

Getting the Query Right: User Interface Design of Analysis Platforms for Crisis Research

Mario Barrenechea, Kenneth M. Anderson, Ahmet Arif Aydin
Mazin Hakeem, Sahar Jambi

@mbarrenecheajr



University of Colorado **Boulder**



Outline

1. Background and Motivation
2. EPIC Analyze
3. User Interface Evaluation
4. Query Data Modeling
5. Demo
6. Conclusions and Future Work

Big Data on the Web

Data-intensive software systems present new challenges with respect to user interface design.

Interface design for these systems must adopt human-centered techniques and be *flexible* and *extensible*



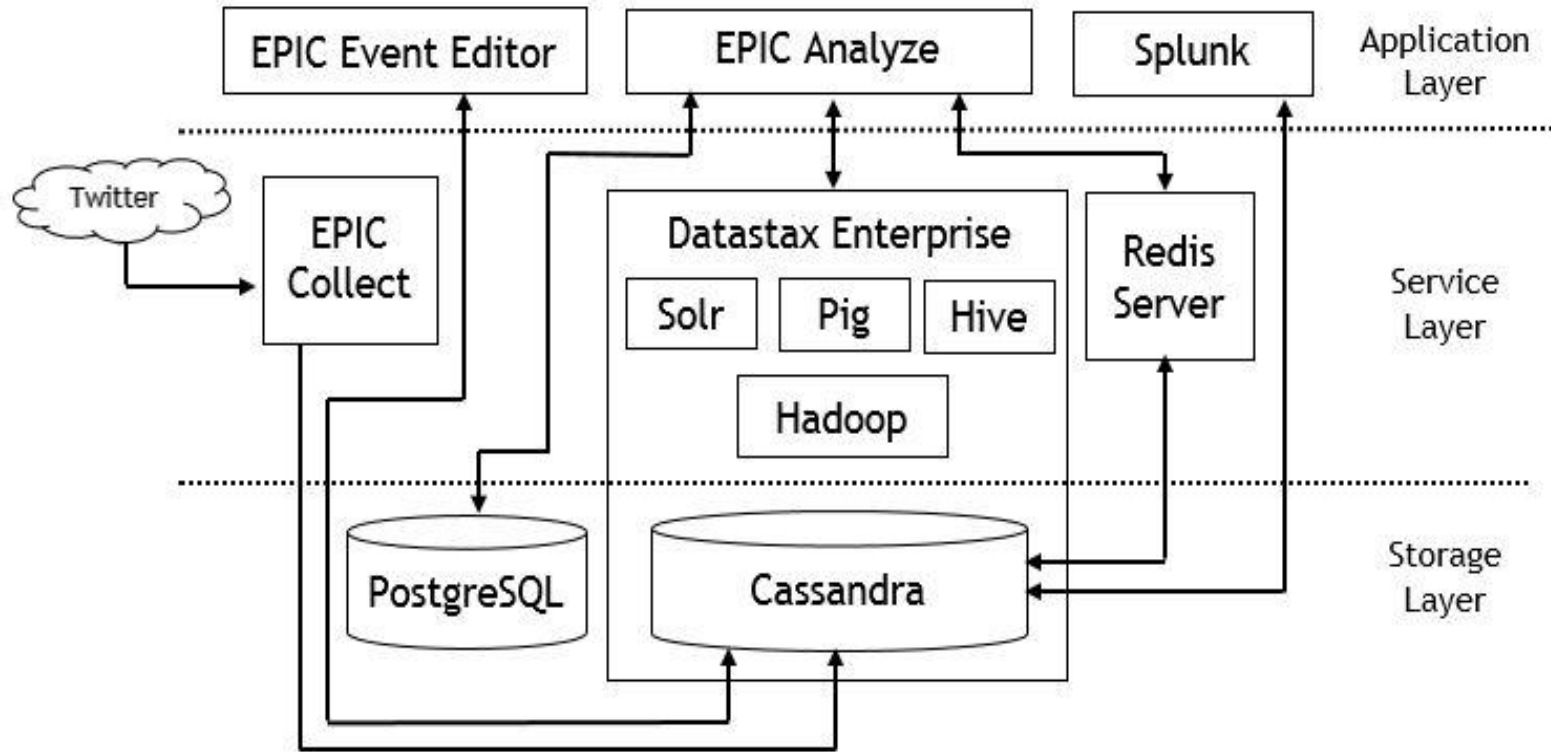
Crisis Informatics

Examines the socio-technical relationships among people, information, and technology during mass emergency events. (Palen, et al., 2010)

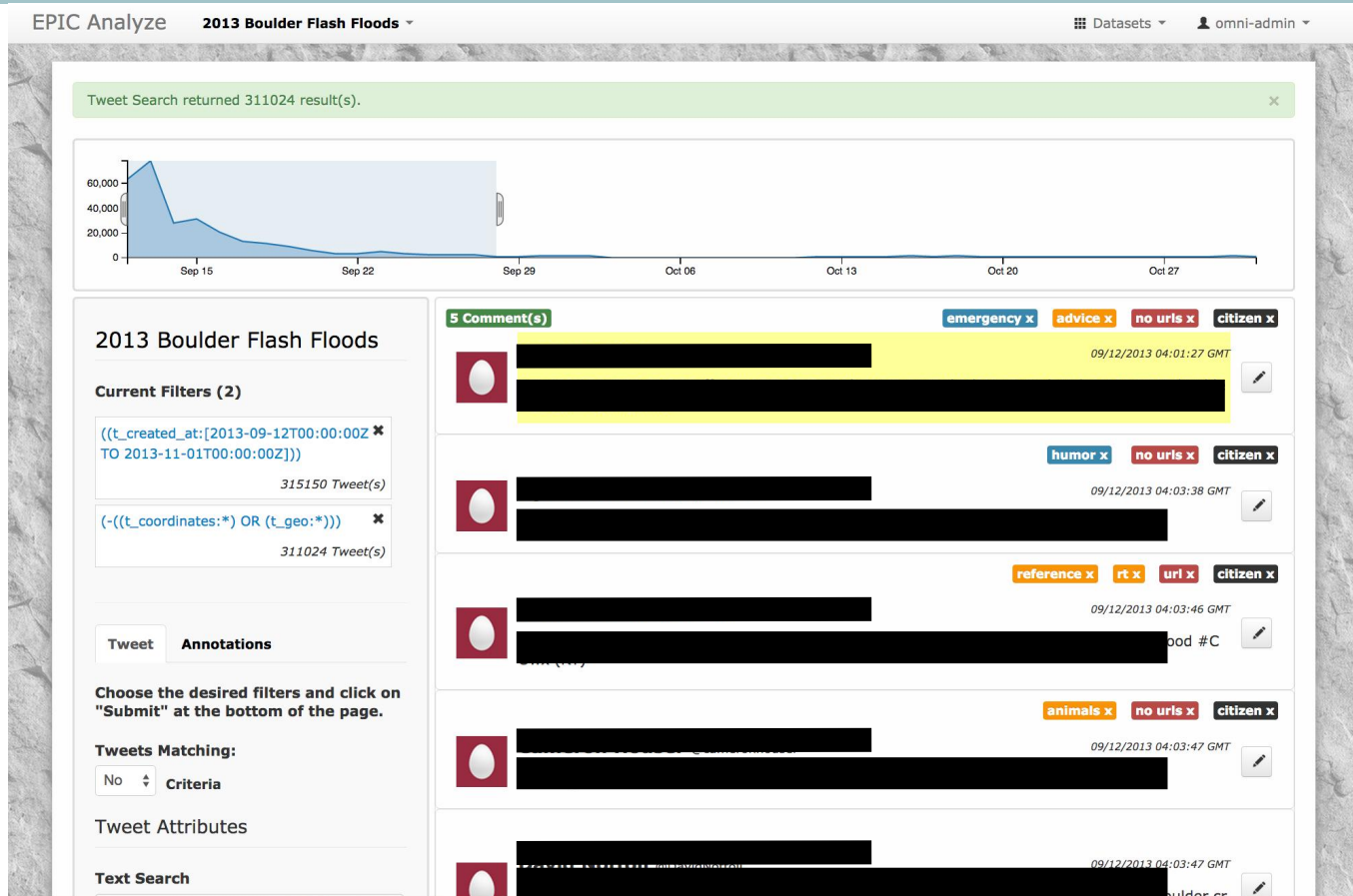
Tools are needed to support data-driven methods for analyzing large-scale social media data in this domain.



EPIC Analyze - Architecture



EPIC Analyze - Interface



Evaluation - The Research Analysts

Research Analysts are information scientists studying crisis informatics.

Programming experience, events of interest, research methods and questions, and aspects of data, all vary from one analyst to another.

EPIC Analyze features were derived from web user interface evaluation.

Evaluation - Think Aloud Protocol

1. Open an event that interests you.
2. Investigate a tweet and look at all of its attributes.
3. How can you view the tweets from only the first or last day of the dataset?
4. How can you get tweets written in Spanish? How about Spanish and French?
5. Can you find tweets that either have the word “earthquake” in the text or are from a specific user? What about both?
6. How can you get back tweets with annotations?
7. How can you comment on a tweet that has already been commented on?

EPIC Analyze 2015 Vail Beaver Creek World Championships

2015 Vail Beaver Creek World Championships

Tweet Annotations

Choose the desired filters and click on "Submit" at the bottom of the page.

Tweets Matching: All Criteria

Tweet Attributes

Text Search: Enter Search Terms Here

Keyword: Choose Keywords Here

Date Tweeted: mm/dd/yyyy to mm/dd/yyyy

Sort By: Date Tweeted Ascending

Geo Enabled? Has URL?

Retweet Count: 1,2,* to 1,2,*

Favorite Count: 1,2,* to 1,2,*

URLs: Enter URL Here

Link to Tweet

ID	Created At	Name	Screen Name	Location	URL	Description	Followers Count	Friends Count	Created At	Favorites Count	Time Zone	Geo Enabled?	Statuses Count
561675858616672257	Sun Feb 01 00:02:14					Water Damage in Vail, Colorado. 24 Hour Emergency Service. 36 Years of Excellence. Serving the Central Colorado Rockies. 800.527.1253	886	891	Fri Mar 05 00:45:07 +0000 2010	None	Mountain Time (US & Canada)	true	2240

Link to User Profile

Source: Facebook

Favorite Count: 0

Lang: en

Retweet Count: 0

Retweeted?: None

Geo/Coordinates: None

Evaluation - Feedback

“Big picture” of the dataset allows analysts to hone in on the data where (and when) it is most important.

Filter data out, rather than qualifying data in.

Integration into crisis work practices such as VOST (Virtual Operations Support Team).

EPIC Analyze Data Modeling

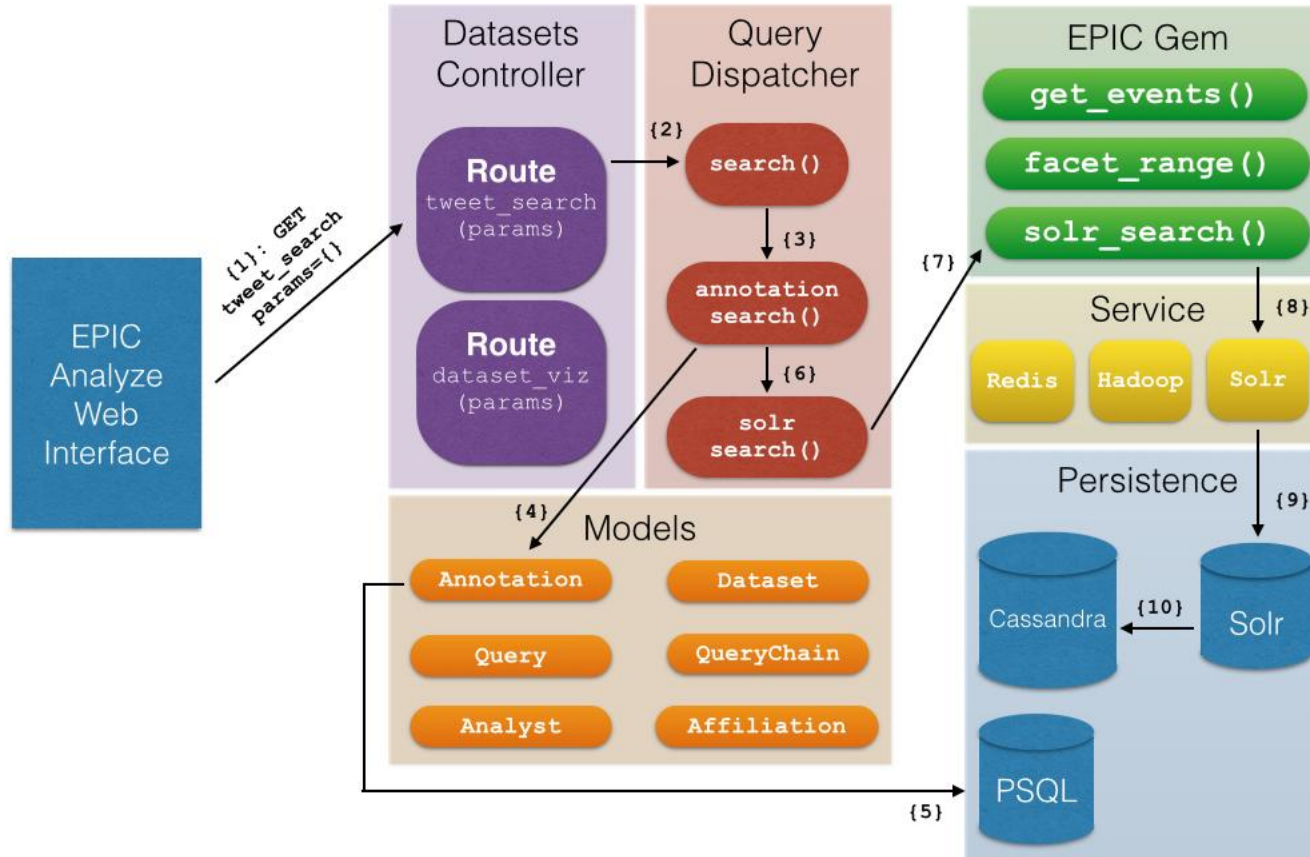
The **Query** object encapsulates information about the parameters used to filter data.

The **QueryChain** object associates and consolidates Query objects.

The **QueryDispatcher** handles the execution flow of a QueryChain.

Tweet Reference = <Tweet ID, Cassandra Row Key>

A Day in the Life of a Query



Demo

Query Visibility, Tangibility, and Expression

Visibility - Queries need to show their impact to the data.

Tangibility - Give users control of their querying power against the data.

Expressivity - Make querying powerful with logical operators and compound querying.

Future Work

Providing more support to our analysts:

- social network analysis
- maps
- automatic tweet labeling

Move to a **microservices-based architecture**

- move functionality currently in EPIC Analyze out into individual services
- allow for other tools to re-use existing functionality
- make it easier to add new services down the line

Conclusions

- Querying Interfaces for Big Data Web Applications
- Queries should be
 - visible
 - tangible
 - expressive
- Use of Human-Centered Design Techniques, an iterative development life cycle, and a commitment to the application domain are all critical to getting the query right!

Acknowledgments

Thank you to:

Ken Anderson

Leysia Palen

EPIC Analyze Team

Project EPIC lab

NSF



University of Colorado **Boulder**

Big Data System Features

Data transformation (Kandel, et al., 2011; Google Refine)

Visualizing and aggregating event data (Wongsuphasawat, et al., 2011; Splunk)

Semantic/Spatial Analysis (Oussalah, et al., 2013)

Analysis via familiar interfaces (Laconich, et al., 2013; NodeXL)

User-generated content on top of big data

Data-driven Collaboration

....