Lucidworks

Understanding Intention

USING CONTENT, CONTEXT, AND THE CROWD TO BUILD BETTER SEARCH APPLICATIONS

Table of Contents

INTRODUCTION

So, You've Got All This Data

YOU'VE PROBABLY HEARD THE STATISTIC: By the year 2020, gigabytes will outnumber humans 5,200:1. Usually accompanying this breathless warning is a Friedmanian mixed metaphor representing the infinite: grains of sand, haystacks, stars in the universe, craft beers, Brett Favre interceptions, you name it.

In this age of information, simply going about our daily routines means creating and collecting a staggering amount of data.

A typical American office worker produces

1.8 MILLION

MEGABYTES OF DATA EACH YEAR

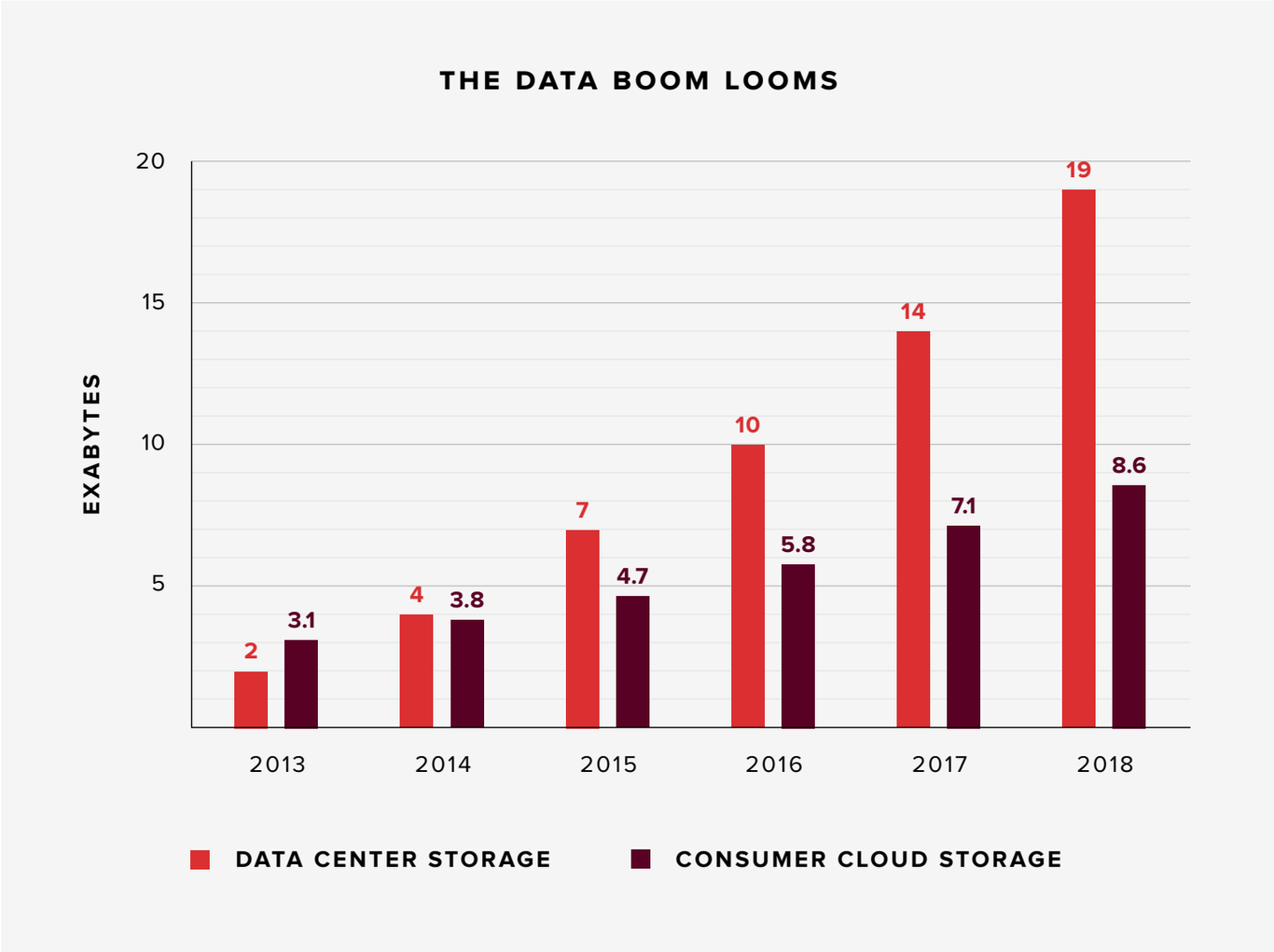
OR ABOUT

5,000

MEGABYTES A DAY.

Source: Technology Review

In just a few years, we've broadly accepted the notion that accumulating and storing all this data (instead of deleting it) is tremendously valuable. But with so much of it, we're now worried we won't be able to capture all of its value.



Source: Cisco

From Data Intrigue to Data Fatigue

"WHERE ARE WE GOING TO PUT ALL THIS STUFF?"
An EMC survey of 800 IT professionals found that 4 of their top 5 anxieties involved the storing, accessing, and security of their data. What's clear is that the stark reality of commodity storage is quickly replacing the fantastical utopian future of big data.

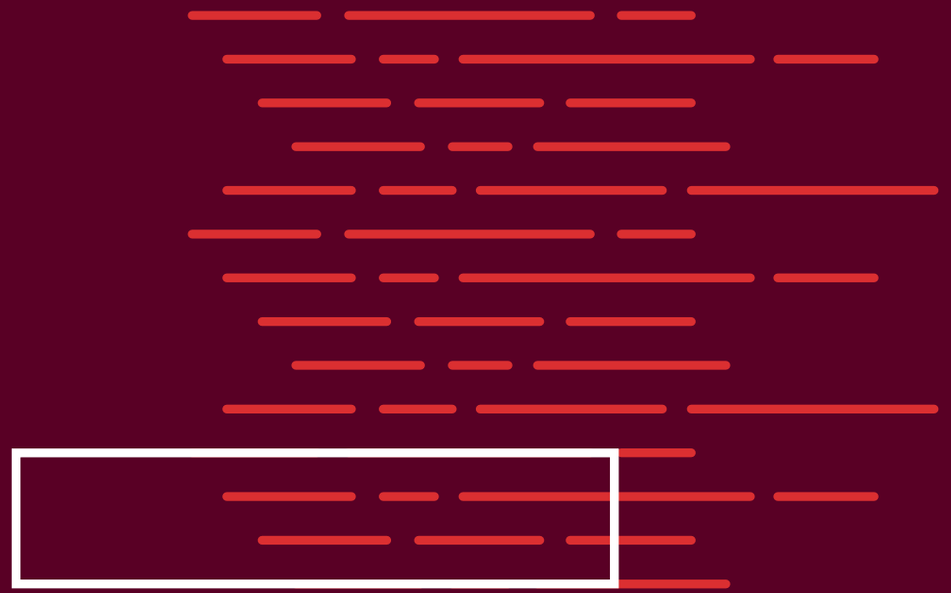
"HOW DO I USE ALL THIS STUFF TO MAKE INFORMED, STRATEGIC, AND BIG-PICTURE DECISIONS?"
Scaling and storage are solved problems. Now what?

The fifth concern voiced in the EMC survey was capturing the supposed game-changing value of data. In other words, I can see all the Brett Favre interceptions, but there are so many that they make my eyes glaze over. How can I pick up on patterns that'll tell me when and why those turnovers are happening?

The promise of big data isn't gone. It's just changed from "Where do we put all of this stuff?" to "How do we make all this data accessible to users in a meaningful way?"

2013/2014 STORAGE ANXIETIES IDENTIFIED BY IT PROFESSIONALS

Source: EMC



CHAPTER 1

Search is Easy

Search: From box to entry point

1,000,000,000,000

IN 2014, GOOGLE ALONE HANDLED MORE THAN 1 TRILLION SEARCHES. Factor in other Web behemoths like Amazon, Facebook, YouTube, and Twitter, and it becomes clear: Search technology touches every part of our daily lives, from how we shop, eat, and date, to how we consume, communicate, and celebrate.

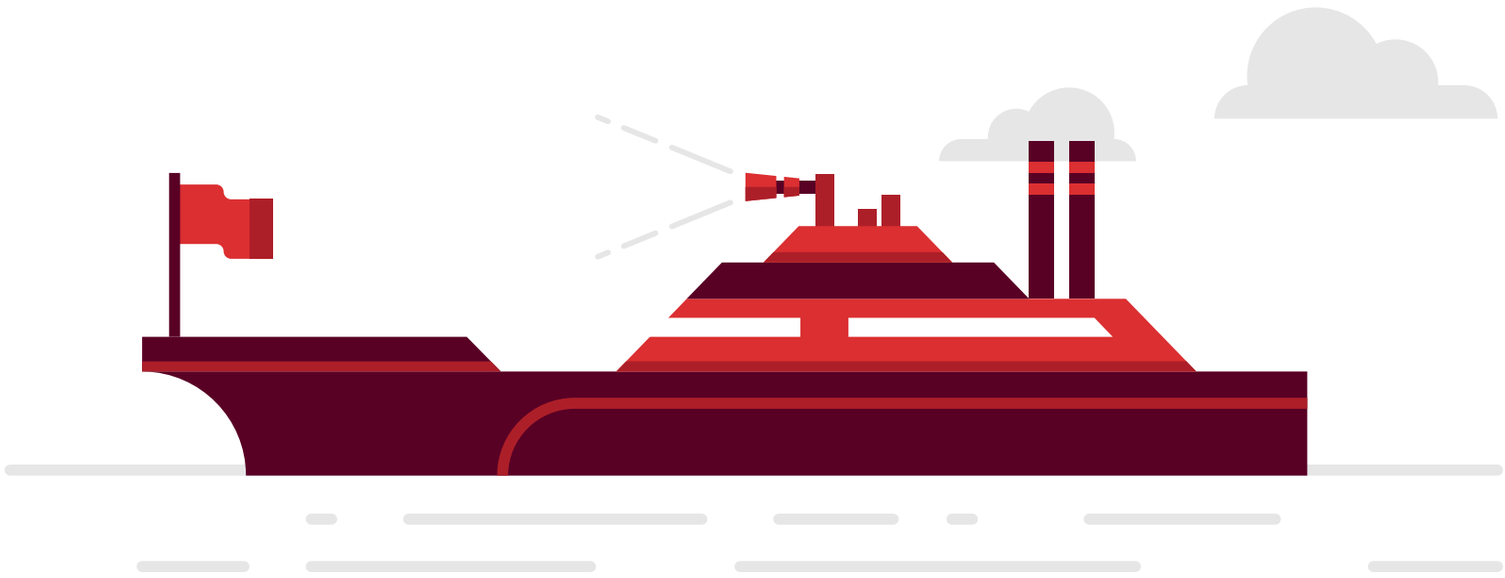
The success of the consumer Web proves that search is the entry point to extracting value and meaning from a nearly infinite (and infinitely growing) amount of data.

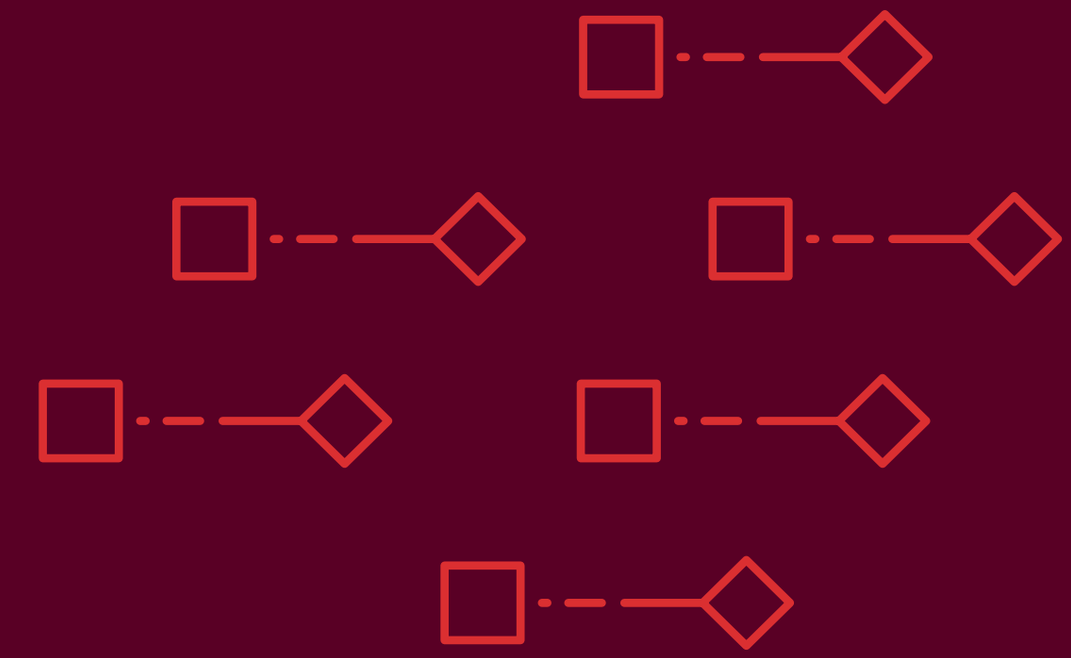
Looking for the search box as our starting point is second nature. We have faith in its ability to discover what we want to do or know and who we want to talk to.

Source: Google

As consumers, we navigate the data deluge aboard the USS Search. Given the obvious benefits, elevated stakes, and advantage of private investment, you'd think it'd be even easier for enterprises.

And you would be wrong.





CHAPTER 2

Enterprise Search Isn't So Easy

Shifting from storage to value means focusing on ease of use

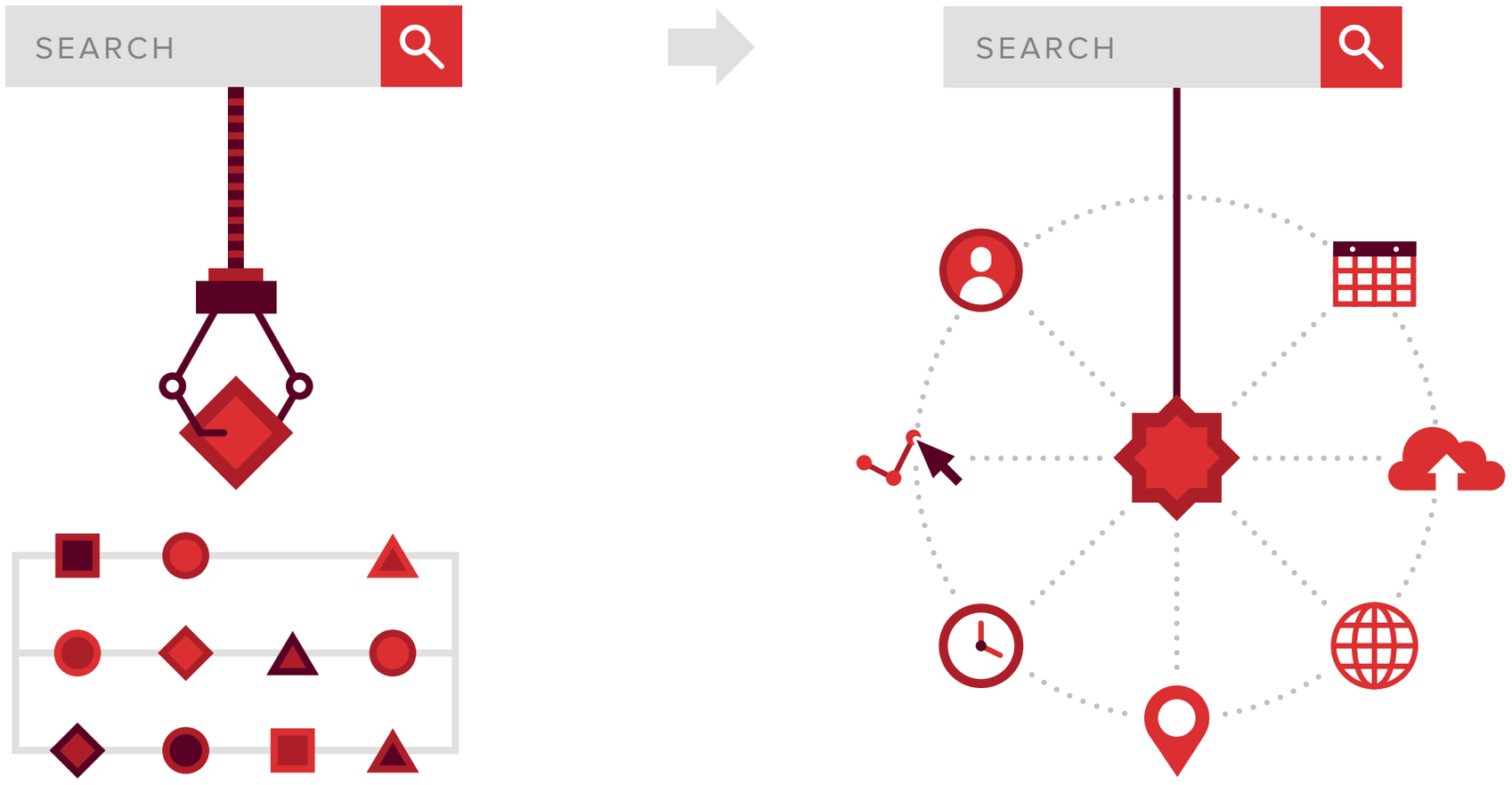
Businesses are People, Too

UNLIKE CONSUMER SEARCH, which has become a seamless part of our everyday lives, the enterprise side might as well still be running Windows 95. Imagine if Amazon, Google, or Facebook treated every user the same, regardless of who they are, where they are, what they're searching for, and what they've clicked.

Your users expect that same sophistication in their enterprise apps.

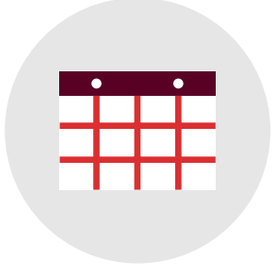
Easy Isn’t Enough

It has to be smarter.

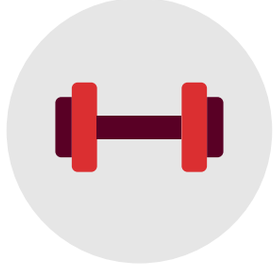


People Want Action

Restaurants where you've made OpenTable reservations in the past are highlighted in your Google Maps.



Premieres of your favorite TV shows pop up on your calendar.



A fitness wearable detects your blood pressure on the rise and schedules a gym visit on your calendar.

While browsing Spotify, a discounted ticket offer shows up for a movie you searched for the day before.

Your friend recommended the lasagna at a particular restaurant.

Relevancy is King

The key to relevancy is understanding your users' intentions.

Relevancy in a search app is comprised of **3 MAIN PARTS.**

Data and documents

Individual history and behavior

Similar users

Content

CONTENT REFERS TO DOCUMENTS AND DATA: all of the stuff you want to index, search, and retrieve.

Content comes in 2 main formats:

1. STRUCTURED DATA: Spreadsheets, databases, lists, network logs, and anything else that looks like a table

2. UNSTRUCTURED DATA: PDFs, documents, presentations, scanned documents, instant messages, emails, webpages, audio, video, and anything that doesn't fit neatly into a tabular format

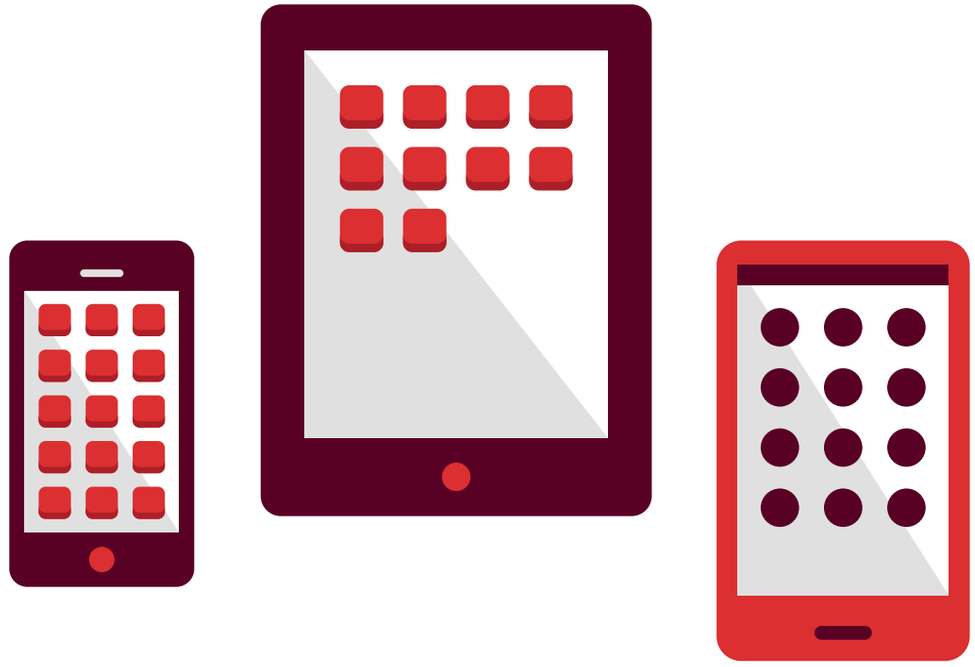
Content comes from many sources, including:

- NETWORK DRIVES
- INTRANET
- WIKIS
- SUPPORT TICKETS
- CLOUD STORAGE
- ON-PREMISE SERVERS
- VENDORS, PARTNERS
- HARD DRIVES
- MOBILE DEVICES
- EMAIL SERVERS
- NEWS SERVICES
- INSURANCE CLAIMS
- BANKING ACTIVITY
- CALL DETAIL RECORDS
- STOCK TICKERS
- NETWORK LOGS
- SOCIAL MEDIA STREAMS
- MEDICAL RECORDS

Ability to access data is enterprise content's primary challenge. To make sense of the data, it needs to be indexed, linked with your search apps, and made accessible to the correct users.

This is what relevancy means for enterprise search. Content can be used to drive relevancy when access controls are enforced and rich metadata is present—such as content classifications, author, and subject fields.

In an ideal world, we'd be able to tap into our data in real time from any device.



Context

OVERALL, THE CONTENT PART OF THE PUZZLE IS SOLVED. Technologies like Hadoop, Solr, and NoSQL have made accessing, indexing, and scaling data easier and more cost-effective than ever. The challenge, then, becomes zooming in. Not only do you need to be able to see the information within the documents and data itself, you need to understand how different files relate to each other and, further, how they relate to you (and other users like you). The second element of the relevancy trio is context.

Along with analyzing how all the bits of data are interrelated, there's the question of relatability. When a search app knows more about you, it can create a relevant search experience that helps you get personal, actionable search results on a consistent basis.

Search apps have solved that problem with signal processing. A signal is any bit of information that tells the app more about who you are. Signals can include your job title, business unit, location, device, and search history, as well as past actions within the search app like clickstream, purchasing behavior, direct reports, upcoming meetings or events, and more.

Crowd

YOU'RE NEARLY GOOD TO GO: All your data is in one place, available at any time, and your app is personalizing search results to provide deeper meaning and context to your requests. But you're not in uncharted waters. No doubt, others before (other users like you) have searched for similar things and navigated to what they needed. So, what about everybody else? Where did their searching take them?

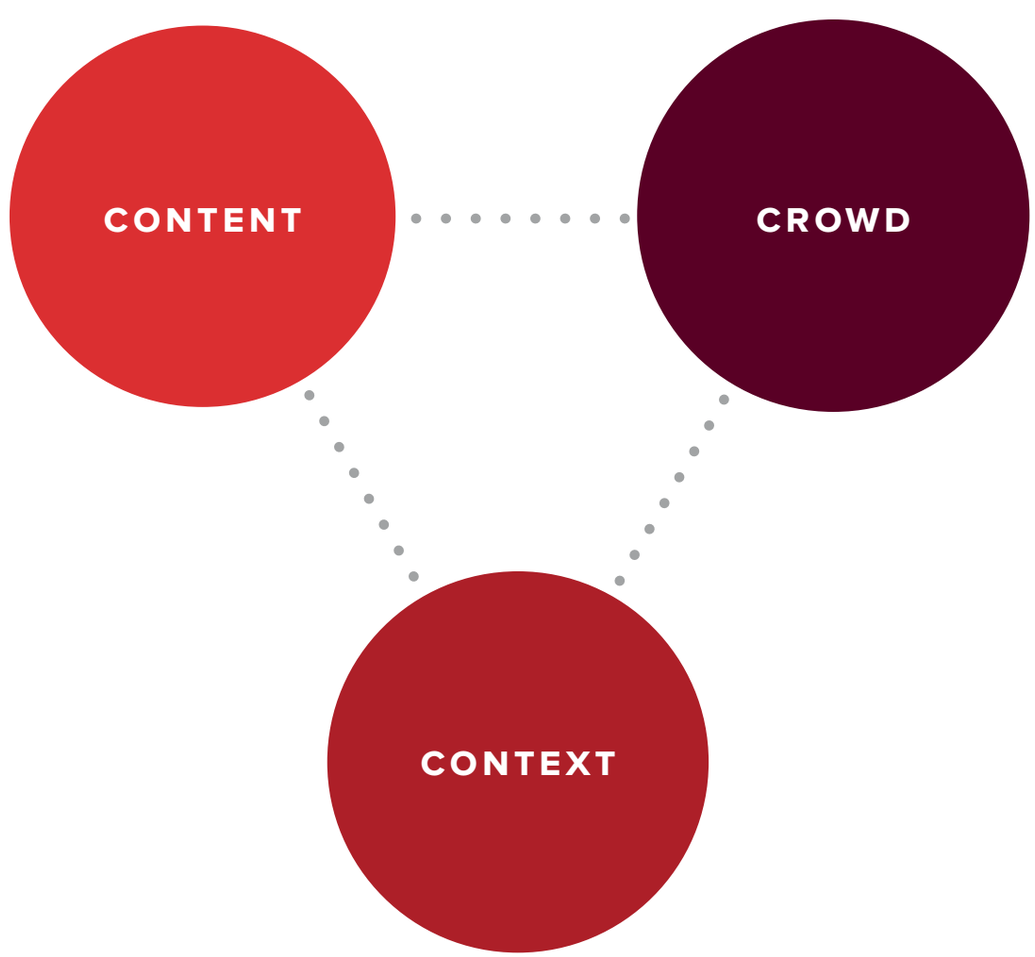
The final point of our relevancy triad is the crowd. When a search app uses the crowd, it goes beyond documents and data, past your specific user profile and relationship, and examines how other users are interacting with the data and information.

A search app knows the behavioral information of thousands—sometimes millions—of other users. By keeping track of every user, search apps can bubble up what you will find important and relevant and what other users like you will want, too. The tech uses its knowledge of your office, role, and demographic to match to the same in other users and make intelligent judgments about what will help you the most.

Content, Context, Crowd

This holy trinity — content, context, and crowd — is already a huge part of the success of consumer-facing search apps.

LET'S TAKE A LOOK at how these 3 ingredients can impact relevancy and importance on the enterprise side.

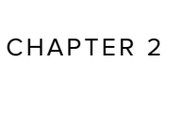


EXAMPLE 1

Ecommerce

CONTENT

Brand names, product descriptions, store availability, product category



CROWD

Cart and purchase patterns, similar customers' purchases

CONTEXT

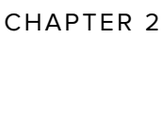
Clickstream, user demographic, loyalty program participation, in-store vs. online behavior, wish list

EXAMPLE 2

Security + Compliance

CONTENT

Audit logs, change requests and work tickets, claims logs, access logs, network logs, emails, instant messages, email attachments



CROWD

Behavioral patterns, user in role/ job function (e.g., lots of failed logins to a payroll system from a marketing user)

CONTEXT

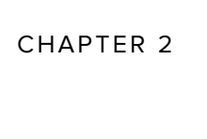
Location, time, sensitivity of content, attempted action, success or failure, associated agent role, business unit

EXAMPLE 3

Enterprise Search

CONTENT

MS Office documents; files in the cloud, intranet, and wiki



CROWD

Other business unit users, regionally popular documents

CONTEXT

Job role, business unit, security access

EXAMPLE 4

Fraud

CONTENT

Transactions, reported travel plans, verifications of previous transactions, insurance claims

CROWD

People with similar customer history; others buying similar items; others in the same location; patterns of fraudulent behavior, users in similar divisions, regions, or business units

CONTEXT

Location, transaction size, in-person vs. online, seller, items purchased, IP address, country of origin

EXAMPLE 5

Customer-Facing Data/Apps

CONTENT

User behavior, purchase history, social signals, referrals, search history, viewing history (Web/app logs), click path, search requests, geo pings

CROWD

People who looked at the same thing, people who bought the same thing, other similar items, similar people (who have bought or looked at the same things), others nearby, aggregated behavior patterns of users

CONTEXT

Items viewed, items abandoned, customer location, time of day, content related to things around your current location; free vs. paid customer

This has been a Lucidworks production.

Lucidworks builds enterprise search solutions for some of the world's largest brands. Fusion, Lucidworks' advanced search platform, provides the enterprise-grade capabilities needed to design, develop, and deploy intelligent search apps—at any scale. Companies across all industries, from consumer retail and health care to insurance and financial services, rely on Lucidworks every day to power their consumer-facing and enterprise search apps. Lucidworks' investors include Shasta Ventures, Granite Ventures, Walden International, and In-Q-Tel.

LEARN MORE AT LUCIDWORKS.COM.

Sources, in order of first use: Technology Review, Cisco, EMC, Google

Share this e-book with a colleague or friend.

Lucidworks

CREATED BY COLUMN FIVE

©2015 Lucidworks. All rights reserved.


Lucidworks