

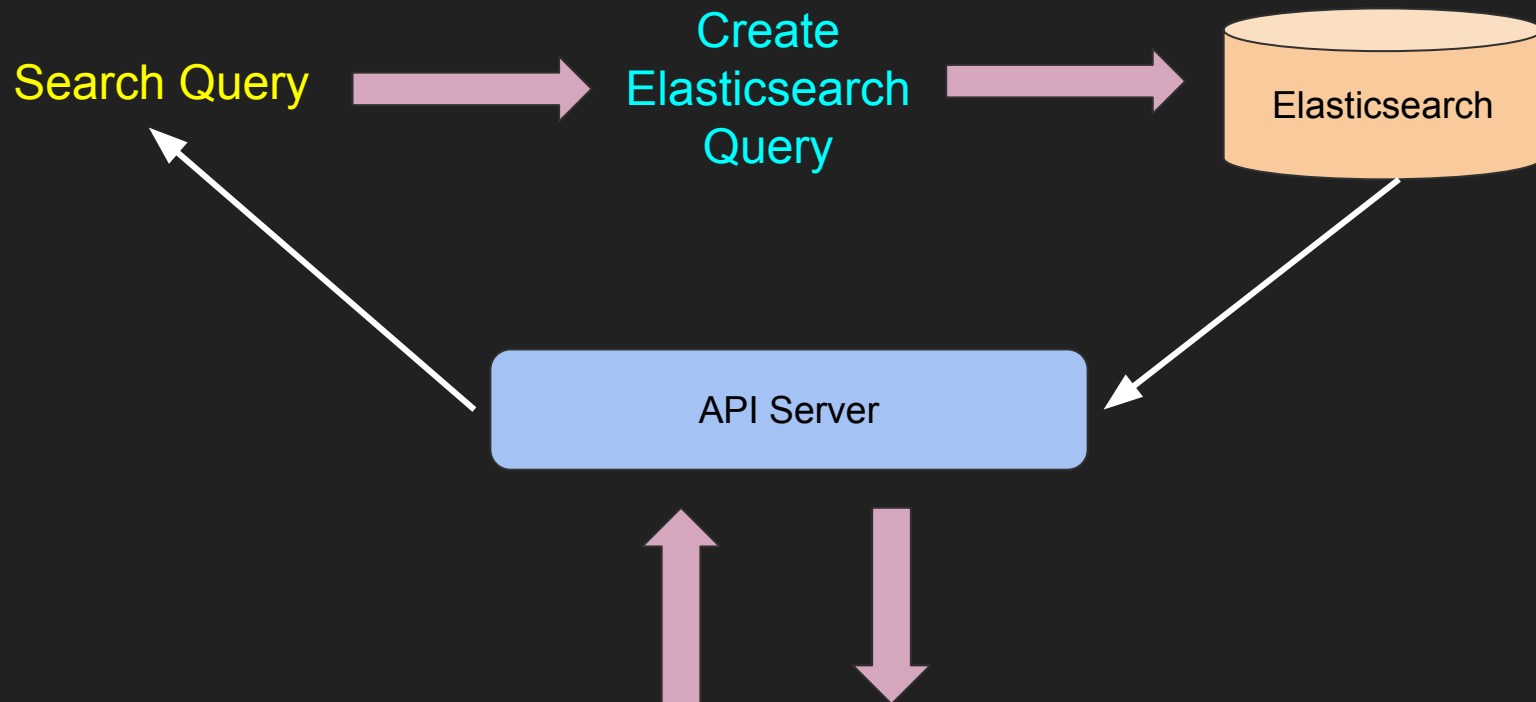


OpenHack

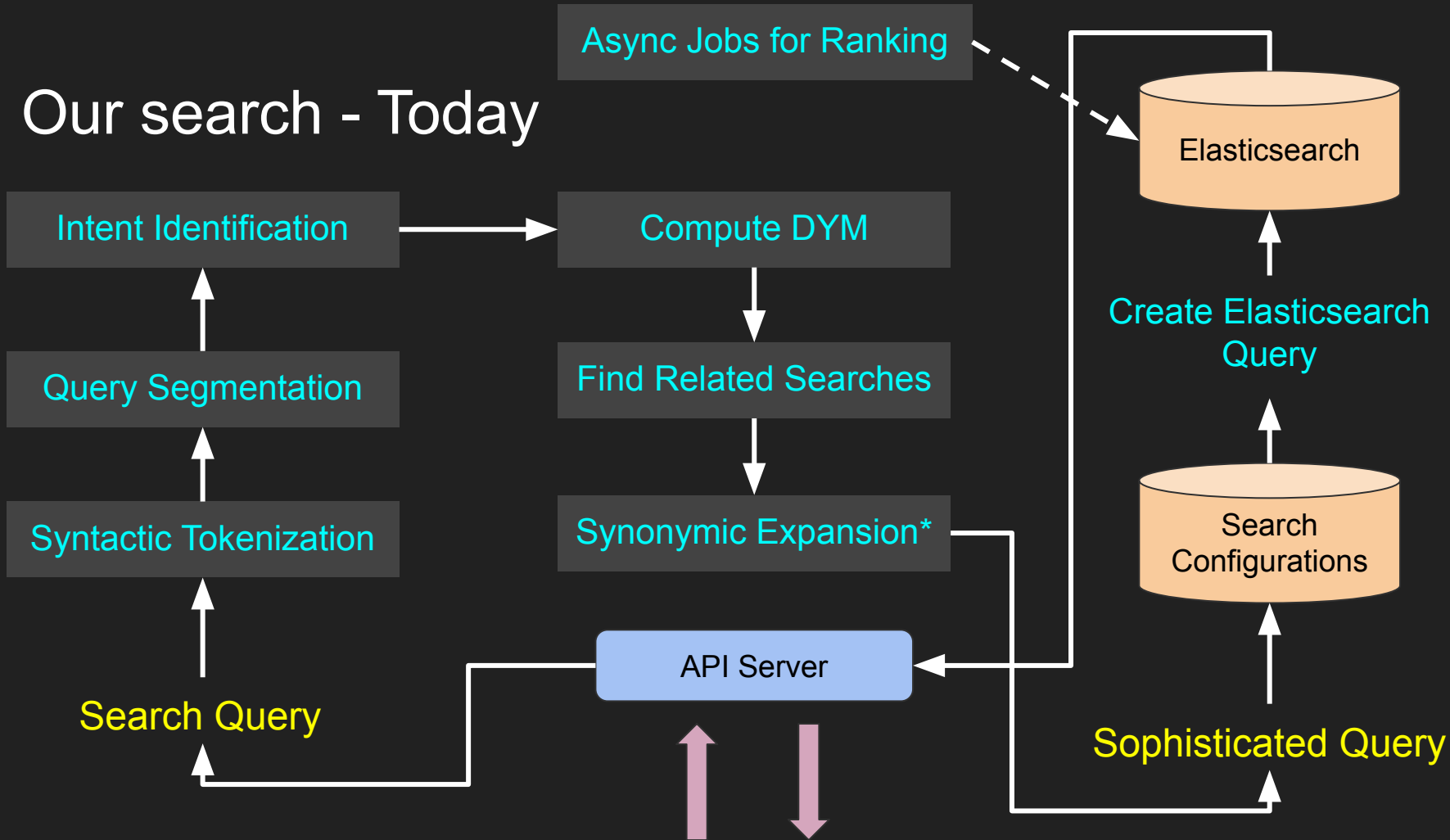
Search and Content Discovery

29/03/2019

Our search - Before



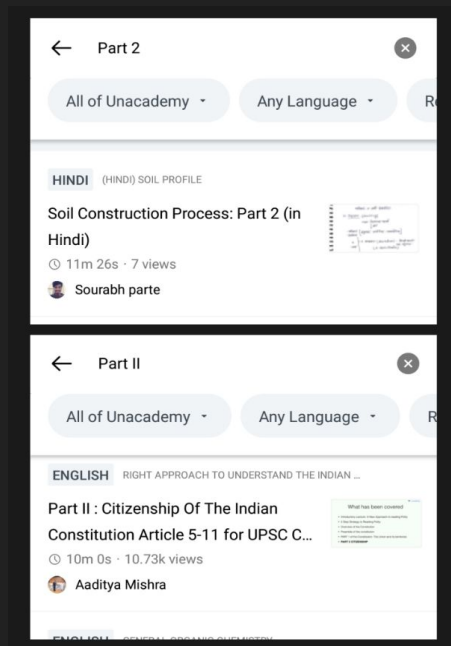
Our search - Today



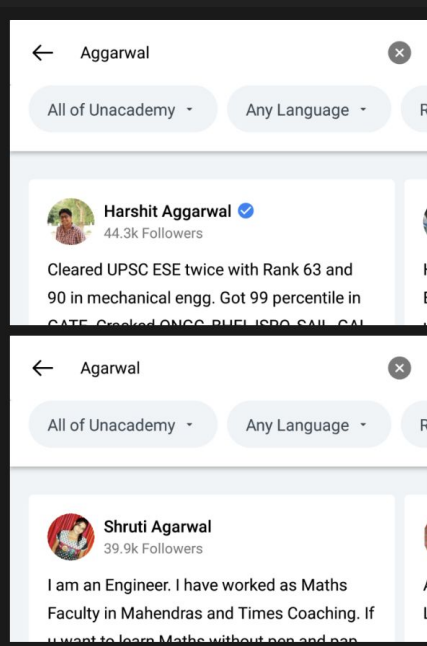
Over to Ayush

Problem Statement

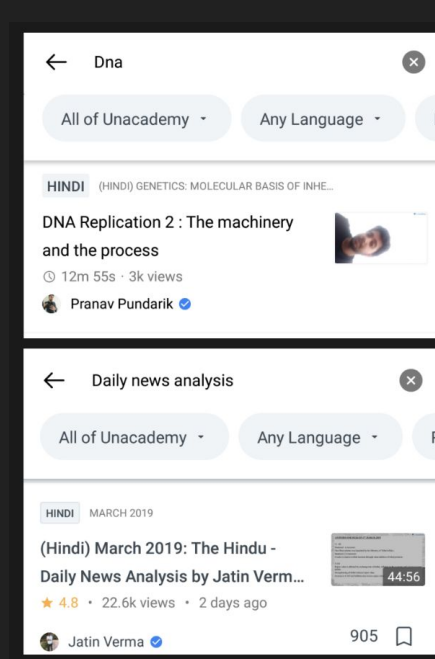
Different words with
same meaning



Similar Sounding
Words



Abbreviations
and Expansions



Semantic Similarity

WordNet and Thesaurus were used to obtain the following results

C	: 100, carbon
100	: hundred, C
one	: 1, 1st, I, start
exam	: examination, test
maths	: mathematics, math
pedagogy	: education, teaching

Phonetic Similarity

Weighted average of Soundex, NYSIIS and Metaphone Encoding were used to obtain the following results

pooja, puja

niharika, neharika

ayush, aayush, ayussh

shaifali, shifali, shefali

chauhan, chouhan

aggarwal, aggarwal, agarwala

bhadauria, bhadouria, bhadoria

rathor, rathaur, rathore, rathour

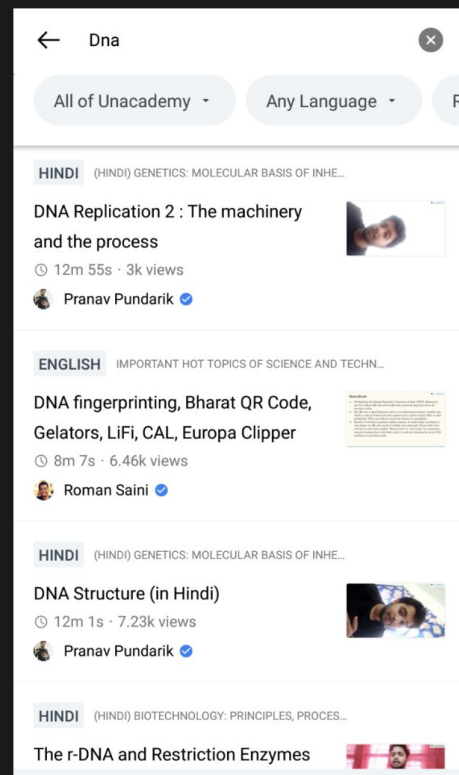
Abbreviation Extraction and Expansion

Methodology

1. Schwartz Hearst Algorithm
2. Custom Algorithm

Dataset

Course and Topology Descriptions



Abbreviation Extraction Results

Examination Names

SSC : Staff Selection Commission
UPSC : Union Public Service Commission
IELTS : International English Language
Testing System

Technical Jargons

IoT : Internet of Things
op-amp : Operational Amplifier
HTML : Hypertext Markup Language
DBMS : Database Management System

Domain Specific

DNA : Daily News Analysis
PYQ : Previous Year Questions
MCQs : Multiple Choice Questions
MCQS : Multiple Choice Question Series

Others

sin : sine
P&C : Permutation & Combinations
UN : United Nations Organization
IPv4 : Internet Protocol Version 4
DI-LR : Data Interpretation and Logical Reasoning

Future Scope

- **Multilingual synonyms**
eg. pur and पुर
- **Extending phonetics** to all proper nouns
e.g. politechnic, polytechnic, pollytecnic
- **Accommodating acronyms** that are
 - infrequent e.g. AFCAT
 - unconventional e.g. PhD, MBBS, CERN

Thank you