KNOWLEDGE GRAPH

What, Why, and How

TABLE OF CONTENTS

01

OVERVIEW

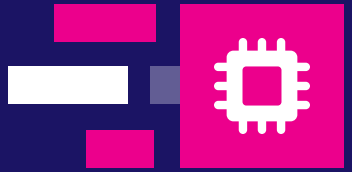
INTRODUCTION

The Knowledge Graph ideology was popularized by Google in 2012 when they publicly attributed their search solution to Knowledge Graphs. Google defined its Knowledge Graph to serve the following objectives: Discoverability, Knowledge Creation, Distinguishability, Speed

“A unit of knowledge can be defined as a piece of information that allows users to reach an outcome when confronted with specific questions. ”

Discoverability

Make it easy for users to navigate billions of data points to discover specific Knowledge



Knowledge Creation

Offer new or unexpected Knowledge to users through new connections or related results. Users are not looking for it, but it adds value to what they are looking for.

Distinguishability:

Intuitive search capability that understands the context around which the user is searching and presents results accordingly. For example, searching 'Apple' should present the Apple Company or the Fruit.

Speed

Surface relevant information within milliseconds.

KNOWLEDGE GRAPH

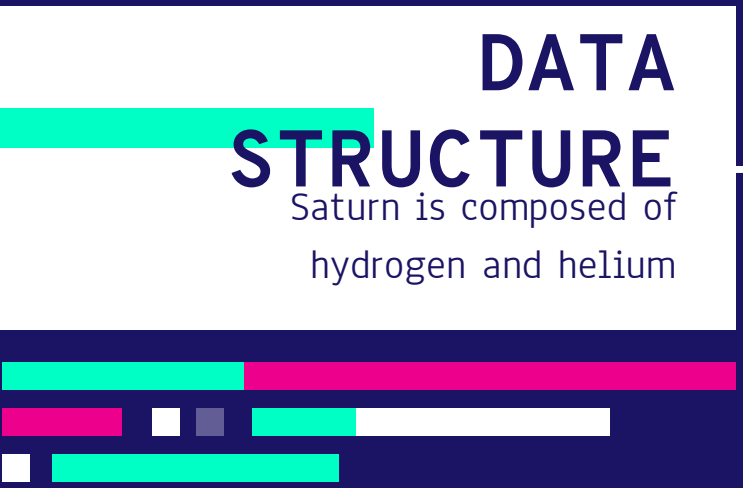
A knowledge graph is a semantic web of entities, relationships, and events. More fundamentally, it is a directed graph where every element is populated with rich information regarding itself and its relationships with other elements.

Every data problem is a knowledge transfer problem, and every knowledge transfer problem can be formalized as a graph. Therefore, every data problem can be formalized as a graph. ~ Stephen Bailey

02

STRUCTURES USED

OVERVIEW DIAGRAM



AVL Tree

Insertion, deletion

Heap

To set priority of data set

GRAPH

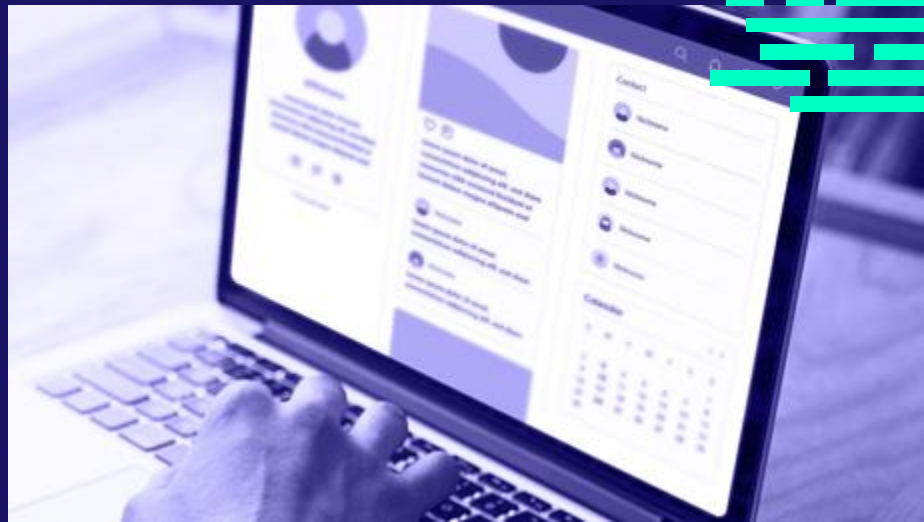
For transitive relation

QUEUE

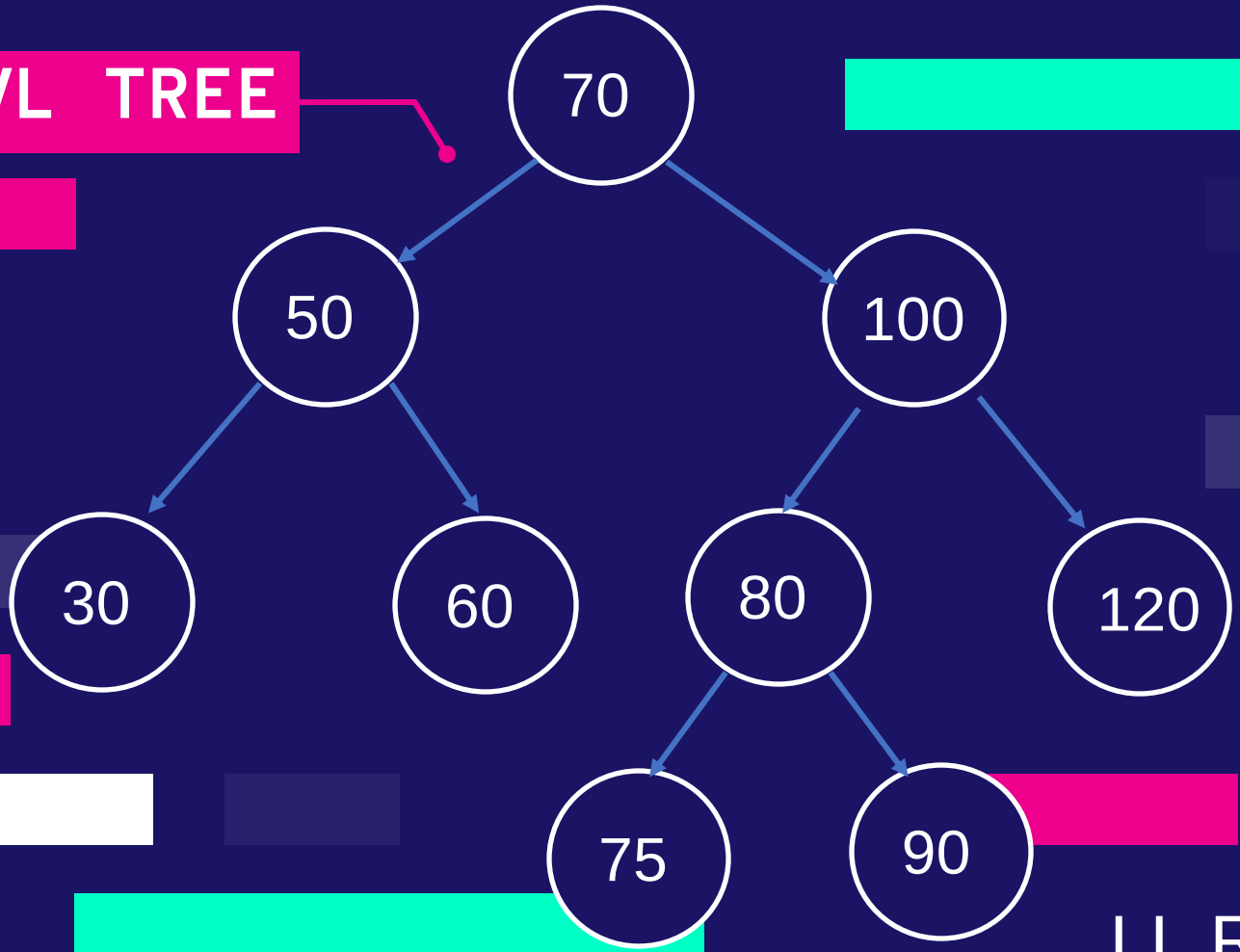
For FIFO , priority transition

DATA STRUCTURE

"In our project, we juggled arrays, danced with linked lists, and played hide-and-seek with hash tables. But let's be real, the most important data structure was the 'Infinite Loop' of our minds, where bugs roamed free and ideas collided like bumper cars."

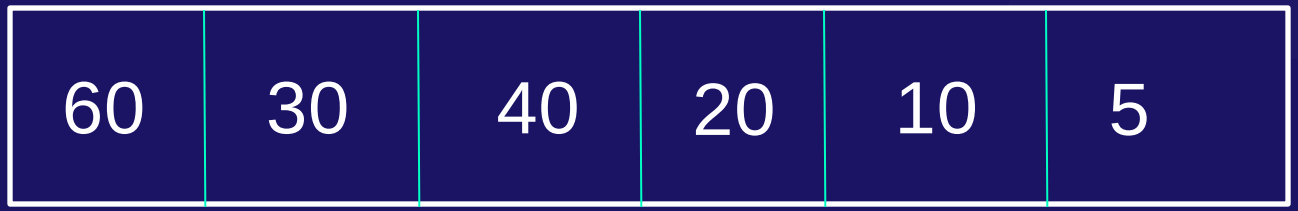


AVL TREE



LL,RR,LR,RL

MAX HEAP



03

OPTIMIZATIONS MADE

AVL

"Our insertion function isn't just about adding nodes; it's a multitasking maestro! With a single swoop, it navigates the tree, searches for existing values, and seamlessly integrates new ones. Say goodbye to O(kn^2) complexity and hello to the streamlined efficiency of O(klnn)."

HEAP

Code works well for input of more than a lakh sentence!

AVL

Code works well for input of more than a lakh sentence!

ARRAY/HEAP

With realloc(), our array grows on demand, sparing us from pre allocating a massive chunk of memory upfront.

AVL

Structure doesn't include concept of parent which makes it the most efficient!

OVERALL

Each part of the code has been crafted with care, each part is primed to handle the heavy lifting, effortlessly managing data sets of over a lakh"

PROCESS

Step 1

Planning the layout of the project

Step 2

Reading various research papers.

Step 3

Deciding the data structures.

.

Step 4

Learning how to write clean, efficient code

Step 5

Compiling code and running it with huge data set

Step 6

Preparing documentation

04

APPLICATIONS

OPERATIONALIZED DATA

Google built applications on top of their Knowledge Graph to add an additional layer for Insights.

For example:

When a user searches “restaurants near me,” it doesn’t just surface the specific detail (restaurant names) the user searched for. It also brings up review data, ratings, directions, and a plethora of well-curated insights that the user can instantly process to choose within seconds.

RELEVANCE

If billions of data points are produced for a specific object, and a user searches that object, which data point should surface?

This is solved through the Knowledge Graph's capability to include peer insights (peer here means peer data assets).

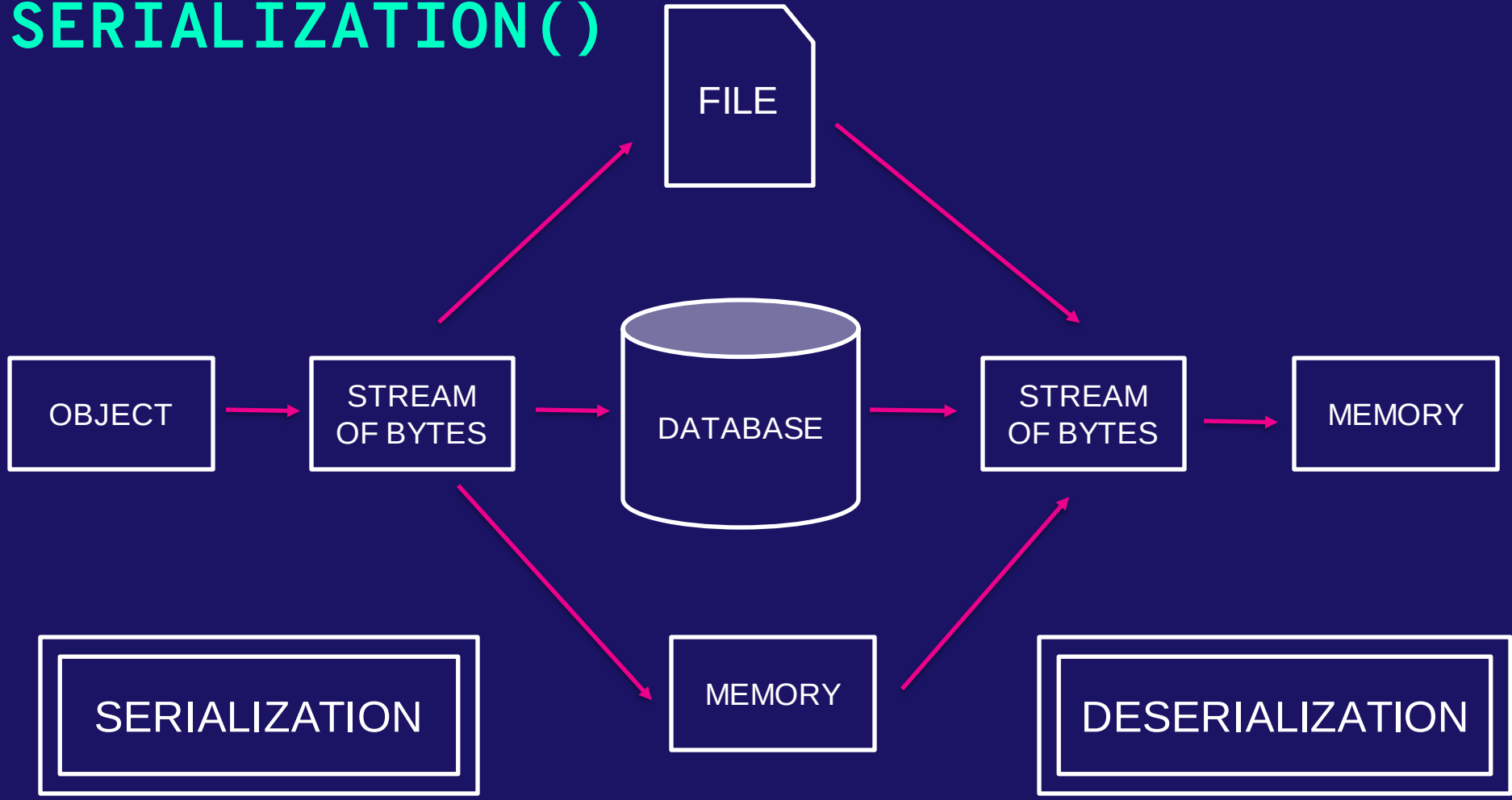
SEO or Search Engine Optimization is the method to curate and surface information that, on a high level, has:

- The right set of keywords (or tags)
- Several data points that point to that data point as a relevant source of information (backlinks- peer validation)

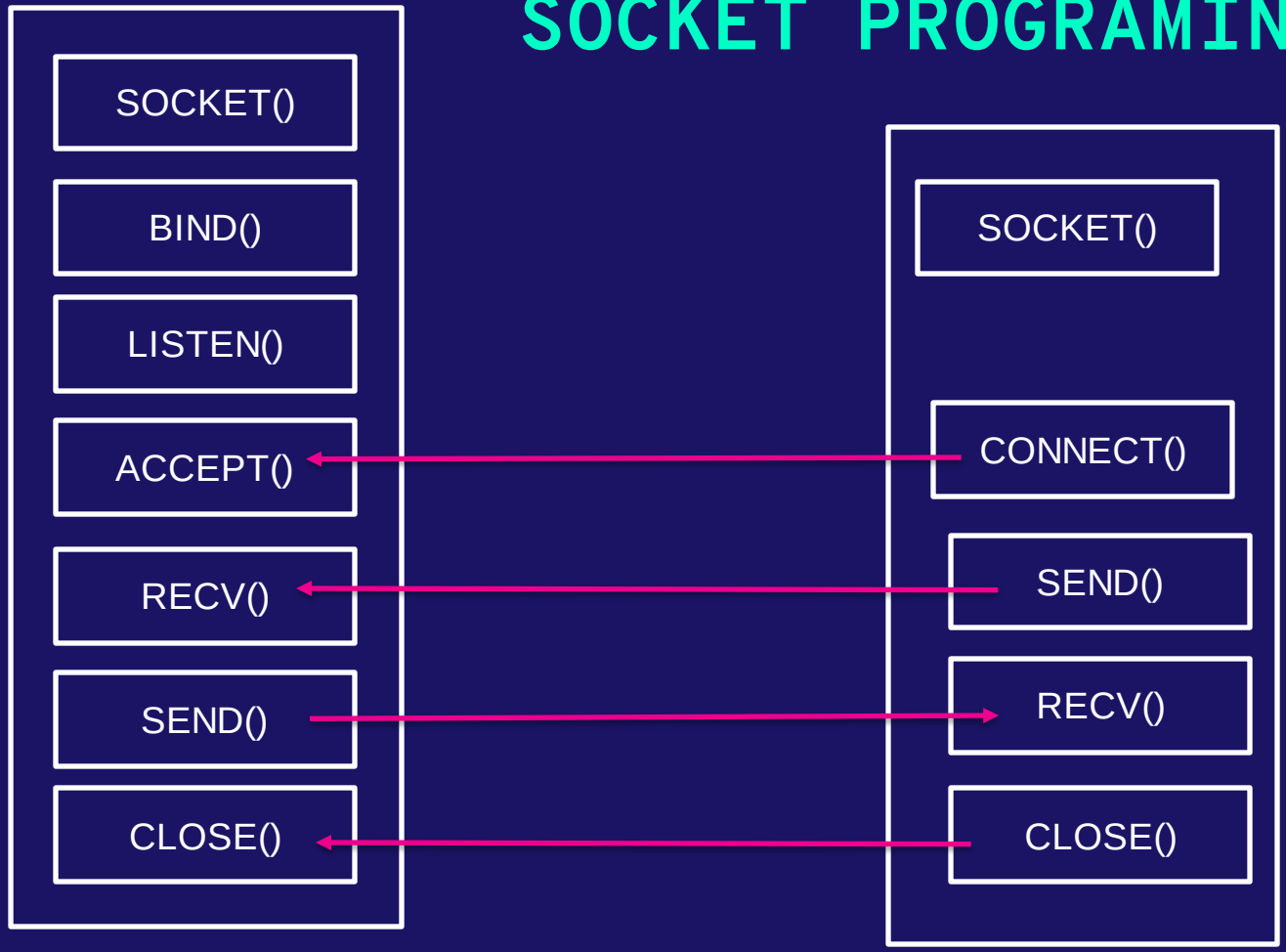
05

FUTURE GOALS

SERIALIZATION()



SOCKET PROGRAMING

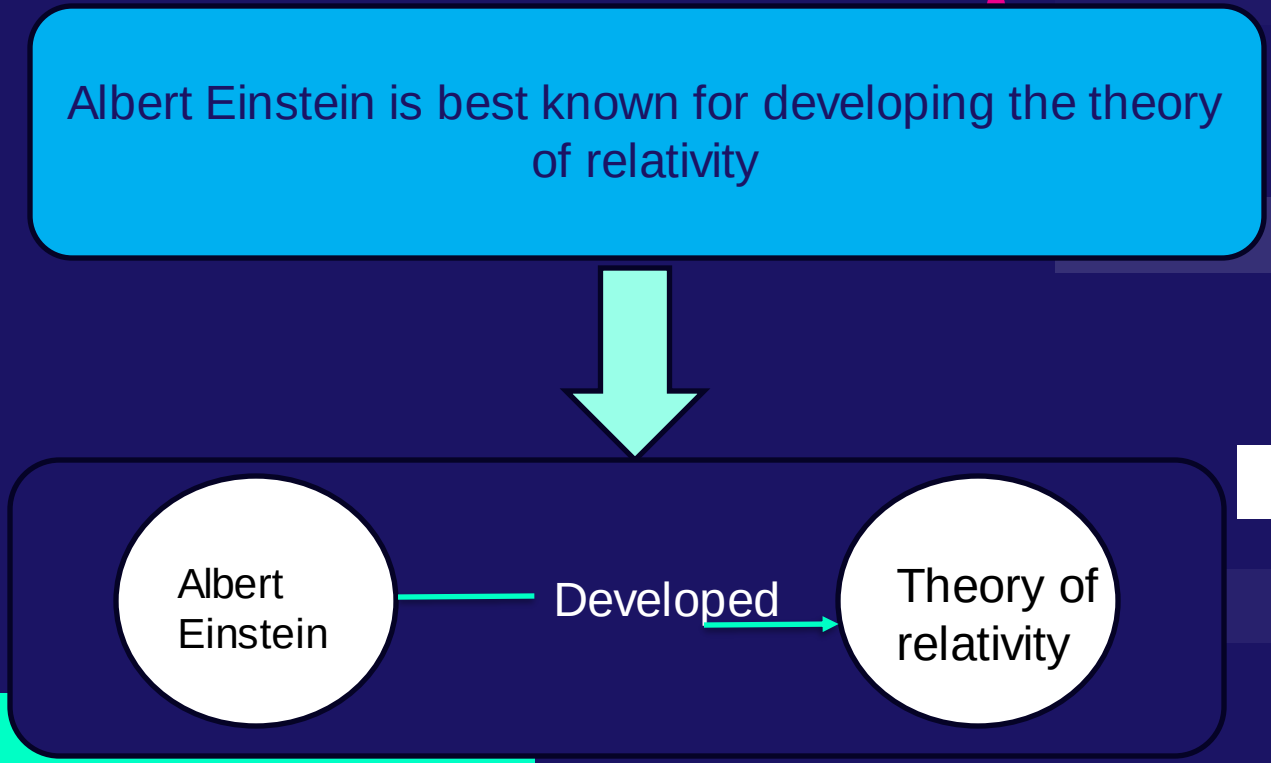


NLP : NATURAL LANGUAGE PROCESSING

OBJECT VERB SUBJECT

Thomas Jefferson was an American statesman, diplomat, lawyer, architect, philosopher and Founding Father who served as the third president of the United States from 1801 to 1809.

The goal of information extraction pipeline is to extract structured information from unstructured text. Image by the author.



THANK YOU!