Precision & Recall

Measuring Search Effectiveness

- After finishing a search:

The nagging question in every searcher's mind is: "*Have I found the most relevant material or am I missing important items?“*

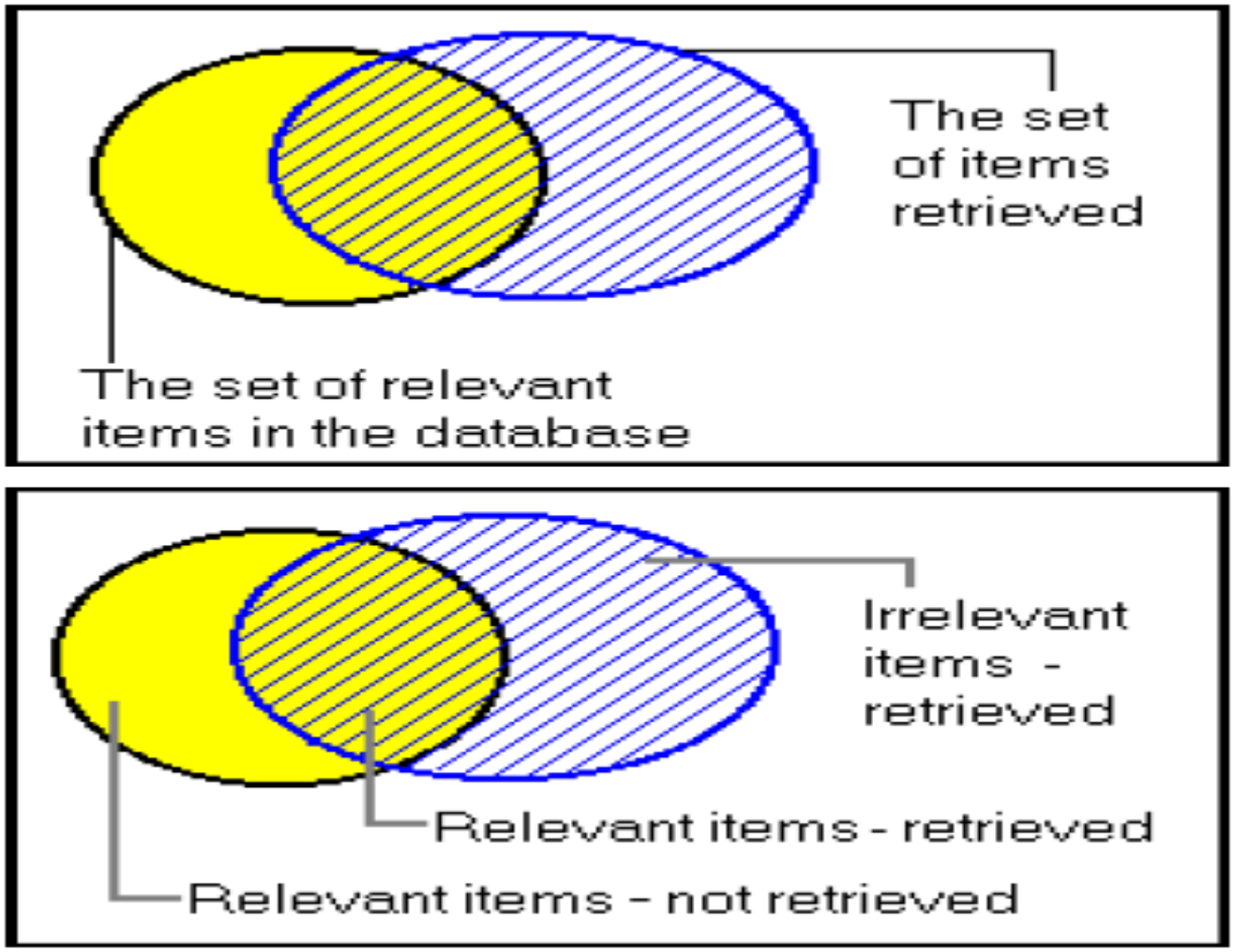
Measuring Search Effectiveness

- In addition every searcher hopes they don't retrieve *"a lot of junk".*

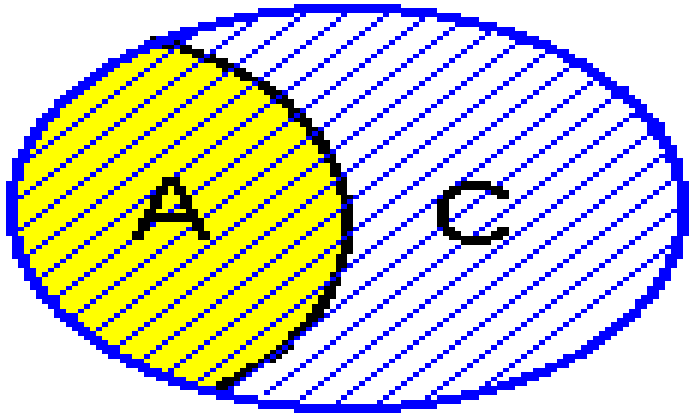
- Unfortunately getting *"everything" while avoiding "junk" is difficult.*
- *Is it* possible to measure how well a search performed with respect to these two parameters. (Correct Result & Incorrect Result)

Precision and recall

- Precision and recall are the basic measures used in evaluating search strategies.
- There is a set of records in the database which is relevant to the search topic
- Records are assumed to be either relevant or irrelevant
- The actual retrieval set may not perfectly match the set of relevant records.


The set of items retrieved
The set of relevant items in the database
Irrelevant items - retrieved
Relevant items - retrieved
Relevant items – not retrieved

PRECISION **is the ratio of the number of relevant records retrieved** to the **total number of irrelevant and relevant records retrieved.** It is usually expressed as a percentage.

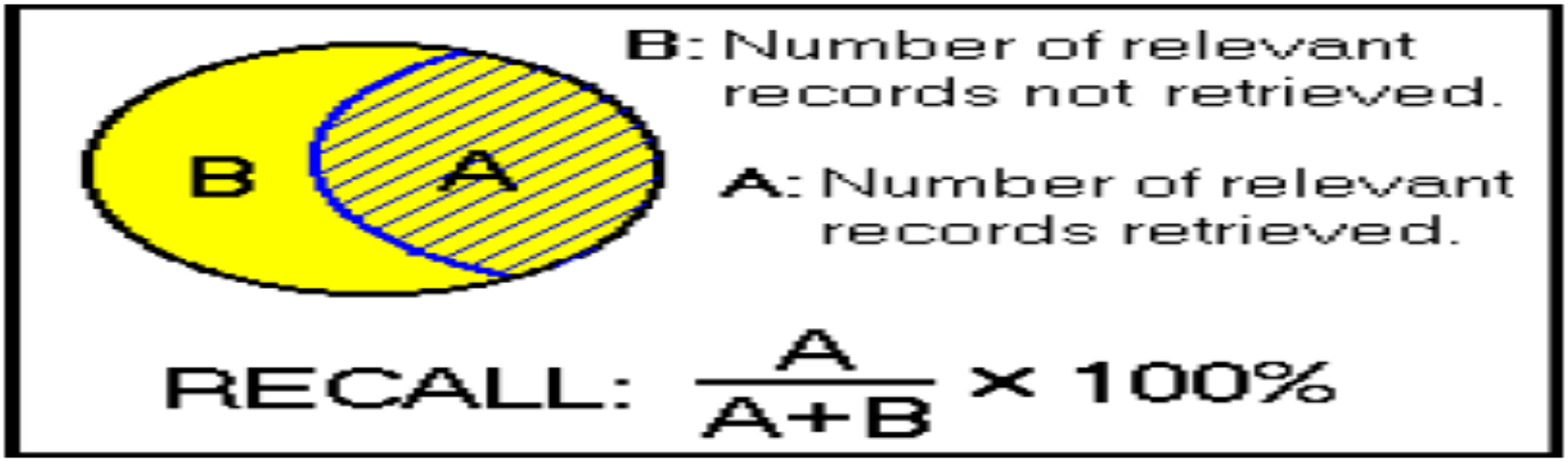


C: No. of irrelevant records retrieved.

A: No. of relevant records retrieved.

$$\text{PRECISION: } \frac{A}{A+C} \times 100\%$$

RECALL is the ratio of the number of relevant records retrieved to the **total number of relevant records in the database. It is usually expressed as a percentage.**



Recall and Precision are Inversely related

As recall ↑ precision ↓

conversely:

As recall ↓ precision ↑

- Example to think of *precision and recall* :
- If someone asked you to list the names of 5 presents you got last Deepawali but you couldn't exactly remember the 5 names — so you randomly guessed seven names. Out of the 7 names you remembered, 5 was recalled correctly while 2 were gifts you received on your birthday.
- Calculate precision and recall for this case

Calculate precision and recall for this case

- Even though you got a 100% recall (5/5)
- your precision was 71.4% (5/7).

- **Let's consider an example:**

- Let's say I searched on Google for *"what is precision and recall?"* and in less than a minute I have about 15,600,000 results.
- Let's say out of these 15.6 million results, the relevant links to my question were about 2 million. Assuming there were also about 6 million more results that were relevant but weren't returned by Google.
- for such system we would say that it has a precision of 2M/15.6M and a recall of 2M/8M.

Question

- Assume the following:
- A database contains **80 records on a particular topic**
- A search was conducted on that topic and **60 records were retrieved.**
- Of the **60** records retrieved, **45 were relevant.**

Calculate the precision and recall scores for the search.

Solution

- **Solution:**
- Using the designations above:
- A = The number of relevant records retrieved,
- B = The number of relevant records not retrieved, and
- C = The number of irrelevant records retrieved.

In this example

A = 45, **B = 35** ***(80-45)*** ***C = 15 (60-45).***

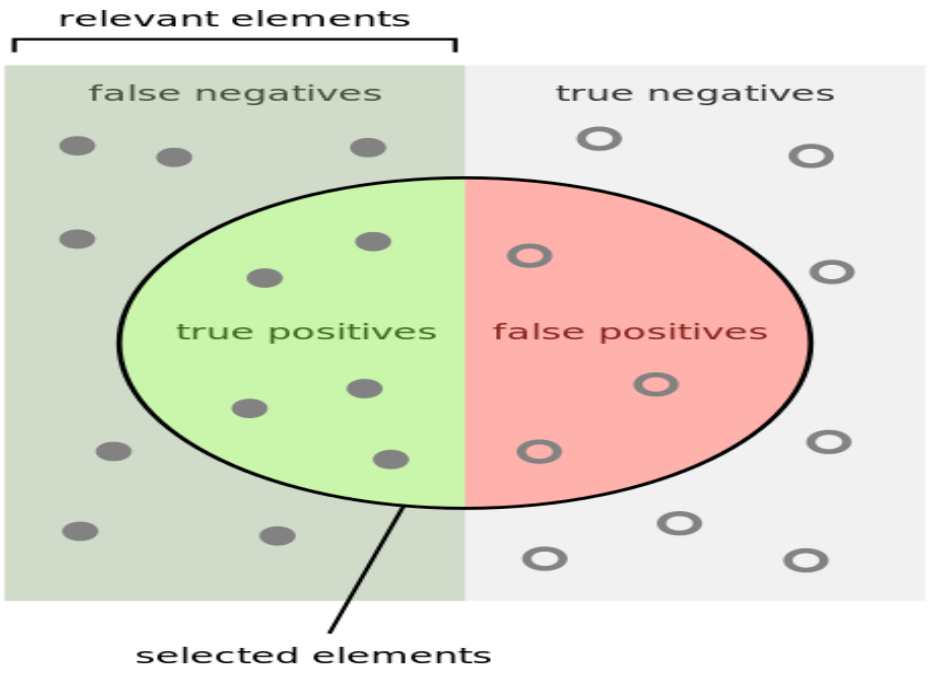
Recall = **(45 / (45 + 35)) * 100% => 45/80 * 100% = 56%**

Precision = **(45 / (45 + 15)) * 100% => 45/60 * 100% = 75%**

Quantifying Precision and recall

- **Precision** can be seen as a measure of exactness or quality, whereas **recall** is a measure of completeness or quantity.
- In simple terms, high **precision** means that an algorithm returned substantially more relevant results than irrelevant ones,
- while high **recall** means that an algorithm returned most of the relevant results.

Left rectangle is what we should have predicted as positive, and **Left half-disc** is what we did predict as positive, correctly.



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

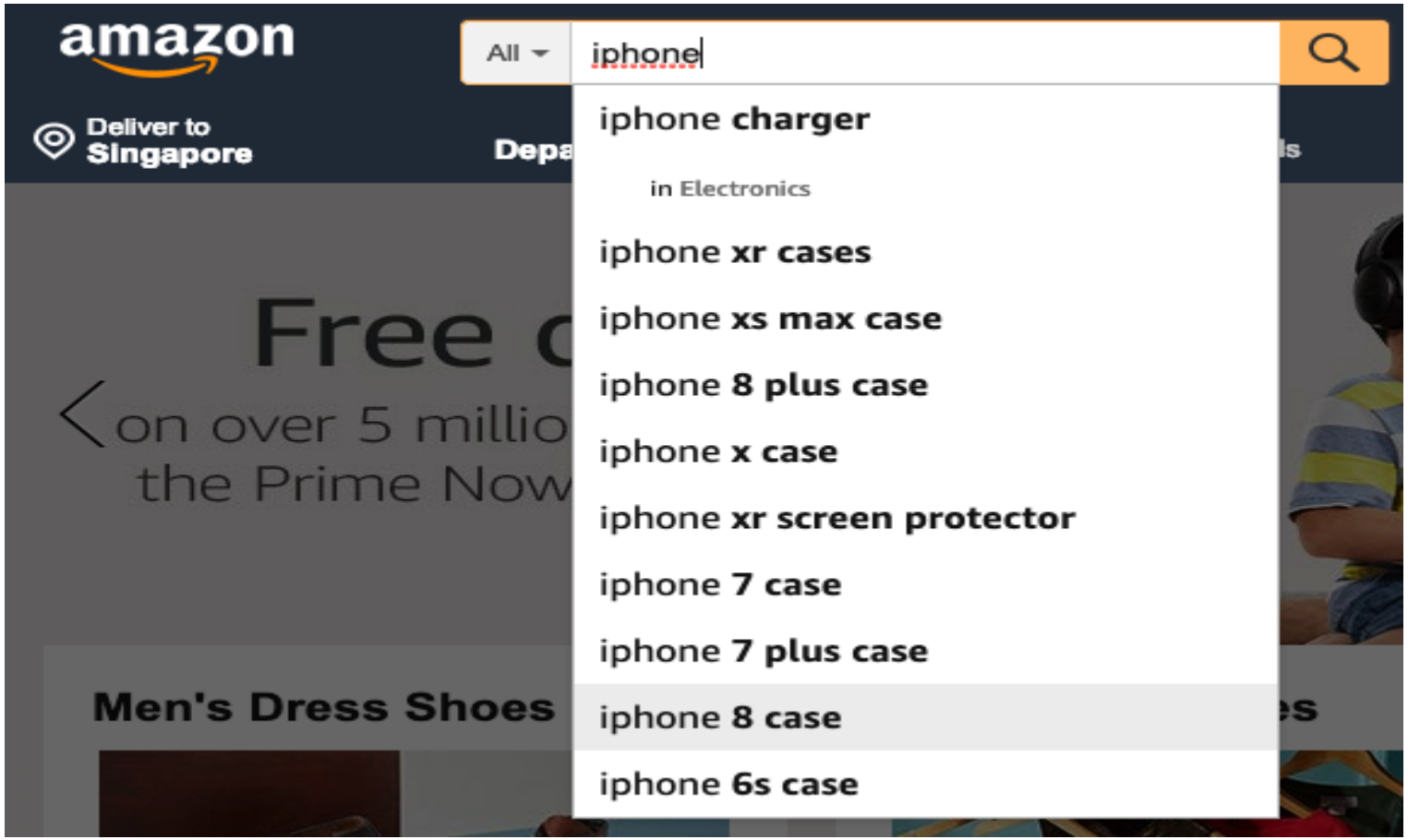
Precision

- **Precision** is to measure the quality of our predictions only based on what our predictor claims to be positive
- All we predicted correctly / All we predicted, correctly or wrongly

Recall

- However, **Recall** is to measure such quality with respect to the mistakes we did (what should have been predicted as positive but we flagged as negative):
- All we predicted correctly / All we should have predicted
-

- To understand *Precision* and *Recall*, let's take an example of Search. Think about the search box on Amazon home page.



- The **precision** is the proportion of relevant results in the list of all returned search results.

- The **recall** is the ratio of the relevant results returned by the search engine to the total number of the relevant results that could have been returned.

F1-Measure

- However, in cases where we want to find an optimal blend of precision and recall we can combine the two metrics using what is called the F1 score.
- The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

Questions?

- **Should we approve a loan?**

We are looking to develop a machine learning algorithm to predict whether someone will pay a loan back or not.

- What is the positive class?
- What would a recall of 75% mean?
- What would a precision of 85% mean?

Answer

1. The positive class are the borrowers that pay back the loans.
2. 75% recall means that 75% of the borrowers that would pay back the loan are approved by our system. We miss 25% of people that would have paid us back by rejecting them. In general, the problem with a low recall is that we are rejecting customers who we would have paid us back (and for whom we would have made interest).
3. 85% precision means that of all the loans we approve, 85% pay us back. The remaining 15% of approved loans go into default. The problem with a low precision is that we are approving loans that are defaulting.

Question

- **Should we unlock a phone?**

We are building a facial recognition algorithm to allow people to unlock their phone. If the phone recognizes the person as the authorized user, it will unlock the phone. If it doesn't recognize the user, it will prompt them to try again or try an alternative method.

- What is the positive class?
- What would a recall of 80% mean?
- What would a precision of 70% mean?

Answer

- The positive class is recognizing the user as authorized.
- 80% recall means 80% of the times the authorized user tries to use this feature, the phone unlocks. The remaining 20% of the time, the authorized user was asked to try again.
- 70% precision means that out of all the times the phone was unlocked using this feature, it was unlocked by an authorized user. The remaining 30% of the times it was unlocked, it allowed in someone that was not authorized.

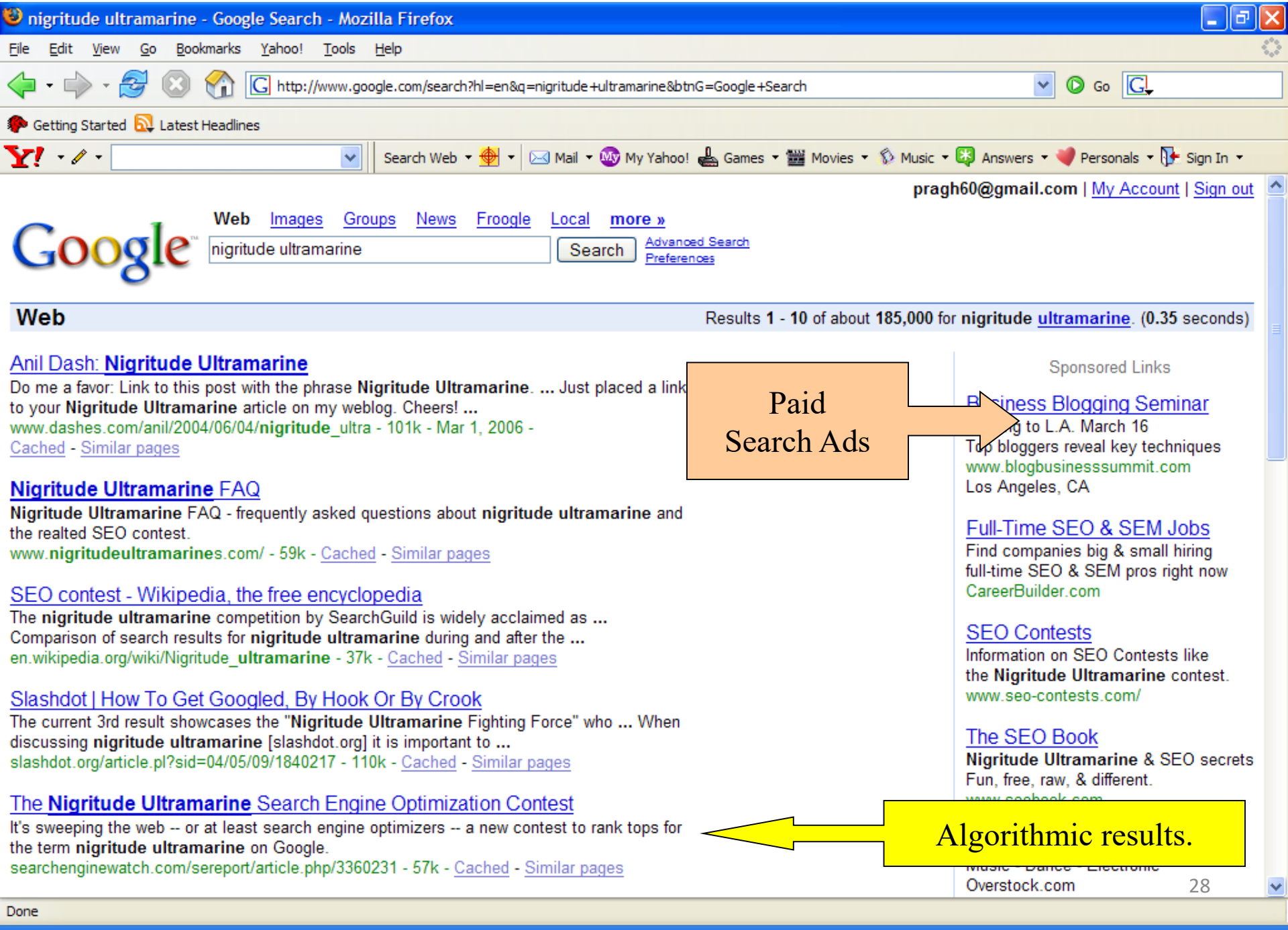
Web Search Basics

Brief (non-technical) history

- Early keyword-based engines ca. 1995-1997
  - Altavista, Excite, Infoseek, Inktomi, Lycos
- Paid search ranking: Goto (morphed into Overture.com → Yahoo!)
  - Your search ranking depended on how much you paid

Brief (non-technical) history

- 1998+: Link-based ranking pioneered by Google
  - Blew away all early engines save Inktomi
  - Great user experience in search of a business model
  - Meanwhile Goto/Overture's annual revenues were nearing $1 billion
- Result: Google added paid search "ads" to the side, independent of search results
  - Yahoo followed suit, acquiring Overture (for paid placement) and Inktomi (for search)
- 2005+: Google gains search share, dominating in Europe and very strong in North America
  - 2009: Yahoo! and Microsoft propose combined paid search offering


nigritude ultramarine - Google Search - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://www.google.com/search?hl=en&q=nigritude+ultramarine&btnG=Google+Search
Go
Getting Started
Latest Headlines
Search Web
Mail
My Yahoo!
Games
Movies
Music
Answers
Personals
Sign In
pragh60@gmail.com | My Account | Sign out
Google
Web Images Groups News Froogle Local more »
nigritude ultramarine
Search
Advanced Search
Preferences
Web
Results 1 - 10 of about 185,000 for nigritude ultramarine. (0.35 seconds)
Anil Dash: Nigritude Ultramarine
Do me a favor: Link to this post with the phrase Nigritude Ultramarine. ... Just placed a link to your Nigritude Ultramarine article on my weblog. Cheers! ...
www.dashes.com/anil/2004/06/04/nigritude_ultra - 101k - Mar 1, 2006 -
Cached - Similar pages
Nigritude Ultramarine FAQ
Nigritude Ultramarine FAQ - frequently asked questions about nigritude ultramarine and the realted SEO contest.
www.nigritudeultramarines.com/ - 59k - Cached - Similar pages
SEO contest - Wikipedia, the free encyclopedia
The nigritude ultramarine competition by SearchGuild is widely acclaimed as ...
Comparison of search results for nigritude ultramarine during and after the ...
en.wikipedia.org/wiki/Nigritude_ultramarine - 37k - Cached - Similar pages
Slashdot | How To Get Googled, By Hook Or By Crook
The current 3rd result showcases the "Nigritude Ultramarine Fighting Force" who ... When discussing nigritude ultramarine [slashdot.org] it is important to ...
slashdot.org/article.pl?sid=04/05/09/1840217 - 110k - Cached - Similar pages
The Nigritude Ultramarine Search Engine Optimization Contest
It's sweeping the web -- or at least search engine optimizers -- a new contest to rank tops for the term nigritude ultramarine on Google.
searchenginewatch.com/sereport/article.php/3360231 - 57k - Cached - Similar pages
Paid Search Ads
Sponsored Links
Business Blogging Seminar
g to L.A. March 16
Top bloggers reveal key techniques
www.blogbusinesssummit.com
Los Angeles, CA
Full-Time SEO & SEM Jobs
Find companies big & small hiring full-time SEO & SEM pros right now
CareerBuilder.com
SEO Contests
Information on SEO Contests like the Nigritude Ultramarine contest.
www.seo-contests.com/
The SEO Book
Nigritude Ultramarine & SEO secrets
Fun, free, raw, & different.
Overstock.com
Algorithmic results.
28
Done

Google

sony laptops

All | Shopping | Images | News | Videos | More | Tools

About 34,60,00,000 results (1.18 seconds)

https://www.sony.co.in › personal-computers-laptop-pc

Laptop PC - Sony

Laptop PC. **Laptop** PC. Search. Where is the model name located on my product.

Laptop PC. All. DownloadsManualsQuestions & Answers. Important Information ...

People also ask

- Does Sony still make laptops?
- Why Sony stopped making laptops?
- Is Sony good company for laptop?
- When did Sony stop making laptop?

Feedback

https://www.sony.co.in › electronics › vaio-laptops

Ads · Shop sony laptops

Samsung - Galaxy Boo...

₹98,990

Samsung.c...

★★★★½ (99)

VAIO - E15 V2IN007P...

₹49,999

Reliance Di...

Free delivery

Samsung - Galaxy Boo...

₹1,19,990

Samsung.c...

Free delivery

Lenovo -

VAIO - SF14

Samsung -

Google

iit delhi

All News Maps Images Videos More Tools

About 10,20,00,000 results (0.68 seconds)

http://www.iitd.ac.in

IIT Delhi: Home Page

Indian Institute of Technology Delhi is one of the Twenty Three IITs created to be Centres of Excellence for training, research and development in science, ...

Results from iitd.ac.in

Departments

IIT Delhi provides science-based engineering education with a ...

Jobs

Welcome to our new careers section. At IIT Delhi, we provide ...

PG Admissions 2022-2023

PG Admissions (2nd Semester 2022-23) for PhD/ MSR ...

PG Admissions

iit delhi pg admission portal.

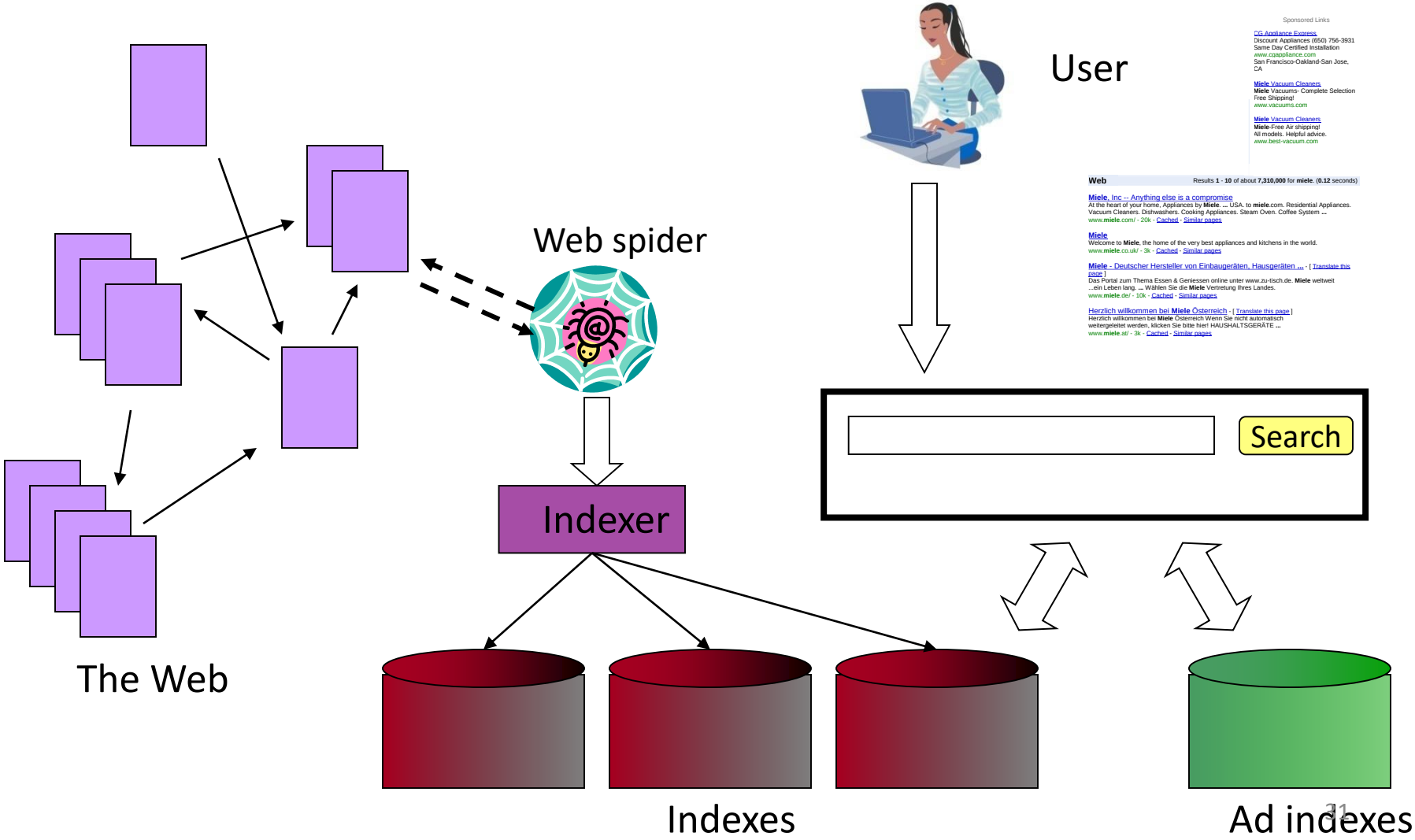
Indian Institute of Technology Delhi

Website Directions Save Call

Research institution in New Delhi

Indian Institute of Technology Delhi is an autonomous public research and technical institute located in New Delhi, India. It is one of the 23 IITs created to be Centres of Excellence for training, research and development in science, engineering and technology in India. Wikipedia

Web search basics-- Web crawler, Spider, Bot



- How does Google Search work?
- **Google follows three basic steps to generate results from web pages:**
- Step 1 Crawling.
- Step 2 Indexing.
- Step 3 Ranking.

- Google searches through indices stored in RAM and Google also uses a map-reduce algorithm to conduct a massively parallel set of operations.

Search queries

- The words and phrases that people type into a search box in order to pull up a list of results – come in different flavours. It is commonly accepted that there are three different types of search queries:
- Navigational search queries
- Informational search queries
- Transactional search queries

Navigational Search Query?

- A **navigational query** is a search query entered with the intent of finding a particular website or webpage.
- For example, a user might enter "youtube" into Google's search bar to find the YouTube site rather than entering the URL into a browser's navigation bar or using a bookmark.
- “facebook” and “youtube” are the top two searches on Google, and these are both navigational queries.
- True navigational queries have very clear intent – the user has an exact site in mind and if you’re not that site, you’re not relevant to their needs.

Informational Search Query?

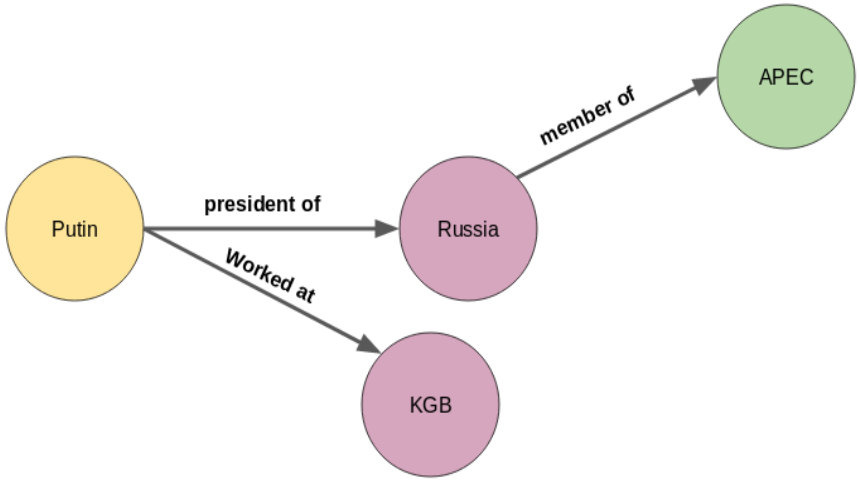
- **Informational search queries** are "Queries that cover a broad topic (e.g., Volcano or *trucks*) for which there may be thousands of relevant results."
- When someone enters an informational search query into Google or another search engine, they're looking for information.
- They are probably not looking for a specific site, as in a navigational query, and they are not looking to make a commercial transaction. They just want to answer a question or learn how to do something.

How Should You Target Informational Search Queries?

- Informational queries are hard to monetize.
- Google knows this, which is why it's been pushing the Knowledge Graph to address these types of queries.
- The best way to target informational searches is with high-quality SEO content that genuinely provides helpful information relevant to the query.

“Knowledge Graph”

- **The term “Knowledge Graph” refers to prepared and compiled search results on certain topics and entities such as people or areas in the Google search and Facebook**.
- A knowledge graph, also known as a semantic network, **represents a network of real-world entities—i.e. objects, events, situations, or concepts—and illustrates the relationship between them**. This information is usually stored in a graph database and visualized as a graph structure, prompting the term knowledge “graph.”



Graph Databses: Neo4j, FlockDB, Microsoft Azure CosmosDB

There are usually multiple statements about the topic: Knowledge Graph representation.

- A name or label (Like “Sachin Tendulkar”).
- A Type or Types (Like “Person”).
- A Description (Say “cricketer/Sportsman”).
- A list of Image URLs (Usually with the associated usage rights).
- A Detailed Description (Usually some Text with a URL of the source).

What Is SEO Content?

- To understand SEO content, it's helpful to break down the phrase into its component parts:
- **"SEO"** refers to search engine optimization, or the process of optimizing a website so that people can easily find it via search engines like Google.
- By **"content,"** we mean any information that lives on the web and can be consumed on the web.So, putting these two concepts together:

SEO content is any content created with the goal of attracting search engine traffic.

SEO - Search engine optimization: **the process of making your site better for search engines**

QUESTION??

- What you'll need to do in order to SEO your web content?
- **SEO Steps**
- Step 1: Find keywords.
- Step 2: Put keywords in the page title.
- Step 3: Put keywords in the page URL.
- Step 4: Put keywords in your meta description.
- Step 5: Put keywords in your H1 text.
- Step 6: Use keywords in the page's content.
- Step 7: Build links to your website.
- Step 8: Monitor your rank.

Answer

- **Keyword Research**:
- **Keyword Optimization**
- **Content Organization**
- **Content Promotion**

Question??

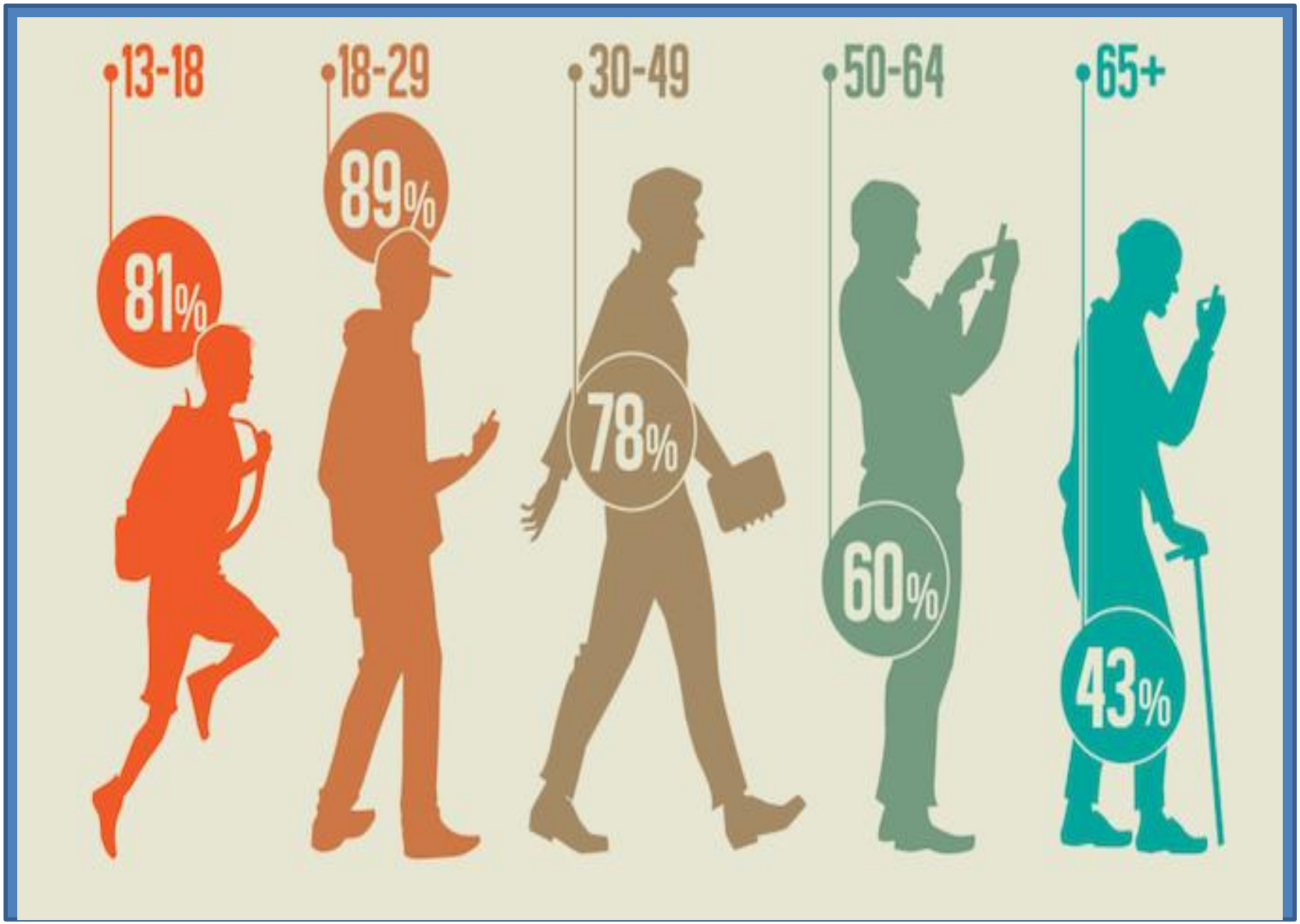
- **Types of SEO Content**

Answer

- **Product Pages** :These are the best for retail e-commerce site. A good product page can serve as both SEO content and a PPC landing page.
- **Articles** – Think news article, interview, or feature piece. This is the main kind of content you'll find on most newspaper- or magazine-style websites.
- Lists – A list is really just a kind of article, but framing it as a list (such as "10 Ways to Lower Your Energy Bill" or "101 Things I Hate About Google") makes it easier to scan. These types of titles also seem to be more clickable .
- Guides – A guide is a longer piece of content that explains in detail how to do something. (Guides are often broken up onto multiple web pages, though it's a best practice to allow users to view long content as a single page if they wish.)

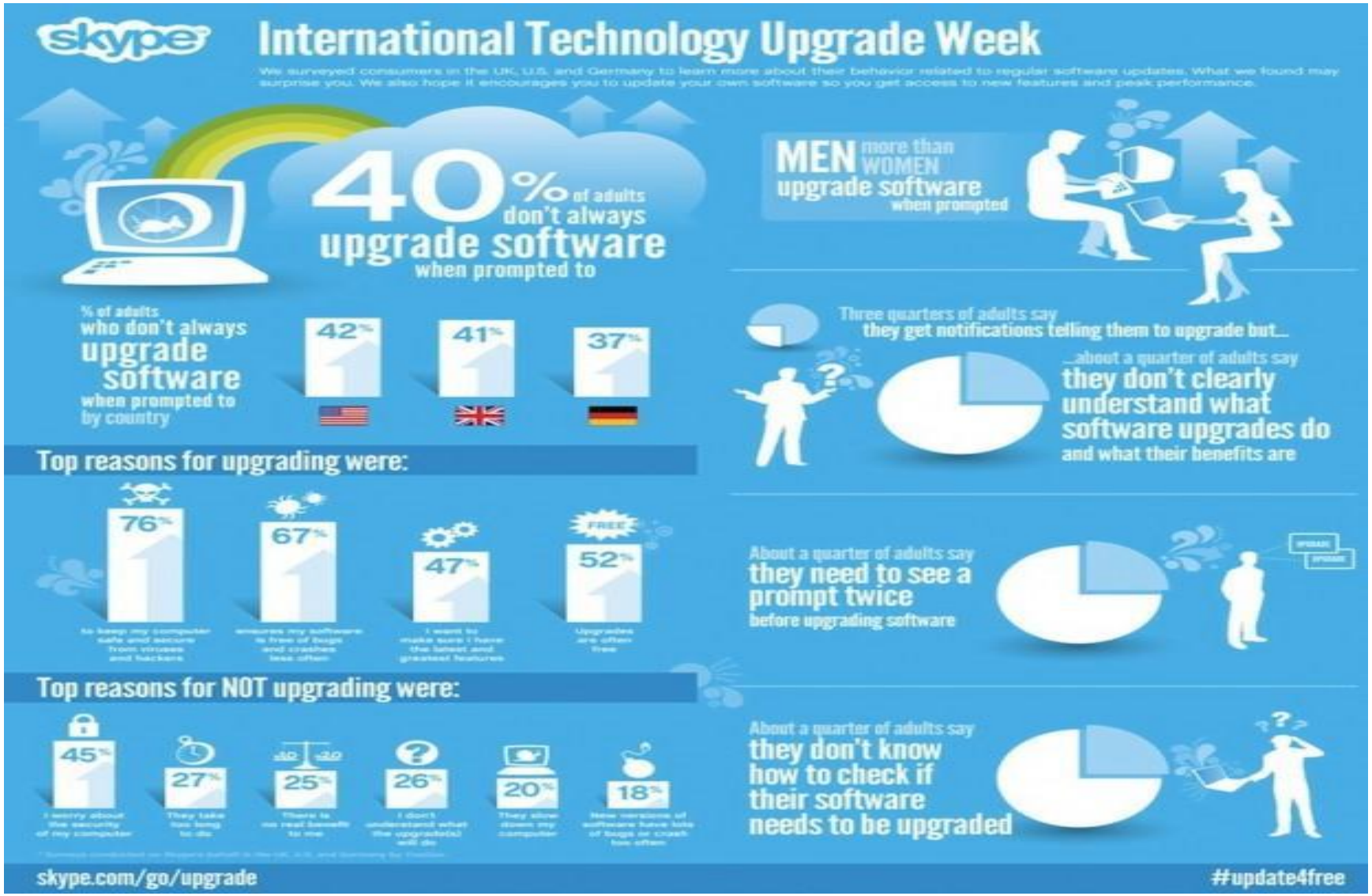
- **Videos** –videos can be a great way to attract and reach an audience. Consider creating video tutorials of how to use your products. Or illustrate a process that is related to your business – for example, a plumber could make a video showing how to unclog a sink.
- **Infographics** – Infographics, or large-format images that contain a lot of data (often in the form of graphs or charts) on a single subject, can rack up a lot of page views and links.
- **Slideshows** – A slideshow is a way to display a series of related images. Sometimes pictures are more important than text .
- **Directories** – A directory is a useful taxonomy of links to sites or resources around a given topic. For example, a perfume blog might create a directory of places to buy perfume, from major department stores to independent shops around the country.
- **Glossaries** –If you work in a specialized industry, a well built-out glossary can be a good way to capture some search traffic. Think cooking terms, medical terms, fashion terms, architectural terms …
- **Consideration of audience:**

Example : **Consideration of audience:**

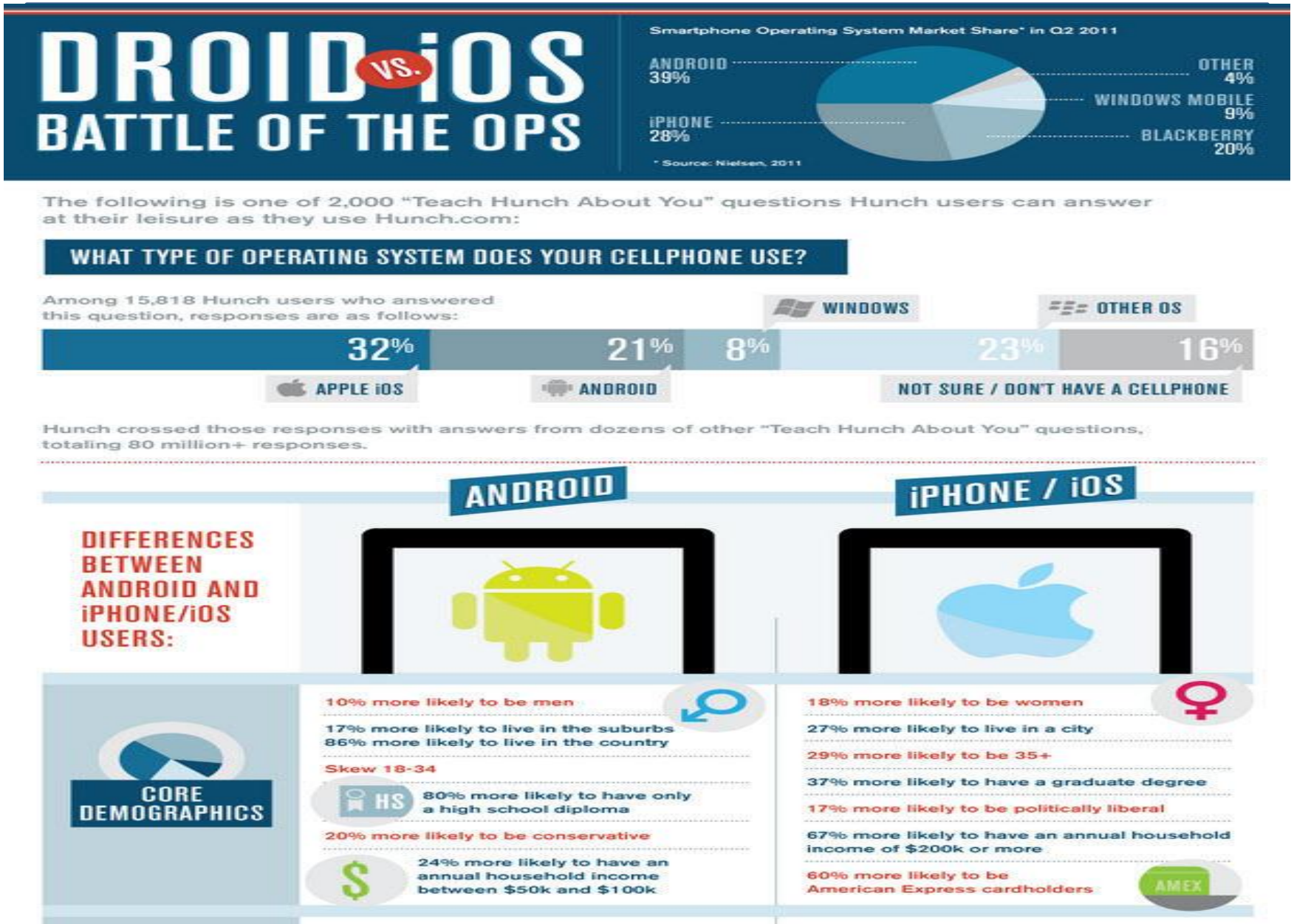


Infographic Example

Example


skype
International Technology Upgrade Week
We surveyed consumers in the UK, U.S. and Germany to learn more about their behavior related to regular software updates. What we found may surprise you. We also hope it encourages you to update your own software so you get access to new features and peak performance.
40% of adults don't always upgrade software when prompted to
MEN more than WOMEN upgrade software when prompted
% of adults who don't always upgrade software when prompted to by country
42%
41%
37%
Three quarters of adults say they get notifications telling them to upgrade but...
...about a quarter of adults say they don't clearly understand what software upgrades do and what their benefits are
Top reasons for upgrading were:
76%
67%
47%
52%
FREE
About a quarter of adults say they need to see a prompt twice before upgrading software
Top reasons for NOT upgrading were:
45%
27%
25%
26%
20%
18%
About a quarter of adults say they don't know how to check if their software needs to be upgraded
skype.com/go/upgrade
#update4free

Example


DROID VS. iOS
BATTLE OF THE OPS
Smartphone Operating System Market Share* in Q2 2011
ANDROID 39%
iPHONE 28%
OTHER 4%
WINDOWS MOBILE 9%
BLACKBERRY 20%
* Source: Nielsen, 2011
The following is one of 2,000 "Teach Hunch About You" questions Hunch users can answer at their leisure as they use Hunch.com:
WHAT TYPE OF OPERATING SYSTEM DOES YOUR CELLPHONE USE?
Among 15,818 Hunch users who answered this question, responses are as follows:
WINDOWS
OTHER OS
32%
21%
8%
23%
16%
APPLE iOS
ANDROID
NOT SURE / DON'T HAVE A CELLPHONE
Hunch crossed those responses with answers from dozens of other "Teach Hunch About You" questions, totaling 80 million+ responses.
DIFFERENCES BETWEEN ANDROID AND iPHONE/iOS USERS:
ANDROID
iPHONE / iOS
CORE DEMOGRAPHICS
10% more likely to be men
17% more likely to live in the suburbs
86% more likely to live in the country
Skew 18-34
HS
80% more likely to have only a high school diploma
20% more likely to be conservative
24% more likely to have an annual household income between $50k and $100k
18% more likely to be women
27% more likely to live in a city
29% more likely to be 35+
37% more likely to have a graduate degree
17% more likely to be politically liberal
67% more likely to have an annual household income of $200k or more
60% more likely to be American Express cardholders
AMEX

Transactional Search Query?

- A **transactional search query** is a query that indicates an intent to complete a transaction, such as making a purchase.
- Transactional search queries may include exact brand and product names (like “Samsung galaxy s3”) or be generic (like “iced coffee maker”) or actually include terms like “buy,” “purchase,” or “order.”

- Transactional search queries: this kind of request indicates that the user wishes to execute a transaction– for example, “buy books” or “purchase an iPhone ”.

User Needs

- **Informational** – want to learn about something
- **Navigational** – want to go to that page
- **Transactional** – want to do something (web-mediated)
  - Access a service
  - Downloads
  - Shop

Google results for search queries

- Search engines like Google present different kinds of content based on assumptions they make about the underlying intentions of a user's query,
- Results for transactional, informational, and navigational queries can therefore look quite different.
- Search results for transactional keywords generally call up the appropriate online store.
- Results for navigational searches usually include the actual brand or company pages.
- Results for informational, Google often displays blog articles, tutorials, and videos as well as knowledge bases such as Wikipedia or Merriam Webster.

Question??

- Why people uses search engines?

Answer

- Search engines are used for three particular reasons: to obtain information, to orientate themselves online, or to perform a transaction.
- These intentions give rise to informational, navigational, or transactional searches respectively.

Question??

- How far do people look for results?

OR

- When you perform a search on a search engine and don’t find what you are looking for , at what point do you typically either revise your search or move on to another search engine.?

Answer

"When you perform a search on a search engine and don't find what you are looking for, at what point do you typically either revise your search, or move on to another search engine? (Select one)"