Computers 5,163
Appliance 4,058
Audio 3,342
Digital Communication 2,414

Info

Query rewrite triggered for spell **laptopo**, replace with **laptop**

Modified query is **samsung black laptop**

Search results are based on **dcommerce_main2_query_rewrite** pipeline

samsung black laptopo ×

SHOWING 1 - 10 OF 34,186 • Debug

Sort by **relevance**

1 2 3 4 5 next

Samsung - 15.6" Laptop - 6GB Memory - 640GB Hard Drive - Black

This laptop features a 640GB hard drive for abundant storage and a built-in high-speed wireless LAN for an easy Internet connection.Windows 7 Home Premium installed Learn more.

Dept **Computers** Brand **Samsung** Color **Black**

ID **3603163**

$646

Score 22.64946

Samsung - 15.6" Laptop - 6GB Memory - 500GB Hard Drive - Black

This Samsung NP305V5A-A09US laptop features a 1.3MP high-definition webcam for video chatting and wireless networking for a quick and easy connection to the Internet. The double-layer DVD±RW/CD-RW drive allows you to create customized CDs and DVDs.Windows 7 Home Premium installed Learn more.

Dept **Computers** Brand **Samsung** Color **Black**

ID **4700152**

$556

Score 22.105953

Search

STATUS
Pending (278)
Approved (1)
Denied (1)

PUBLISHED
Yes (2)
No (278)

CONFIDENCE
0 1

Synonym Detection

SIMULATOR >>

278 unpublished changes

PUBLISH

Synonyms are used to expand queries so that they include all known synonyms of the query terms. Schedule the Synonym and Similar Queries Detection job to get AI-generated synonyms based on your signals collection.

Search Term	Synonyms	Confidence	Edited On ↓	Status	Published
harmony	universal remote, remote, logitech	0.02843...	04/19/19 10:21	Approved	Yes
music	reciever, piano	0.16679...	04/16/19 16:13	Denied	Yes
speaker	wireless speaker	0.02403...		Pending	No
blueray	complete, movies	0.02130...			No
literacy	logictech	0.1875			No
headset	wireless headset	0.02892...		Pending	No
hdr	hdrcx	0.01718...		Pending	No
invisibleshield	zagg	0.01199...		Pending	No
floor	vacuum	0.06024...		Pending	No
screen	screen protector	0.17502...		Pending	No
audio cable	cable	0.01482...		Pending	No
gps	navigation	0.01358...		Pending	No
cam	camcorder, video camera, camra	0.01665...		Pending	No
computer speaker	promedia	0.05882...		Pending	No
clean	lens	0.03759...		Pending	No
ark	arkhangelsk, arkam, arkhan	0.10057...		Pending	No

Approved
Denied

Displaying 1 - 20 of 280 results

< Previous 1 2 3 Next >

DIJIHUB

harmony × Search Clear

Trends | Search | Search with Query Rewrite | Search with Signals Aggr | Compare | Personalization Aggregation

Brand

Search...

- Celestial Harmonies 178
- Harmony 45
- Black Sun/Celestial Harmonies 34
- MTV Games 21
- Logitech 18
- Ruthless 14
- GE 13
- Collectables 12
- Primarily A Cappella 12
- Able Planet 10
- Foreign Media Group 10
- BCI Music 9

Show more

Department

Search...

- Video/Compact Disc 620
- Interactive Software 85
- Audio 27
- Accessories 21
- Appliance 16
- Computer Software 14
- Computers 3

Color

- Black 9
- White-On-White 5
- Dark Platinum Color 3
- Metallic Silver 3

Suggested Searches

harmony logitech | remote controls | logitech remote | logitech remote control | logitech remotes | harmony remote | harmony remotes | rf remote | logitech harmony remote | harmony 900

harmony ×

SHOWING 1 - 10 OF 791 • Debug

Sort by relevance

1 2 3 4 5 next

Antares - Harmony Engine Evo Vocal Modeling Harmony Generator Software

Create realistic, natural-sounding harmonies from a single vocal line with this Antares Harmony Engine Evo vocal modeling harmony generating plug-in that features Evo Voice Processing technology for fast and accurate pitch detection.

Dept Audio Brand Antares

ID 4947502
$199
Score 11.661513

Dreaming Awake - CD

Dept Video/Compact Disc Brand Massacre Records
Artist Harmony

ID 14532561
$24
Score 11.361046

Chapter II: Aftermath - CD

Dept Video/Compact Disc Brand Ulterium Records
Artist Harmony

ID 9167616
$14
Score 11.361046

Roland - BOSS Harmonist Stompbox

This stompbox features 4 effect modes, including harmony, pitch shifter, detune and s-bend for a wide variety of tonal effects. Easily create 2- or 3-voice harmonies with selectable key and voicing for customized sound.

Dept Audio Brand Roland

ID 1196757
$159
Score 10.594049

Harmony - CD

Dept Video/Compact Disc Brand Rhino Artist Never Shout

Harmony - CD

Dept Video/Compact Disc Brand The Orchard Artist Frank

DIJIHUB

harmony × Search Clear

Trends | Search | Search with Query Rewrite | Search with Signals Aggr | Compare | Personalization Aggregation

Brand

Search...

- Universal 3,457
- Universal Music Latino 2,801
- Universal Special Products 1,210
- Universal Records 583
- Logitech 416
- Universal International 312
- Universal Republic 224
- Universal Music 217
- Celestial Harmonies 178
- Universal Import 131
- Universal Motown 113
- Uptown/Universal 88

Show more

Department

Search...

- Video/Compact Disc 10,923
- Interactive Software 1,142
- Computers 484
- Photo/Commodities 396
- Accessories 310
- Audio 257
- Mobile Audio 248
- Appliance 211
- Digital Communication 148
- Computer Software 118
- Magnolia Home Theatr 102
- Video 48

Show more

Color

Info

Query rewrite triggered for synonym **harmony**, expand with **universal remote,remote,logitech**

Modified query is **harmony universal remote remote logitech**

Search results are based on **dcommerce_main2_query_rewrite** pipeline

Suggested Searches

harmony logitech | remote controls | logitech remote | logitech remote control | logitech remotes | harmony remote | harmony remotes | rf remote | logitech harmony remote | harmony 900

harmony ×

SHOWING 1 - 10 OF 14,696 • Debug

Sort by relevance

1 2 3 4 5 next

Logitech - Harmony 700 - 8 Device Universal Remote

This universal remote features a color LCD screen and lets you control up to 8 A/V devices.

Dept Accessories Brand Logitech

ID 9365402

$89

Score 49.655983

Logitech - Harmony 900 - 15 Device Universal Remote with RF

This universal remote features a LCD touch screen and lets you control up to 15 A/V devices.

Dept Accessories Brand Logitech

ID 9365395

$243

Score 49.002407

Logitech Harmony 1100 Universal Remote & Wireless Extender Package

Take command of your expansive home entertainment center with the Logitech Harmony 1100 15-device universal remote. The package also includes a free Logitech wireless extender, which gives your remote a 100' range through cabinets and walls. Plus, save $50 on the remote.

Logitech - Harmony 670 Advanced Universal Remote with LCD Screen

Stop fumbling with multiple remotes. This universal remote lets you take control of your home entertainment system with the help of a clear, interactive LCD screen. Learn about the advantages of Harmony Setup and Programming Resources. Click here to see more products from Logitech.

Dept Accessories Brand Logitech

Example Query:

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "sort":"popularity desc",
      "matchText":"true",
      "json.nl":"map",
      "field":"name_tag",
      "echoParams":"all",
      "fl":"id,type,canonical_form,name,country:countrycode_s,admin_area:admin_code_1_s,popularity,*_p,command_function"}},
  "tagsCount":2,
  "tags":[{
      "startOffset":4,
      "endOffset":8,
      "matchText":"near",
      "ids":["5",
        "1",
        "6"]},
    {
      "startOffset":9,
      "endOffset":20,
      "matchText":"haystack eu",
      "ids":["27"]}],
  "response":{"numFound":4,"start":0,"docs":[
      {
        "type":"command",
        "popularity":10.0,
        "command_function":"cmd_text_distance(quer
        "canonical_form":"{text_distance}",
        "id":"5"},
      {
        "type":"command",
        "popularity":90.0,
        "command_function":"cmd_location_distance(
        "canonical_form":"{location_distance}",
        "id":"1"},
      {
        "type":"event",
        "popularity":100.0,
        "canonical_form":"haystack eu conference",
        "id":"27"},
      {
        "type":"command",
        "popularity":10.0,
        "command_function":"cmd_text_distance(query, position)",
        "canonical_form":"{text_distance}",
        "id":"6"}]
  }}
```


bbq near haystack eu

TAGGED: bbq {near} {haystack eu}

PARSED: { type:keyword, known: false, surface_form: "bbq"}
{"type":"command","popularity":90,"command_function":"cmd_location_distance(query, position)","canonical_form":"{location_distance}","id":"1"}
{"type":"event","popularity":100,"canonical_form":"haystack eu conference","id":"27"}

TO SOLR: bbq^0.91032 ribs^0.65674 brisket^0.63386 doc_type:"restaurant" {!geofilt d=50 sfield="coordinates_pt" pt="52.554930,13.383930"}

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "sort":"popularity desc",
      "matchText":"true",
      "json.nl":"map",
      "field":"name_tag",
      "echoParams":"all",
      "fl":"id,type,canonical_form,name,country:countrycode_s,admin_area:admin_code_1_s,popularity,*_p,command_function"}},
  "tagsCount":2,
  "tags":[{
      "startOffset":4,
      "endOffset":8,
      "matchText":"near",
      "ids":["5",
        "1",
        "6"]},
    {
      "startOffset":9,
      "endOffset":17,
      "matchText":"activate",
      "ids":["26"]}],
  "response":{"numFound":4,"start":0,"docs":[
      {
        "type":"command",
        "popularity":10.0,
        "command_function":"cmd_text_distance(query
        "canonical_form":"{text_distance}",
        "id":"5"},
      {
        "type":"command",
        "popularity":90.0,
        "command_function":"cmd_location_distance(
        "canonical_form":"{location_distance}",
        "id":"1"},
      {
        "type":"event",
        "popularity":100.0,
        "canonical_form":"activate conference",
        "id":"26"},
      {
        "type":"command",
        "popularity":10.0,
        "command_function":"cmd_text_distance(query, position)",
        "canonical_form":"{text_distance}",
        "id":"6"}]
  }}
```


bbq near activate

TAGGED: bbq {near} {activate}

PARSED: { type:keyword, known: false, surface_form: "bbq"}
{"type":"command","popularity":90,"command_function":"cmd_location_distance(query, position)","canonical_form":"{location_distance}","id":"1"}
{"type":"event","popularity":100,"canonical_form":"activate conference","id":"26"}

TO SOLR: bbq^0.91032 ribs^0.65674 brisket^0.63386 doc_type:"restaurant" {!geofilt d=50 sfield="coordinates_pt" pt="38.916120,-77.045220"}

Demo Data

bbq near activate

[Web crawl of restaurant and product reviews sites]

Places (also includes geonames database)

```
[{
        "id": "1",
        "type": ["event"],
        "name": "Activate Conference",
        "description": "The Search and AI Conference",
        "has_location": true,
        "venue": "Washington Hilton"
        "popularity": 96
    },
    {
        "id": "1",
        "type": ["event"],
        "name": "Haystack  Conference",
        "description": "The Search Relevance Conference",
        "has_location": true,
        "venue": "Violet Crowne",
        "popularity": 95
    },
        {
        "id": "13",
        "type": ["hotel"],
        "name": "Washington Hilton",
        "description": "Hilton in Washington, DC",
        "has_location": true,
        "coordinates": "38.916120,-77.045220",
        "address": "1919 Connecticut Ave. NW, Washington DC, 20009",
        "popularity": 82
    },
    {
        "id": "3",
        "type": ["restaurant"],
```

Entities (includes search commands)

```
id,surface_form,canonical_form,type,popularity,command,function
1,near,"{location_distance}",command,90,"cmd_location_distance(query, position)"
2,in,"{location_distance}",command,100,"cmd_location_distance(query, position)"
3,around,"{location_distance}",command,100,"cmd_location_distance(query, position)"
4,by,"{location_distance}",command,90,"cmd_location_distance(query, position)"
5,by,"{text_within_one_edit_distance}",10,"cmd_text_within_one_edit_distance(query, position)"
6,near,"{text_distance}",command,10,"cmd_text_distance(query, position)"
7,popular,{popular},command,100,"cmd_popularity(query, position)"
8,top,{popular},command,100,"cmd_popularity(query, position)"
9,best,{popular},command,100,"cmd_popularity(query, position)"
10,good,{popular},command,100,"cmd_popularity(query, position)"
11,hilton,hilton,brand,100
12,washington hilton,washington hilton,hotel,100
13,washington dc hilton,washington hilton,brand,100
14,washington d.c. hilton,washington hilton,hotel,100
15,activate,activate conference,event,100
16,activate conf,activate conference,event,100
17,activate conference,activate conference,event,100
18,activat,activate conference,event,100
19,activat conf,activate conference,event,100
20,activat conference,activate conference,event,100
21,lucidworks activate,activate conference,event,100
22,violet,violet,color,100
23,violet crowne,violet crowne,brand,100
24,violet crowne charlottesville,violet crowne charlottesville,movie_theater,100
25,violet crown,violet crowne,brand,100
26,violet crown charlottesville,violet crowne charlottesville,movie_theater,100
27,haystack,haystack conference,event,100
28,haystack conf,haystack conference,event,100
29,haystack conference,haystack conference,event,100
30,heystack,haystack conference,event,100
31,heystack conf,haystack conference,event,100
32,heystack conference,haystack conference,event,100
33,around,"{location_distance}",command,100,"cmd_location_distance(query, position)"
34,is,is,keyword,100
```

Solr Knowledge Graph Traversal Query

```
{
  params: {
    unknown_phrase_1: "bbq",    location_keyword_1: "activate",
    content_field: "text", doc_type_field: "type", name_field: "name_s", name_search_field: "name_t",
    venue_field: "venue", address_field: "address", coordinates_field: "coordinates",
    all_docs: "*:*", echoParams: none
  },
  query: "{!edismax v=$all_docs}", limit: 0,
  facet: {
    $unknown_phrase_1: {
        type: query, query: "{!type=edismax qf=$content_field v=$unknown_phrase_1}",
        facet: {
            doc_type : {
              type : terms, field : ${doc_type_field}, limit: 1, sort: { "r1": "desc" },
              facet : {
                r1 : "relatedness($unknown_phrase_1,$all_docs)",
                related_terms : {
                  type : terms, field : "${content_field}", limit: 4, sort: { "r2": "desc" },
                  facet : {
                    r2 : "relatedness($unknown_phrase_1,$all_docs)"
                  }
                }
              }
            }
        }
    },
    location_1: {
        type: query,
        query: "{!type=graph from=${name_field} to=${venue_field} returnOnlyLeaf=true}{!type=edismax qf=${name_search_field} v=${location_keyword_1}}",
        facet: {
            "venue": {
                type:terms, field:${name_field}, limit:1,
                facet: {
                  "address": {type:terms, field:${address_field}},
                  "coordinates": {type:terms, field:${coordinates_field}}
                }
            }
        }
    }
  }
}
```

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "sort":"popularity desc",
      "matchText":"true",
      "json.nl":"map",
      "field":"name_tag",
      "echoParams":"all",
      "fl":"id,type,canonical_form,name,country:countrycode_s,admin_area:admin_code_1_s,popularity,*_p,command_function"}},
  "tagsCount":3,
  "tags":[{
      "startOffset":0,
      "endOffset":3,
      "matchText":"top",
      "ids":["8"]},
    {
      "startOffset":8,
      "endOffset":10,
      "matchText":"in",
      "ids":["2"]},
    {
      "startOffset":11,
      "endOffset":17,
      "matchText":"berlin",
      "ids":["2950159",
        "4930431",
        "4556518",
        "5083330",
        "5245497",
        "4348460",
        "4500771",
        "3614789",
        "3587266"]}],
  "response":{"numFound":11,"start":0,"docs":[
      {
        "location_p":"52.52437,13.41053",
        "type":"city",
        "popularity":0.15354161869920396,
        "name":["Berlin"],
        "canonical_form":"Berlin",
        "id":"2950159",
        "country":"DE",
        "admin_area":"16"},
      {
        "type":"command",
        "popularity":100.0,
        "command_function":"cmd_popularity(query, position)",
        "canonical_form":"{popular}",
        "id":"8"},
      {
        "type":"command",
        "popularity":100.0,
        "command_function":"cmd_location_distance(query, position)",
        "canonical_form":"{location_distance}",
        "id":"2"},
      {
        "location_p":"42.3812,-71.63701",
        "type":"city",
        "popularity":1.0853455319837705E-4,
        "name":["Berlin"],
        "canonical_form":"Berlin",
        "id":"4930431"
```


top bbq in berlin

TAGGED: {top} bbq {in} {berlin}

PARSED: {"type":"command","popularity":100,"command_function":"cmd_popularity(query, position)","canonical_form":"{popular}","id":"8"}{ type:keyword, known: false, surface_form: "bbq"}{"type":"command","popularity":100,"command_function":"cmd_location_distance(query, position)","canonical_form":"{location_distance}","id":"2"} {"location_p":"52.52437,13.41053","type":"city","popularity":0.15354161869920396,"name": ["Berlin"],"canonical_form":"Berlin","id":"2950159","country":"DE","admin_area":"16"}

TO SOLR: {!func}mul(min(popularity,1),100) bbq^0.91032 ribs^0.65674 brisket^0.63386 doc_type:"restaurant" {!geofilt d=50 sfield="location_p" pt="52.52437,13.41053"}

Why this Semantic Nuance Matters

bbq near activate

TAGGED: bbq {near} {activate}

PARSED: { type:keyword, known: false, surface_form: "bbq"}
{"type":"command","popularity":90,"command_function":"cmd_location_distance(query, position)","canonical_form":"{location_distance}","id":"1"}
{"type":"event","popularity":100,"canonical_form":"activate conference","id":"26"}

TO SOLR: bbq^0.91032 ribs^0.65674 brisket^0.63386 doc_type:"restaurant" {!geofilt d=50 sfield="coordinates_pt" pt="38.916120,-77.045220"}

bbq grill

TAGGED: bbq {grill}

PARSED: { type:keyword, known: false, surface_form: "bbq"} {type: product_category, canonical_form: "grill"}

TO SOLR: ("bbq" OR "grill"^0.722 "propane"^0.646 OR "charcoal"^0.613) AND type:"outdoor appliance" product_category:"grill"

Other Knowledge Graph Search examples:

popular barbeque near Haystack EU
(popular same as "good", "top", "best")

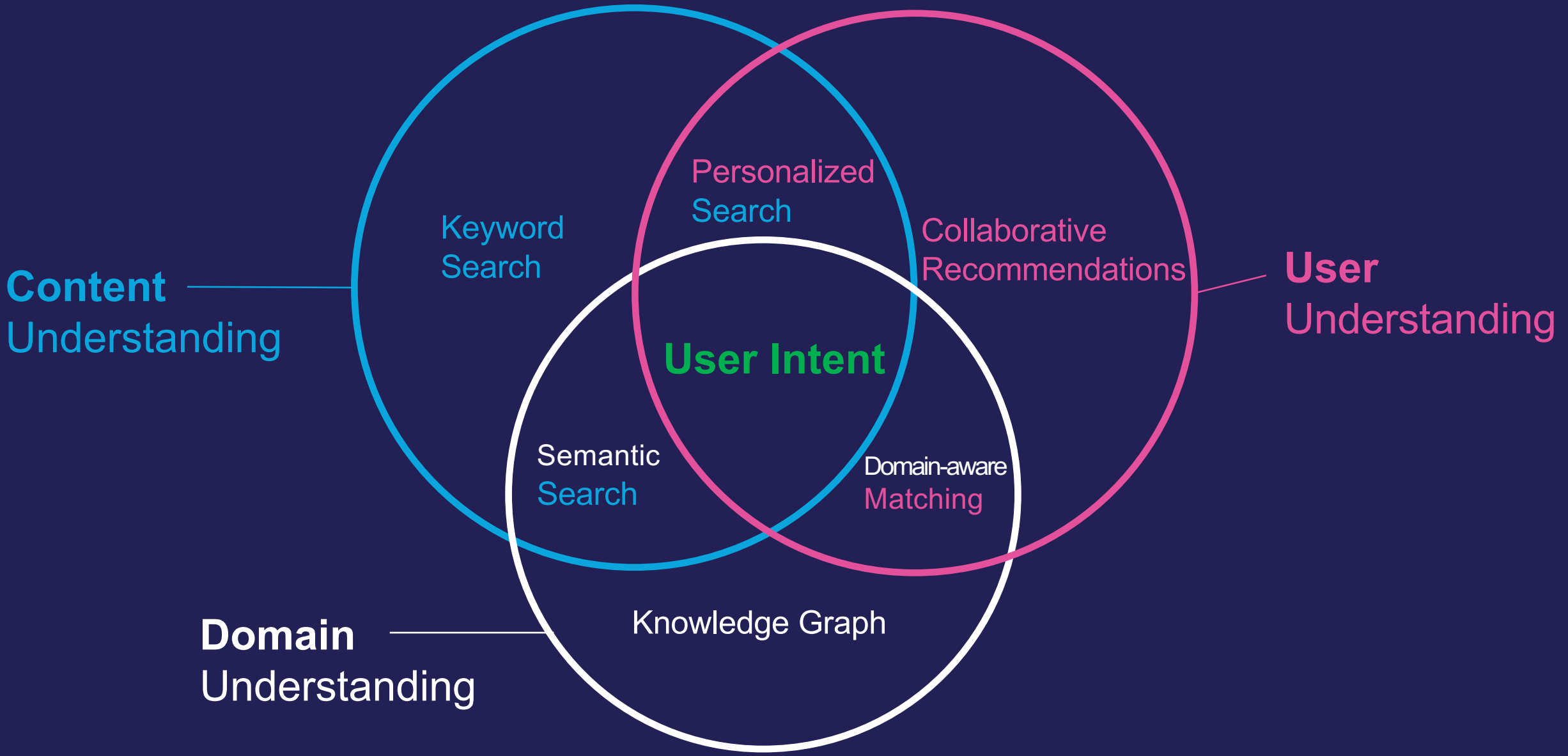
hotels near Haystack EU

hotels near popular BBQ in Berlin

BBQ near airports near Berlin

hotels near movie theaters in Berlin …

Dimensions of User Intent



The right ranking algorithm is domain and context-dependent

News Search : popularity and freshness drive relevance

Restaurant Search: geographical proximity and price range are critical

Ecommerce: likelihood of a purchase is key

Movie search: More popular titles are generally more relevant

Job search: category of job, salary range, and geographical proximity matter

Example Combining Content + Domain + User Context

News website:

```
/select?
    fq={!cache=false v=$keywords}&
    q=  {!func}scale(query($keywords),0,25)
        {!func}scale(geodist(),0,25)
        {!func}recip(rord(publicationDate),1,25,0)
        {!func}scale(popularity,0,25)&
    keywords="fall festival"&
    sfield=location&
    pt=33.748,-84.391
```

25%

25%

25%

25%

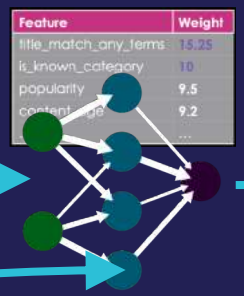
*Example from chapter 16 of *Solr in Action*

But how do we figure out the right balance of weights?

Learning to Rank

ipad

Initial Results:
1) doc1
2) doc2
3) doc3



Final Results:
1) doc3
2) doc1
3) doc2

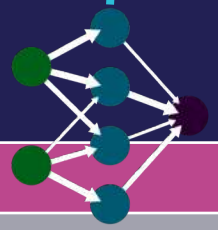
User	Query
Alonzo	**ipad**
Elena	**printer**
Ming	**ipad**
…	…

User Searches

User Sees Results

User takes an action

Users' actions inform system improvements

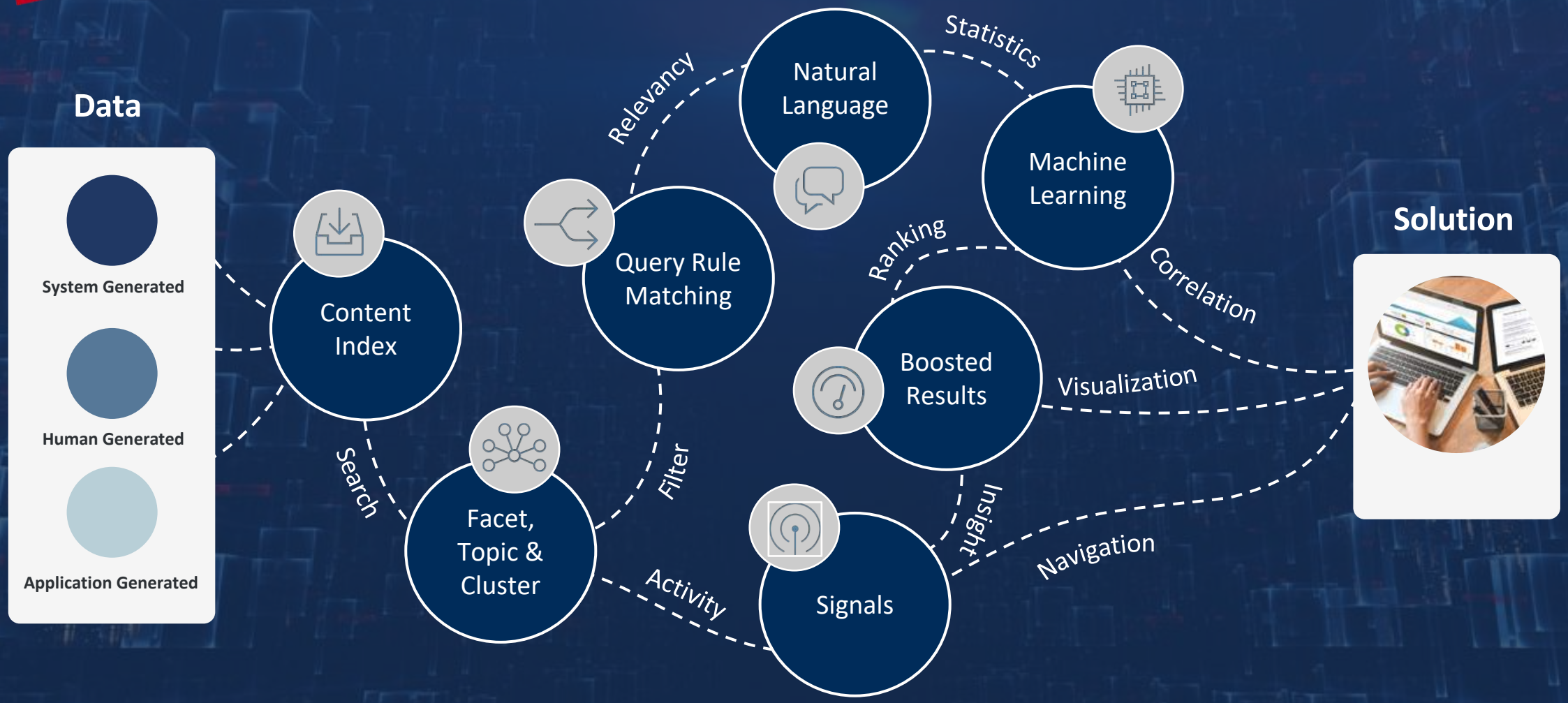


Feature	Weight
title_match_all_terms	**15.25**
exact_phrase_match	**10**
signal_boost	**9.5**
content_age	**9.2**
user_geo_distance	6.5
personalization_cat_1	2.8
doc_popularity	2.75
…	…

User	Action	Document
Alonzo	**click**	doc22
Elena	**click**	doc17
Ming	**click**	doc12
Alonzo	**purchase**	doc22
Ming	**click**	doc22
Ming	**purchase**	doc22
Elena	**click**	doc2
…	…	…

Build Ranking Classifier
(from Implicit Relevance Judgements)

Lucidworks


Data
System Generated
Human Generated
Application Generated
Content Index
Query Rule Matching
Natural Language
Machine Learning
Facet, Topic & Cluster
Boosted Results
Signals
Relevancy
Statistics
Ranking
Correlation
Search
Filter
Visualization
Insight
Navigation
Activity
Solution

Lucidworks

We operationalize AI for the largest businesses on the planet.


Questions?
HAYSTACK EU
Lucidworks

Thank You!

40% Discount code: **ctwhay19**

Trey Grainger

trey@lucidworks.com
@treygrainger
Lucidworks

Other presentations:

http://www.treygrainger.com

Books:

*http://*aiPoweredSearch*.com*
*http://*solrinaction*.com*