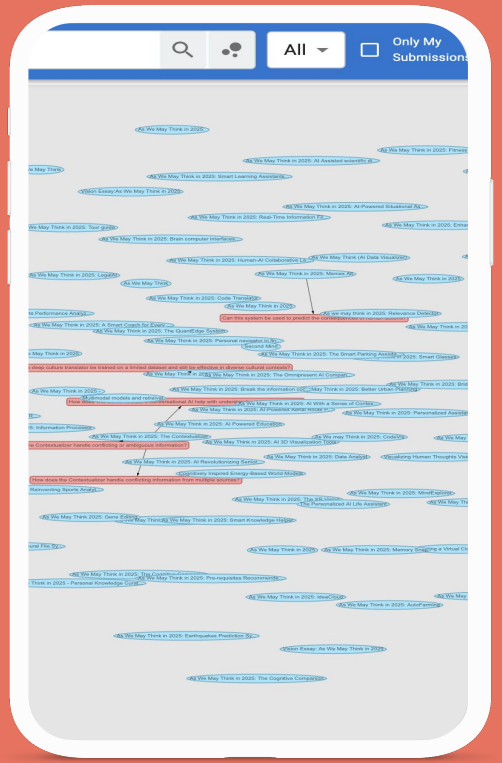
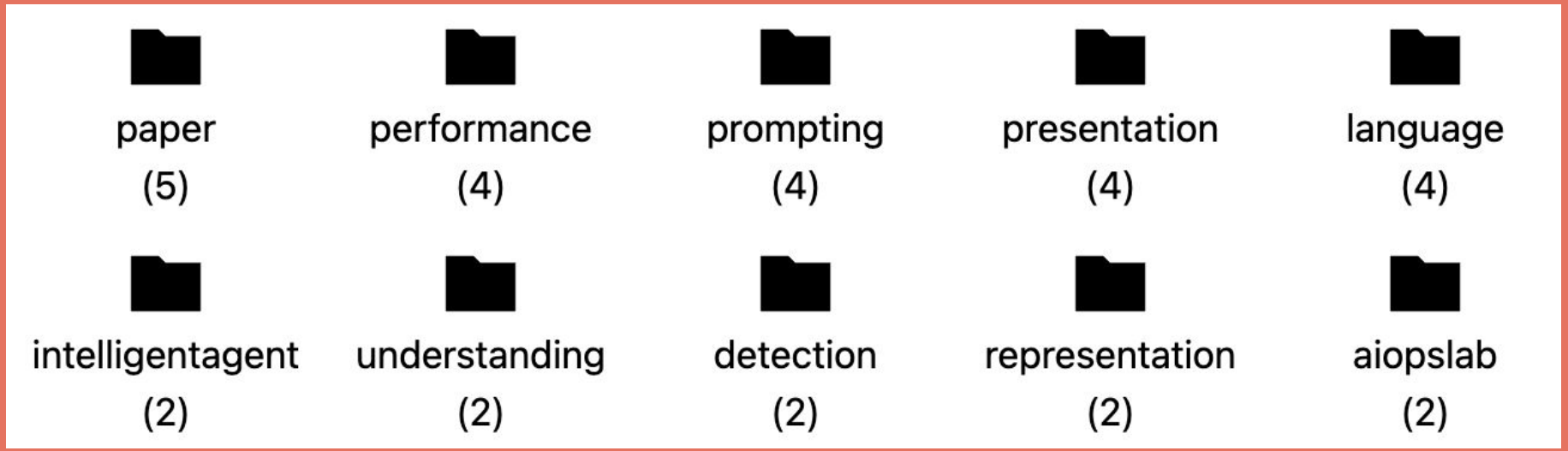
TextData Visual Search

Dhruv Saligram
Group 7

Motivations

- Engineering a potential improvement on the current system
- Creating a more intuitive search mechanism
- Increasing ease of exploration
- Making submissions more navigable
- Centralizing information into one area





New Visual Search

Key Goals

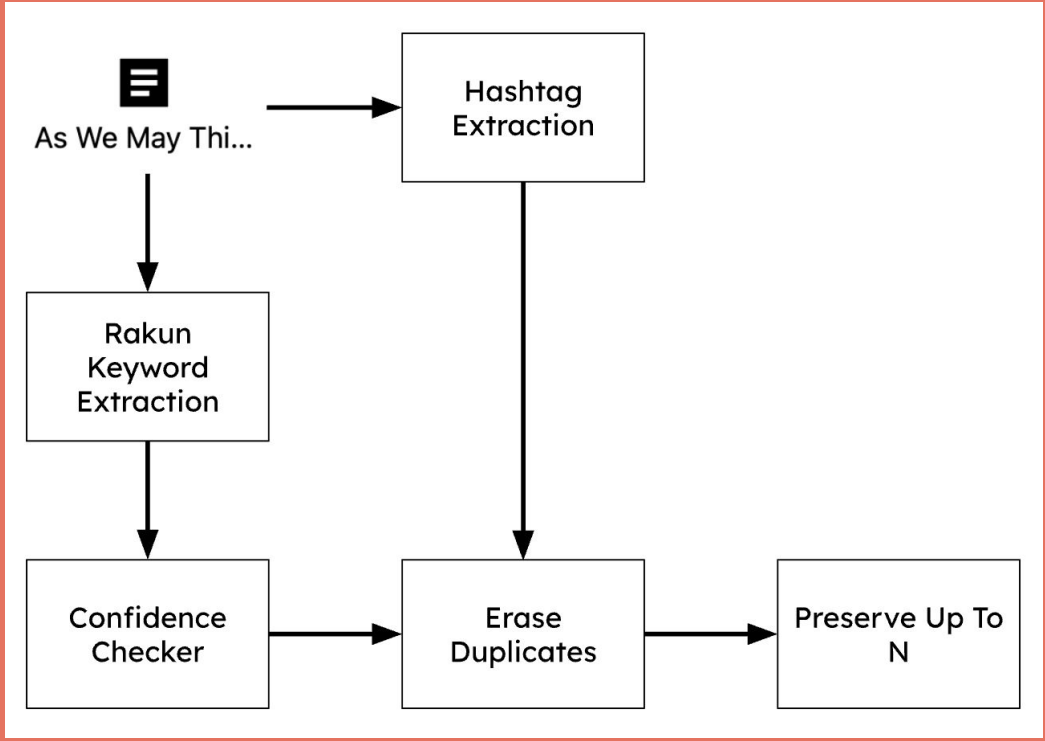
To successfully create a useful tool, there were a few key goals that had to be achieved:

1. Get meaningful keywords from each submission
2. Merge keywords to reduce clutter and improve quality
3. Organize documents hierarchically by keyword
4. Visualize the tree to users for free and intuitive navigation

Key Goal #1

Get meaningful keywords from each submission

- Extract user hashtags
- Keyword extractors
  - KeyBert, Rakun, RAKE, TextBlob
- Extract keywords
- Preserve by confidence
- Ensure no duplicates
- Limit amount

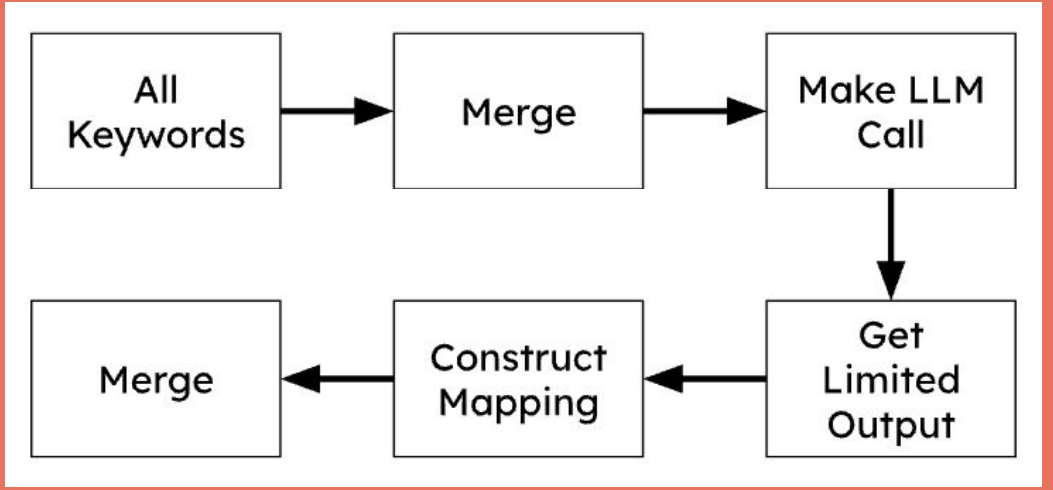


“#vision #sports A sports data analysis tool could be in the future. It would take in data from sports games and offer analysis.”
→ [vision, sports, data, analysis]

Key Goal #2

Merge keywords to reduce clutter and improve quality

- Merge keywords on
  - Plurals, suffixes, & synonyms
- LLM pass over keywords
- Limit output
  - For time & garbage
- Extract any dict structure and merge again



(1) [ai, artificial intelligence, llm, llms, language models, design, designer]

(2) [ai, artificial intelligence, llm, language models, design]

(3) {artificial intelligence: ai, language models: llm}

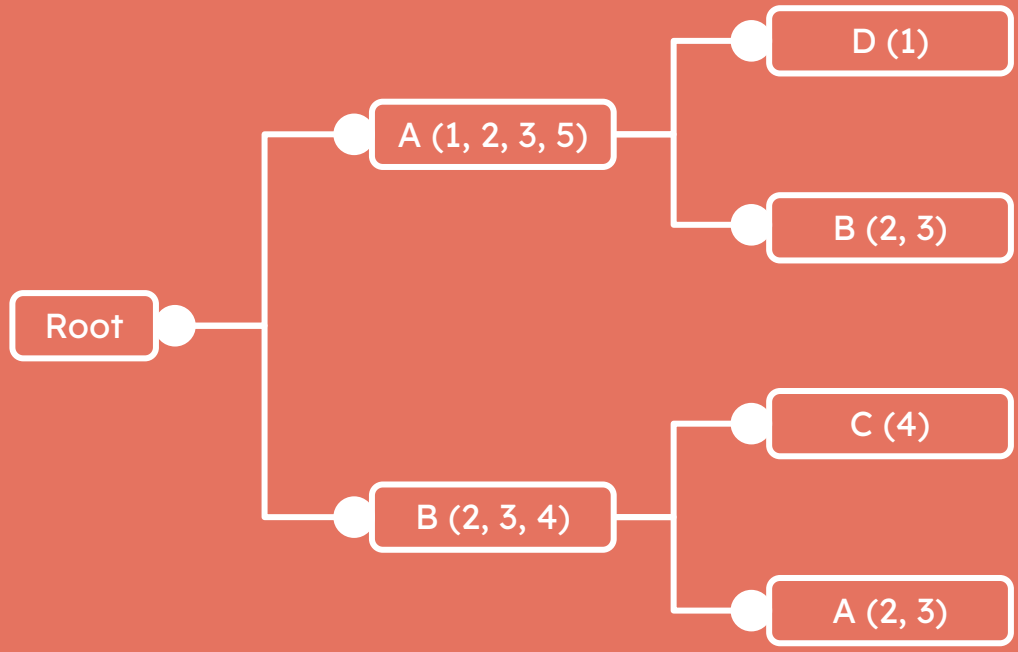
(4) [ai, llm, design]

Key Goal #3

Organize documents hierarchically by keyword

- Tree recursion + greedy set cover
- Correctly sharing documents between children
- Documents exist at many levels → each doc's path depth varies
- Misc. folder

Document 1: A, D
Document 2: A, B
Document 3: A, B
Document 4: B, C
Document 5: A

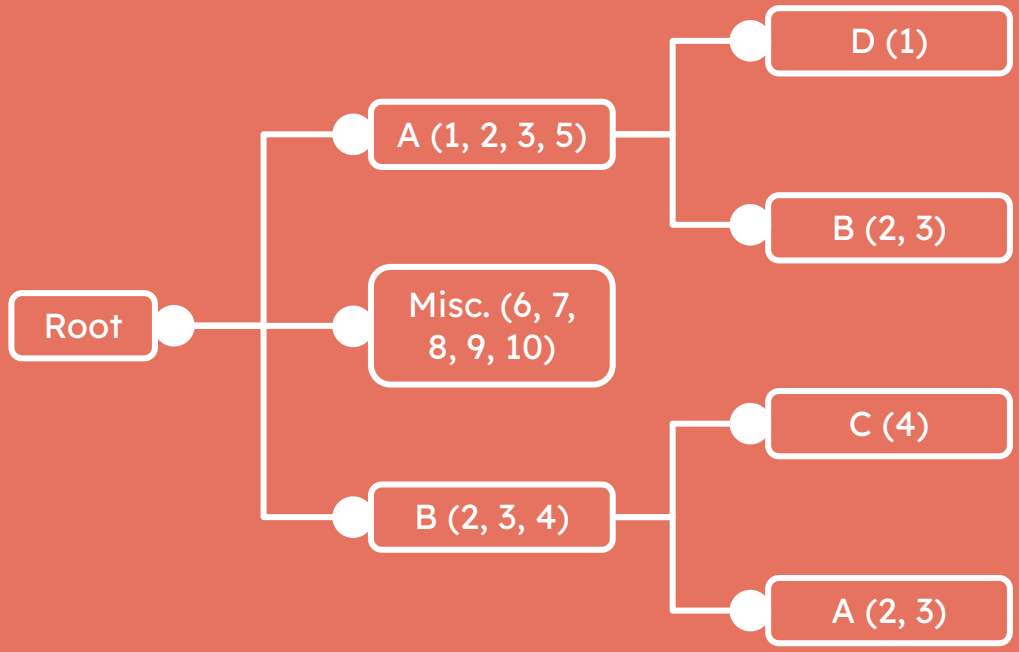


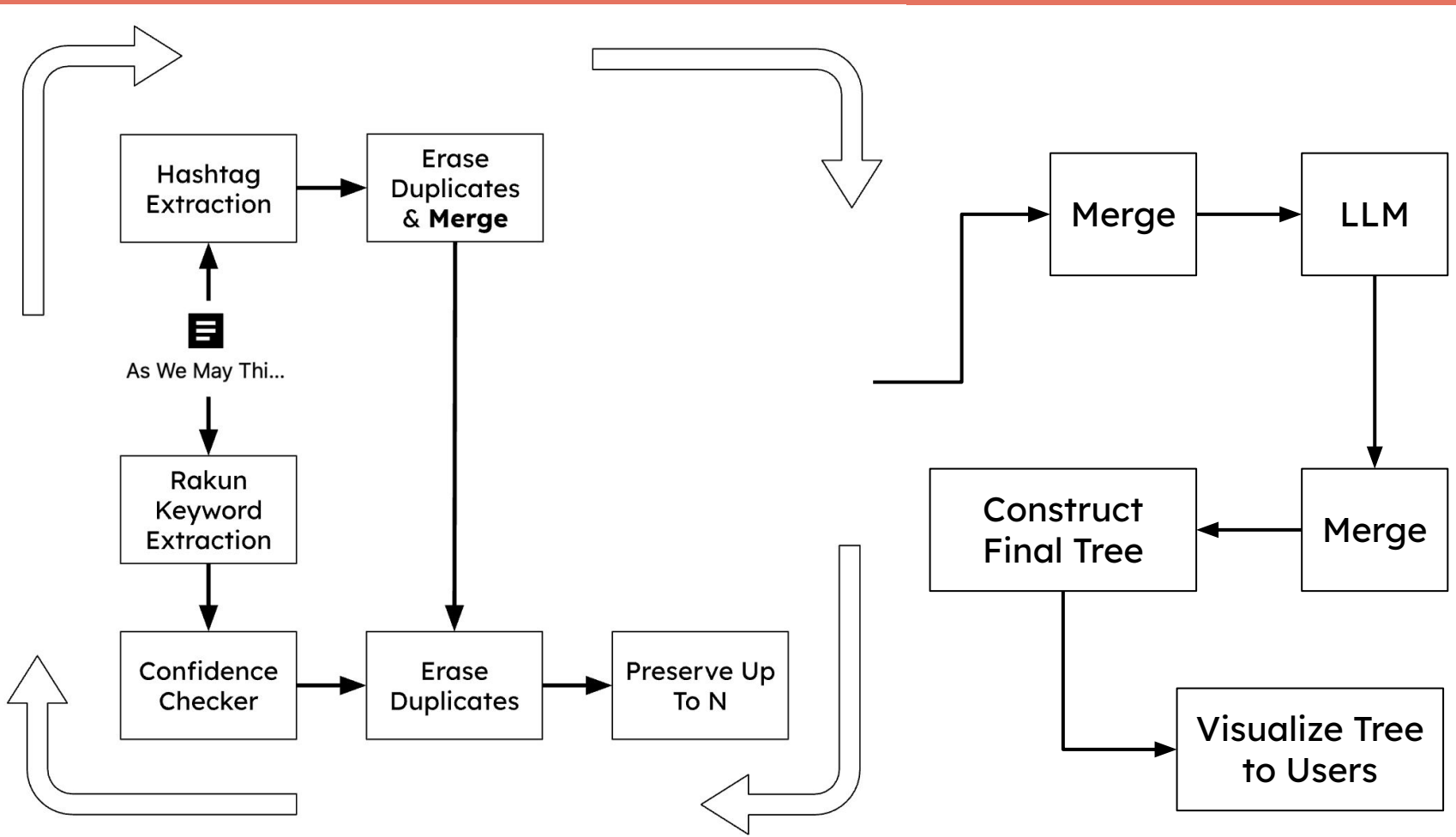
Key Goal #3

Organize documents hierarchically by keyword

- Tree recursion + greedy set cover
- Correctly sharing documents between children
- Documents exist at many levels → each doc's path depth varies
- Misc. folder

Document 1: A, D
Document 2: A, B
Document 3: A, B
Document 4: B, C
Document 5: A
Document 6: V
Document 7: W
Document 8: X
Document 9: Y
Document 10: Z

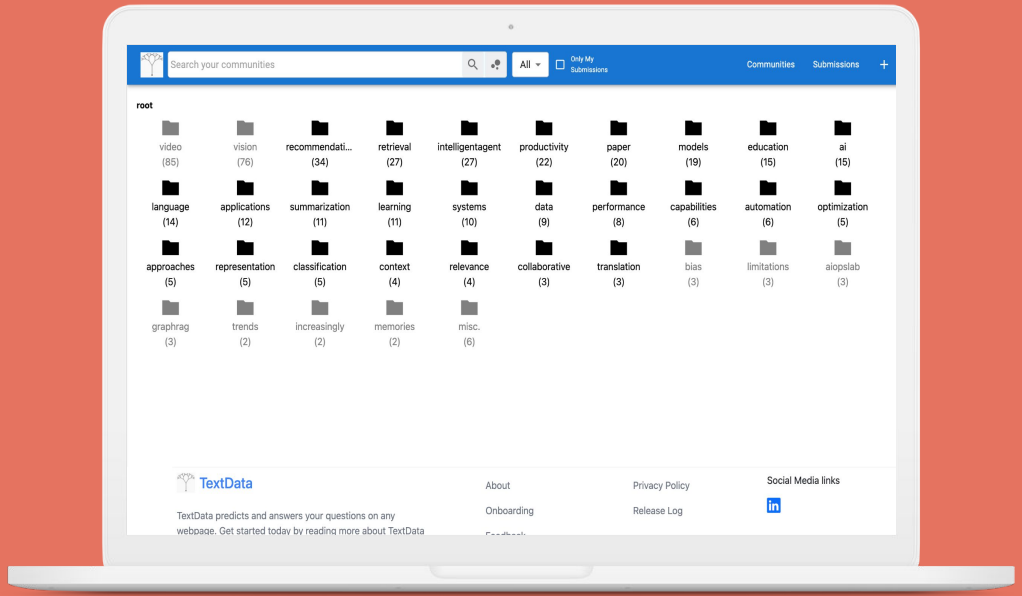



Hashtag Extraction
Erase Duplicates & Merge
As We May Thi...
Rakun Keyword Extraction
Confidence Checker
Erase Duplicates
Preserve Up To N
Merge
LLM
Merge
Construct Final Tree
Visualize Tree to Users

Key Goal #4

Visualize the tree to users for free and intuitive navigation

- React frontend
- Navigating through folders
- Breadcrumb trail
- Opening documents
- Visited documents
- Visited paths (not keywords)



Intended Users & Impacts

Community Members

- Help users get a better understanding of a community's entire range
- Allow users to immediately find closely related documents
- Enable exploration and searching without any pre-established direction

Community Administrators

- Increase engagement in communities
- Allow communities to be more inclusive of various topics (i.e. no need for separate Frontier Topics & Base communities)
- Find clusterings of documents and/or users for matching + grouping

Thank You!