System Design: Search – Session 2

Pooja Gada

Agenda

- Speaker Introduction
- Why is search important
- Part 1 - Design a search auto-complete system
- Part 2 - Advanced Search

Speaker Intro

Pooja Gada

Current: Engineering Leader @ 1Password

Previously: Slack, Digit, Qventus

Education: Carnegie Mellon University

Motto: *Lift as we climb*

pooja.s.gada@gmail.com

Who is this meetup for

- Preparing for interviews
  - Software Engineering
  - Engineering Management
  - Technical Product/Project Manager

- Build on System Design Skill set
  - Early in career
  - Expand knowledge on types of challenges
  - Learn how to build scalable systems

Caveats...

- Most System Design Interviews require a bit of prep
    - You have 1 hour to impress the interviewer
    - Instagram / Youtube / Meetup - wasn’t build in an hour - you kinda have to….
- There is no one-perfect-solution
    - There are many ways of designing a system - Leverage your background!
- You will hear names of products - they are just for references

I want you go back to first principles…

What we will and will not cover

We will

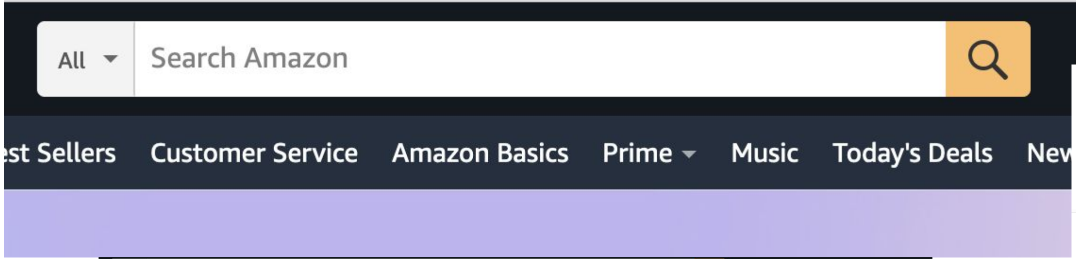
- We will explore some interesting questions centered around Search
- We will explore most popular tech stacks
- Provide useful resources for future

We will not

- We will not be going into algorithms around search such as edit distance
- We will not cover all possible tech stacks(its a lot)

Why is Search Important

Yep, users need to find stuff.



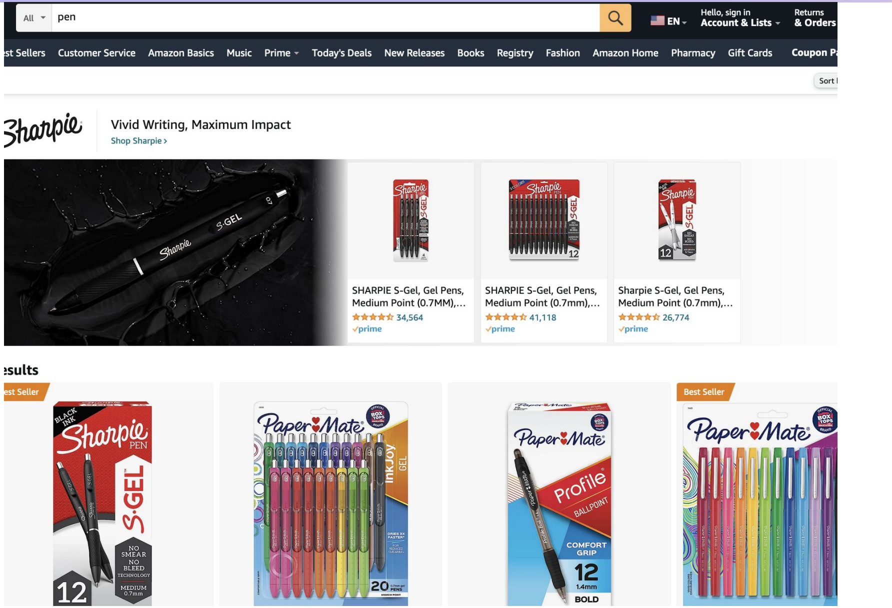
All

Search Amazon

EN

He

st Sellers Customer Service Amazon Basics Prime Music Today's Deals New



All pen

EN

Hello, sign in
Account & Lists

Returns
& Orders

st Sellers Customer Service Amazon Basics Music Prime Today's Deals New Releases Books Registry Fashion Amazon Home Pharmacy Gift Cards Coupon P

Sort

Vivid Writing, Maximum Impact

Shop Sharpie ›

SHARPIE S-Gel, Gel Pens, Medium Point (0.7MM),… 34,564 prime

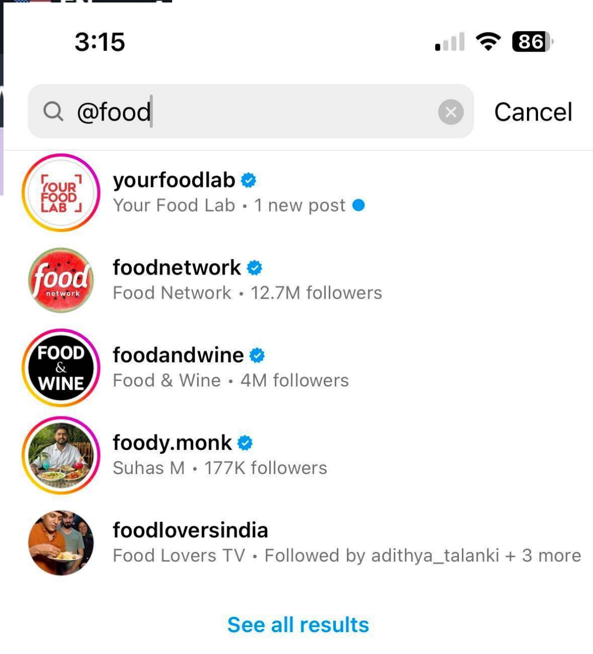
SHARPIE S-Gel, Gel Pens, Medium Point (0.7mm),… 41,118 prime

Sharpie S-Gel, Gel Pens, Medium Point (0.7mm),… 26,774 prime

esults

est Seller

Best Seller



3:15

86

@food

Cancel

yourfoodlab
Your Food Lab • 1 new post

foodnetwork
Food Network • 12.7M followers

foodandwine
Food & Wine • 4M followers

foody.monk
Suhas M • 177K followers

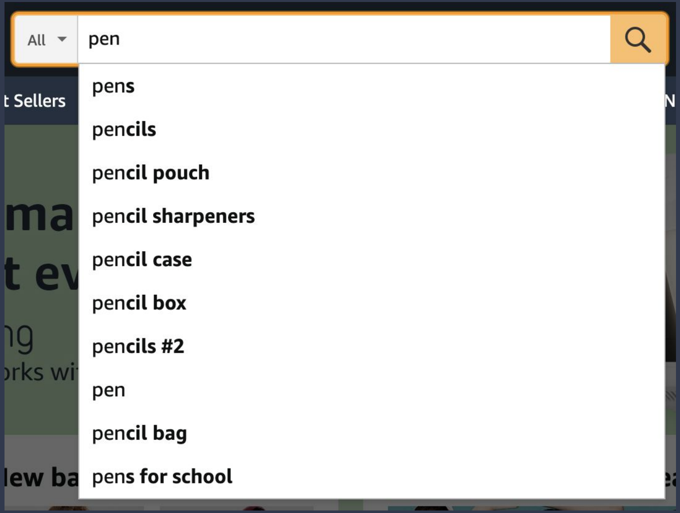
foodloversindia
Food Lovers TV • Followed by adithya_talanki + 3 more

See all results

Search is everywhere!

Let’s Build Something!

Part 1 – Design a search auto-complete system



pen

pens

pencils

pencil pouch

pencil sharpeners

pencil case

pencil box

pencils #2

pen

pencil bag

pens for school

System Design Interview Blueprint

- **Phase 1 - Understand & Define Scope**
  - Functional / Non-Functional Requirements
- **Phase 2 - Plan & Get Buy-in**
  - Focus of time
- **Phase 3 - Deep Dive**
  - Algorithms
  - System Design / Data Flow
  - Why / What / How - with hint of trade-offs
- **Phase 4 - Summarize**

Phase 1 - Understand & Define scope

What are we matching against?

Product names

What is the matching criteria?

Only beginning letters, 2 char min

Spell check or localization or casing support?

No need to support

Scale - users/product names to search?

1 million DAU / 10 million product names

What type of results should be returned?

Top 5 most **frequently** searched - *frequency is important!*

Phase 2 - Plan & Get Buy in

Database Search

Select * from db_table where name ‘like_txt%’;

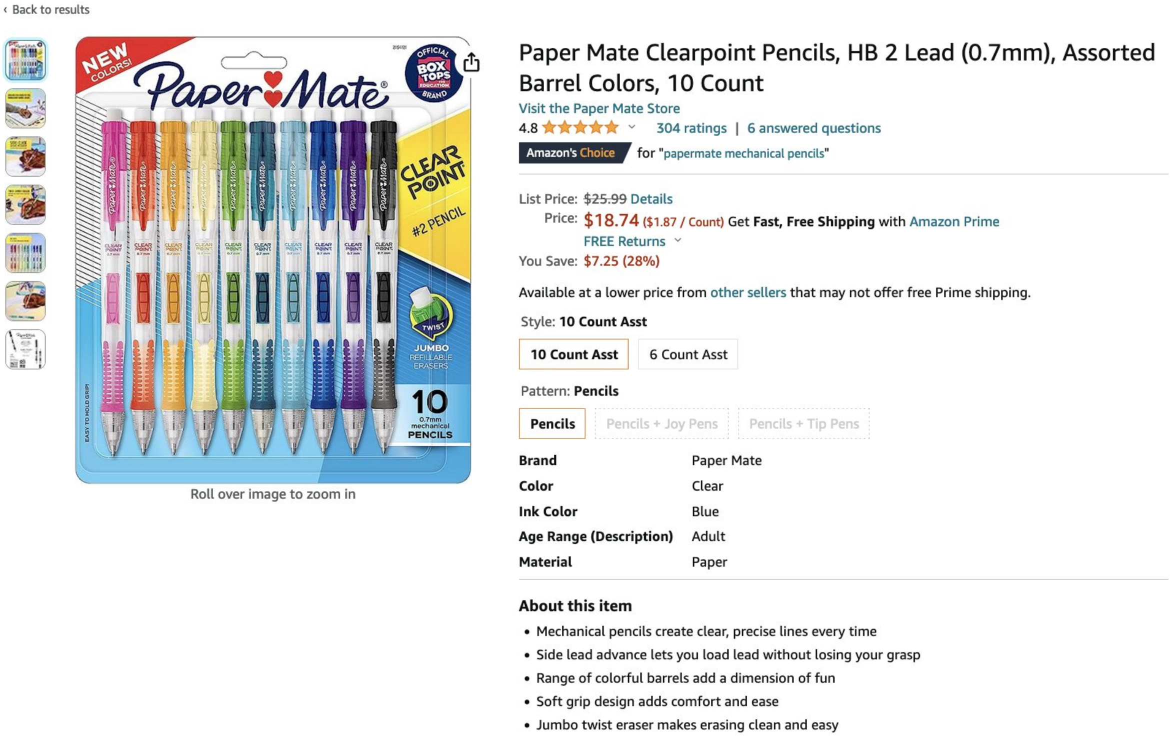
- Begins with ‘a’
  - Select * from db_table where name ‘a%’;
- Ends with ‘a’
  - Select * from db_table where name ‘%a’;

Why can’t we just shoot off this query for every character entered?

Don’t forget scale & performance!

I can just add an index right?

Product Details



Paper Mate Clearpoint Pencils, HB 2 Lead (0.7mm), Assorted Barrel Colors, 10 Count

Visit the Paper Mate Store

4.8 ★★★★★ 304 ratings | 6 answered questions

Amazon's Choice for "papermate mechanical pencils"

List Price: ~~$25.99~~ Details

Price: $18.74 ($1.87 / Count) Get **Fast, Free Shipping** with Amazon Prime

FREE Returns

You Save: $7.25 (28%)

Available at a lower price from other sellers that may not offer free Prime shipping.

Style: **10 Count Asst**

10 Count Asst | 6 Count Asst

Pattern: **Pencils**

Pencils | Pencils + Joy Pens | Pencils + Tip Pens

Brand	Paper Mate
Color	Clear
Ink Color	Blue
Age Range (Description)	Adult
Material	Paper

About this item

- Mechanical pencils create clear, precise lines every time
- Side lead advance lets you load lead without losing your grasp
- Range of colorful barrels add a dimension of fun
- Soft grip design adds comfort and ease
- Jumbo twist eraser makes erasing clean and easy

DB model products

products_table

name	text
seller	FK
price	float
description	text

Indexing in Databases

Most databases use B-Trees for indexing

- May work if size of DB is small enough
- Indexes are expensive
- Indexes are not sufficient
    - For the scope of the question - frequency of the words is important

Keywords + Frequency

keyword_frequency_table

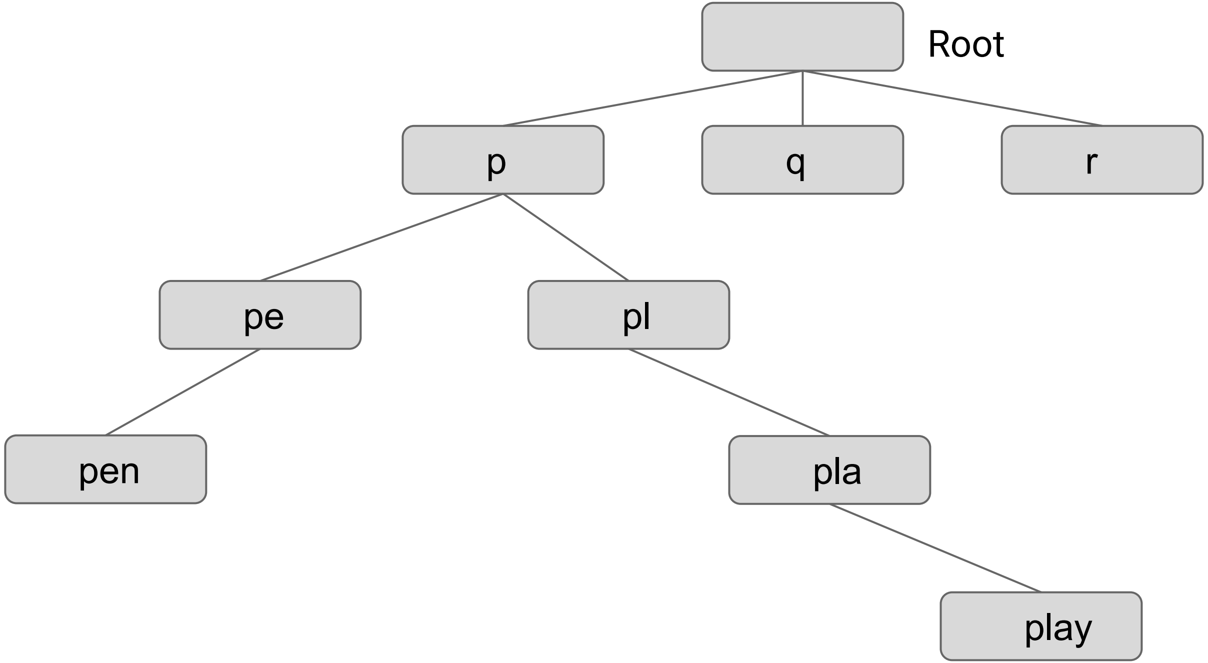
keyword	text
frequency	int

DB Model

keyword	**frequency**
pencil	45
pen	30
pencil case	20
pencil pouch	10
pencil bag	5
pencil eraser	3

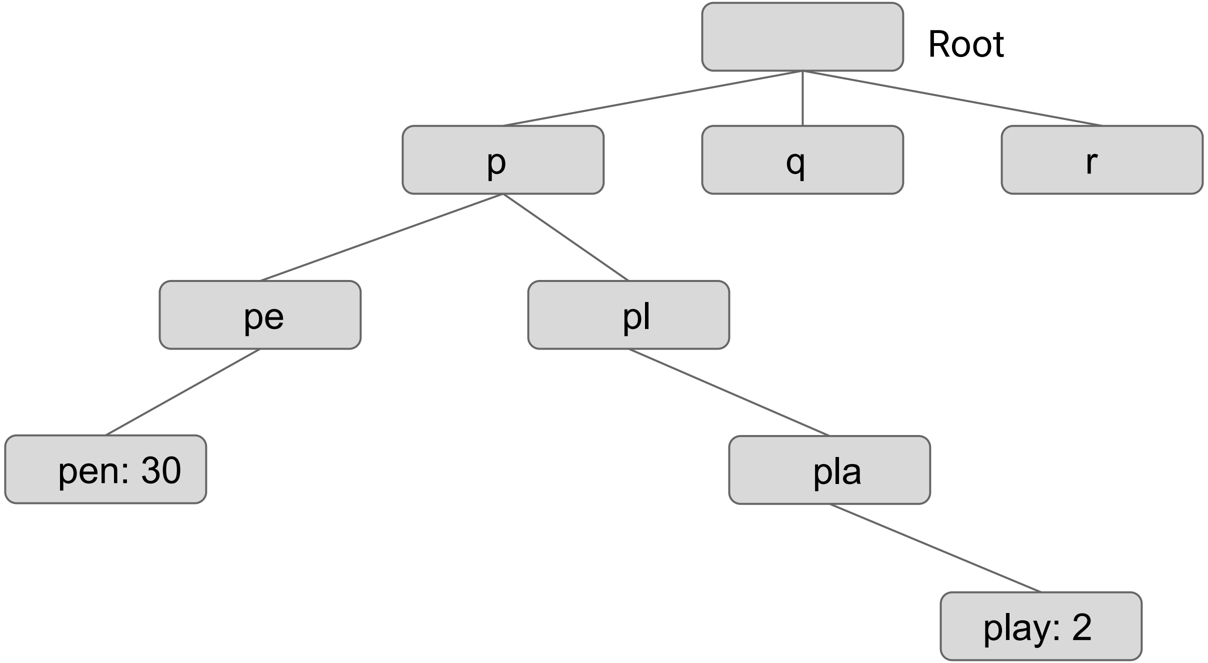
How do we build faster lookup & factor in frequency?

Trie to the rescue



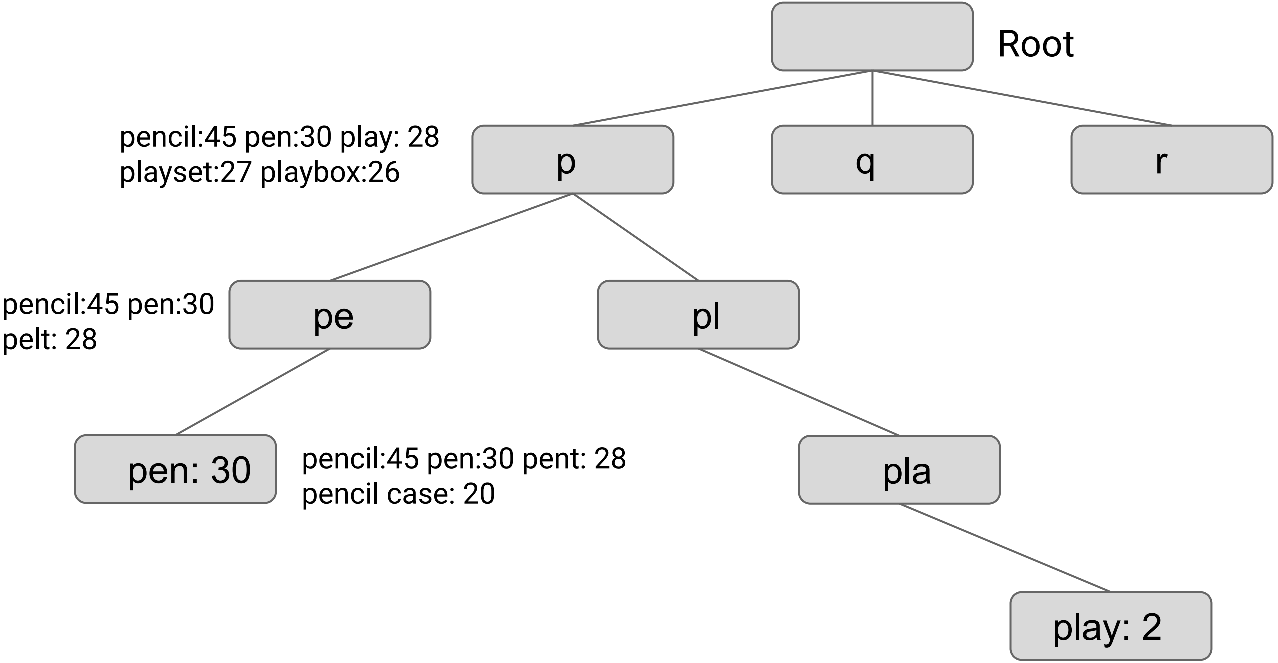
But what about the frequency info?

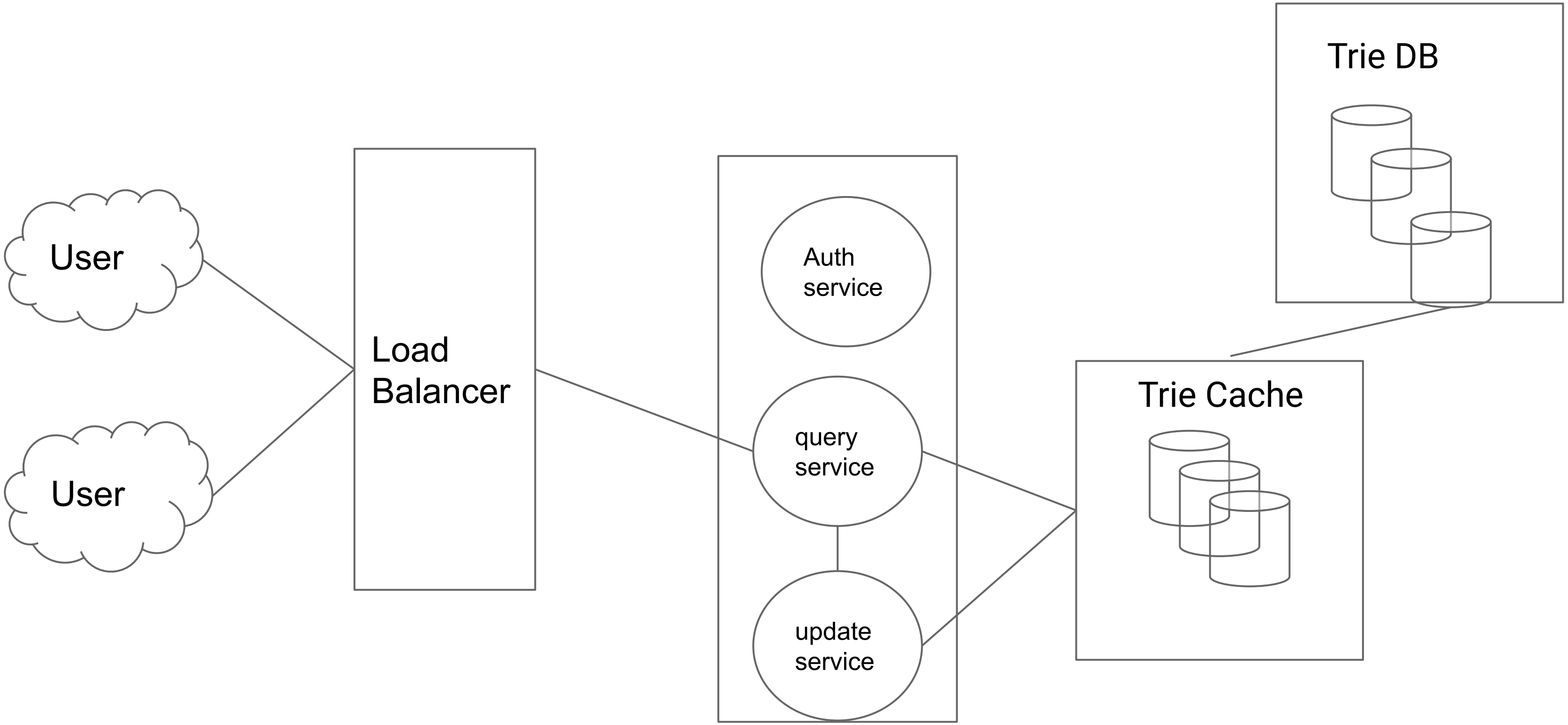
Add frequency to the child nodes!



But wouldn’t traversing the nodes to find top 5 results still take time?

Cache upto top 5 results at each node





High level system design

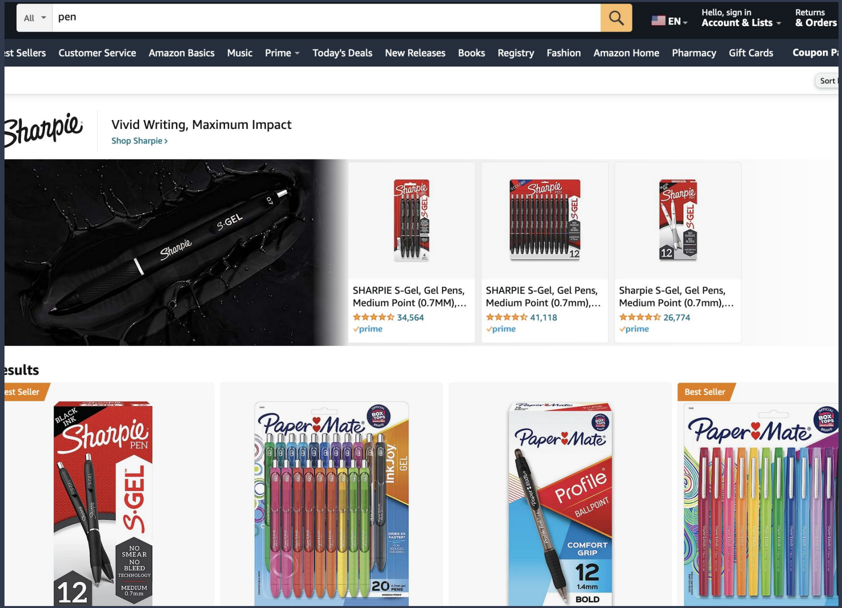
Trie DB

- Can either be
  - Document store
    - MongoDB
  - Key-value store
    - DynamoDB

Other potential questions/follow-ups

- Browser caching
- In-memory trie-db
- Sharding strategies
- Localization
    - Vary auto-complete suggestions by country
- Recency bias
    - Trending topic

Part 2 – List all products that match keyword



Vivid Writing, Maximum Impact

SHARPIE S-Gel, Gel Pens, Medium Point (0.7MM),... 34,564

SHARPIE S-Gel, Gel Pens, Medium Point (0.7mm),... 41,118

Sharpie S-Gel, Gel Pens, Medium Point (0.7mm),... 26,774

esults

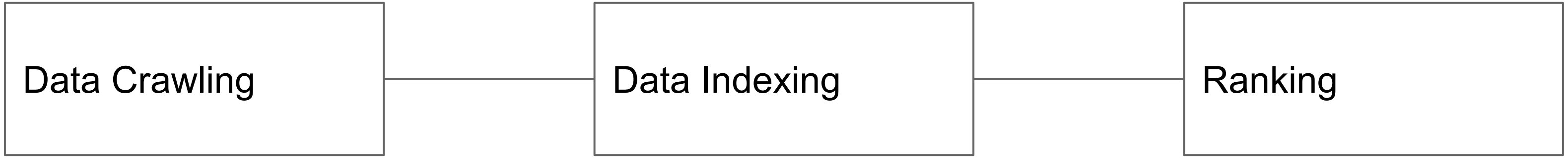
We need a way to tie back keywords to product

keyword	prod1	prod2	prod3
pen	x		x
pencil		x	x
pencil case		x	

- Inverted index!
- This is the algorithm that solr/elastic search actually use under the hood

Other potential questions/follow-ups

- How would you build a inverted index
    - Database table
    - Binary bits 01101



References

- System Design Interview Volume 1 & 2
  - Alex Xu & Sahn Lam
- Educative.io
- Tryexponent

What would you like to see in the next session?

Upcoming sessions:

- News Feed System
- Chat application
- Nearby

What else would you like to learn?