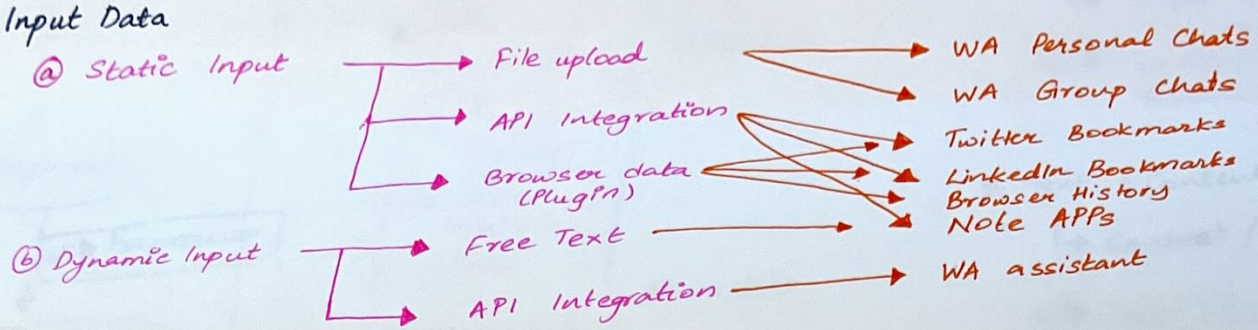
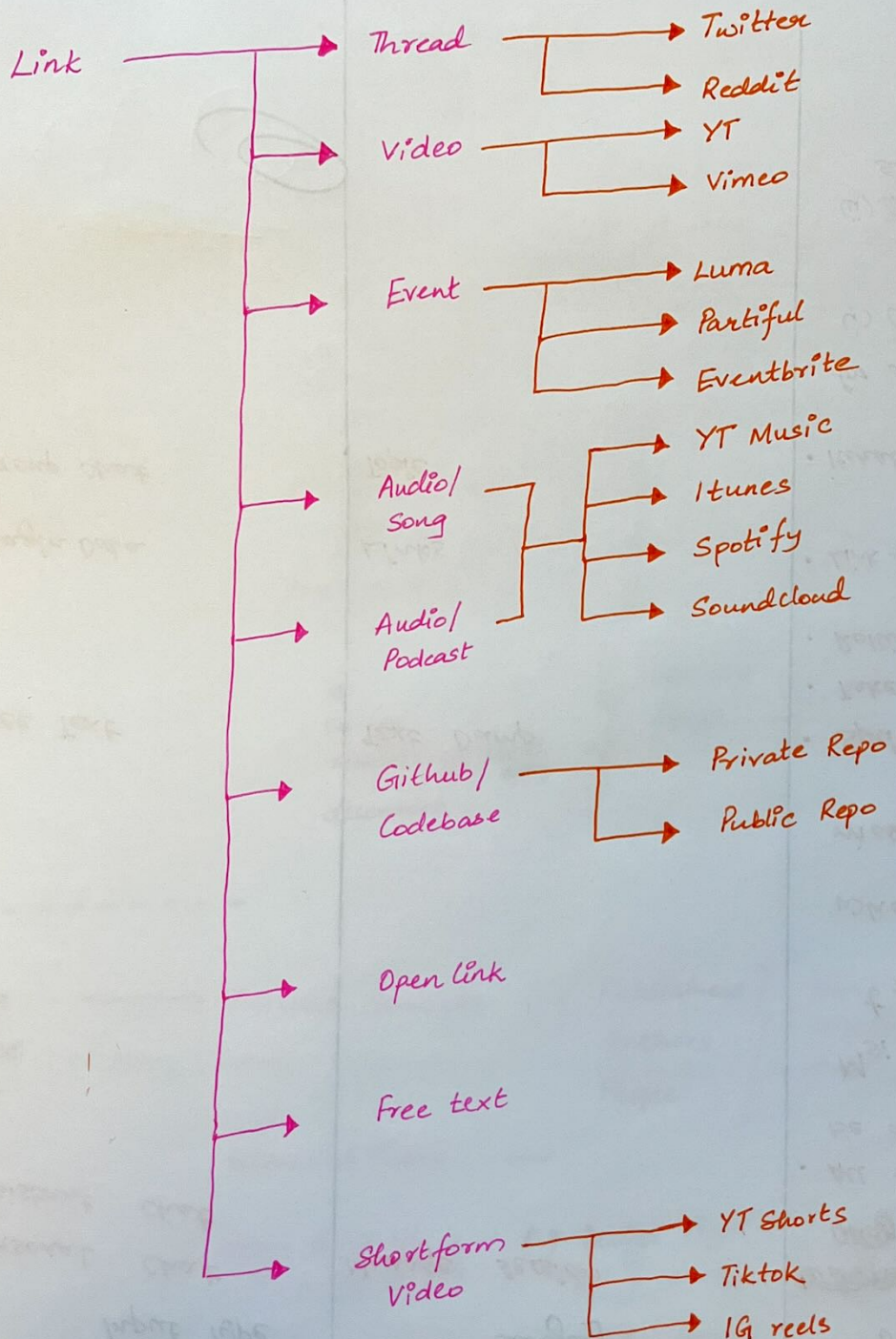
## Input Data

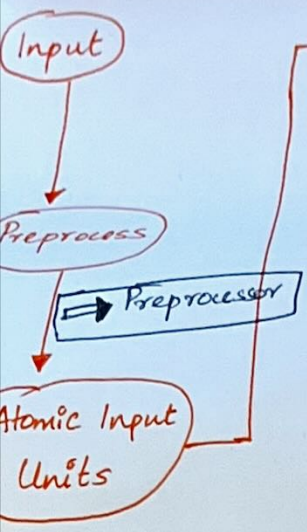



## Input Processing



Atomic Unit for Input →

| Input Type | Category | Definition |
|---|---|---|
| (a) Personal chat<br>(b) Assistant chat | Message session | ~~Segment Messages sequence using log~~<br>• All messages in a session will be defined by<br>$M_{si} = \{m_n\}$ where<br>$t_{m_{n+1}} - t_{m_n} \leq 5$ mins<br>where $t_{m_{n+1}}$ is time at which message $n+1$ is sent |
| (c) Free Text | Text Dump | • Split on new line<br>• Take as whole<br>• Rolling window |
| (d) Plugin Data | Links | • Link + Tweet \| Post \| Reddit thread info |
| (e) Group chat | Topic | • Iteratively .... explo req<br>for all $m$ in $m_t$ of $x$ day chunks:<br>(i) Categorize messages based on tags/labels<br>(ii) ~~Thk~~ Coalesce adjoining Topic - Message sets |



Input

Preprocess

⇒ Preprocessor

Atomic Input Units

⇒ Rawdata Parser

* Text ?
\+
* Links ?

User Generated !

\+

Timestamp
- created
- updated !

Sender
- User/other !
- Name
- Id

Source !

Type of Input !

\# Topic → Display, BE ?

\# Labels → Display, BE !

\# Descriptions ?

Metadata

Read this
- Source → User/System
- Date (Action)
- Updated On

Favourite this
- Date (Action)

Recommend this
- List

Snooze this

User Action Data.

Id
source
relation_type (one/two)
description

Network Graph.

* Text + Content from links
↳ Corpa → Embed corpa

⇒ Content extractor from links

* Text + Content from links
↳ Content /

⇒ Topic | Subtopic generator

⇒ Label generator

⇒ Refiner (Topics) → Agentic

⇒ Refiner (Label) → Agentic

Ap 1:

$L'_i$

↓

Similar labels

↓

$Sim_{max} \geq$ Threshold

Yes → Use

Generate

---

User Space → Trusted Sources → Publishers, Authors, People

name
keyword
link

Interested Topics →

Other Derrieved Attributes ?

e.g. Design Engineer
Product Mgr
Tech Strategy
Coder

current interests
past interests

Generation Tone

Features:

① Search → Search Results View → Navigation to Content → Easy reader
Remind me → Assistant (n
Read this out
Read these out → Podcast (

② Content Space → Translate Content to a 2D Chart → ⓐ Give me suggestions
(i) Agentic Search
↓
- Generate key words using "Trusted Sources"
- Balance → shouldn't become an echo chamber
(ii) Other users' reco list

③ About Me / Profile Me → ⓐ Public Reccos ("This is what I'm reading)
ⓑ Private Shared Lists (e.g. Tech articles on LLMs)
ⓒ To do list
↳ ⓓ Synced with calendar
ⓓ Create a tinder profile (? Hehehe)
ⓔ Create a portfolio website
ⓕ Tweets / LinkedIn Posts / Shareable WA

Ⓐ Help me cite → Track browser history, Saved / Bookmarked links } ⓐ ↔ Uploaded info to create citations for → Blogpost