NLP Workbench: Efficient and Extensible Integration of State-of-the-art Text Mining Tools

Peiran Yao, Matej Kosmajac, Abeer Waheed, Kostyantyn Guzhva, Natalie Hervieux, Denilson Barbosa

Department of Computing Science, University of Alberta, Canada

UNIVERSITY OF ALBERTA

NLP Workbench is ...

A web-based platform for text mining
- Targeting non-expert users
- Semantic understanding of large-scale corpora
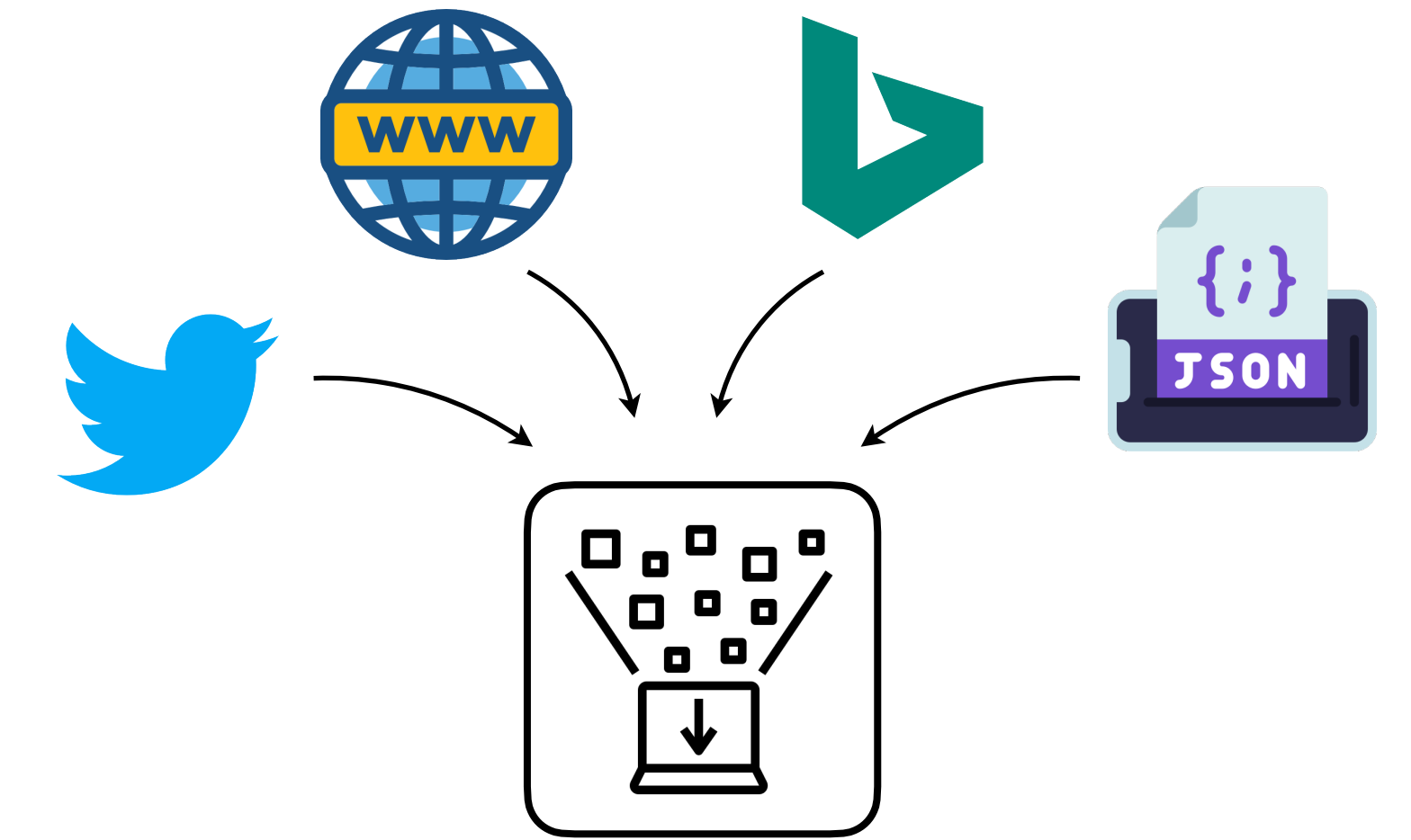- Built upon latest pre-trained models and open source systems

Providing semantic understanding
- Entity linking
- Customizable relation extraction
- Semantic parsing
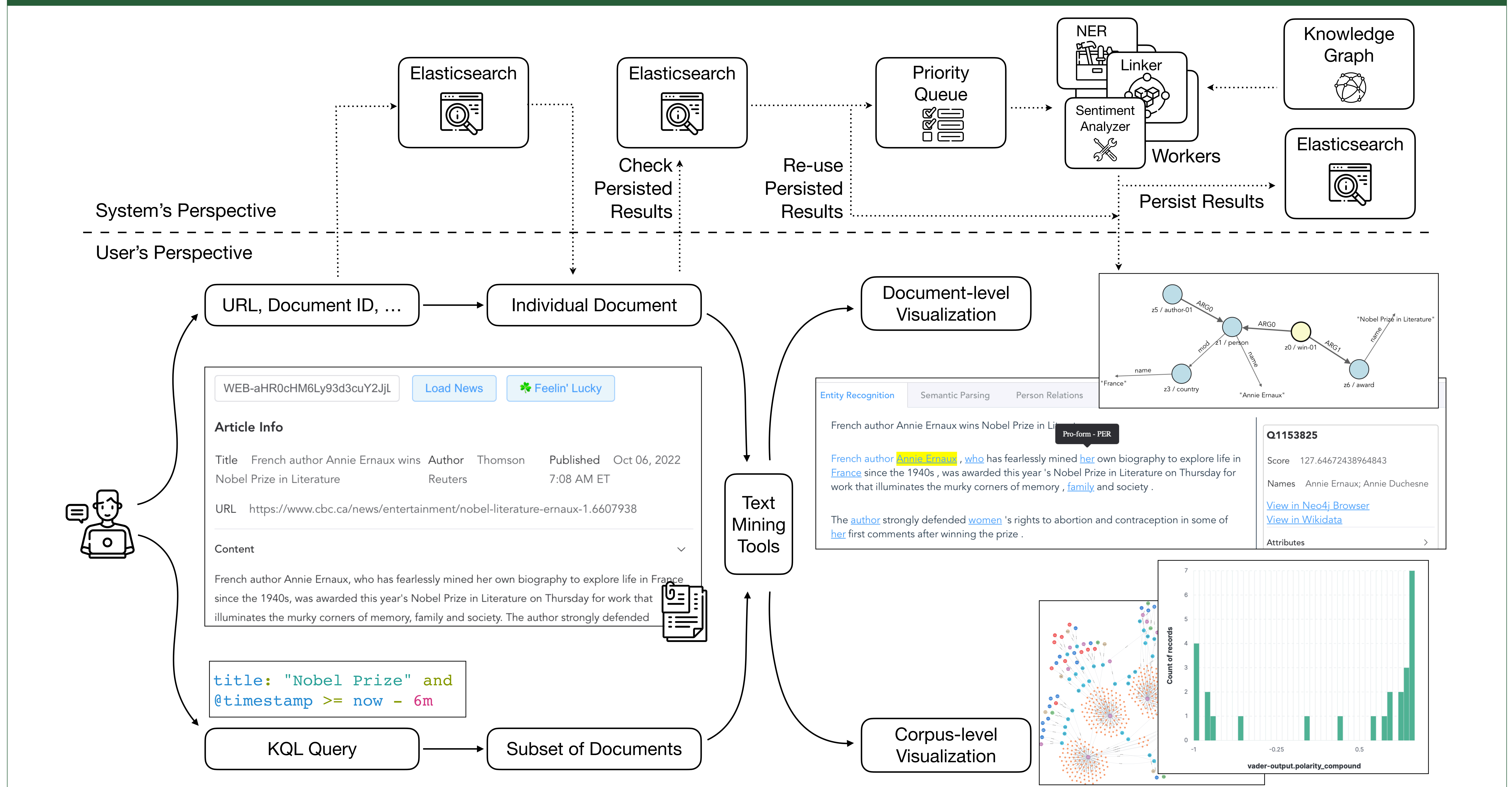- Classification
- Easily extensible to include more!

Corpus Building

A document collection can be created
- From many sources
- By custom queries
- With social network interactions associated

Workflow: Apply Models & Visualize

Elasticsearch — Elasticsearch — Priority Queue — NER / Linker / Sentiment Analyzer / Workers — Knowledge Graph — Elasticsearch

Check Persisted Results — Re-use Persisted Results — Persist Results

System's Perspective

User's Perspective

URL, Document ID, ... → Individual Document → Document-level Visualization

```
WEB-aHR0cHM6Ly93d3cuY2JjL...      Load News      🍀 Feelin' Lucky
```

Article Info

Title: French author Annie Ernaux wins Nobel Prize in Literature   Author: Thomson Reuters   Published: Oct 06, 2022 7:08 AM ET

URL: https://www.cbc.ca/news/entertainment/nobel-literature-ernaux-1.6607938

Content: French author Annie Ernaux, who has fearlessly mined her own biography to explore life in France since the 1940s, was awarded this year's Nobel Prize in Literature on Thursday for work that illuminates the murky corners of memory, family and society. The author strongly defended

```
title: "Nobel Prize" and
@timestamp >= now - 6m
```

KQL Query → Subset of Documents

Text Mining Tools

Corpus-level Visualization

Entity Recognition | Semantic Parsing | Person Relations

French author Annie Ernaux wins Nobel Prize in Li... [Pro-form - PER]

French author Annie Ernaux, who has fearlessly mined her own biography to explore life in France since the 1940s, was awarded this year's Nobel Prize in Literature on Thursday for work that illuminates the murky corners of memory, family and society.

The author strongly defended women 's rights to abortion and contraception in some of her first comments after winning the prize.

Q1153825
Score 127.64672438964843
Names Annie Ernaux; Annie Duchesne
View in Neo4j Browser
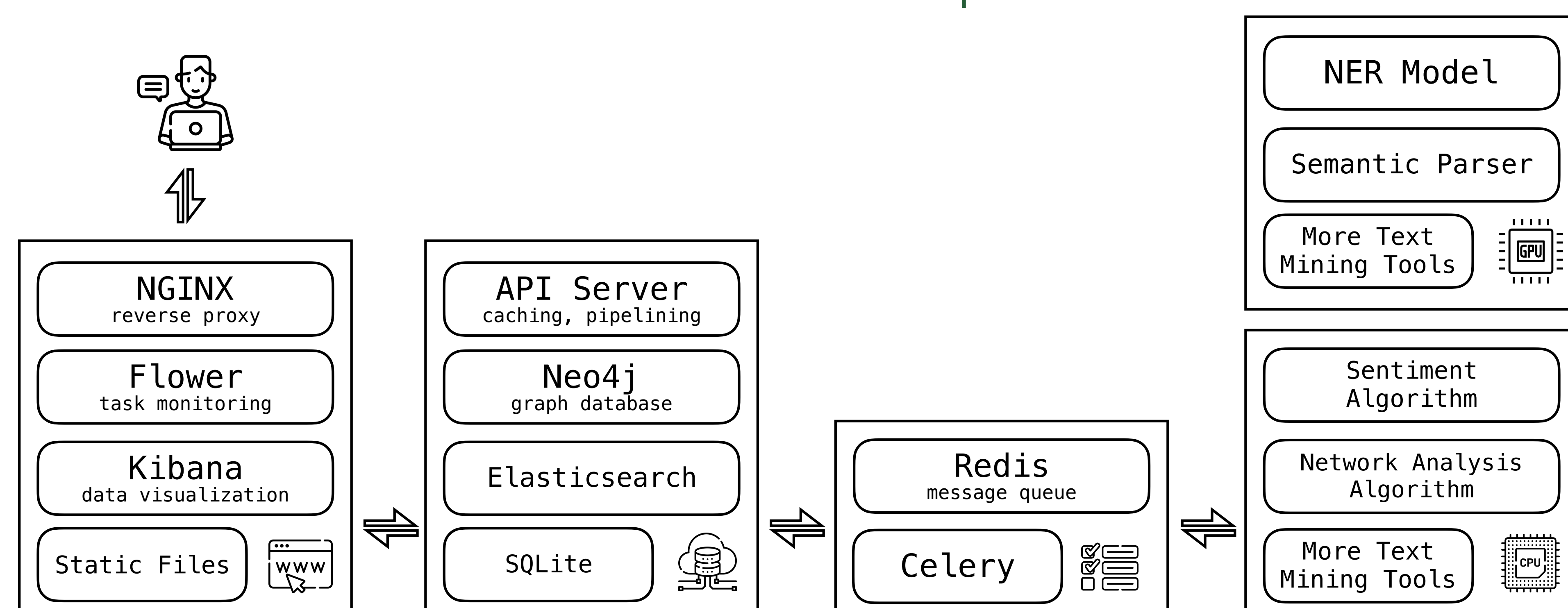View in Wikidata
Attributes

Design

Non-experts face challenges dealing with
- ✗ Heterogeneous software stacks required to run models
- ✗ Lack of user-friendly interfaces to apply models and visualize results
- ✗ Constraints on computation and networking resources
- ✗ Non-standardized, inconsistent input and output formats

Our design solves these by having
- ✔ Individual model and its dependencies runs in its standalone container
- ✔ Web-based interface for non-expert users, RESTful and Python API for developers
- ✔ Containerized microservices enable distributed deployment of models on machines with varying capabilities
- ✔ Corpus building from web search, news, social media and more platforms

NGINX (reverse proxy) / Flower (task monitoring) / Kibana (data visualization) / Static Files

API Server (caching, pipelining) / Neo4j (graph database) / Elasticsearch / SQLite

Redis (message queue) / Celery

NER Model / Semantic Parser / More Text Mining Tools

Sentiment Algorithm / Network Analysis Algorithm / More Text Mining Tools

Use Cases

Digital Humanities
Generate linked data from cultural heritage texts; analyze how extremists use social media to promote agendas

Business Analytics
"Are recent news reports about Apple Inc. positive or negative?"

NLP Research
Access models as a service without setting up development environments

Read Our Paper

bit.ly/nlpworkbench