

Program Next Level (PNL)

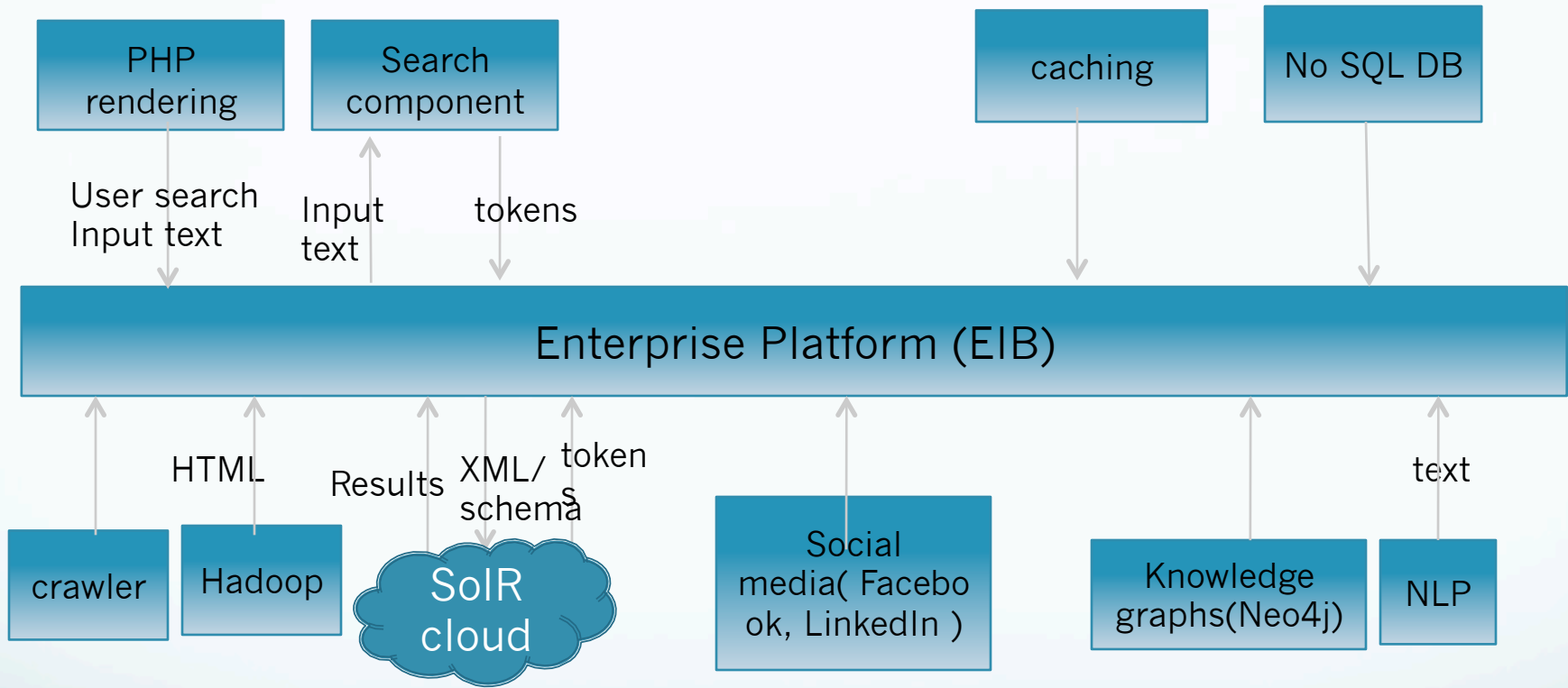
Architectural Overview and Workflow

PNL : A powerful answer engine which creates simplified, interactive and one page stories of every search query.

Capabilities

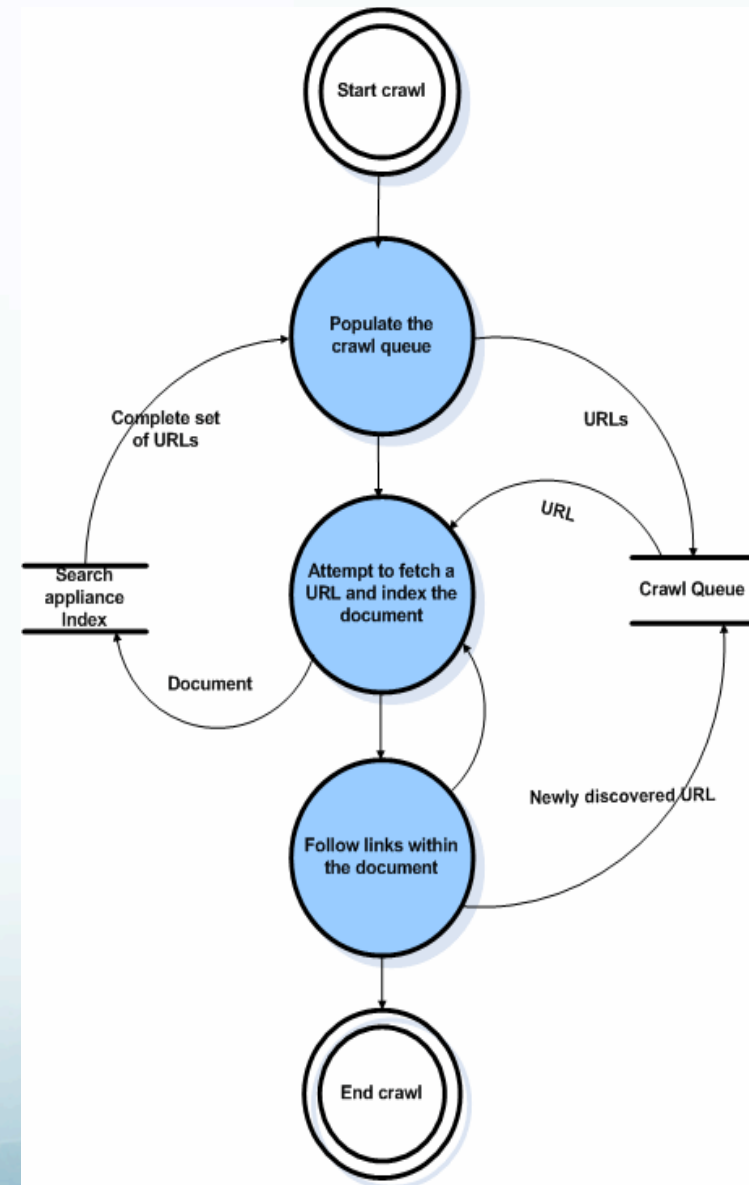
- Quickly search the web to produce right and precise information.
- Add context to your search queries using your personalized data from social networking sites and blogs resulting in re-ranked pages.
- Add meaning to your search queries using Natural Language Processing tool.
- Storing the indexed data in graphical format using highly scalable ,robust native Neo4j graph database.

Overall Solution Architecture



Stage 1: Data crawling Workflow

1. Populate the crawl queue with the seed and complete set of URLs
2. Fetch the URL from the crawl queue and start crawling the data
3. Process the documents using Hadoop Map Reduce paradigm
4. Follow links within document for crawling.
5. Store the processed and indexed data in SOLR

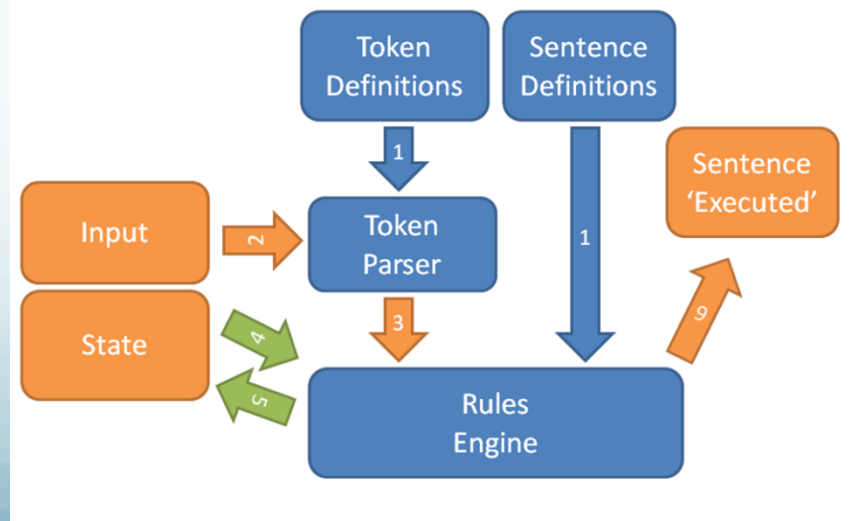
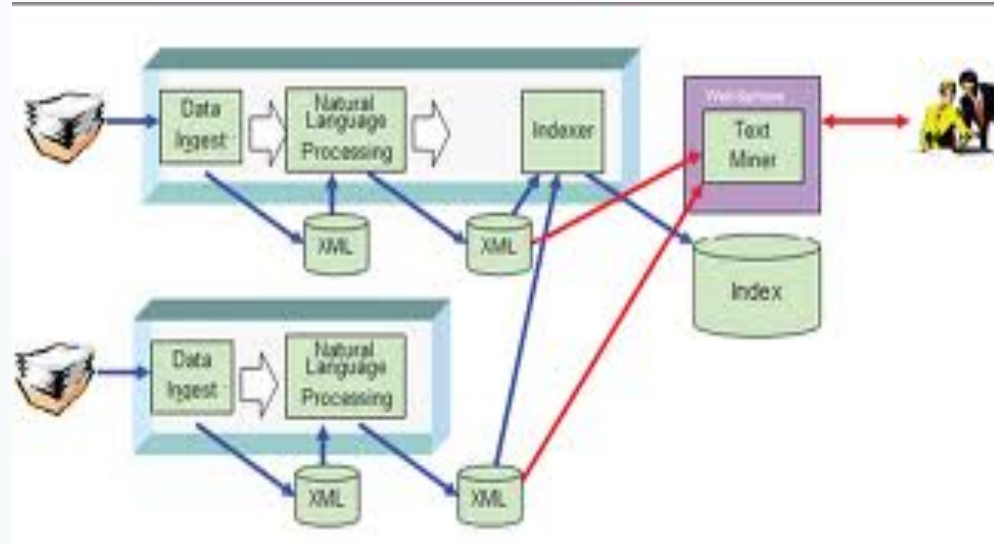


Stage 1(b): Natural Language Processing

This involves

1. Classify the content into different categories

We are working on this step & defining ontologies

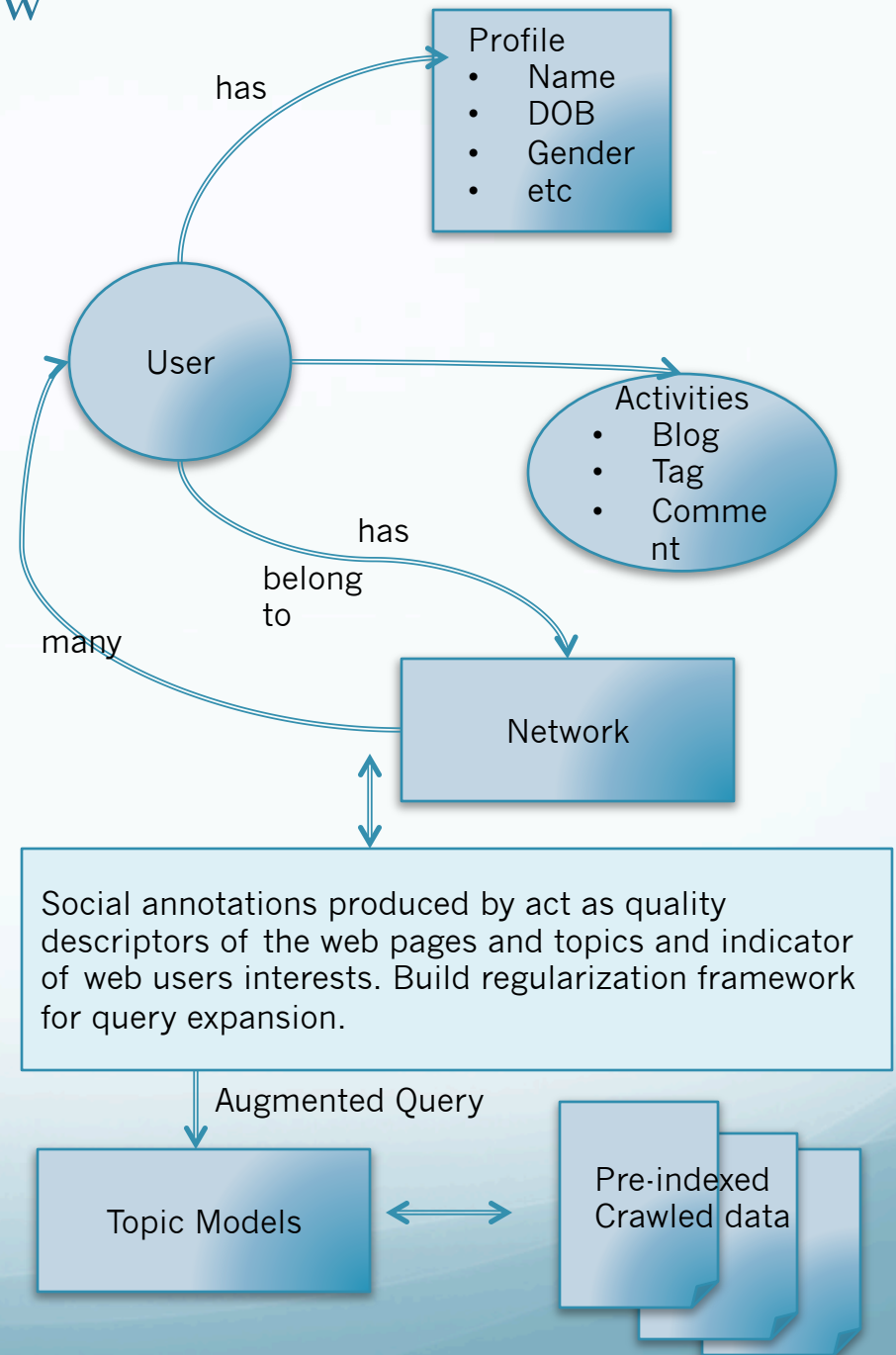


Stage 2: Personalized Data Workflow

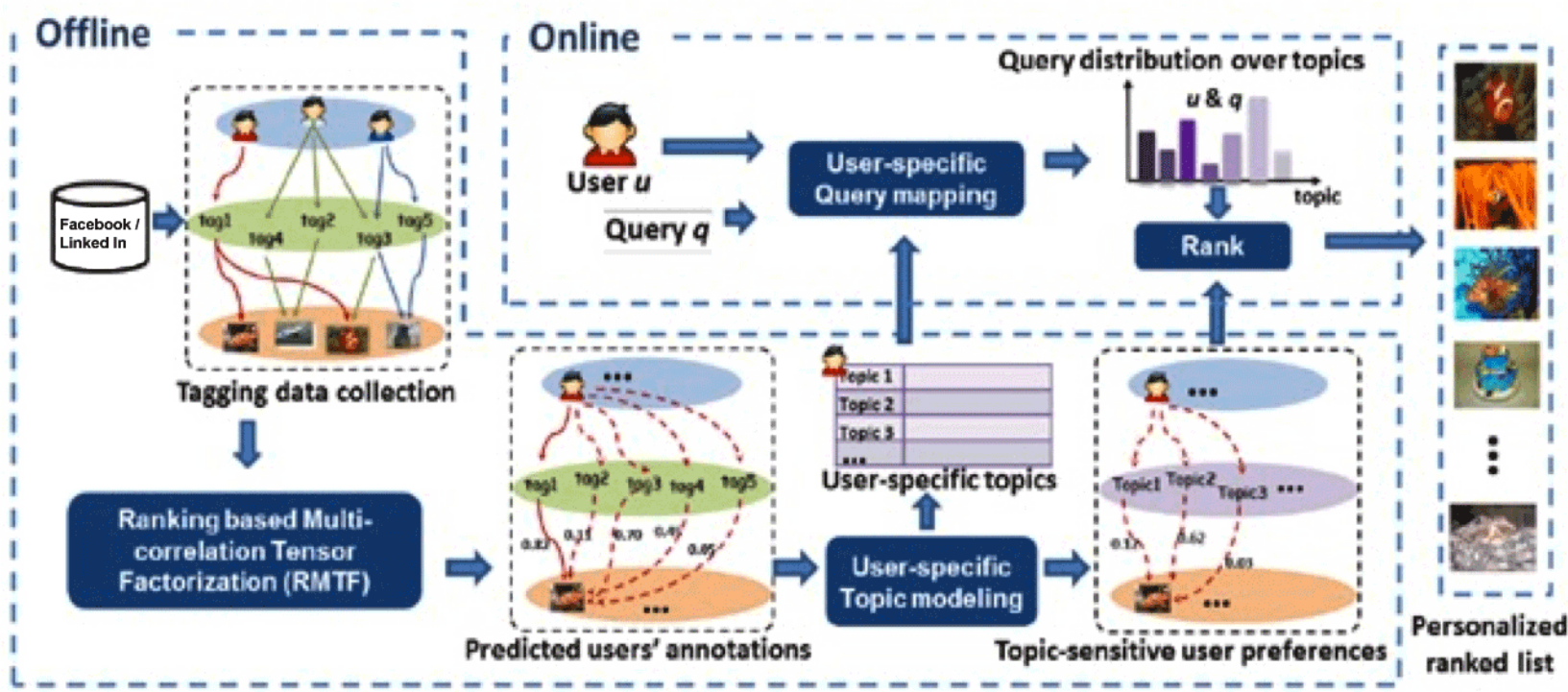
Embed the user preference and query related search intent into user-specific topic spaces.

This involves

1. **User specific information** i.e. name, date of birth , gender, language ,country, will be considered to distinguish the exact intentions of the user queries.
2. **User 's social activities** such as rating, tagging and commenting will be considered to indicate the user's interest and preference in a specific document.
3. **User's social network** i.e. friend list **social activities'** to indicate the affinity between user and user's friends for better decision making for every search query.
4. Finally , storing this data in Neo4j graphical database.



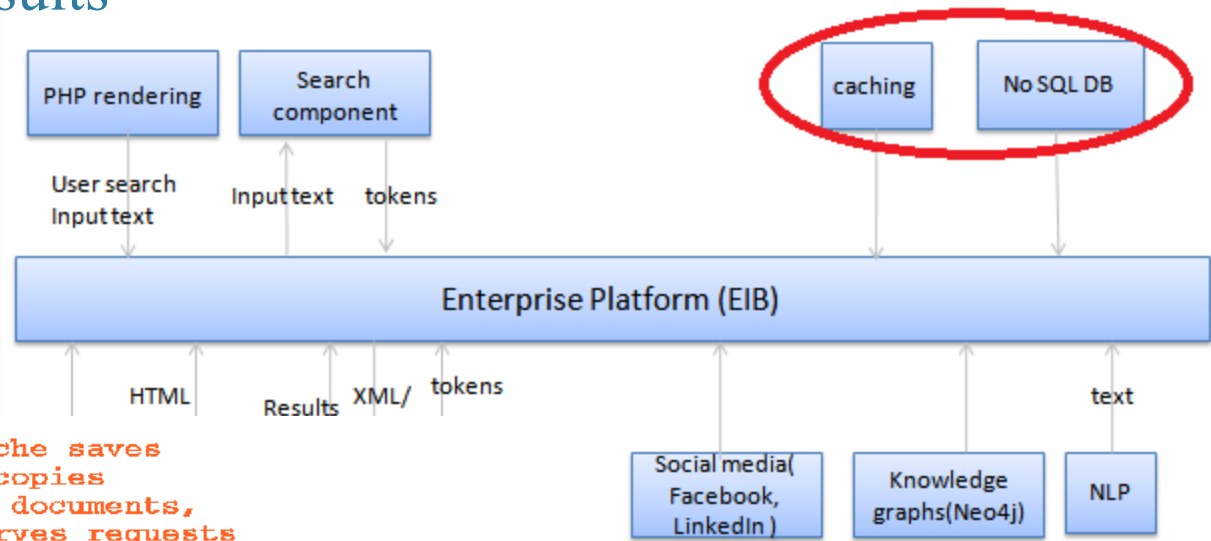
Stage 2 (b): Personalized Data Workflow



Search



Stage 4: Preparing Results



The cache saves local copies of Web documents, and serves requests when a copy is available

