

Vuk Vuković,[‡] Akhil Arora,[‡] Huan-Cheng Chang,[‡] Andreas Spitz,[§] and Robert West[‡]

[‡]Data Science Laboratory (DLAB), EPFL

[§]University of Konstanz

“Quotations will tell the full measure of meaning, if you have enough of them.”

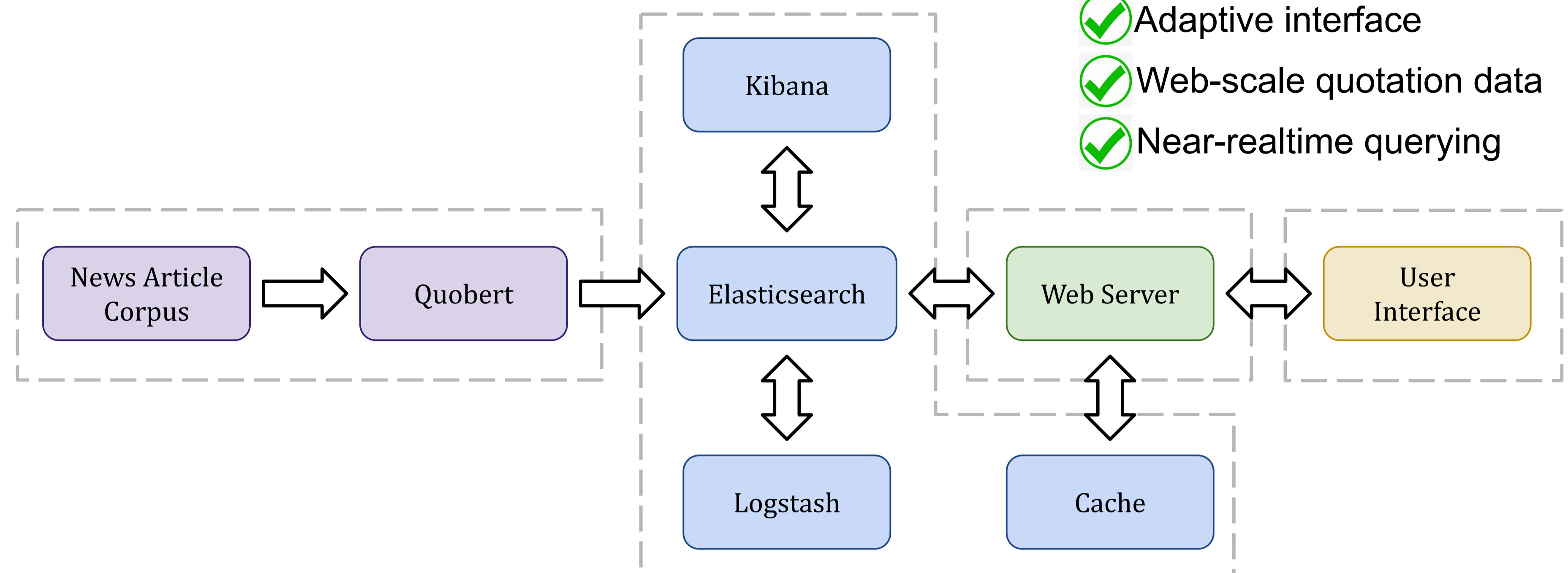
— James Murray

Central to **conception**, **reception**, and **analysis** of news stories

- Pathway of **information propagation**
- Identifying **source of news** information
- Fact checking and **credibility** assessment
- Opinion mining

The Quotebank System

- ✓ Faceted search
- ✓ Adaptive interface
- ✓ Web-scale quotation data
- ✓ Near-realtime querying



The Quotebank User Interface

Quotebank

Keywords of quotations...

I'm feeling lucky

Select data set to search:

Quotation-centric

Article-centric

Enable exact match



Search with context



From

Select a date

To

Select a date

Select data set to search:

Quotation-centric

Article-centric

Enable exact match



Name of speaker

Search for a name

Nationality of speaker

Search for a country

Gender of speaker

Search for a type of gender

Occupation of speaker

Search for an occupation

URL domain

Input a URL domain

Minimum number of occurrences

Input a number

From

Select a date

To

Select a date

A sample search engine result page (SERP)

Found 134 quotations in 0.028 seconds. Share Save results

Number of quotations published between 2019-12-01 and 2020-04-17:
Note: Some time periods were affected by data outages and have missing data. Such periods are marked with a dotted line on the horizontal timeline.

2019-11-25 2020-04-13

ⓘ The speakers are attributed by our machine learning model, which uses the textual context. As with all machine learning models, some attributions may be erroneous. For details, please refer to [our paper](#).

Sort by Content matching score (descending)

combat the spread of coronavirus misinformation
Fri, 27 Mar 2020 16:35:00 GMT

Possible speakers	Unknown speaker	80.84%
Meghan Markle	Q3304418	19.16%

Takeaways

- 👍 **First-ever** system for exploration of **quotes in news**
- 👍 **Wider accessibility**: journalists, social scientists etc
- 👍 **Faceted search** based on speaker attributes
- 👎 Extraction of attributed quotes far-from-perfect
 - Requires speaker **disambiguation**
 - Marred with **dirty/unclean** Web data

Broader Impact

Enable exploration of quotes by **end users** who **lack** the computational **background** to work directly with the **raw Quotebank data**

Quote Erat Demonstrandum



Quote-level exploration

Identify all quotes from **Donald Trump** containing the text **“great again”** that appeared in at least 500 distinct articles



Article-level exploration

Identify all articles published on May 19, 2018 that contain the text **“gdp”** in either the **quotes** or in their surrounding **context**



Quotebank

<https://quotebank.dlab.tools>



Contact: akhil.arora@epfl.ch

