

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



EACL 2023



UNIVERSITY
OF ALBERTA

NLP Workbench is

A web-based platform for text mining

NLP Workbench is

A web-based platform for text mining

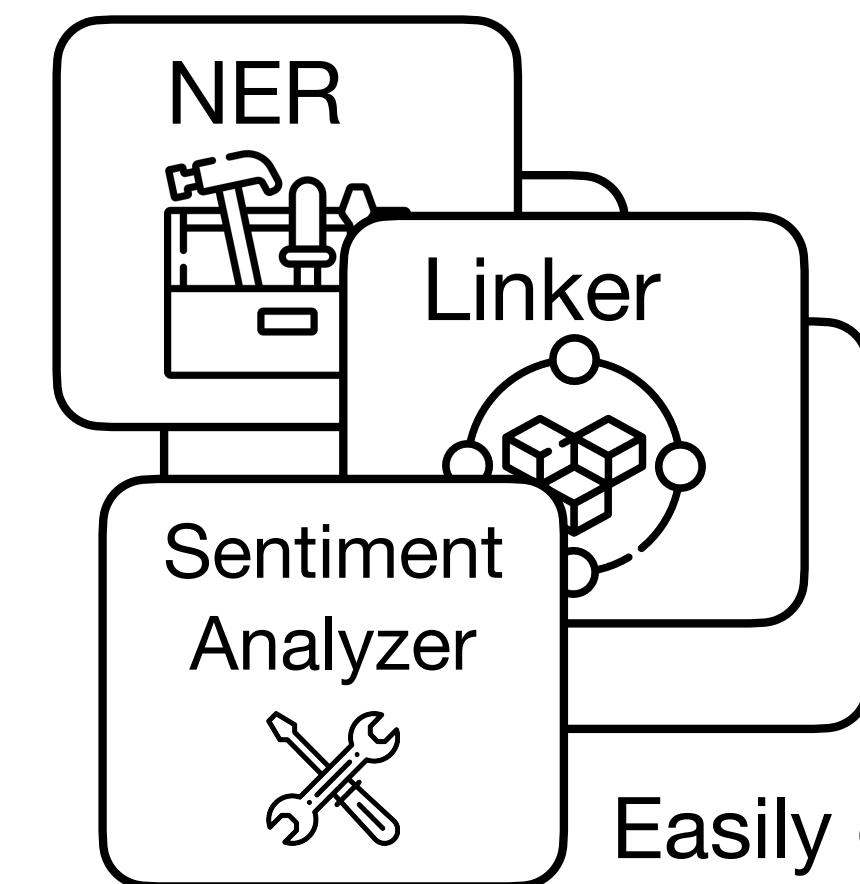
Targeting **non-expert** users

NLP Workbench is

A web-based platform for text mining

Targeting **non-expert** users

Providing **semantic understanding**
of large-scale corpora



Easily extensible to
add more!

NLP Workbench is

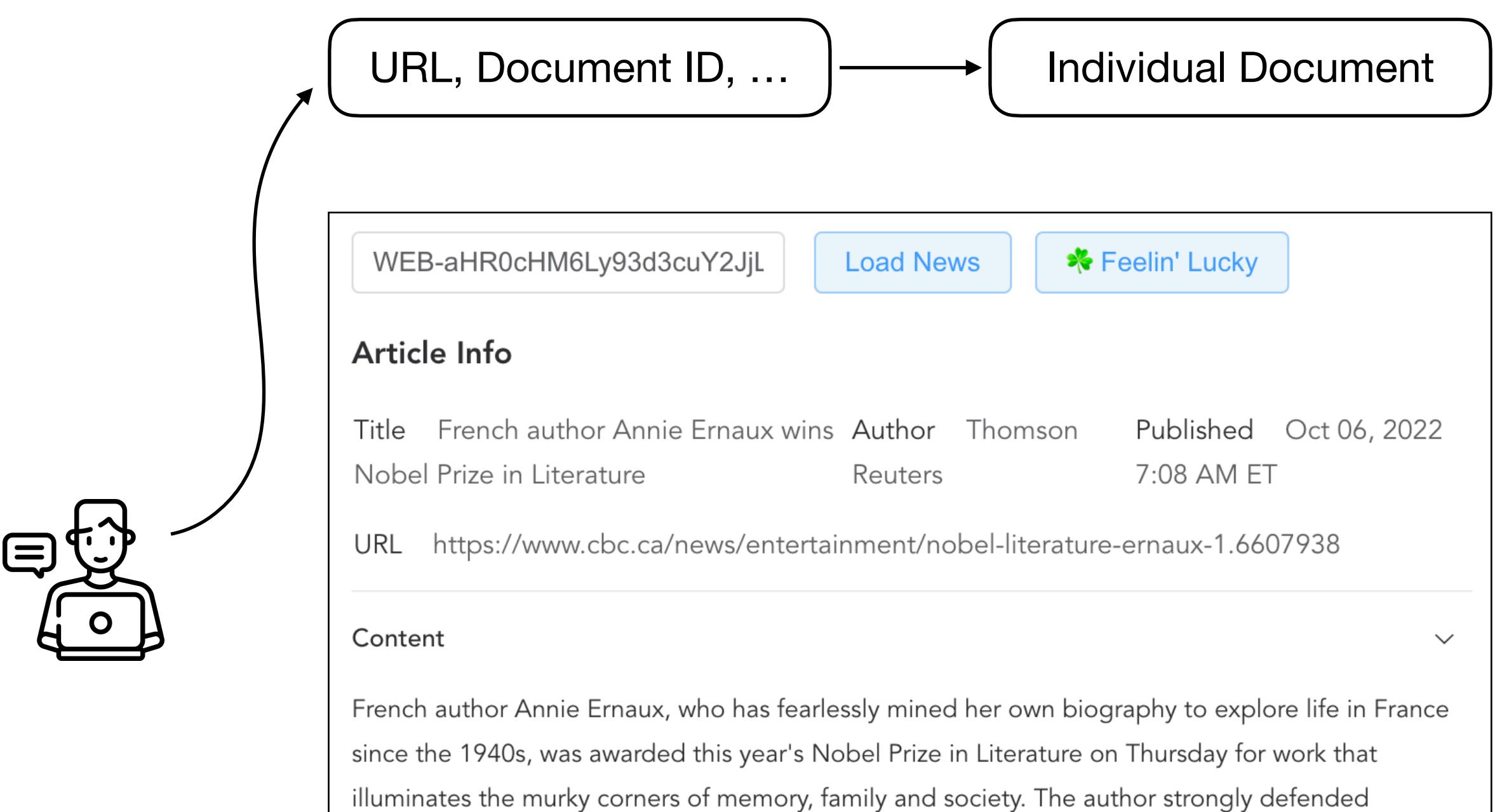
A web-based platform for text mining

Targeting **non-expert** users

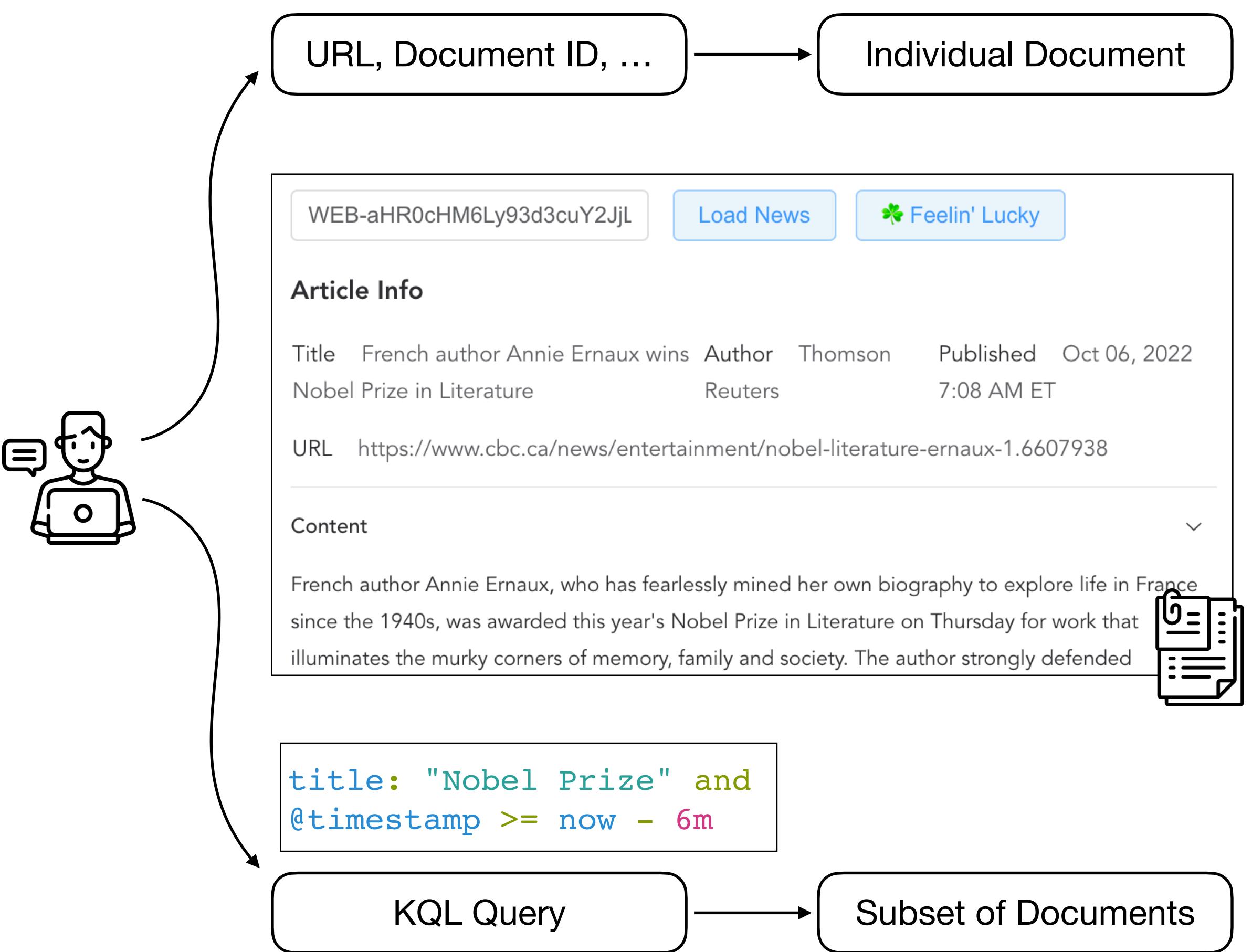
Providing **semantic understanding**
of large-scale corpora

Built upon **latest** pre-trained models
and **open source** systems

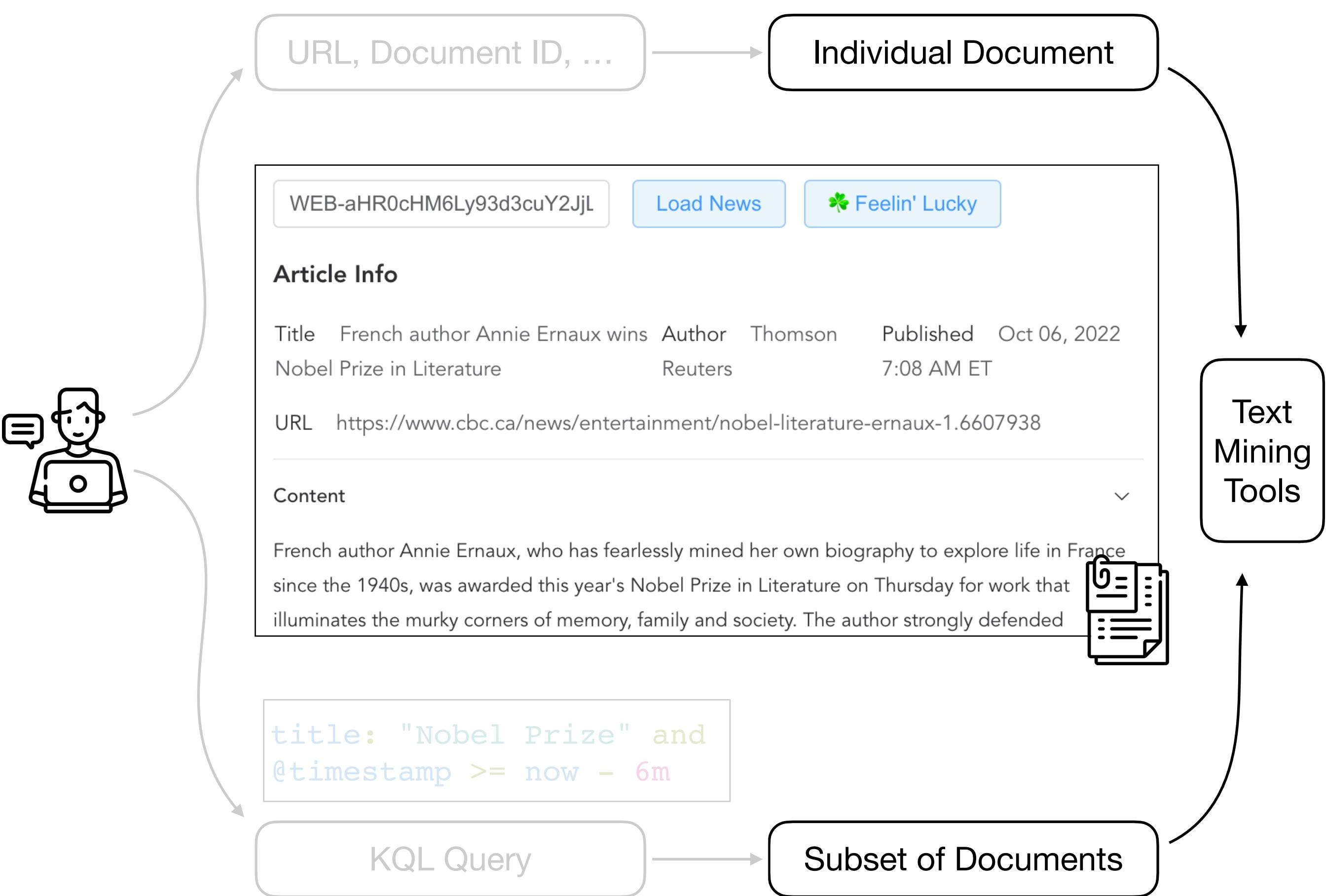
User's Perspective



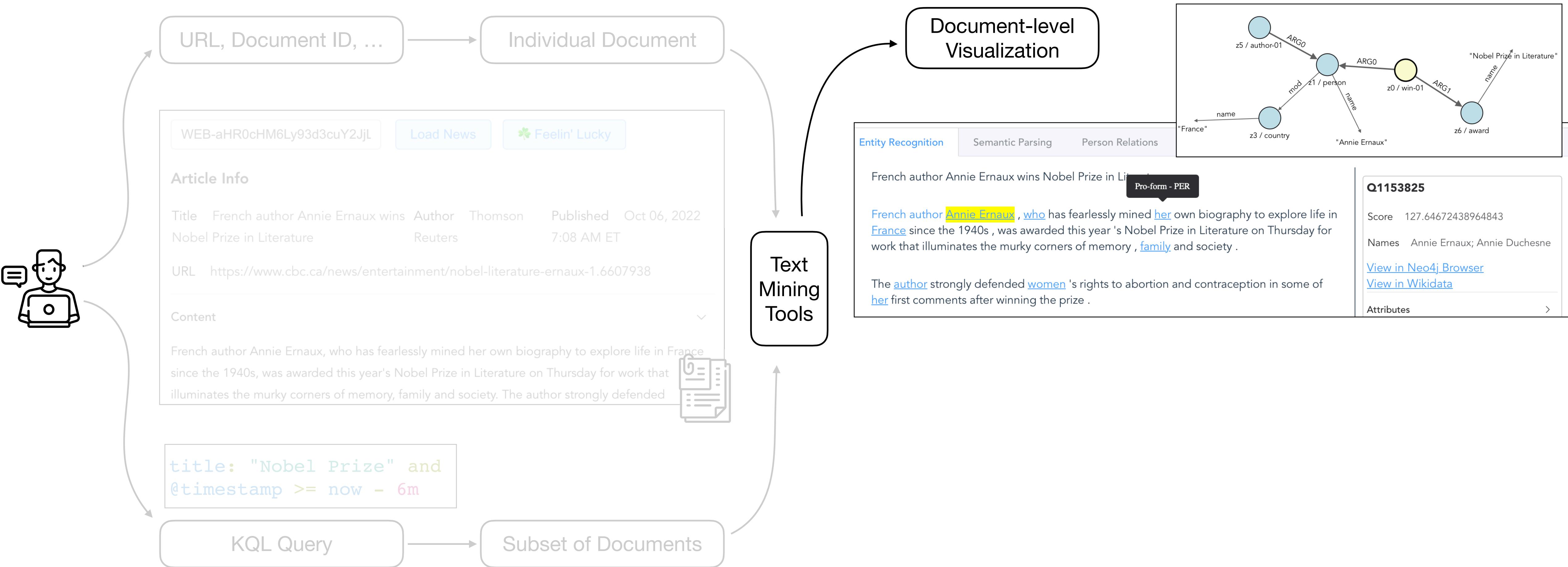
User's Perspective



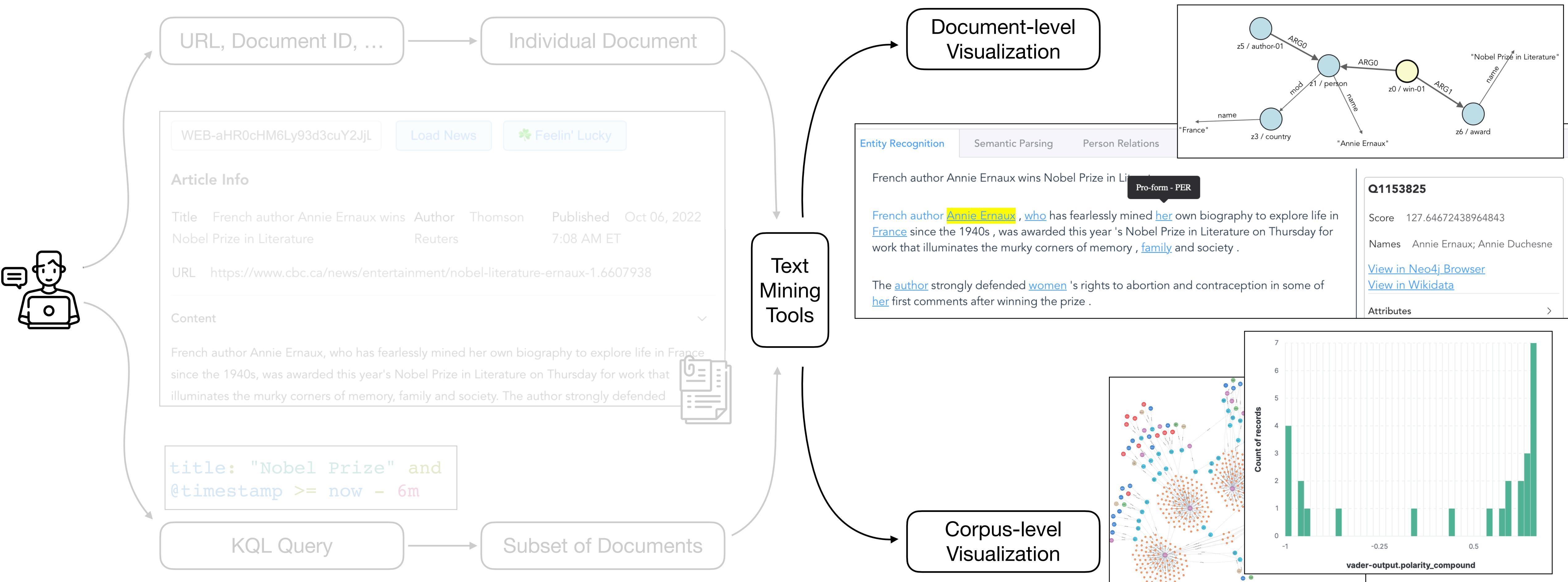
User's Perspective

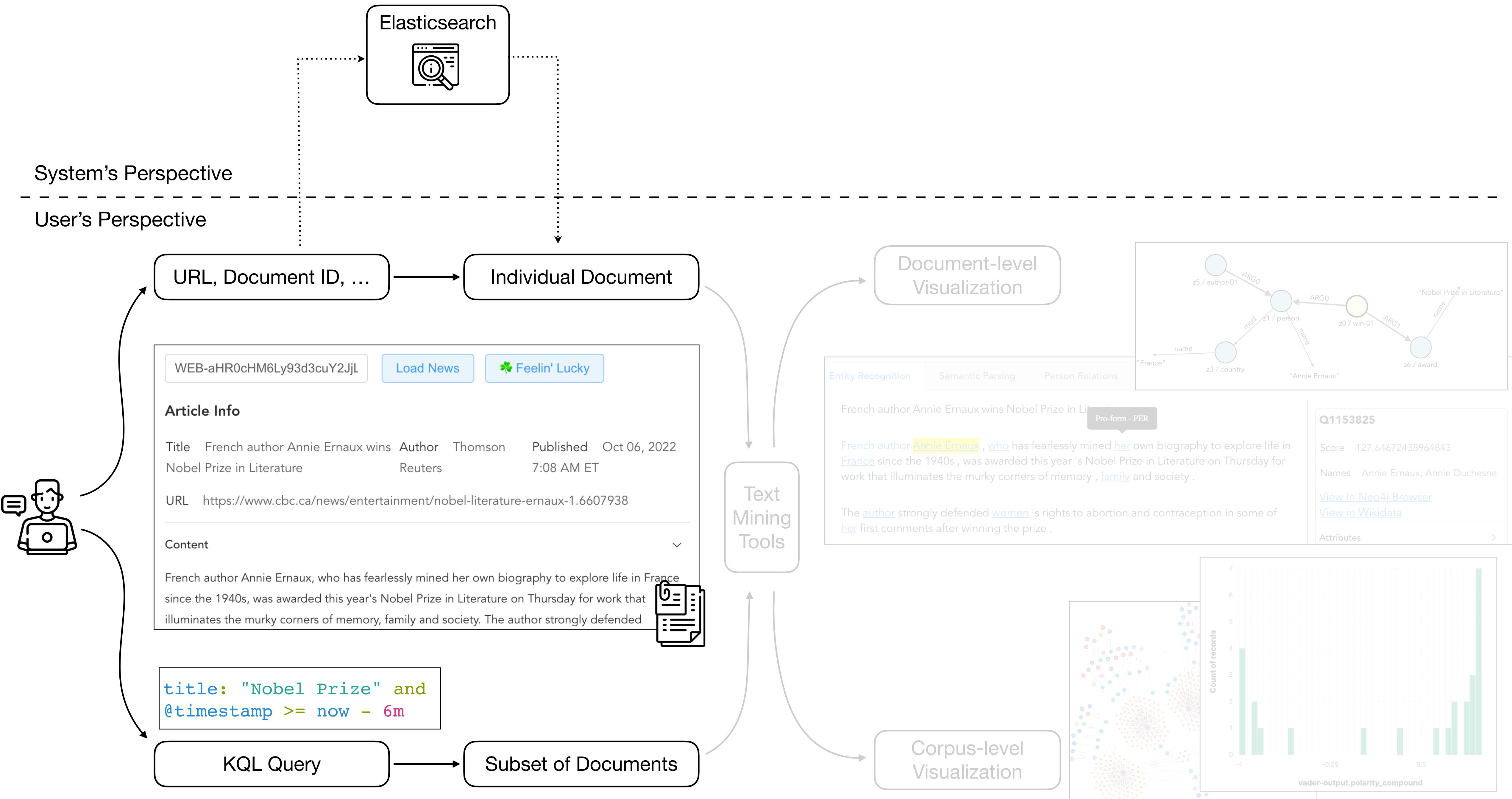


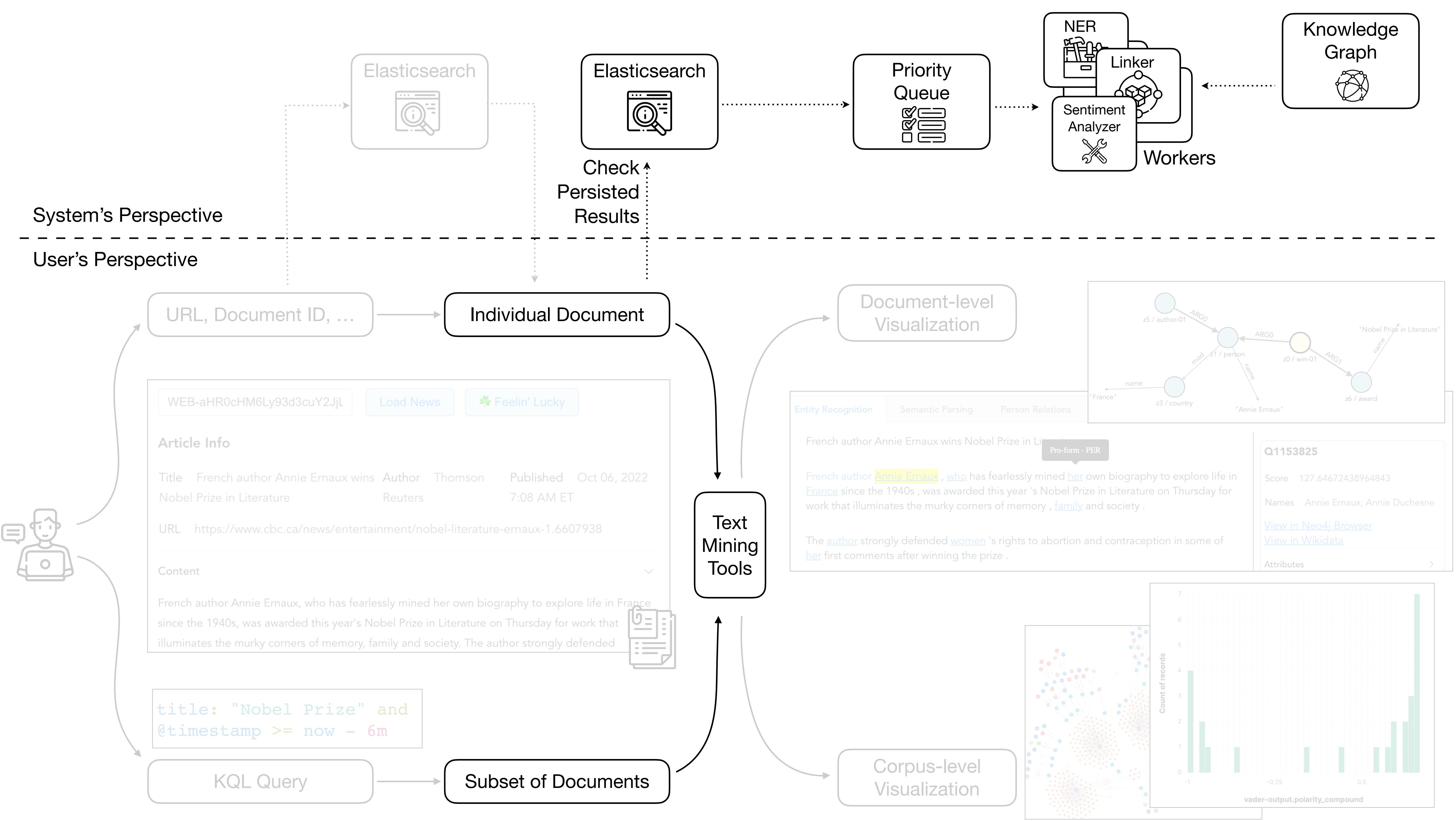
User's Perspective

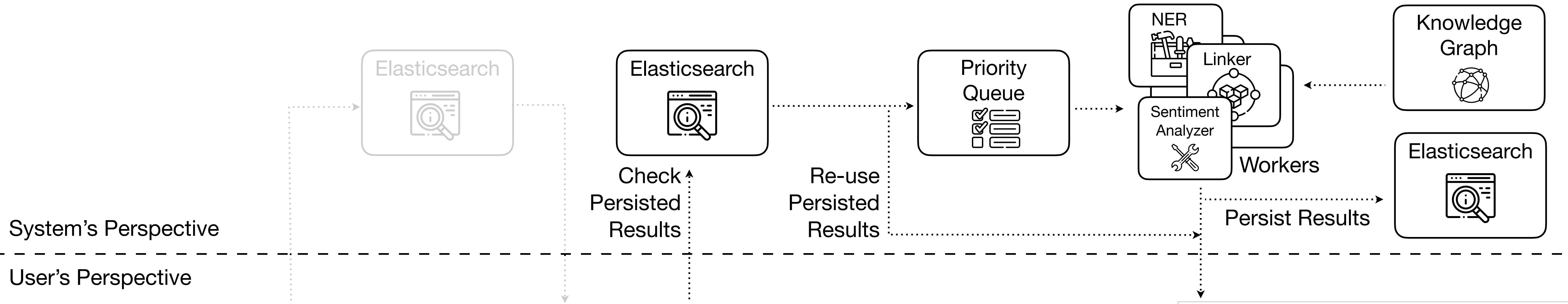


User's Perspective

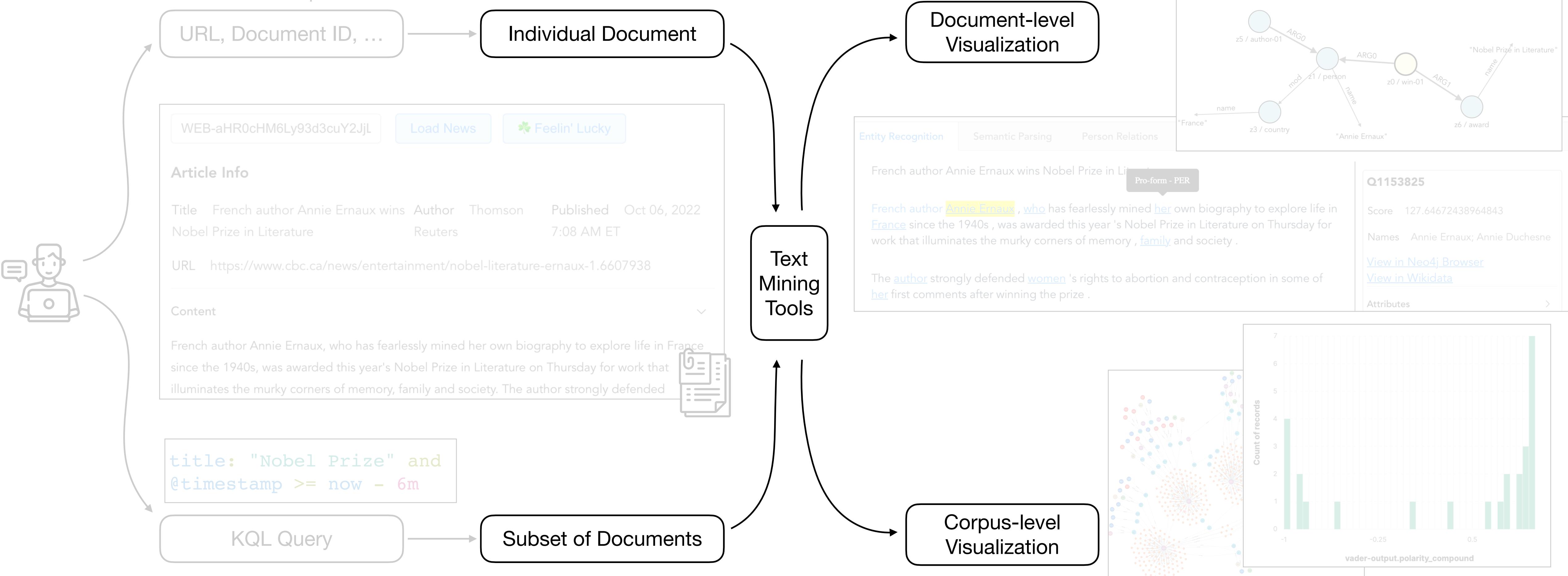








User's Perspective



Non-experts face challenges in

✗ Heterogeneous software stacks

required to run models

Non-experts face challenges in

✗ Heterogeneous software stacks

required to run models

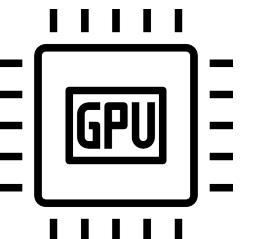
We solve this by having

✓ Individual model and its dependencies runs in its standalone **container**

NER Model

Semantic Parser

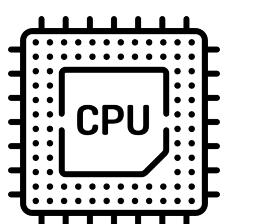
More Text
Mining Tools



Sentiment
Algorithm

Network Analysis
Algorithm

More Text
Mining Tools



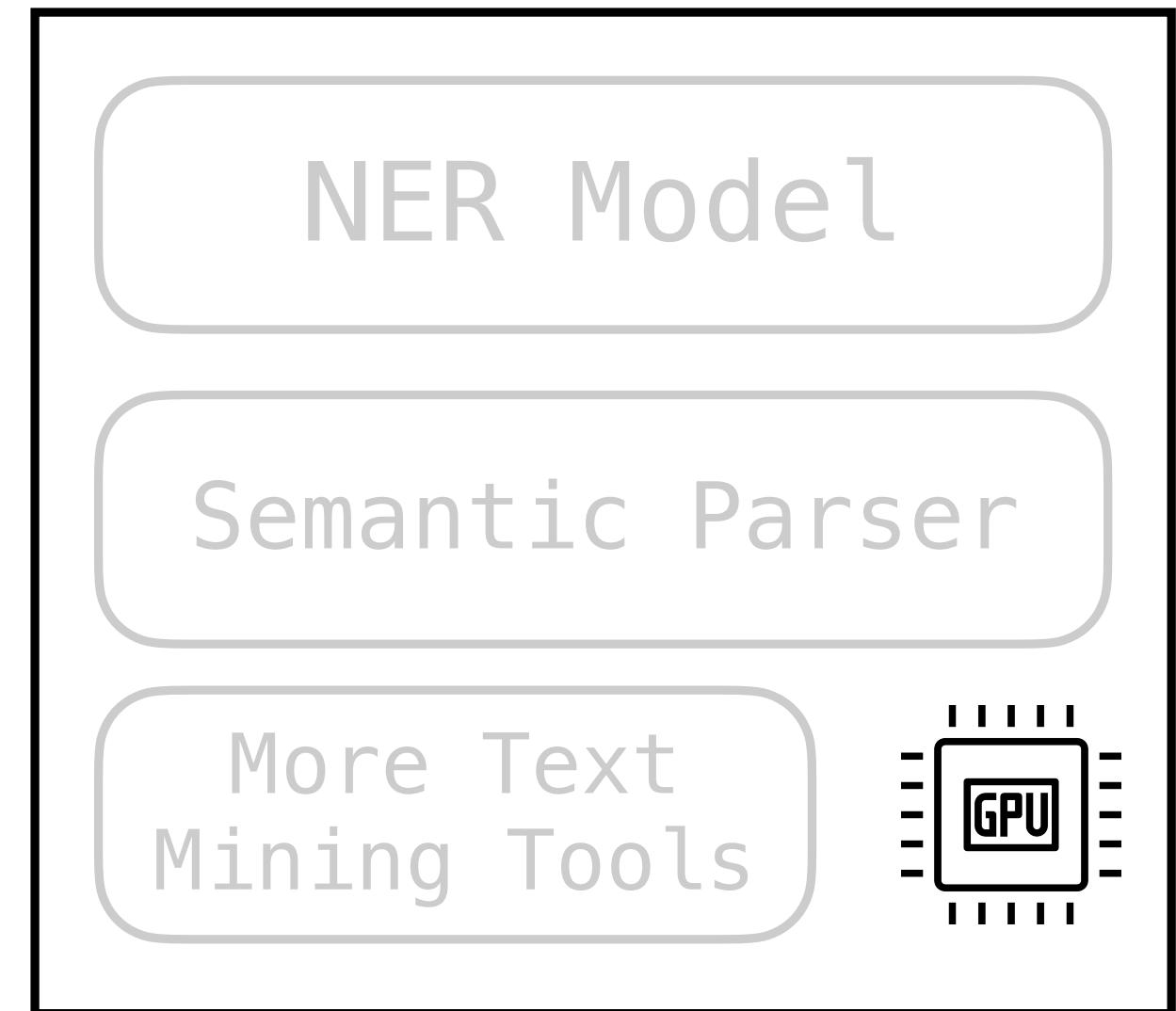
Non-experts face challenges in

✗ Constraints on computation and

networking resources

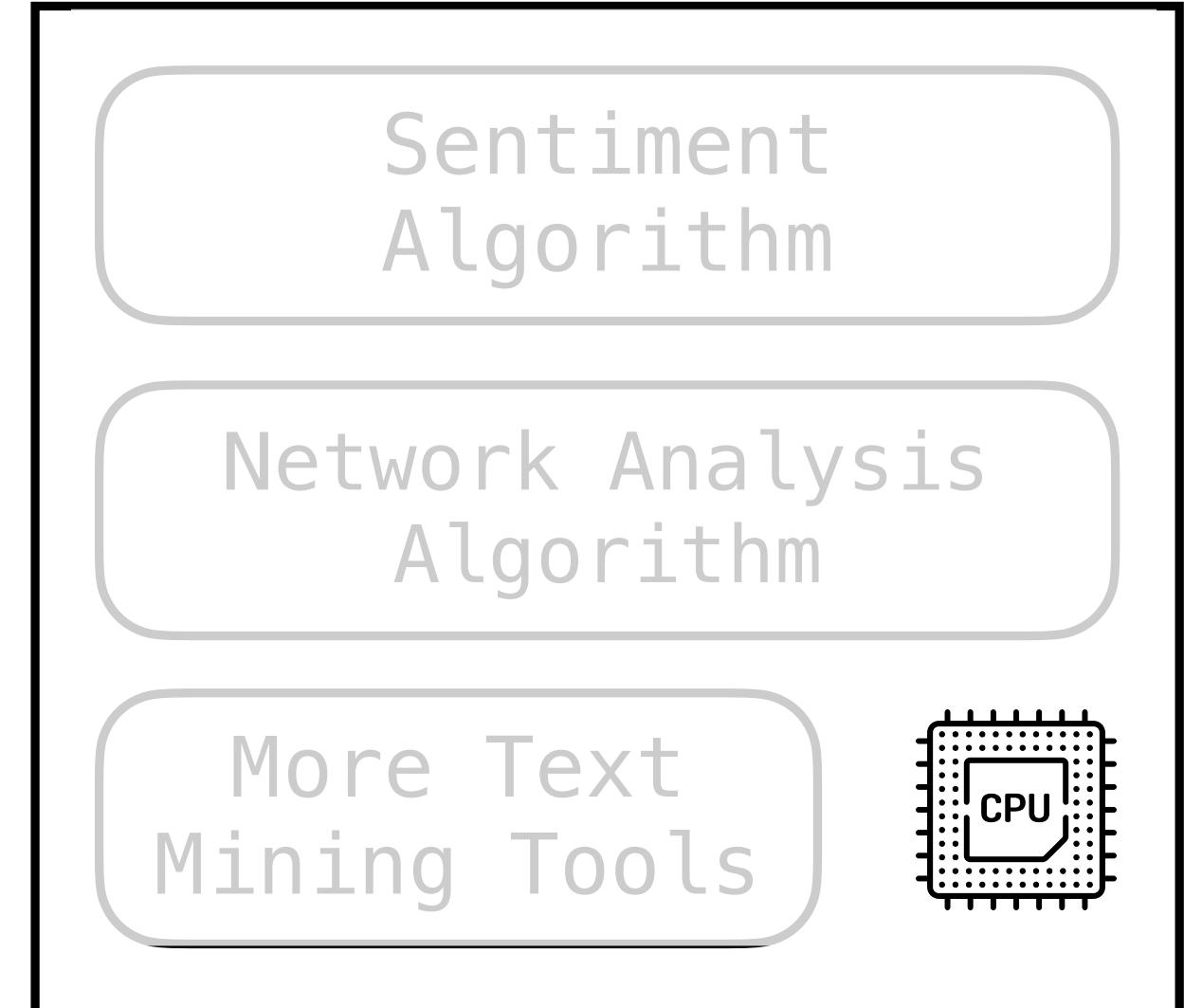
Non-experts face challenges in

✗ Constraints on computation and
networking resources



We solve this by having

✓ Containerized microservices
enable **distributed deployment** on
machines with varying capabilities



Non-experts face challenges in

✗ Non-standardized, inconsistent

input and output formats

Non-experts face challenges in

✗ Non-standardized, inconsistent

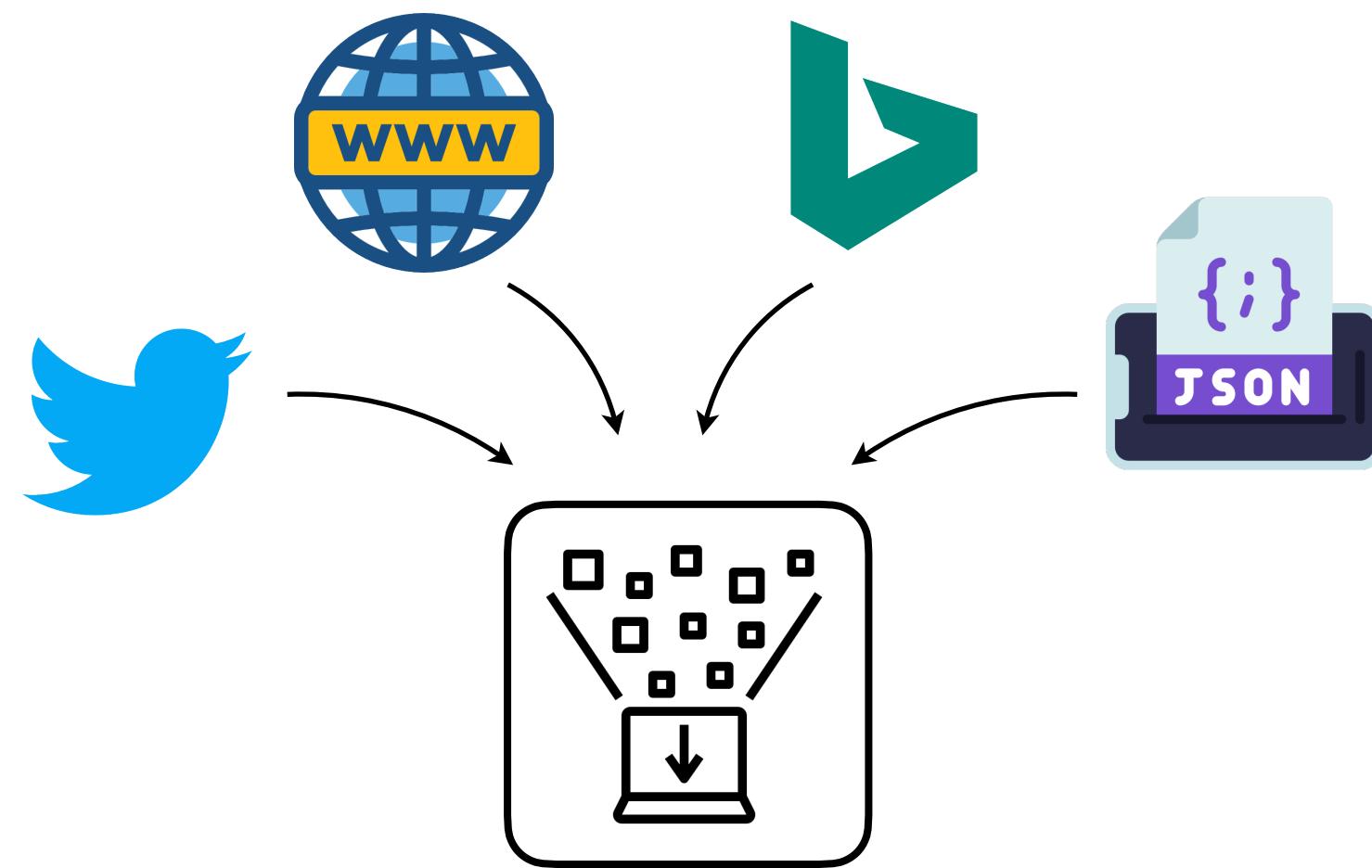
input and output formats

We solve this by having

✓ Corpus building from web search,

news, social media and more

platforms



Use Cases

Digital humanists generate linked data from heritage texts, and analyze how extremists use social media to promote agendas

Business analysts want to learn “*are recent news reports about Apple Inc. positive or negative?*”

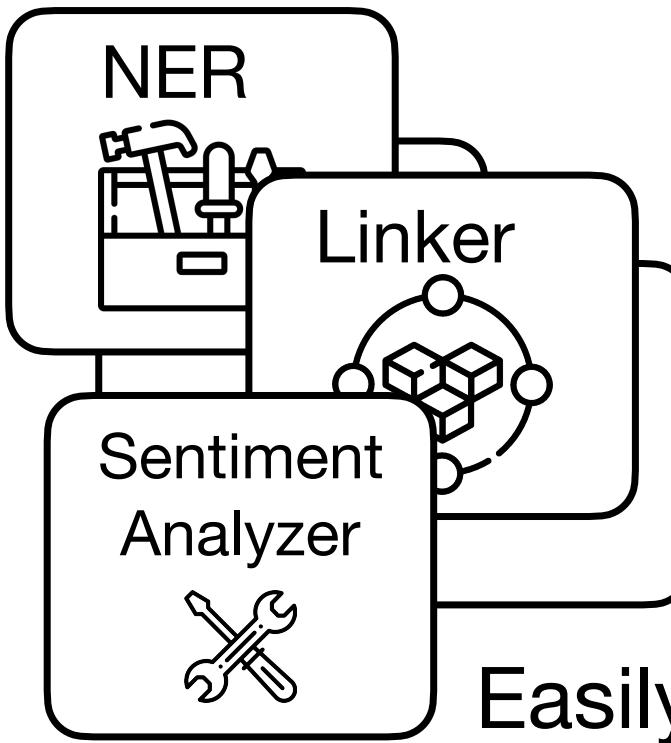
NLP researchers can access models without configuring development environments

Demo



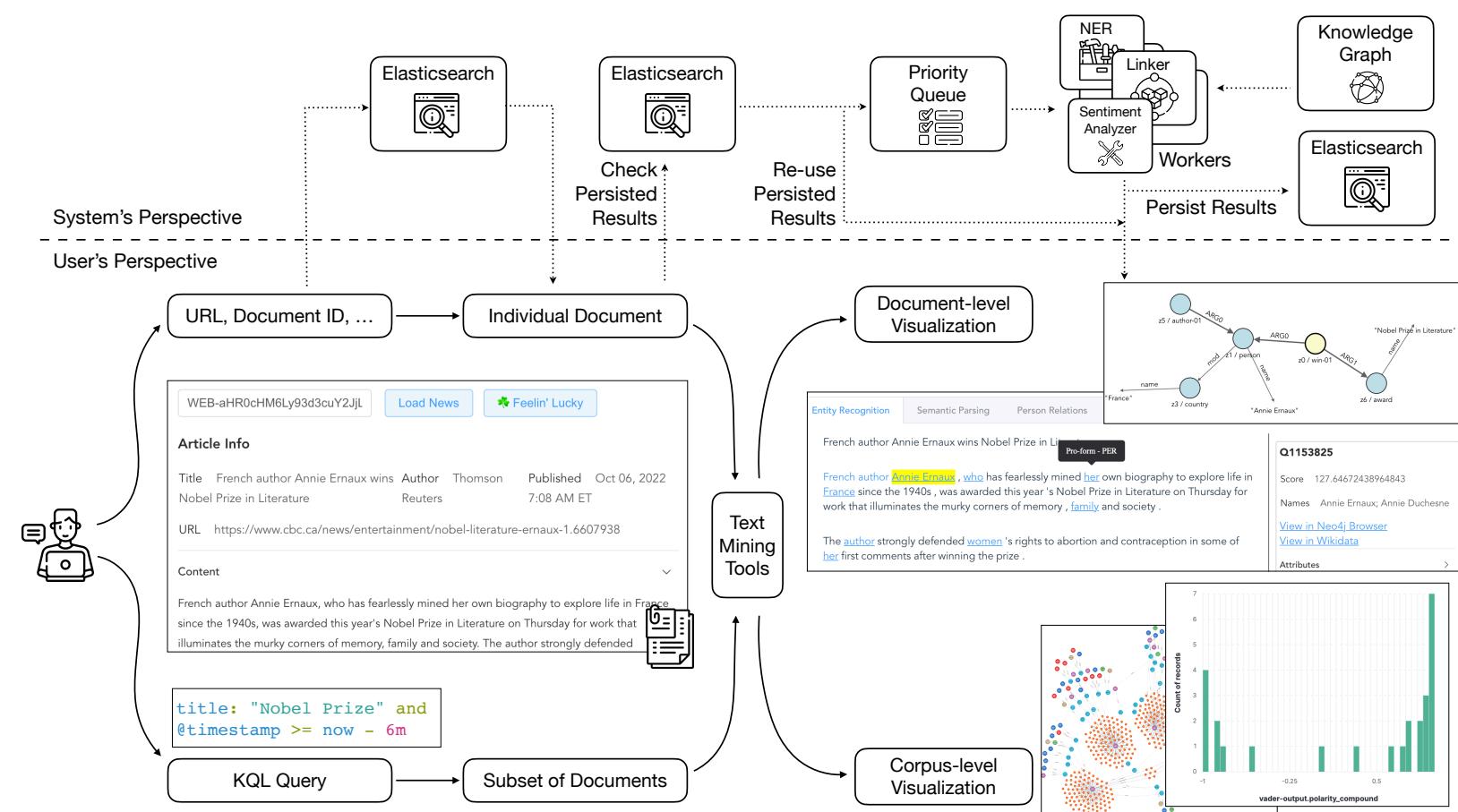
<https://vimeo.com/801006908>

Recap

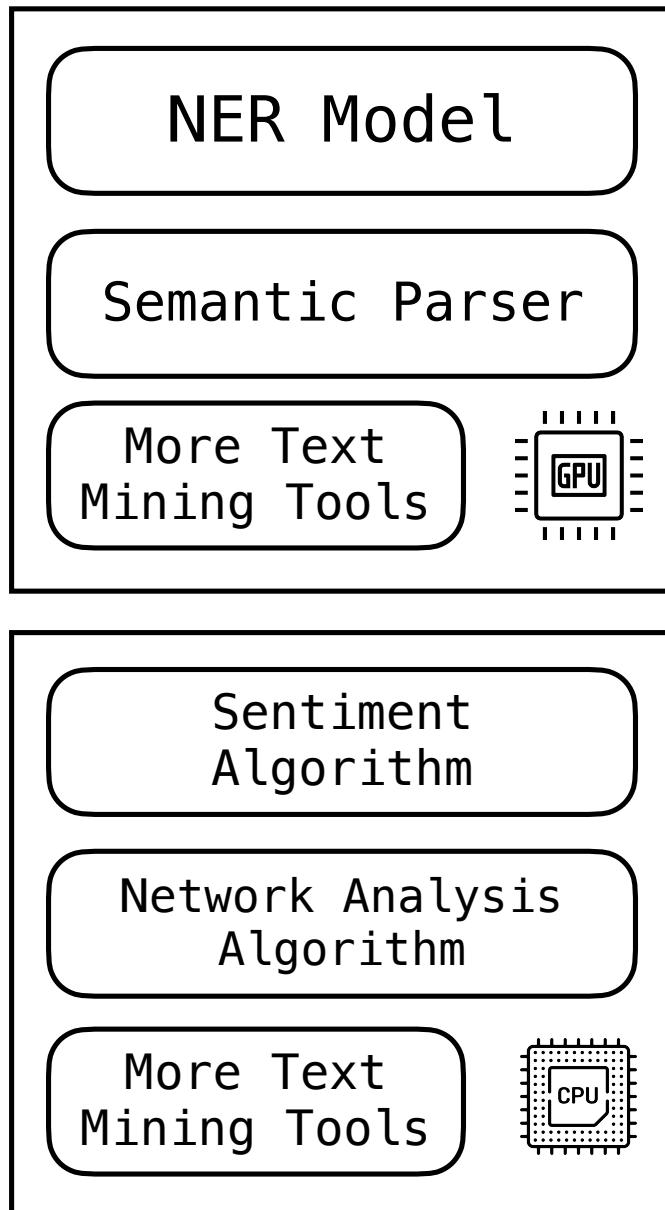


Easily extensible
to add more!

Integration of state-of-the-art models



Web-based platform



Containerized
microservices

[Code](#) | [Issues](#) | [Preprint](#)



U-Alberta / NLPWorkbench