

Using NLP in Clinical Data Analytics

Tomasz Oliwa, PhD | tolwa@bsd.uchicago.edu

Brian Furner | bfurner@bsd.uchicago.edu

Center for Research Informatics

Biological Sciences Division

University of Chicago

Agenda

1. Goals
2. Lifecycle of a clinical note
3. Natural language processing (NLP)
 - a. Named-entity recognition
 - b. Information extraction
 - c. Search
 - d. Classification
4. Contact
5. References

Goals

Explain what **types of problems** can be solved with NLP

Show **examples** of applied NLP for **clinical research** / **notes**

Provide an understanding of what **applications we can build** for you to **aid you in your research**

Lifecycle of a clinical note

Order Comment: ...

Exams: ...

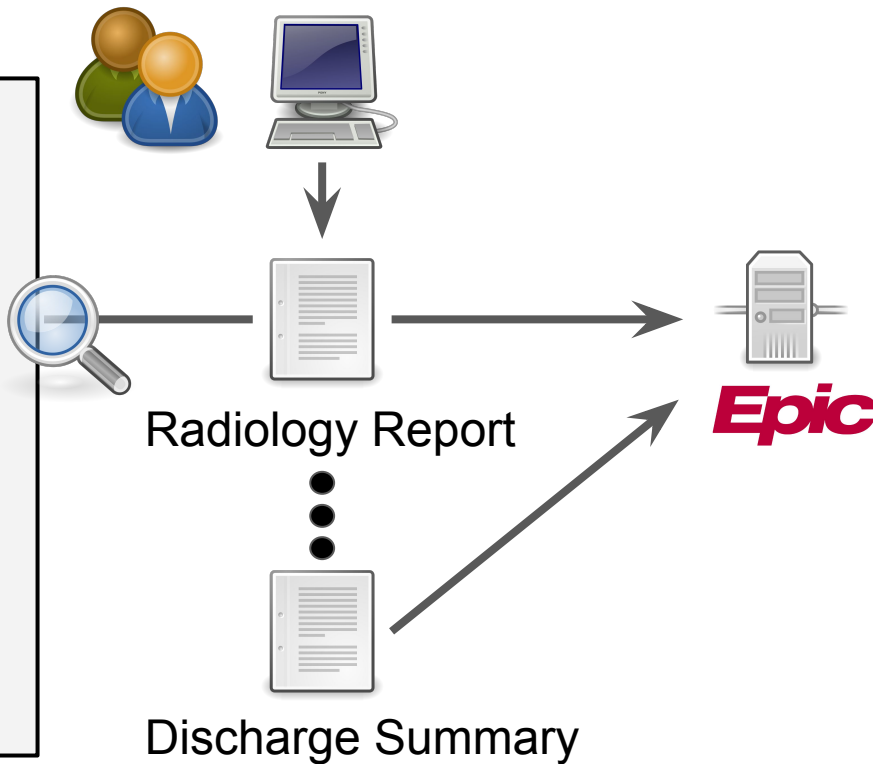
CLINICAL DATA: ...

COMPARISON: ...

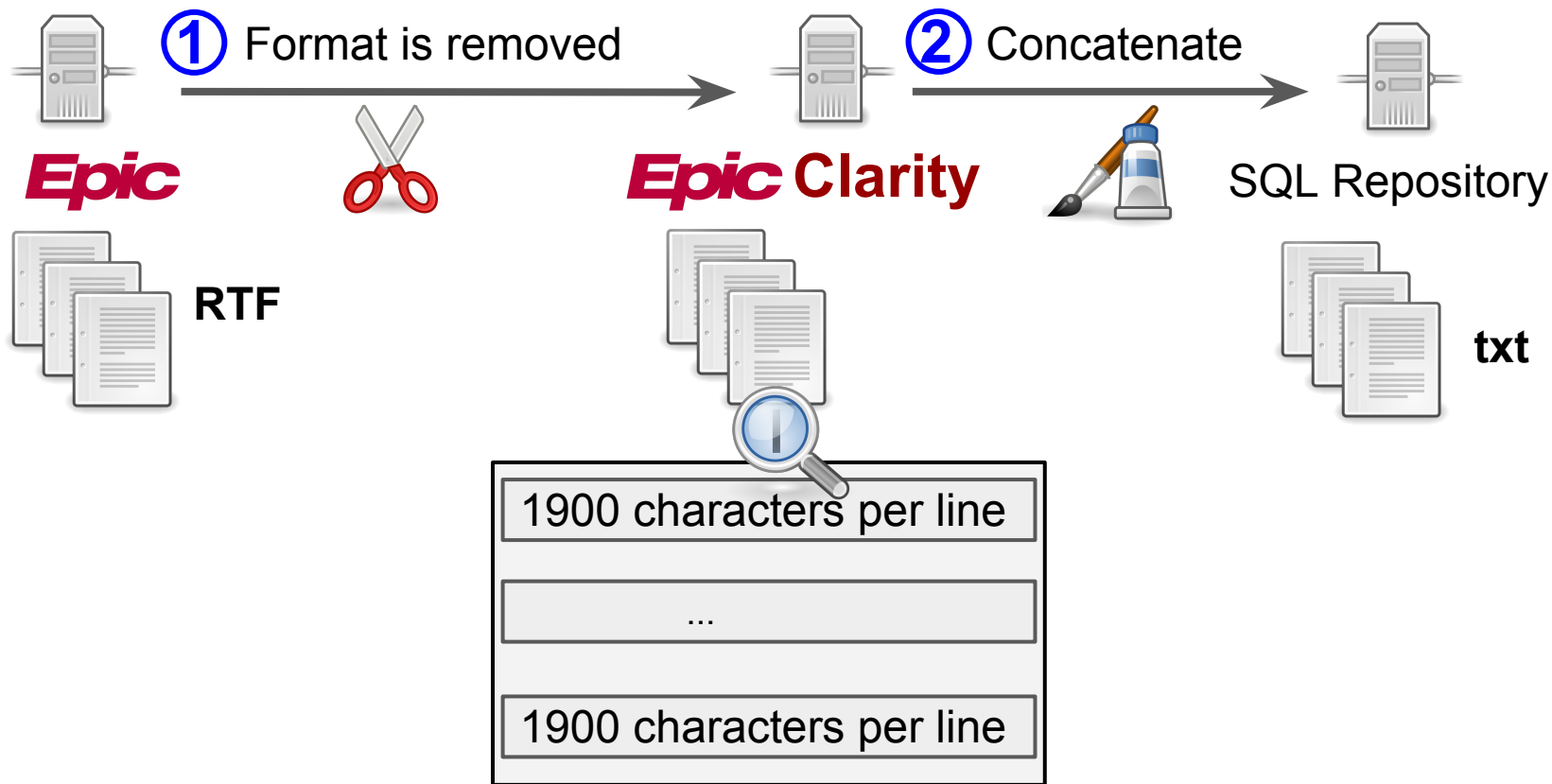
FINDINGS: ...

RESULT ID / ADDENDUM: ...

Ordering Physician: ...



Lifecycle of a clinical note



Natural language processing - NLP

Why NLP ?

Vast and **growing** number of **unstructured clinical notes**

NLP enables **computations** with natural (human) **languages**

- Retrieve **hidden information** and turn it into **knowledge**
- Harness **untapped** project-specific textual data sources

Named-entity recognition

Supervised machine learning or rule-based approach

Identify and classify named entities in text

- PHI (names, dates, locations, ...)
- Document sections

In general, **find** project-specific tokens/phrases in unstructured text which might not all necessarily be known **a priori**

Named-entity recognition

Identify and classify

Synthetic note (fabricated patient info) named entities in text

History of Present Illness

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital, RI [...]

History and Physical conducted by: Jeff York, MD

Named-entity recognition

Identify and classify

Identify: PHI (names, locations ...) named entities in text

History of Present Illness

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital, RI [...]

History and Physical conducted by: Jeff York, MD

Named-entity recognition

Identify and classify

Identify: PHI (names, locations ...) named entities in text

History of Present Illness

In practice, de-identification systems can mark all ages, not only > 89

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

Not likely to be found in a dictionary

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital RI [...]

History and Physical conducted by: Jeff York MD

Same token, different NE

Named-entity recognition

Redact named-entities

Identify and classify
named entities in text

History of Present Illness

****NAME<AAA>** is a ****AGE<in 60s>** year old white male from ****LOCATION** with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization. post-MI.

****DATE<[**2015-07-08**]>** at ****HOSPITAL** ****LOCATION** [...]

History and Physical conducted by: ****NAME<CCC>** MD

Named-entity recognition

Identify and classify
named entities in text

History of Present Illness

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital, RI [...]

History and Physical conducted by: Jeff York, MD

Named-entity recognition

Identify: Document sections

Identify and **classify**
named entities in text

History of Present Illness

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital, RI [...]

History and Physical conducted by: Jeff York, MD

Named-entity recognition

Identify: Document sections

Identify and classify
named entities in text

History of Present Illness

Peter Miller is a 65 year old white male from New York with a past medical history significant for an MI and depression who presents today complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen. [...]

PAST MEDICAL HISTORY

Surgeries/procedures Cardiac catheterization, post-MI, 10/11/2012 at Famous Hospital, RI [...]

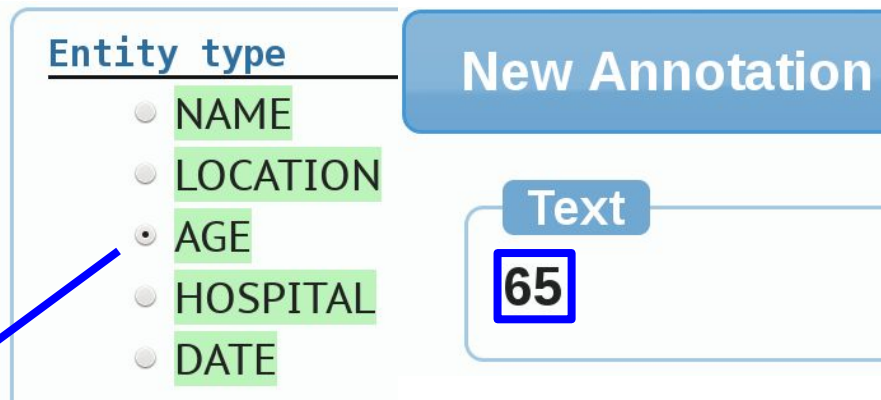
History and Physical conducted by Jeff York, MD

Ambiguity, NEs are project-specific



Named-entity recognition - Annotate

Browser based tool to
accelerate the
project-specific manual
annotation process



Entity type

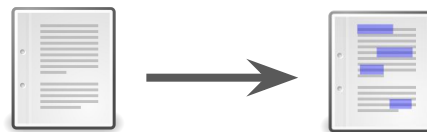
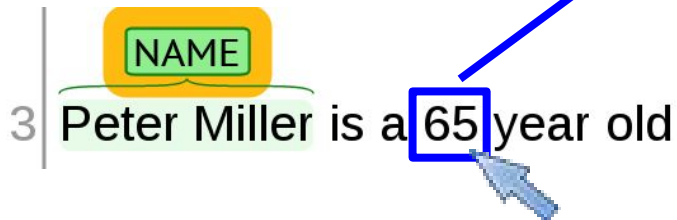
- ☐ NAME
- ☐ LOCATION
- ☒ AGE
- ☐ HOSPITAL
- ☐ DATE

New Annotation

Text

65

3 Peter Miller is a 65 year old



Named-entity recognition - Generate training data

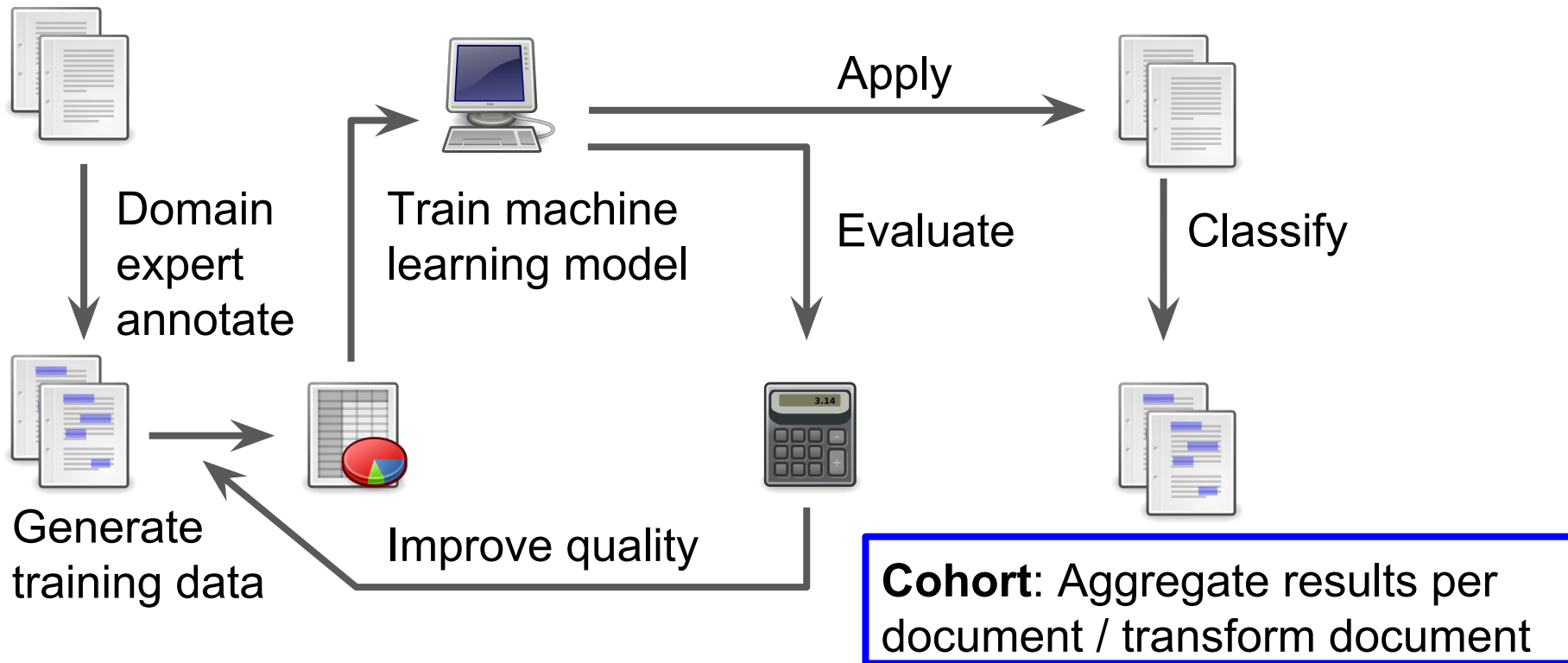
NER-TYPE	BEGIN END INDICES	NER-TEXT
NAME	28 40	Peter Miller
AGE	46 48	65
LOCATION	74 82	New York
DATE	410 420	10/11/2012
HOSPITAL	424 439	Famous Hospital
LOCATION	441 443	RI
NAME	487 496	Jeff York
...

Example features generated for each **token**

- The token itself
- The previous token
- The next token
- The last k-character token suffixes
- The first k-character token prefixes
- The token shape (upper or lower cases of each token character)
- The token gazetteer membership
- ...

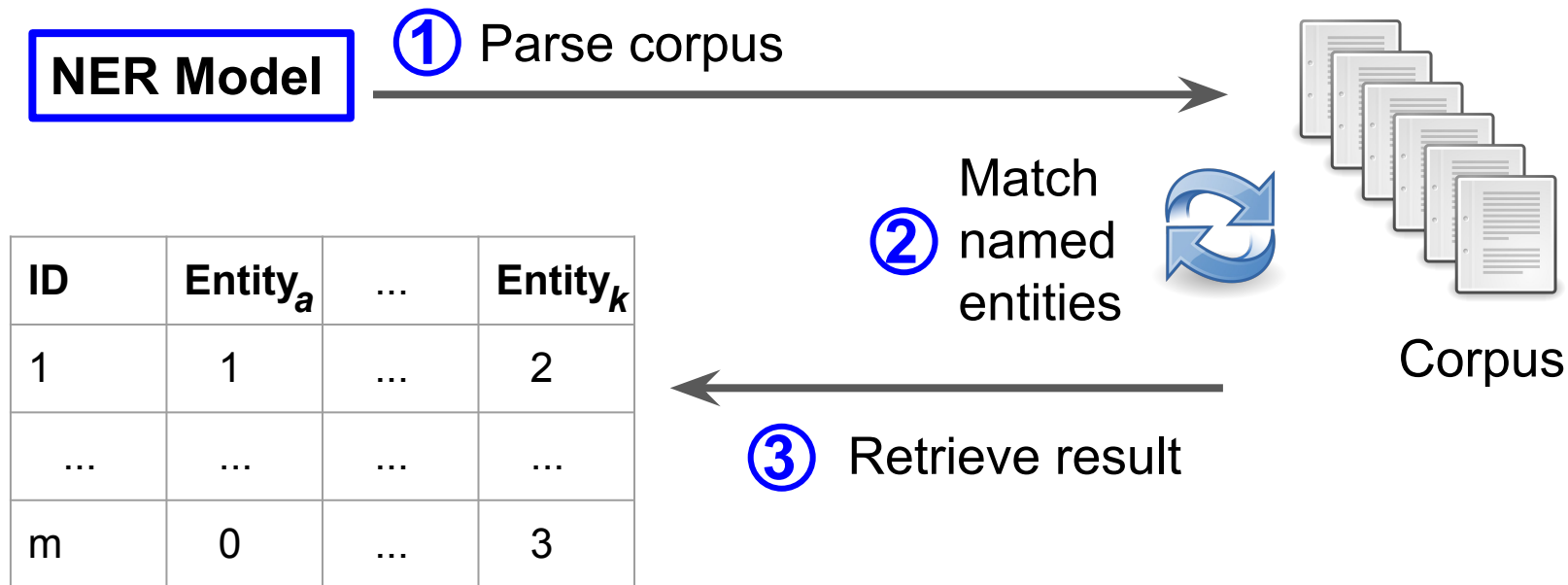


Named-entity recognition - Train and apply



Information Extraction

Generate document-entity count matrix from a corpus



Information Extraction - Rule-based example

Rule-based approach for **risk-factor** extraction

List of risk-factors

Social factors	Substance abuse	...
jail	alcohol abuse	...
lonely	(use abuse) ... substance	...
financial assistance	alcoholic	...
isolation



Clinical documents


Regex
NER

Information Extraction - Rule-based example

Rule-based approach for **risk-factor** extraction:

List of risk-factors

Social factors	Substance abuse	...
jail	alcohol abuse	...
lonely	(use abuse) ... substance	...
financial assistance	alcoholic	...
isolation


Clinical documents

**Regex
NER**

Lowercasing, stemming / lemmatization:

abuse, abuses, Abuse

abuse

Negation detection:

He is not lonely.

He is not lonely_**NEG**.

Information Extraction - Rule-based example



Token matching **pitfalls**:

List of risk-factors

Social factors	Medical factors	Others
aid	AIDS	hearing aids



Clinical documents



Lowercasing, stemming:
Looking to match financial aids but specifying only **aid**, falsely match AIDS.

**Regex
NER**

Lowercasing, stemming / lemmatization:

AIDS, the acquired immune deficiency syndrome

... **aid** ...

Financial **aids** ...

Information Extraction - Rule-based example



Token matching **pitfalls**:

List of risk-factors

Social factors	Medical factors	Others
aid	AIDS	hearing aids



Lowercasing: Looking to match **AIDS** and falsely match hearing aids.



Clinical documents

**Regex
NER**

Lowercasing, stemming / lemmatization:

AIDS, the acquired immune deficiency syndrome

... **aid** ...

hearing **aids** ...




Information Extraction - Rule-based example

Document-term/phrase **risk-factor** matrix:

Document-ID	Text	abuse alcohol	abuse_NEG alcohol_NEG	...
1		2	0	...
2		0	1	...
3		0	1	...
...

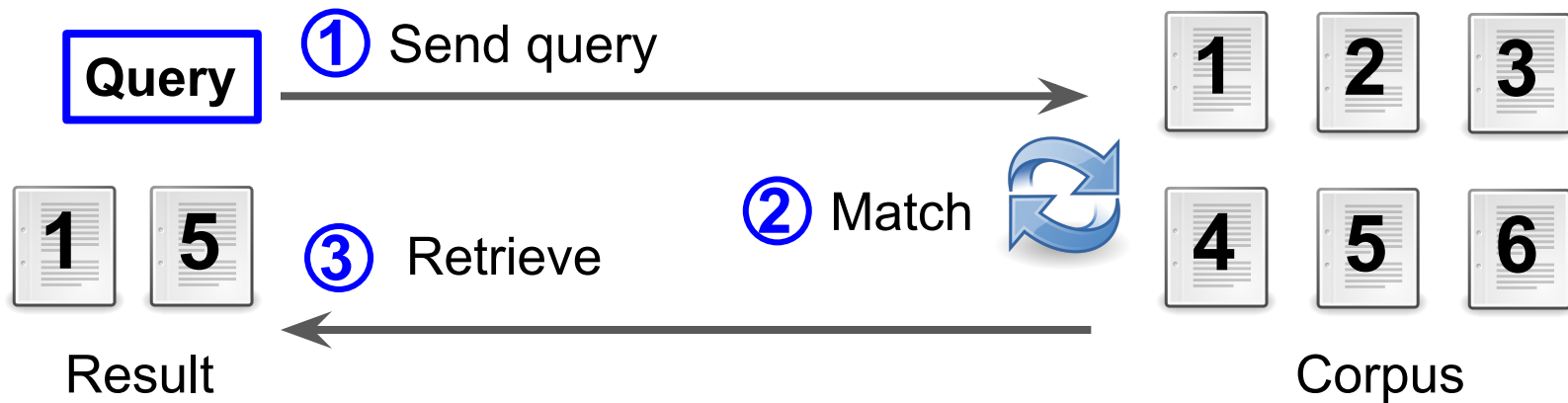
Information Extraction - Hybrid example with NER

Document-term/phrase **risk-factor** matrix:

Document-ID	Text	abuse alcohol	abuse_NEG alcohol_NEG	...	Named entities of class-1
1		2	0	...	0
2		0	1	...	2
3		0	1	...	0
...

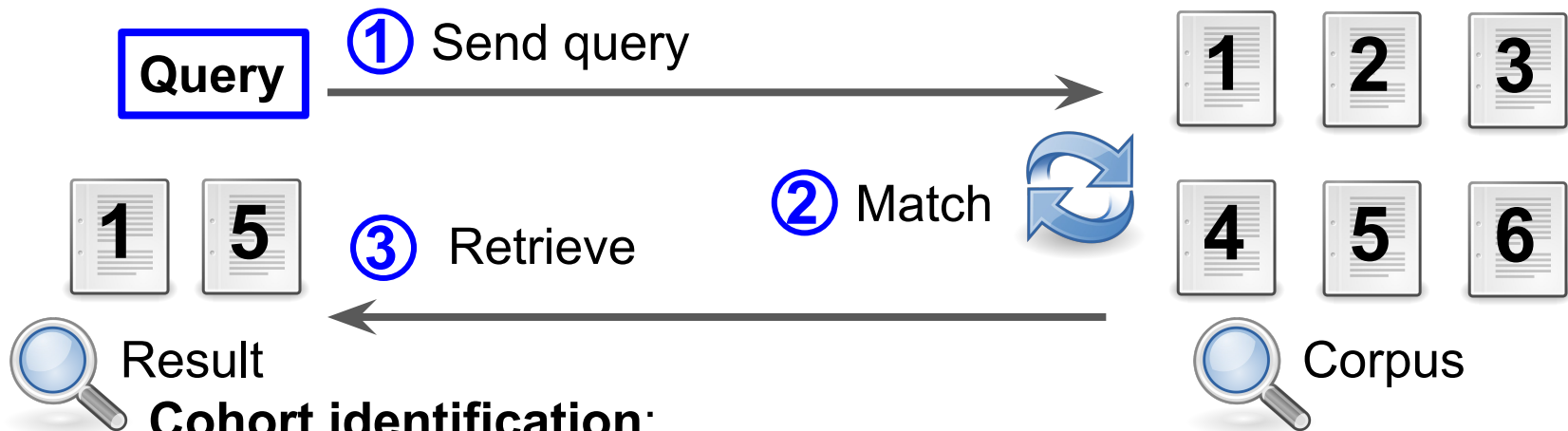
Search

Query arbitrary corpus to **retrieve matching** documents



Search

Query arbitrary corpus to **retrieve matching** documents



Cohort identification:

- Full document text
- Anonymized patient ID
- Associated discrete attributes

- Radiology reports
- Discharge summaries
- External project-specific corpus

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search

Search

Query arbitrary corpus to **retrieve matching** documents

Synthetic note (fabricated patient info)

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search

History of Present Illness

Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY

Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: pain

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: leg

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- **Boolean operators**
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: pain **AND** epigastric

History of Present Illness
Pt is complaining of sharp, **epigastric** abdominal **pain** of 3-4 months duration. The **pain** is located in the **epigastric** region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- **Boolean operators**
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: pain **AND** leg

History of Present Illness
Pt is complaining of sharp, epigastric abdominal **pain** of 3-4 months duration. The **pain** is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- **Boolean operators**
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: pain **OR** leg

History of Present Illness
Pt is complaining of sharp, epigastric abdominal **pain** of 3-4 months duration. The **pain** is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- **Boolean operators**
- Proximity phrase match
- Word stem match
- Section search
- Concept search



History of Present Illness
Pt is complaining of sharp, epigastric abdominal **pain** of 3-4 months duration. The **pain** is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Query: pain **AND** (**NOT** leg)



**Absence, not negation of
“leg”**

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- **Boolean operators**
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: pain **AND** (**NOT**
duration)

History of Present Illness
Pt is complaining of sharp,
epigastric abdominal **pain** of
3-4 months **duration**. The **pain**
is located in the epigastric
region and left upper quadrant
of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac
catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity **phrase match**
- Word stem match
- Section search
- Concept search



Query: “**abdominal pain**”

History of Present Illness
Pt is complaining of sharp, epigastric **abdominal pain** of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: “abdominal pain” AND region

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity **phrase match**
- Word stem match
- Section search
- Concept search



Query: “**epigastric pain**”

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- Concept search



Query: “epigastric pain”~1

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- **Word stem match**
- Section search
- Concept search



Query: complaining

History of Present Illness
Pt is **complaining** of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- **Word stem match**
- Section search
- Concept search



Query: complain

History of Present Illness
Pt is **complaining** of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- **Word stem match**
- Section search
- Concept search



History of Present Illness
Pt is **complaining** of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Query: complains

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- **Section search**
- Concept search



Query:

PAST_MEDICAL_HISTORY:
cardiac AND surgery

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- **Section search**
- Concept search



History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Query:

History_of_Present_Illness:
cardiac AND surgery

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- **Concept search**

Query: pain AND Age:[40 TO 60] AND Gender:Male



Patient
Data

History of Present Illness
Pt is complaining of sharp, epigastric abdominal **pain** of 3-4 months duration. The **pain** is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- **Concept search**

Query: pain AND Age:[20 TO 30] AND Gender:Male



Patient
Data

History of Present Illness
Pt is complaining of sharp, epigastric abdominal **pain** of 3-4 months duration. The **pain** is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: Cardiac catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- **Concept search**



Query: cardiac OR heart OR
coronary OR cor

UMLS
hand-
craft

History of Present Illness
Pt is complaining of sharp,
epigastric abdominal pain of
3-4 months duration. The pain
is located in the epigastric
region and left upper quadrant
of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: **Cardiac**
catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Token match
- Boolean operators
- Proximity phrase match
- Word stem match
- Section search
- **Concept search**



Query: concepts:"C0018787"

UMLS
cTAKES

History of Present Illness
Pt is complaining of sharp, epigastric abdominal pain of 3-4 months duration. The pain is located in the epigastric region and left upper quadrant of the abdomen.

PAST MEDICAL HISTORY
Surgeries/procedures: **Cardiac** catheterization.

Search

Query arbitrary corpus to **retrieve matching** documents

- Custom-build **browser-based search applications**
- Allow **interactive** data-driven exploration of project texts
- **Expose** NLP and machine learning results for **queries**

Search - Web Demo

ISEAR corpus - Survey to report situations of **emotions**:

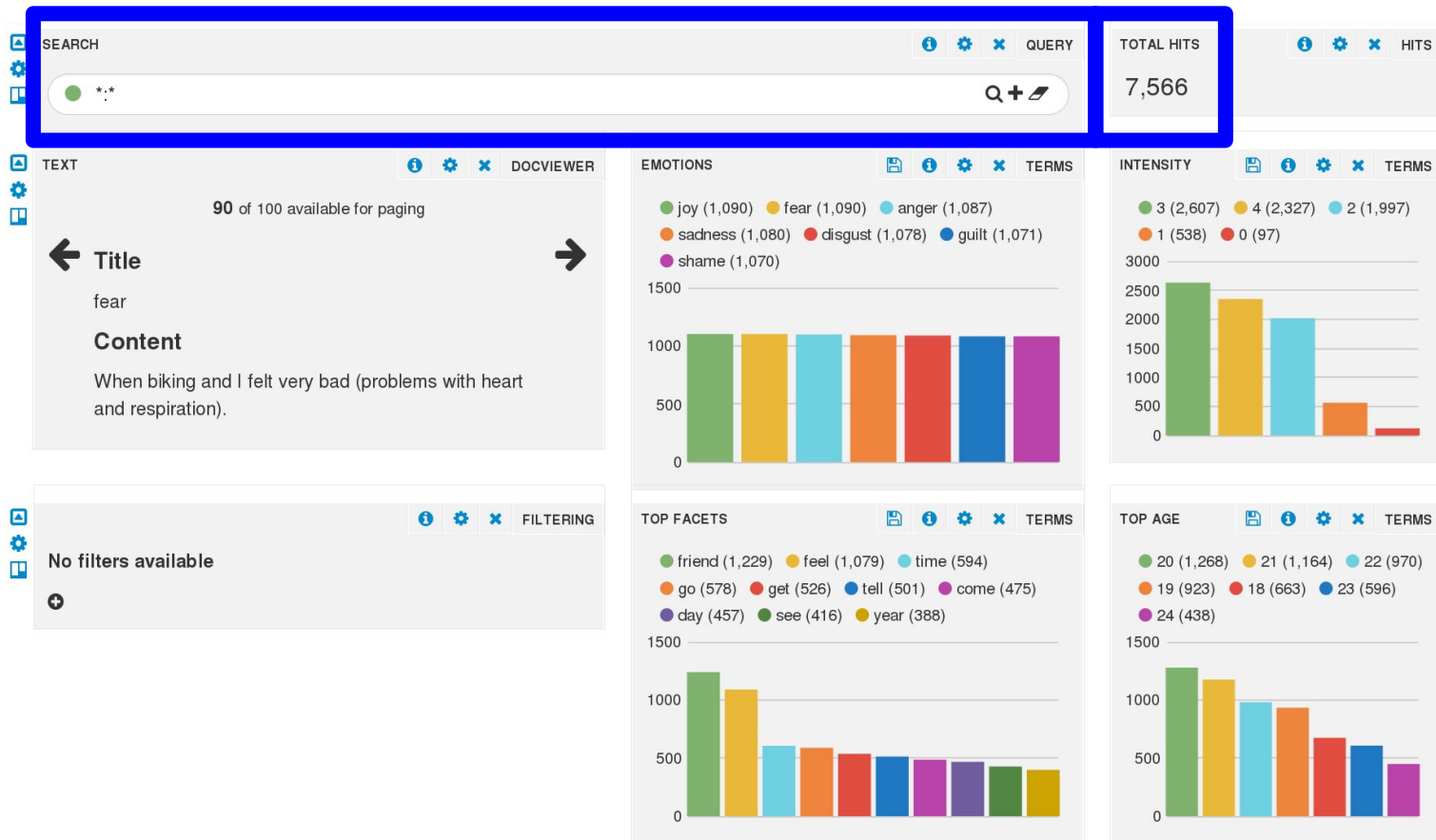
- joy, fear, anger, sadness, disgust, shame, and guilt

Additionally reported:

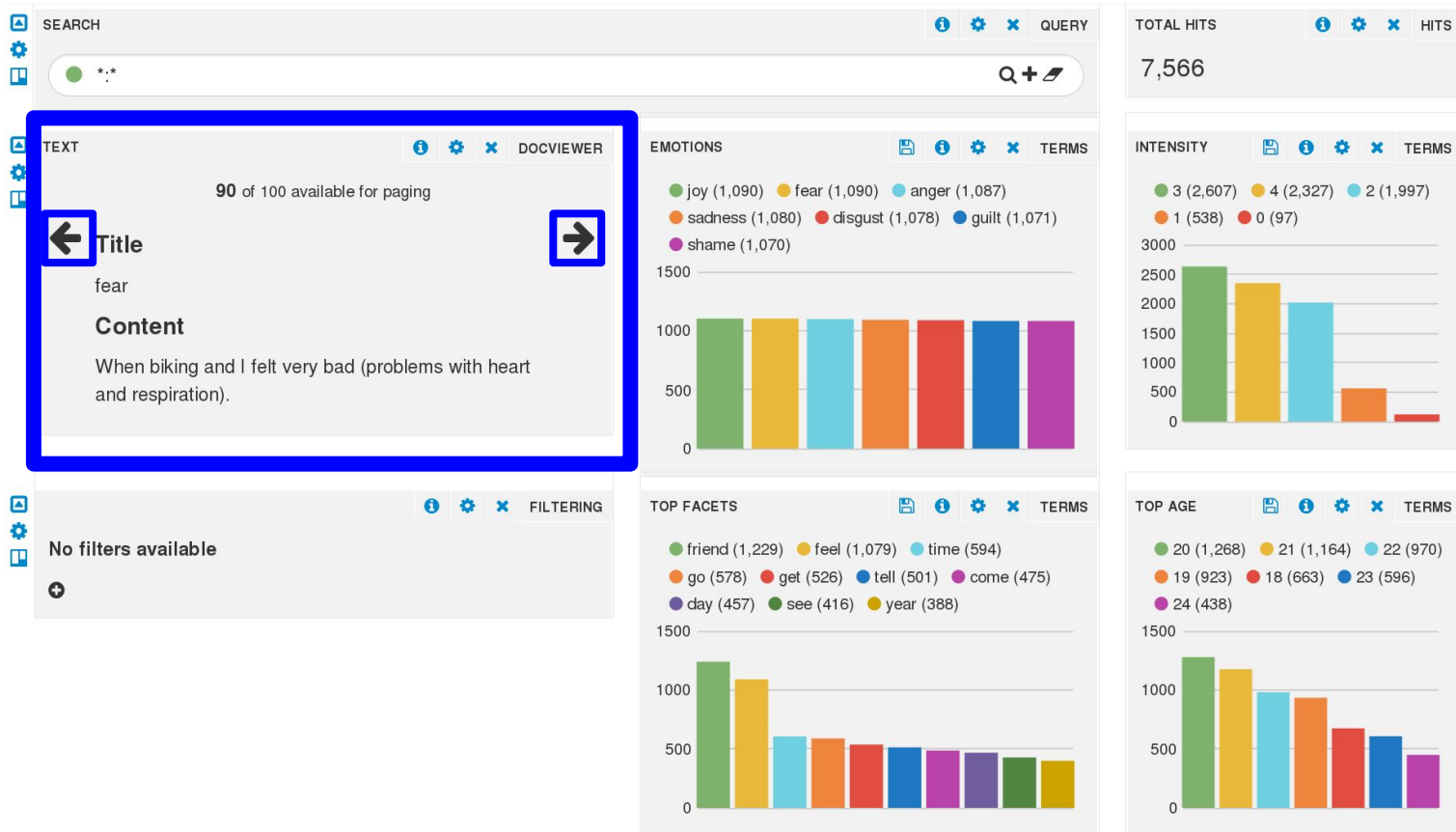
- intensity, ergotropic arousal, coping, expected, fairness, ...

Id	Text	Emotion	Intensity	...
1	I am the secretary of an association, and during the last meeting I forgot to take the minutes.	guilt	2	...
2	Walking in the dark and thinking about ghost stories.	fear	3	...
3	Fighting with my father while drunk.	shame	4	...
...

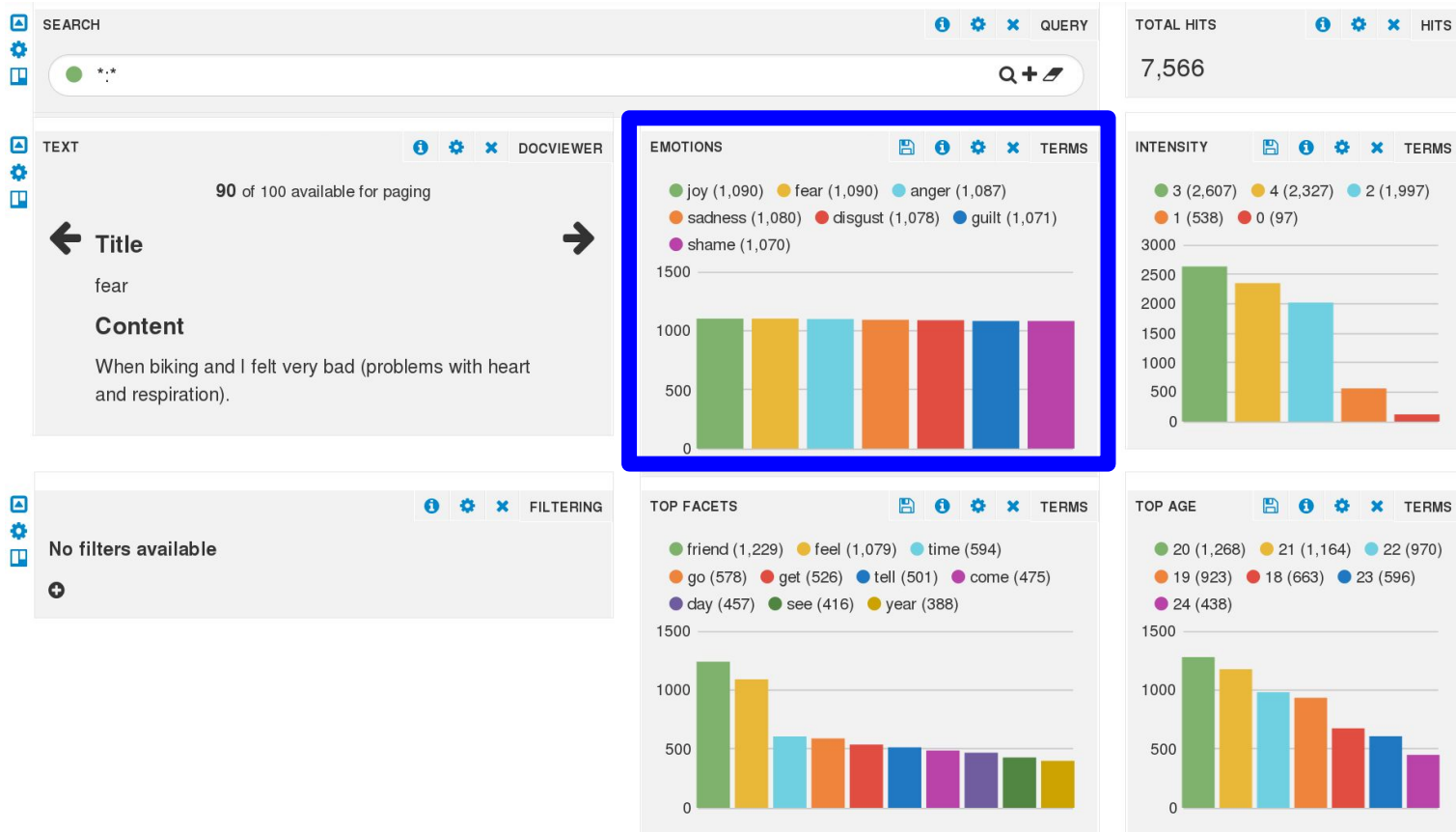
Search - Web Demo



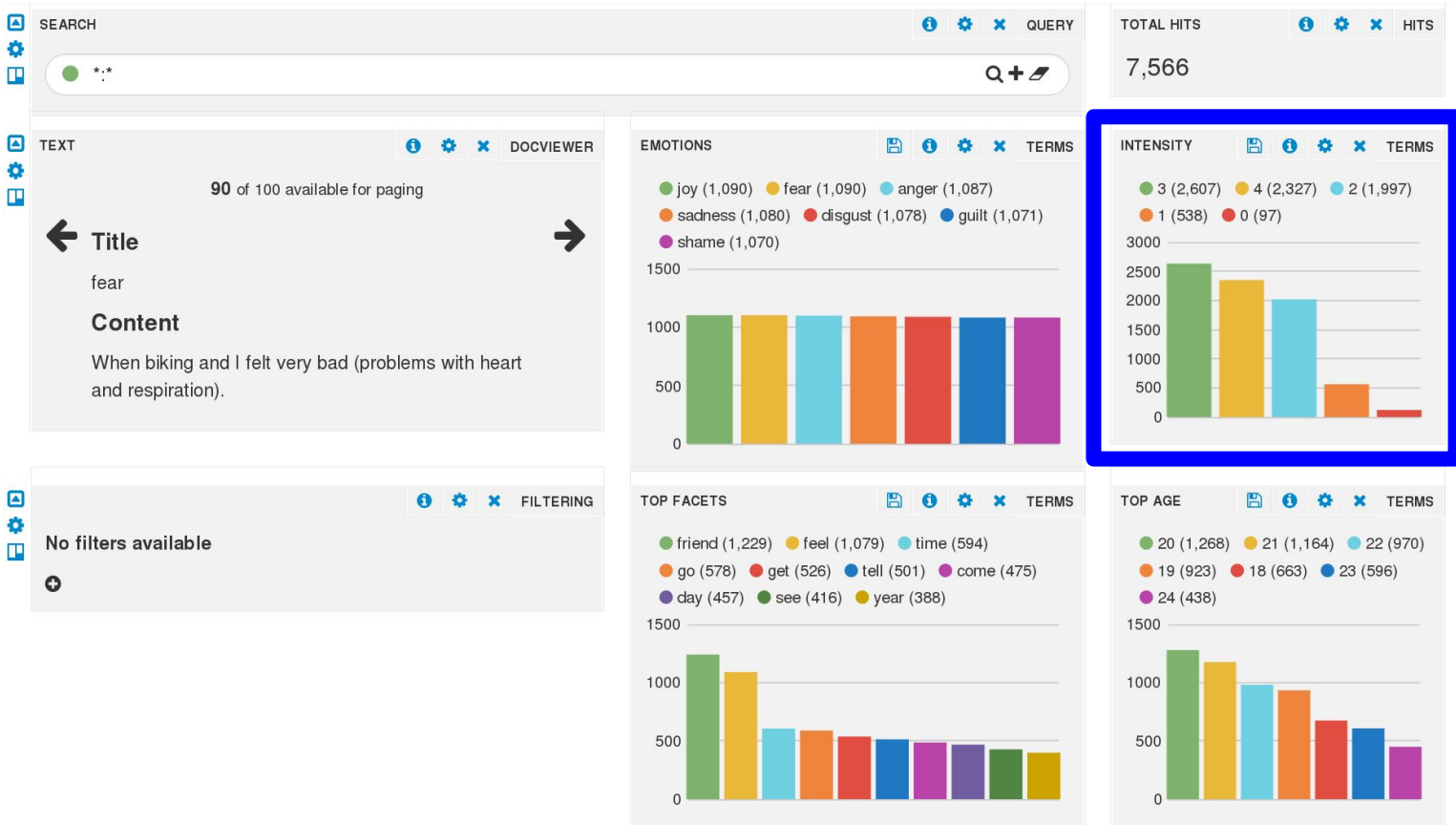
Search - Web Demo



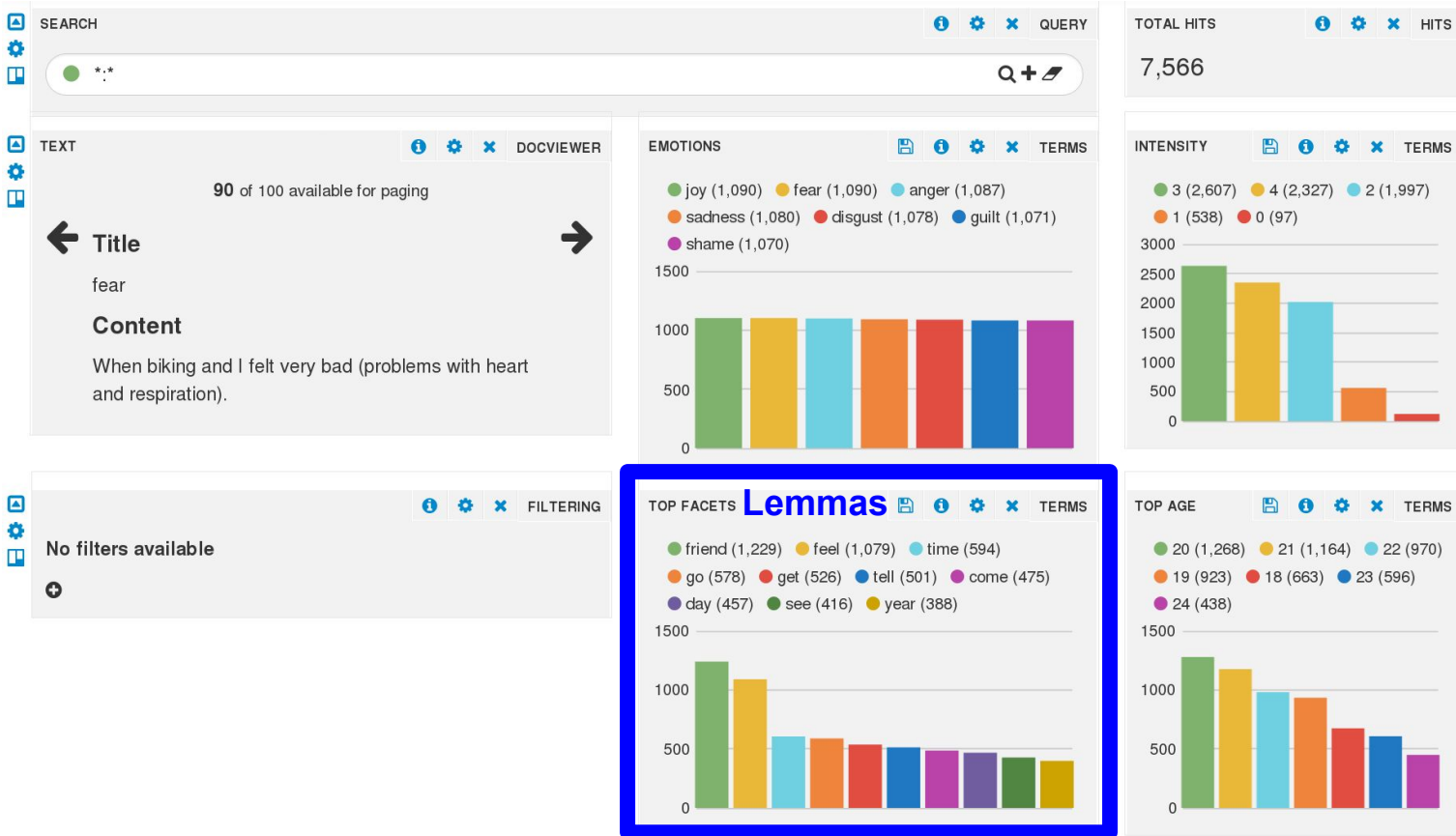
Search - Web Demo



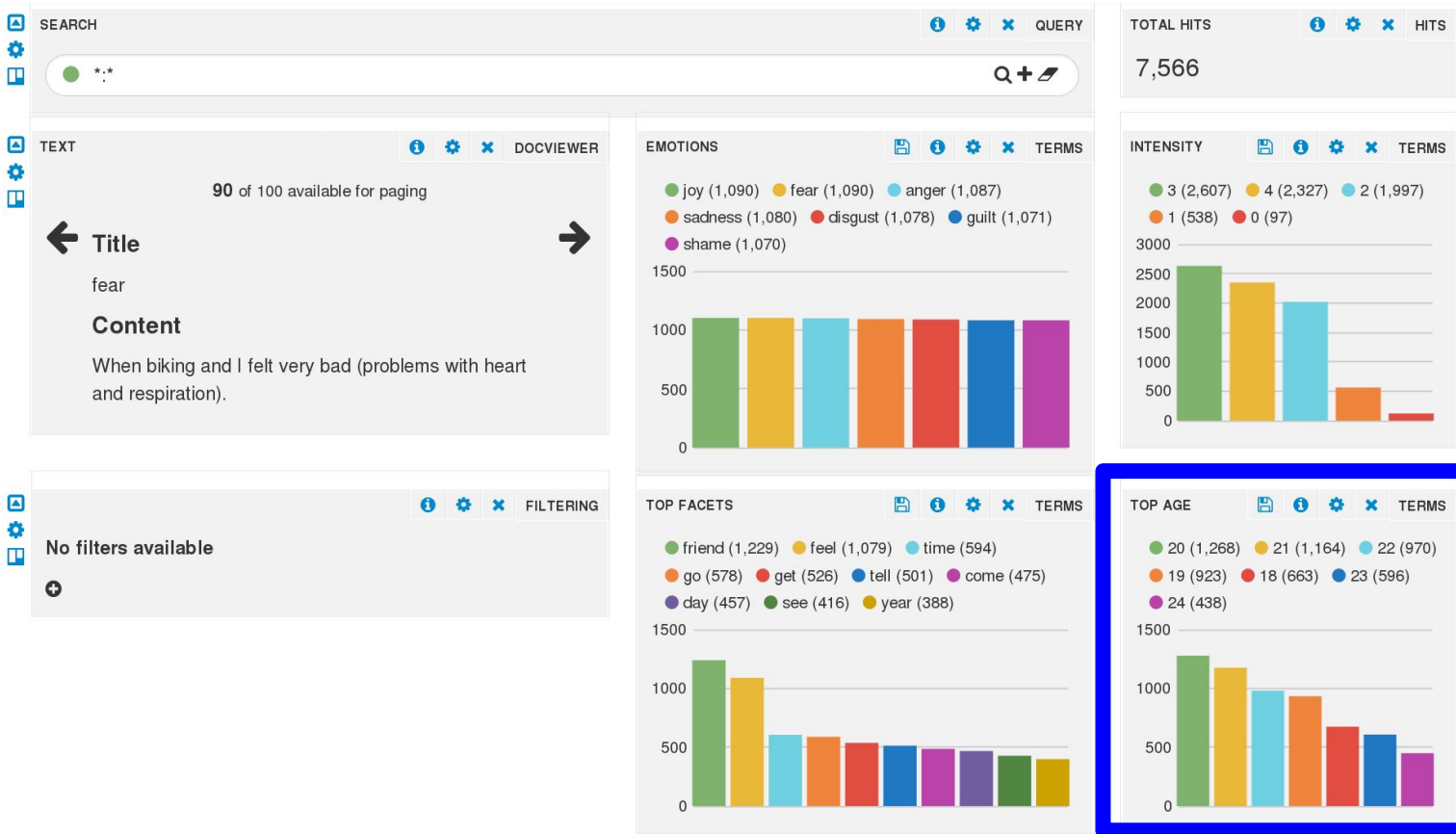
Search - Web Demo



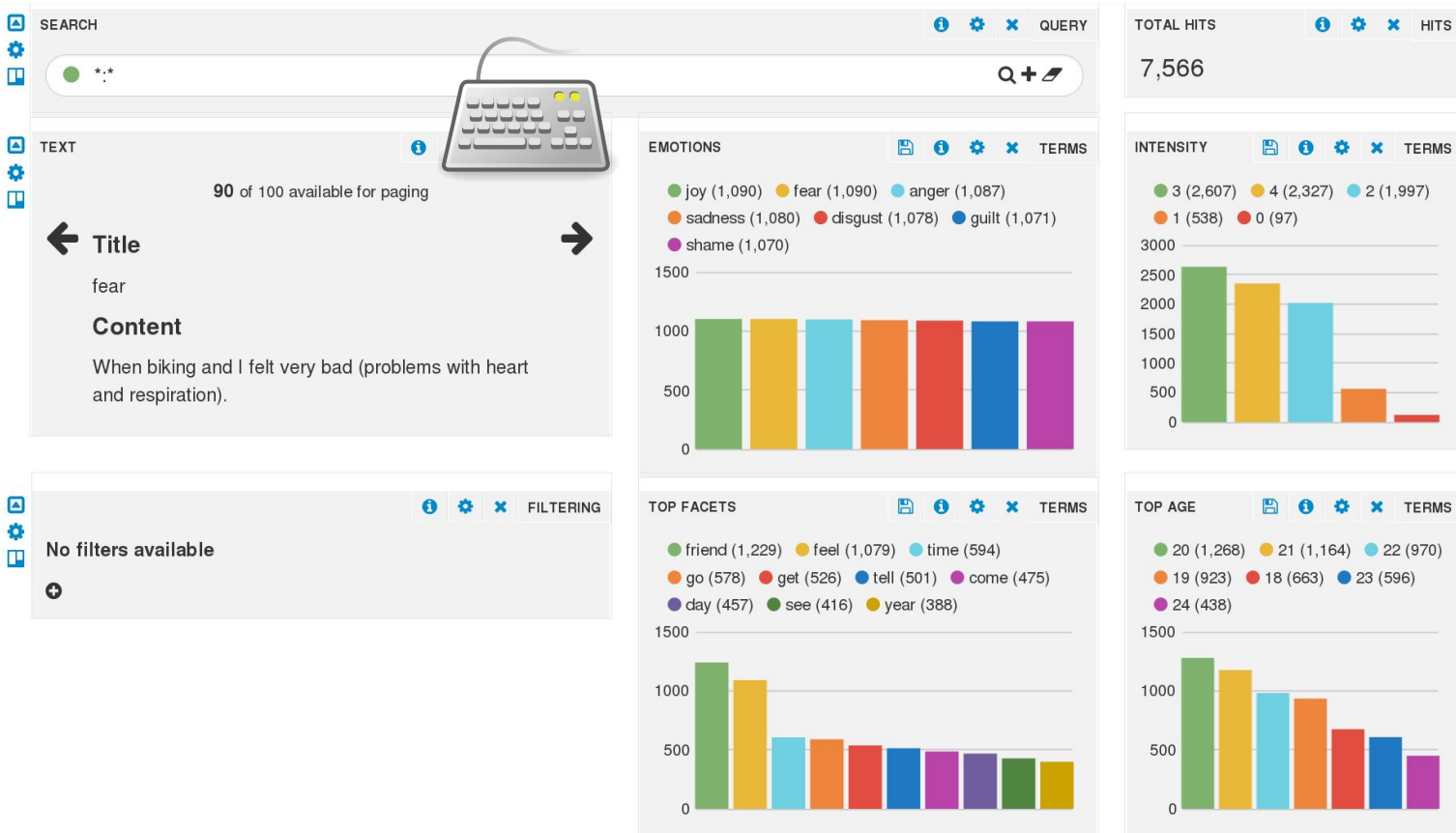
Search - Web Demo



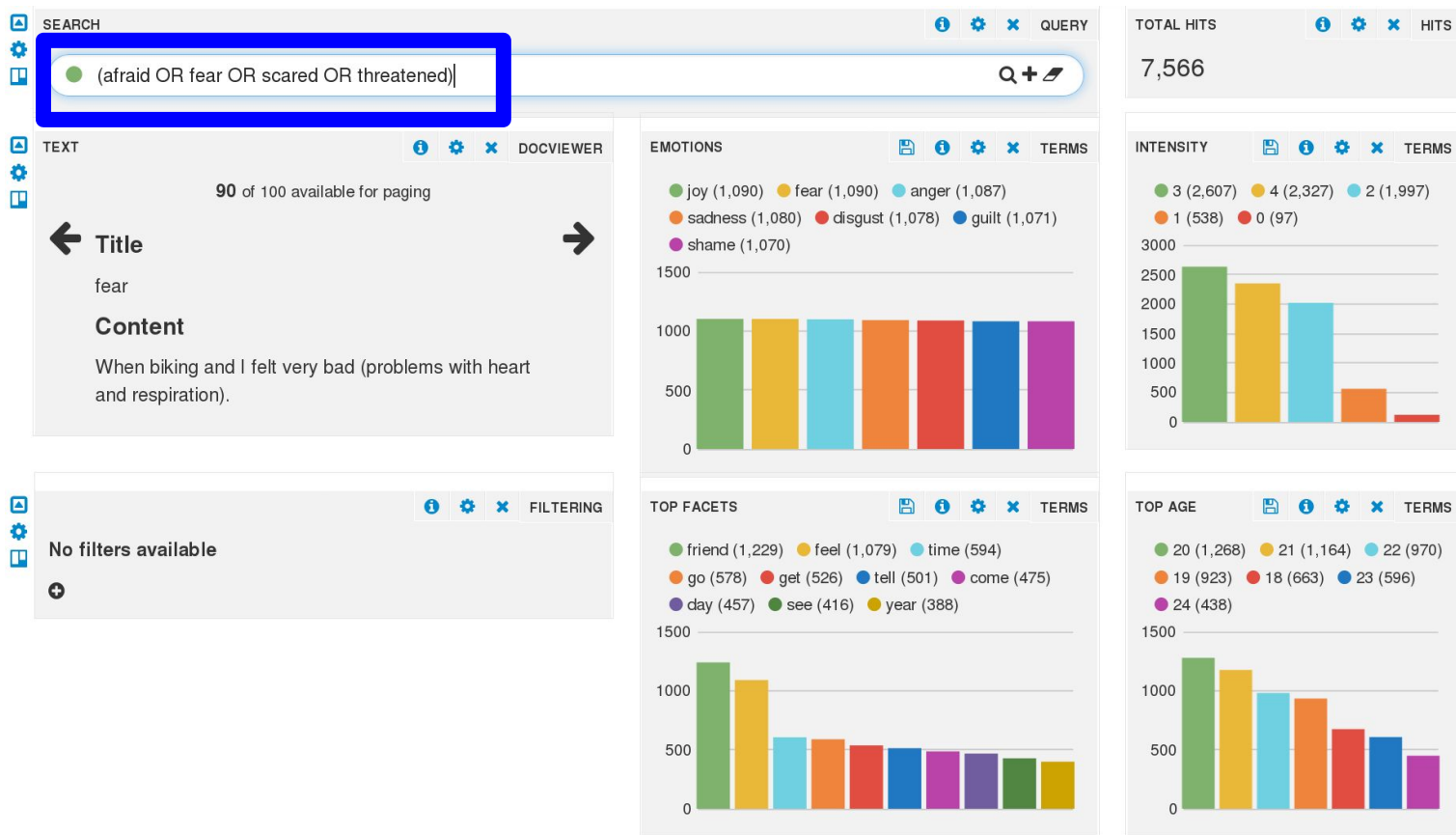
Search - Web Demo



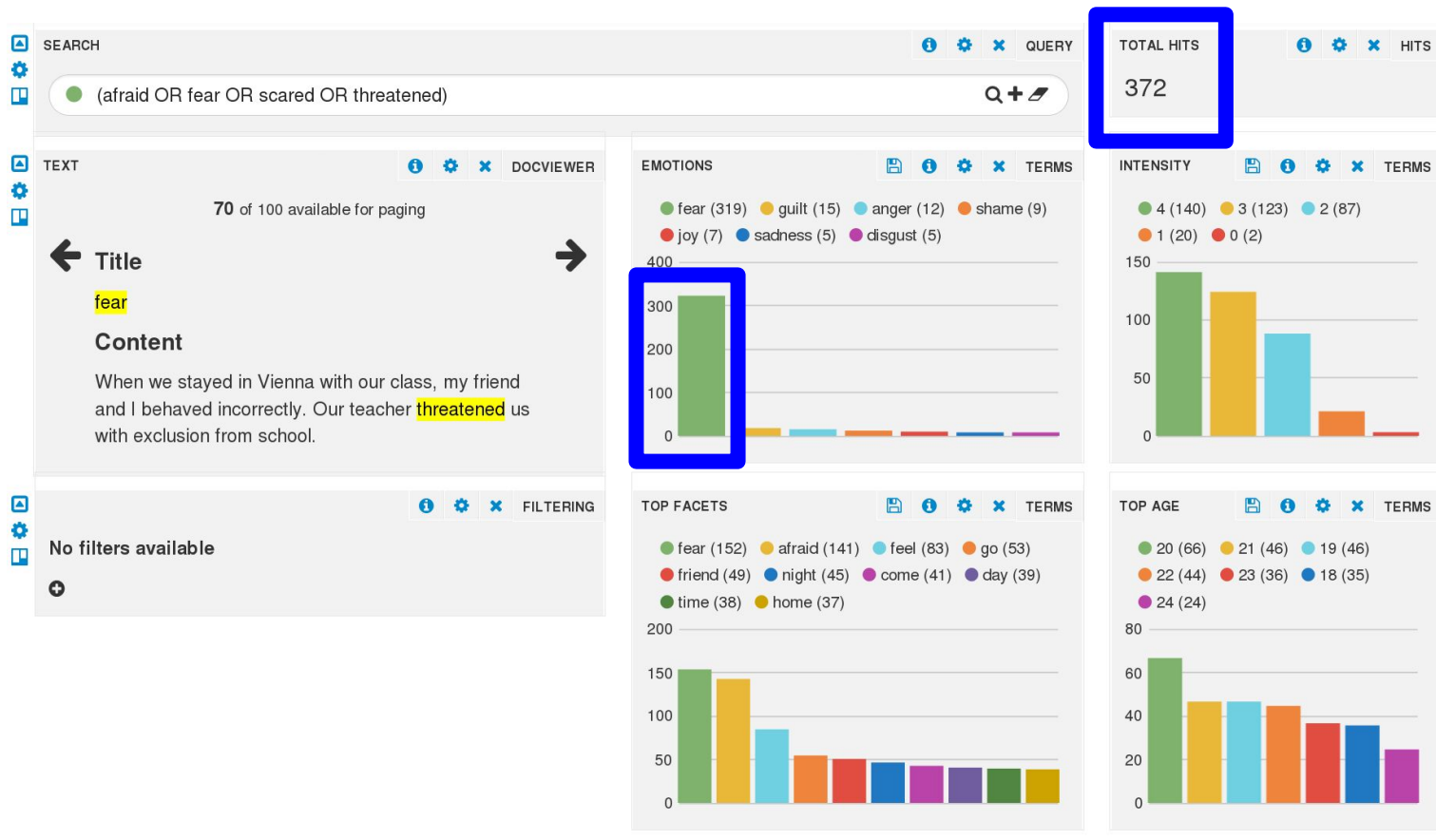
Search - Web Demo



Search - Web Demo



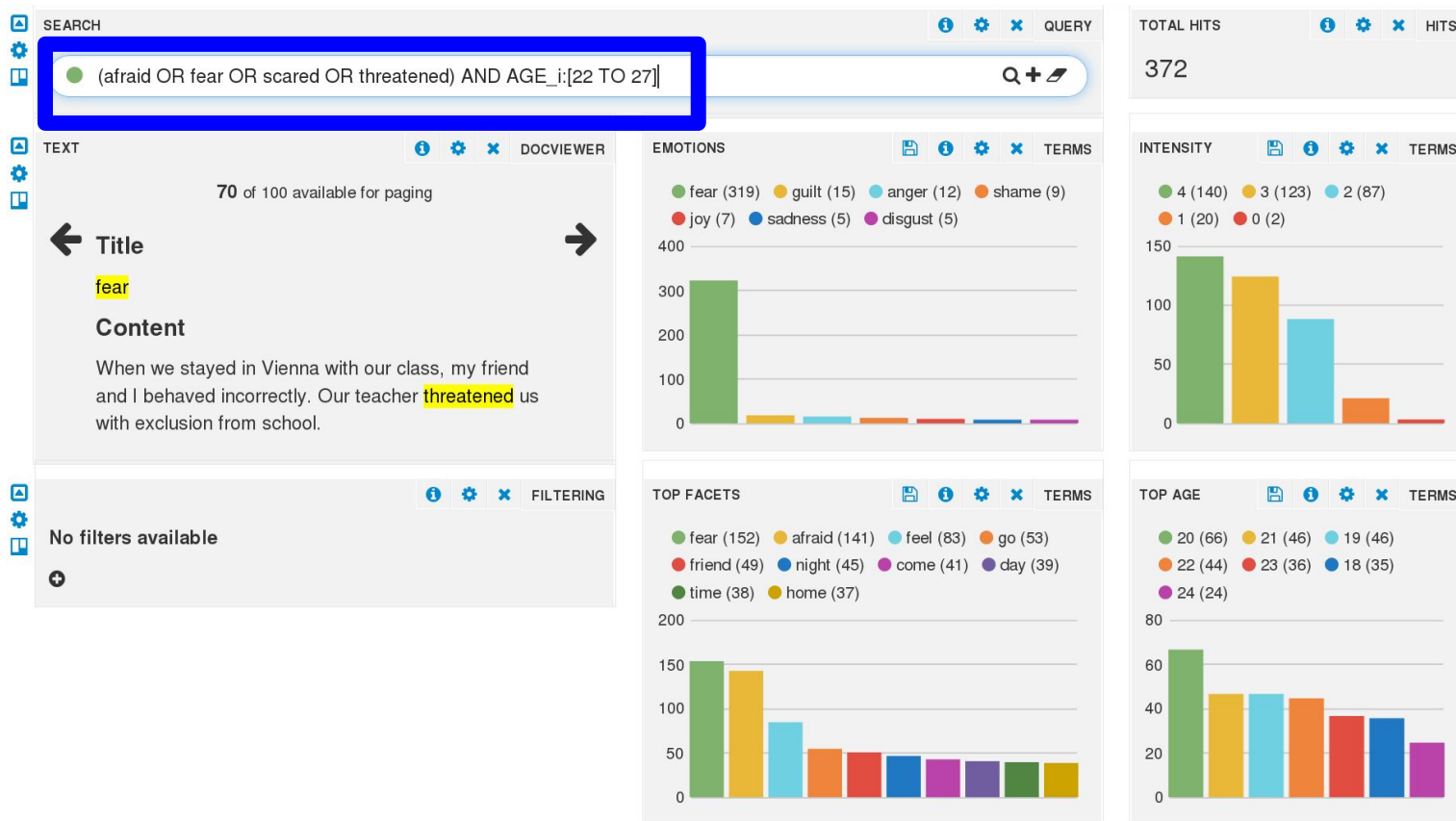
Search - Web Demo



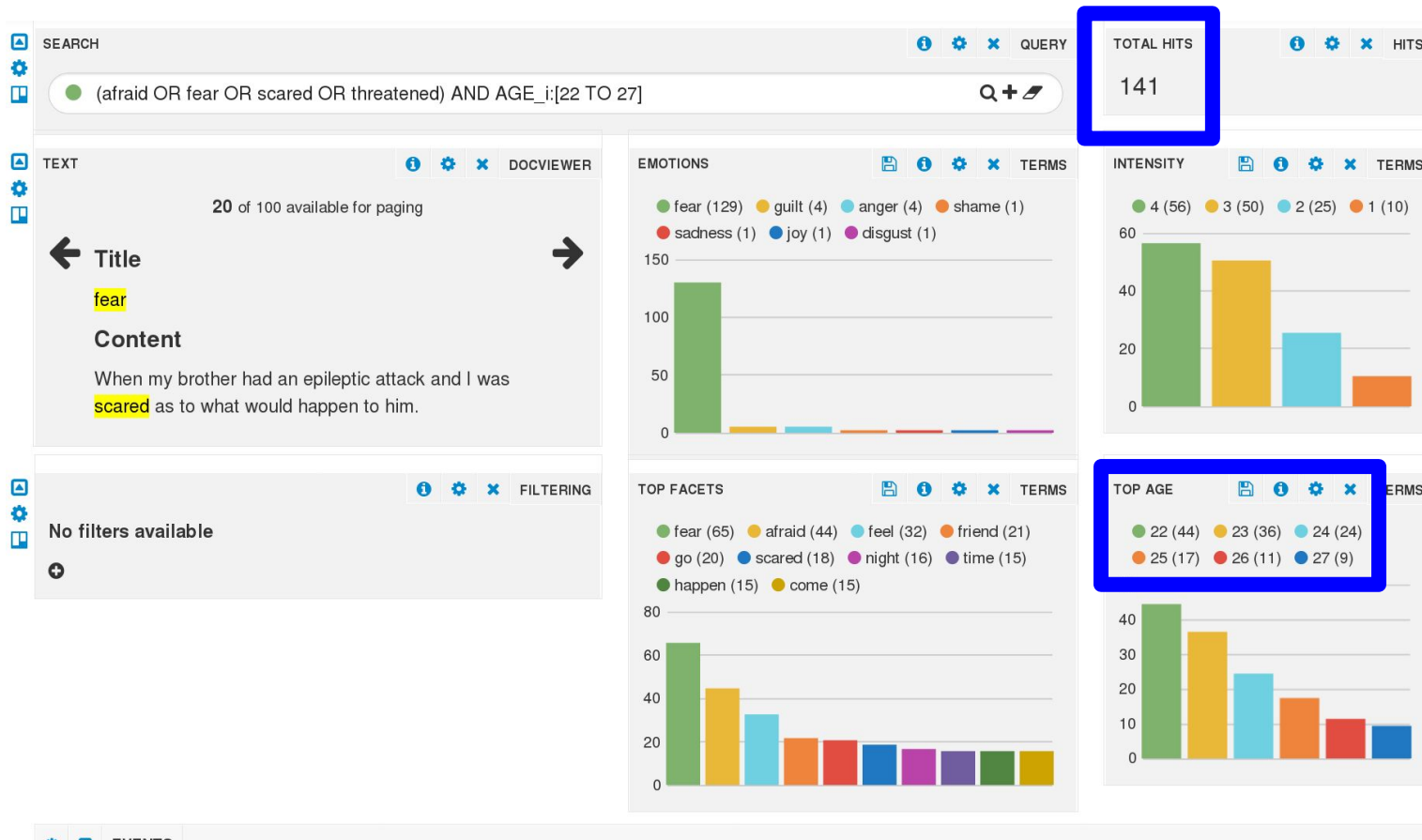
Search - Web Demo



Search - Web Demo



Search - Web Demo



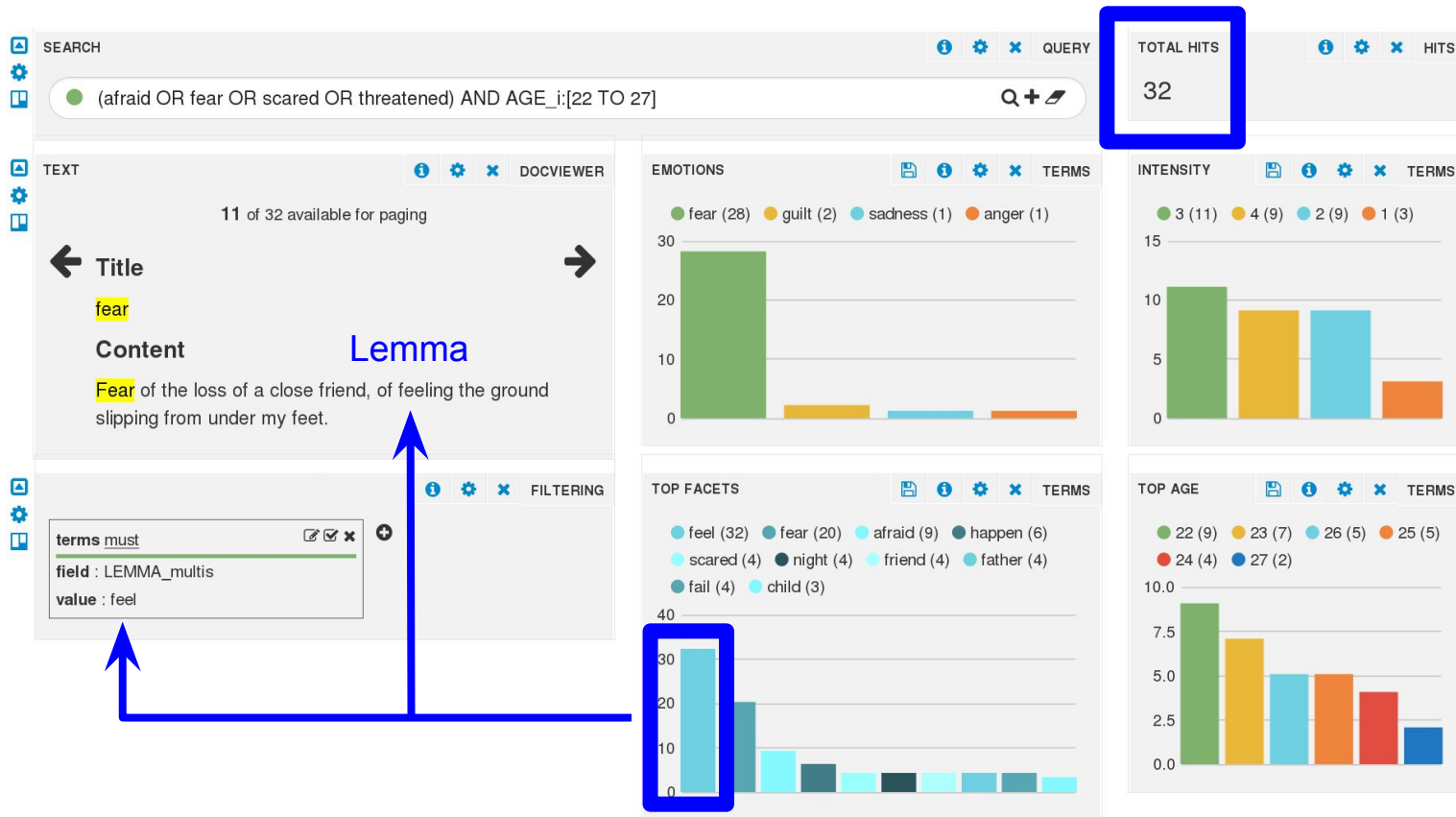
Search - Web Demo



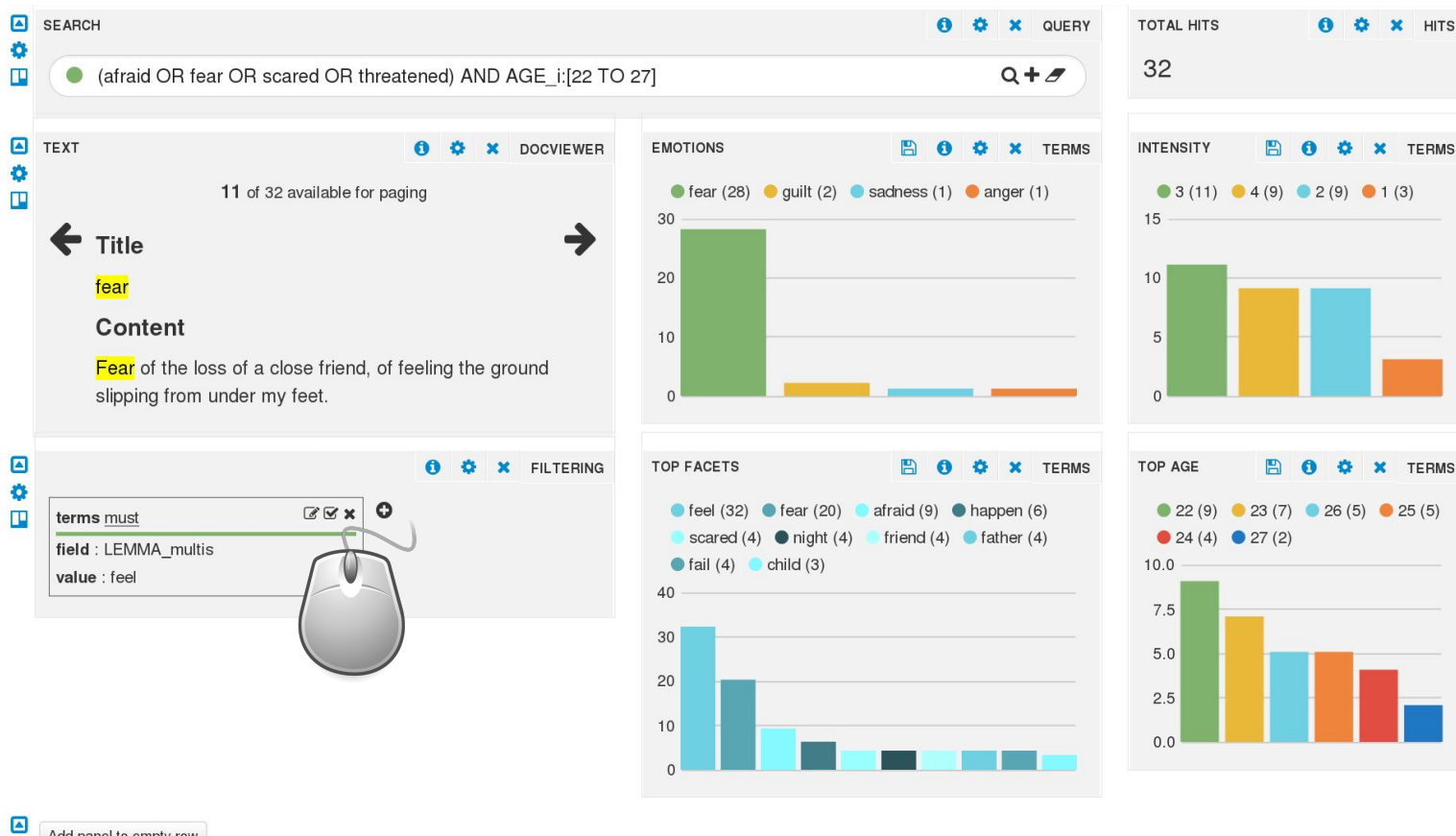
Search - Web Demo



Search - Web Demo



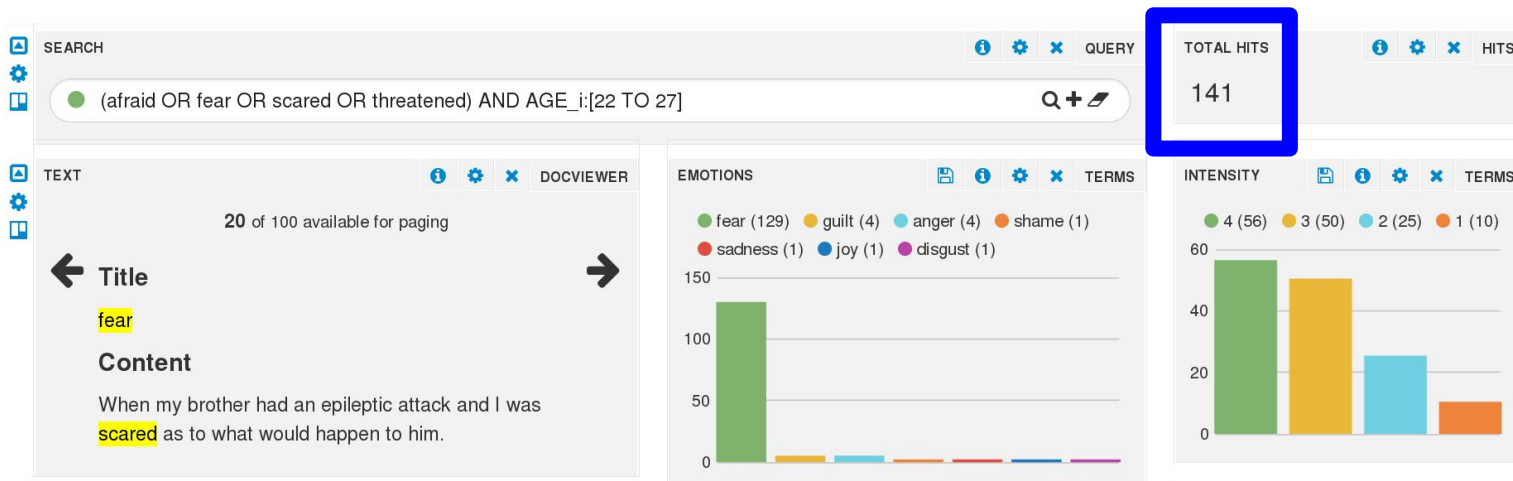
Search - Web Demo



Search - Web Demo



Search - Web Demo

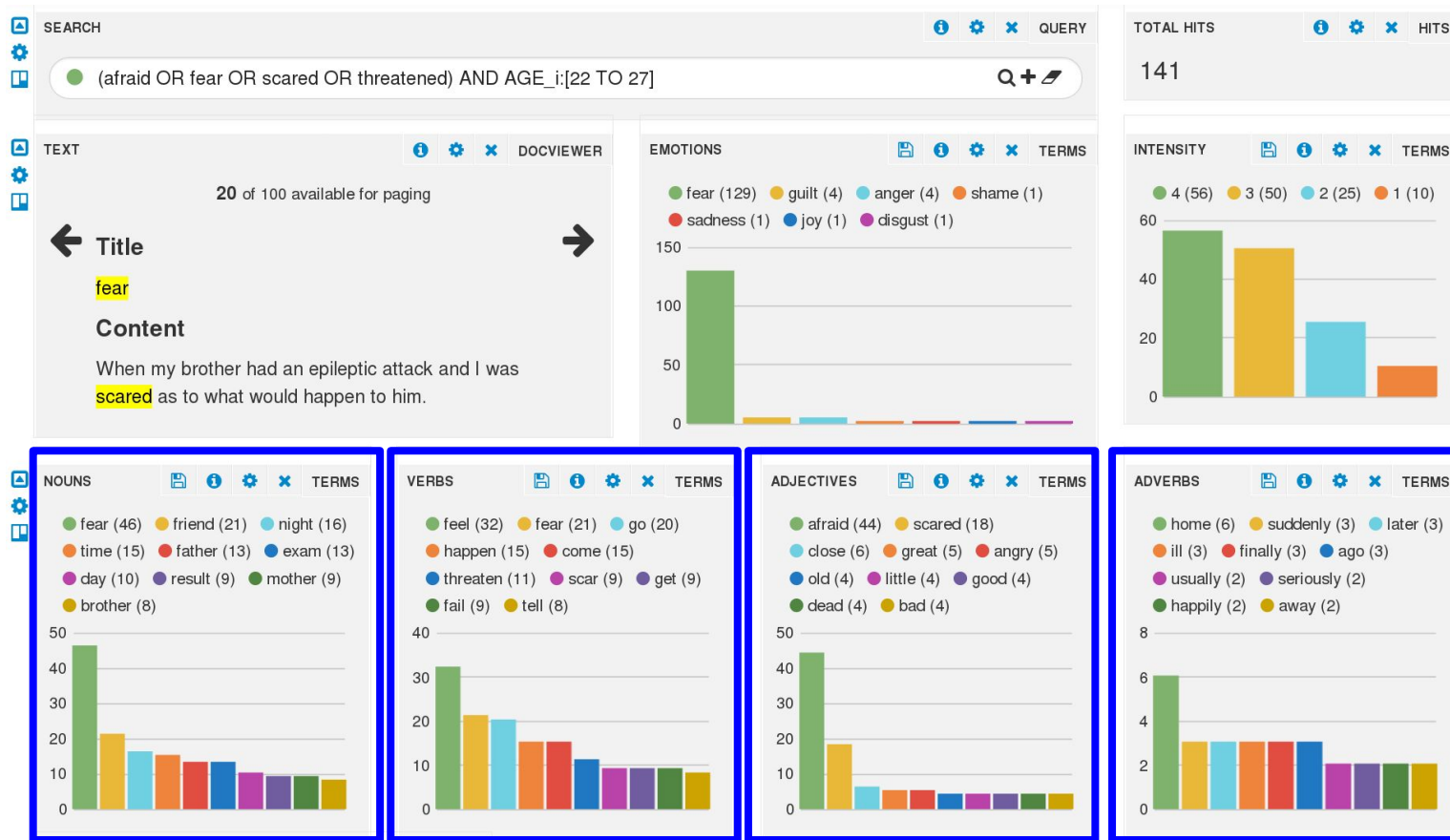


Remove previous panels

Add **part-of-speech** lemma facet panels

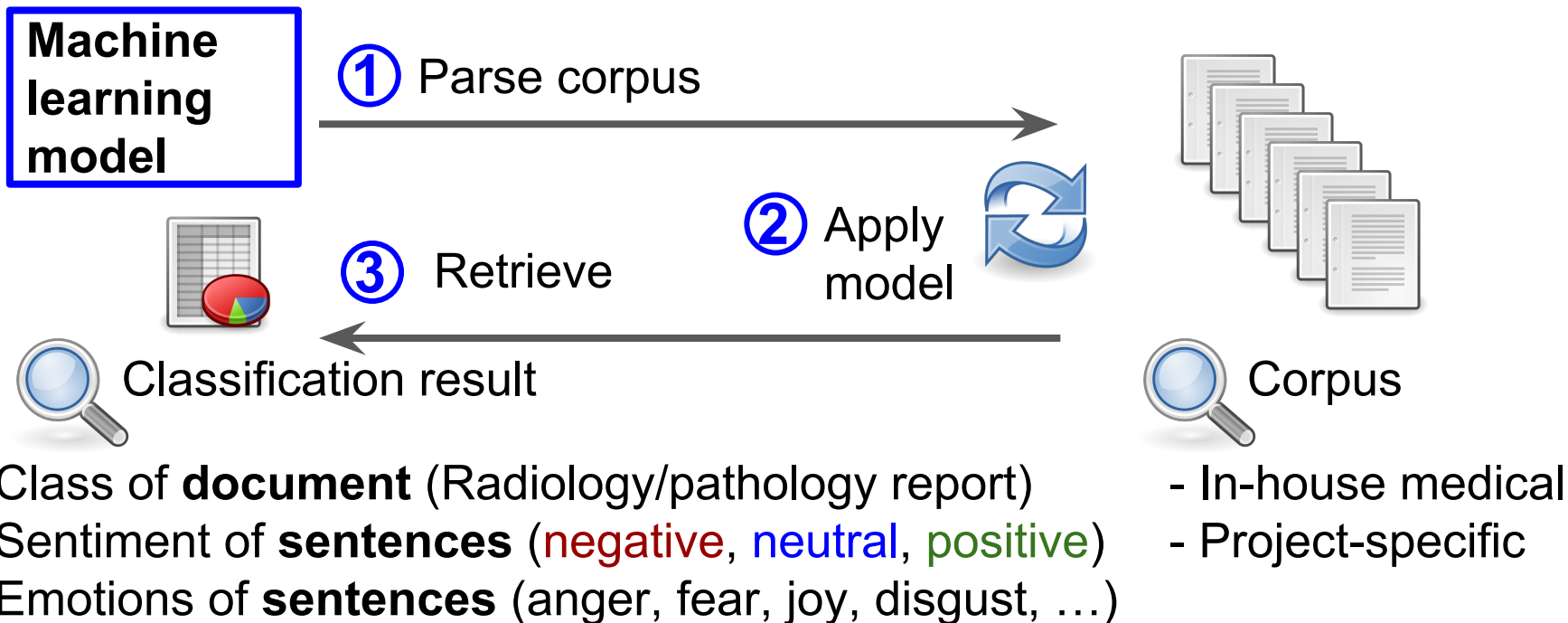


Search - Web Demo

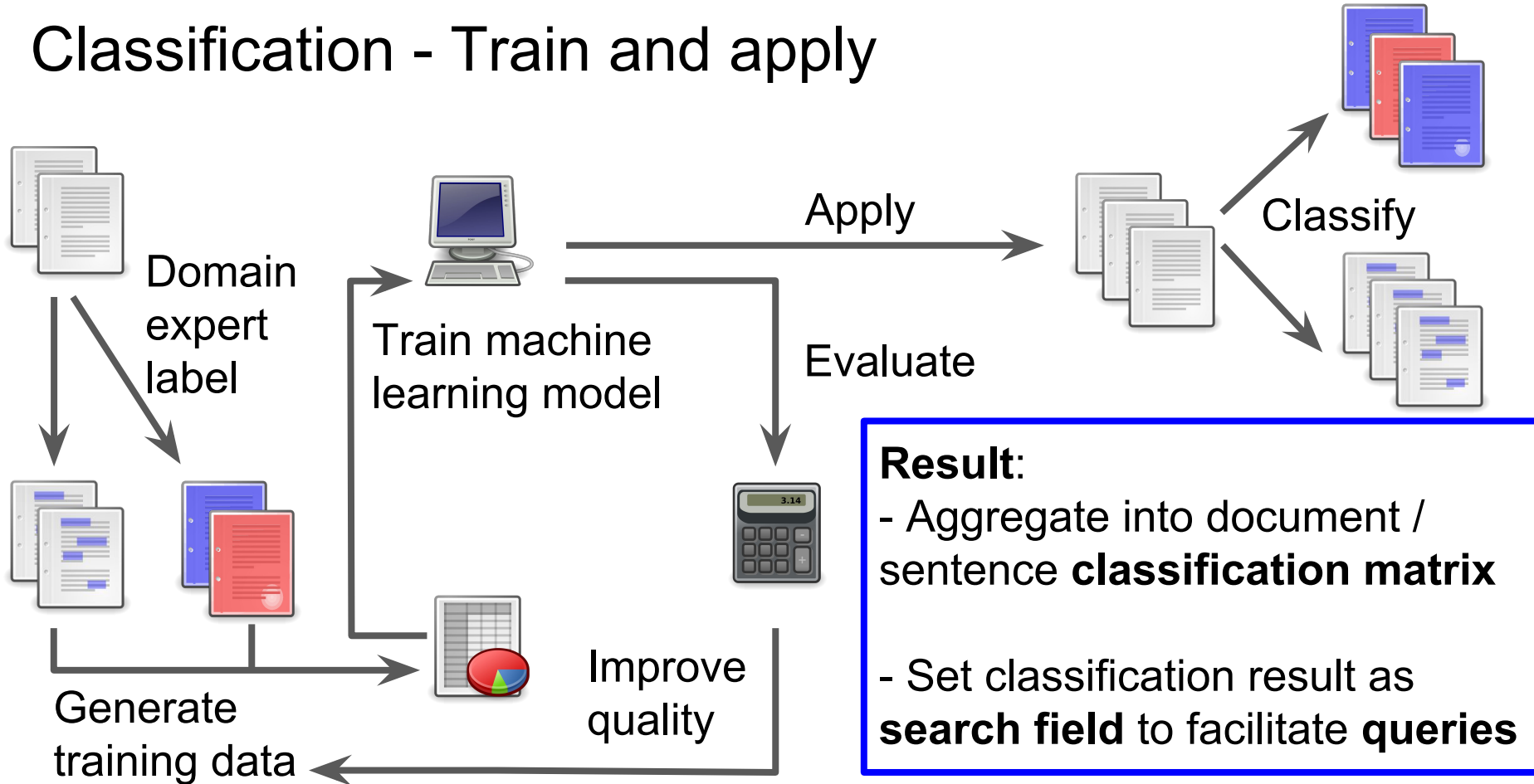


Classification

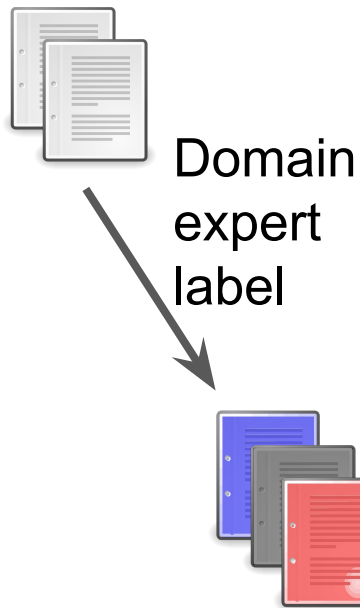
Assign labels to sentences/documents from a corpus



Classification - Train and apply



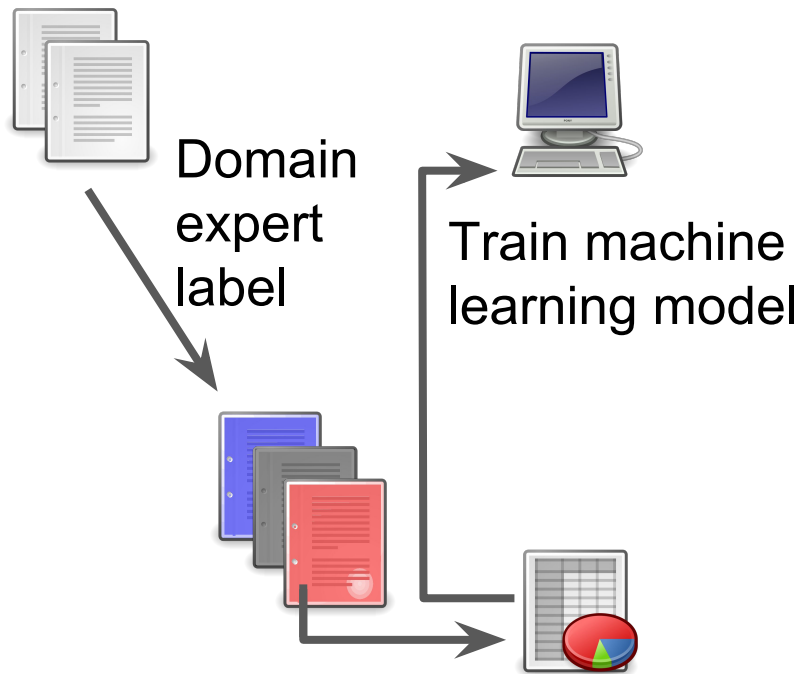
Classification - Example



Ohsumed (medical abstracts from *MeSH* categories) collection subset:

- Musculoskeletal Diseases
- Nutritional and Metabolic Diseases
- Eye Diseases

Classification - Example



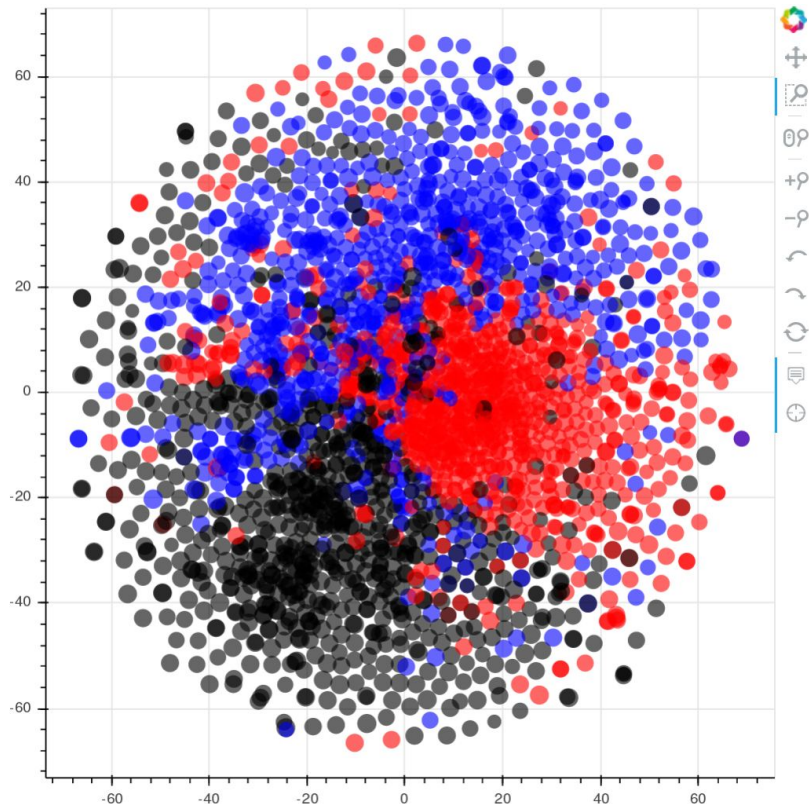
Ohsumed (medical abstracts from *MeSH* categories) collection subset:

- Musculoskeletal Diseases
- Nutritional and Metabolic Diseases
- Eye Diseases

Train machine learning model:

- Data: Abstract text
- Labels: Category

Classification - Example



Ohsumed (medical abstracts from *MeSH* categories) collection subset:

- Musculoskeletal Diseases
- Nutritional and Metabolic Diseases
- Eye Diseases

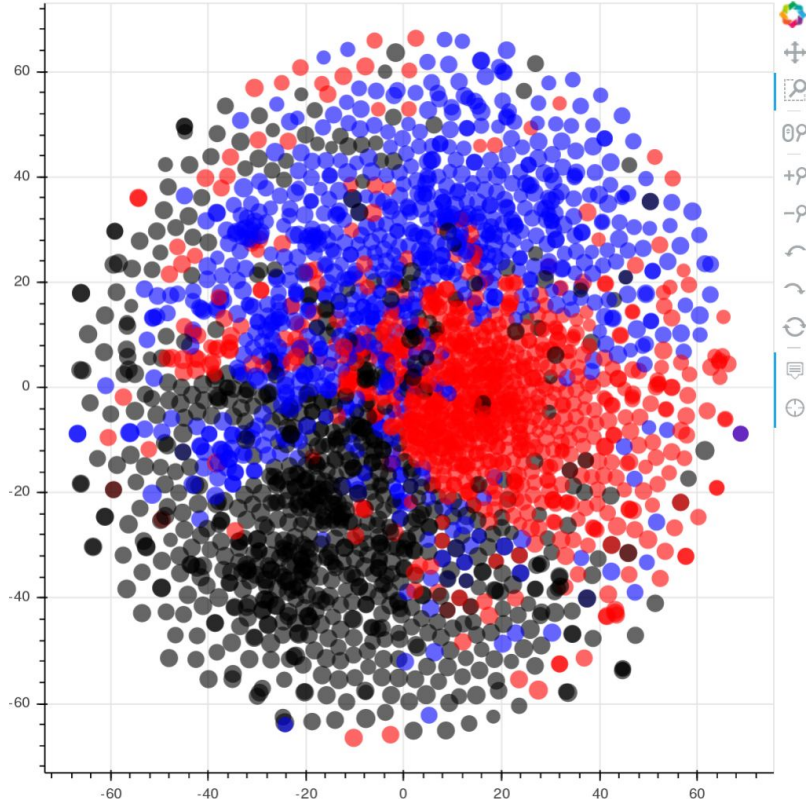
Train machine learning model:

- Data: Abstract text
- Labels: Category

Reduce the dimensionality:

- t-SNE

Classification - Example



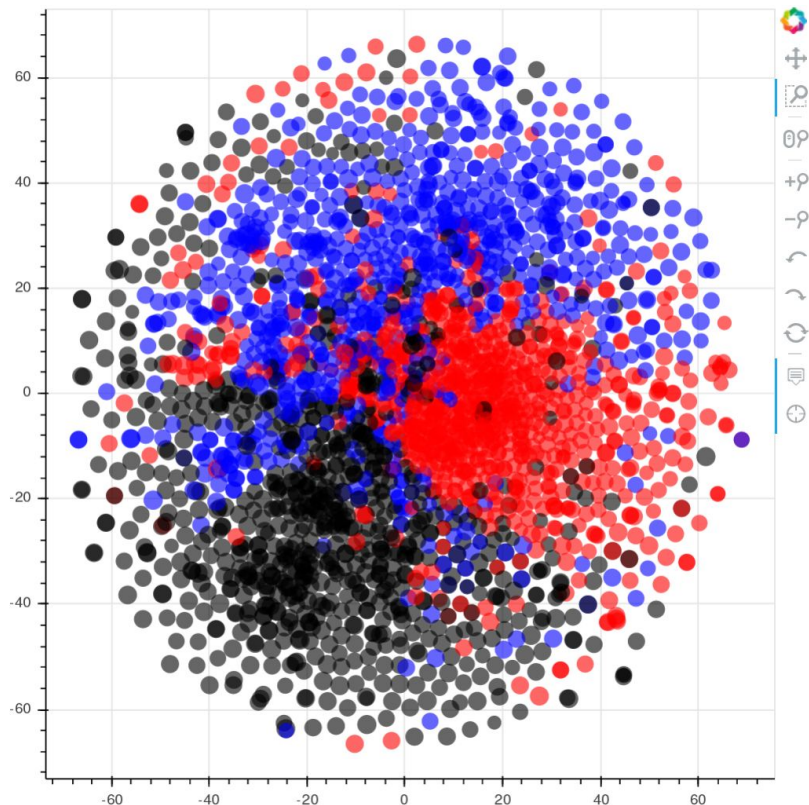
Top predictive features per category:

Musculoskeletal Diseases: bone arthritis
osteomyelitis synovial pain lumbar
myopathy scoliosis joint spine

Nutritional and Metabolic Diseases:
diabetic diabetes insulin obese glucose
malnutrition nutritional coronary renal
cholesterol

Eye Diseases: ocular retinal eye corneal
eyes uveitis graves glaucoma visual
cataract

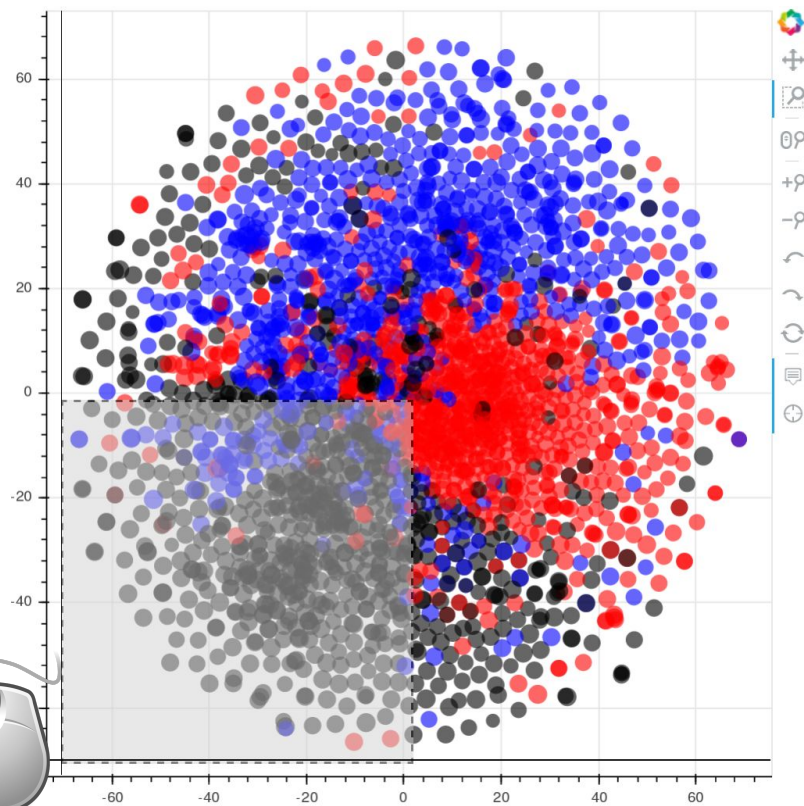
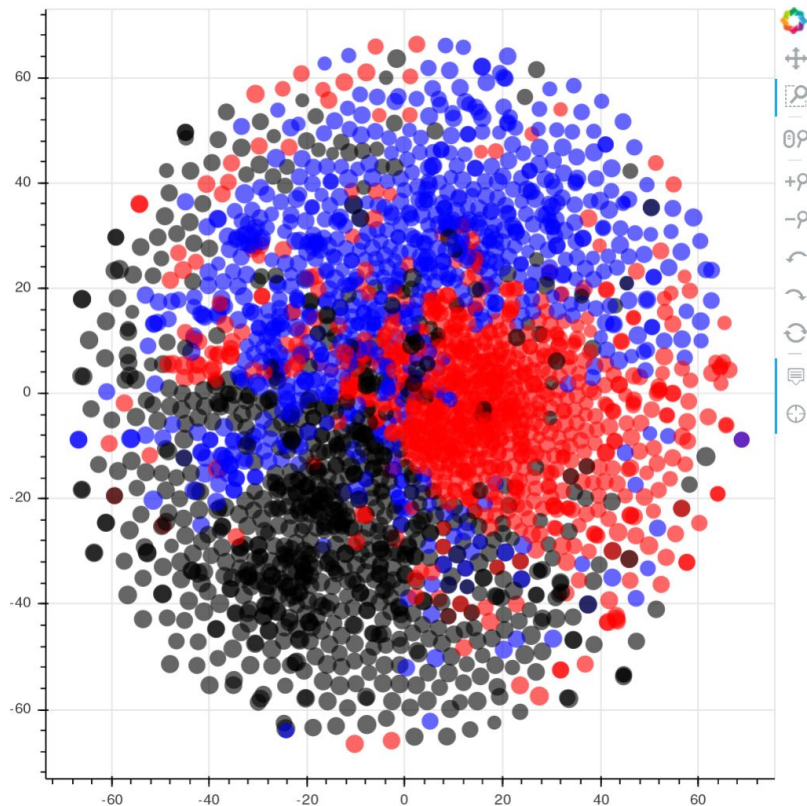
Post-classification - Interactive corpus exploration



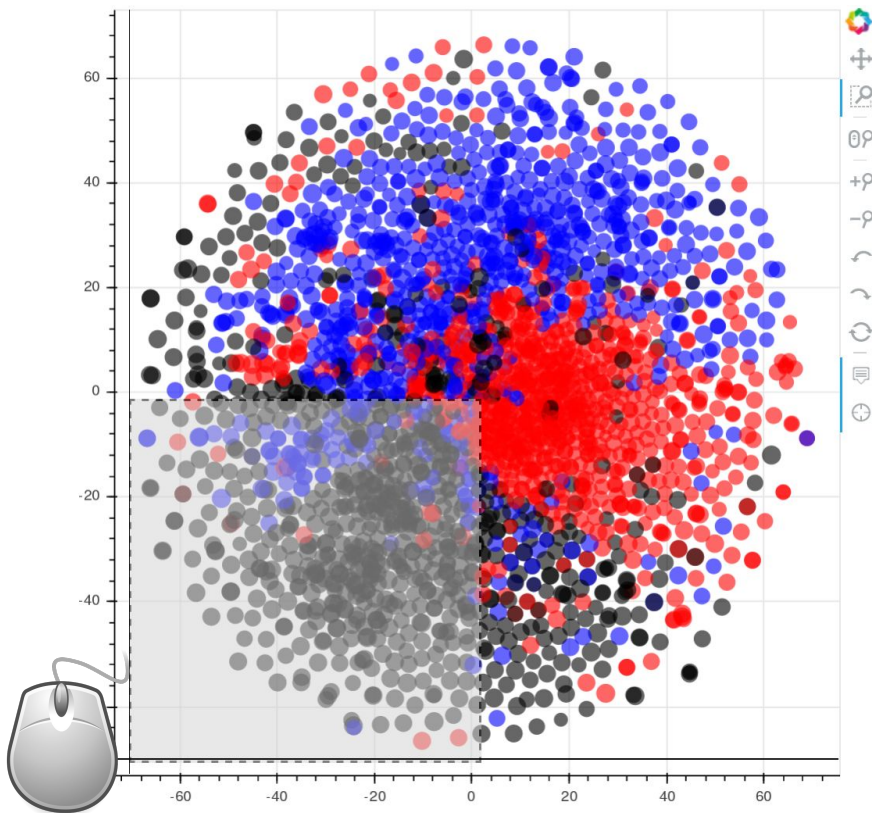
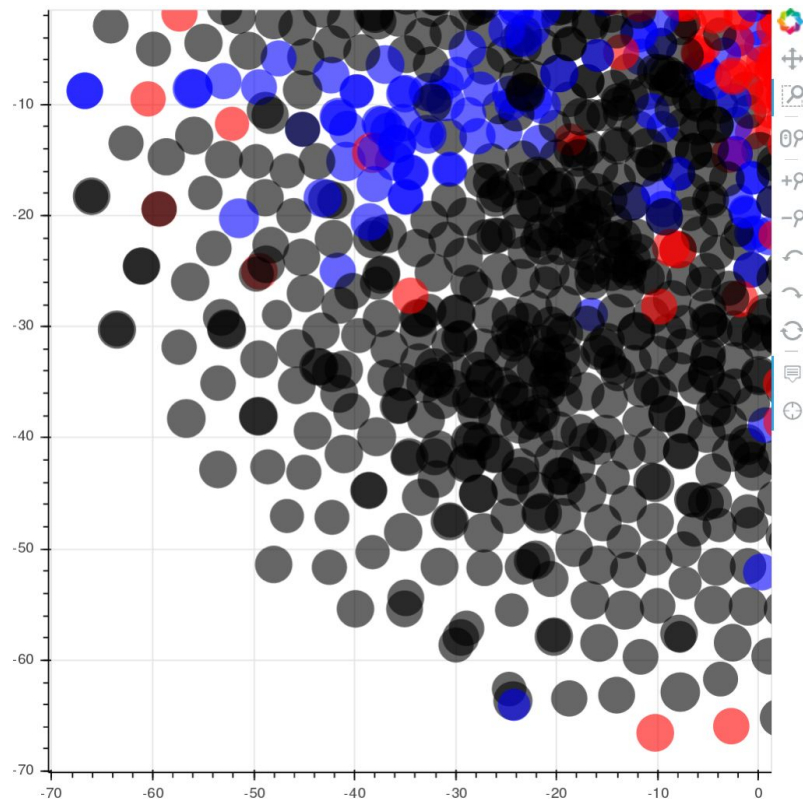
**Ohsumed (medical abstracts from
MeSH categories) collection subset:**

- Musculoskeletal Diseases
- Nutritional and Metabolic Diseases
- Eye Diseases

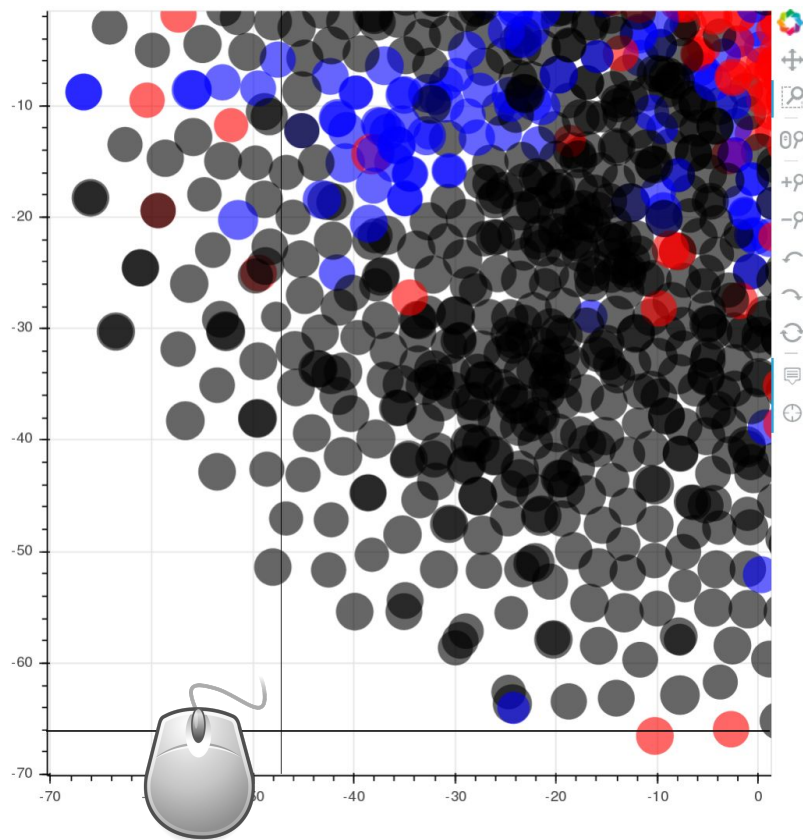
Post-classification - Interactive corpus exploration



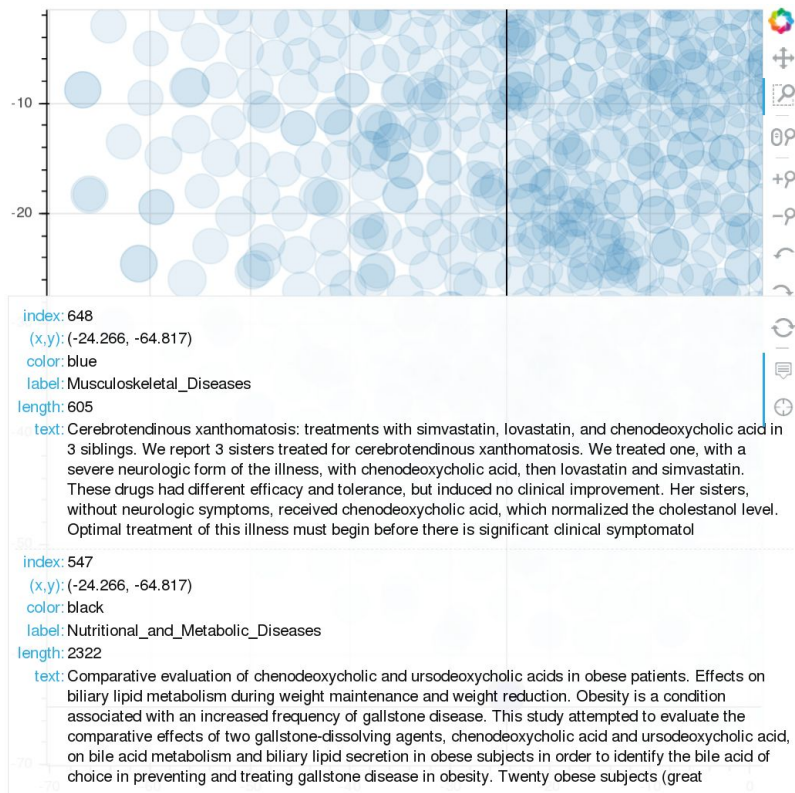
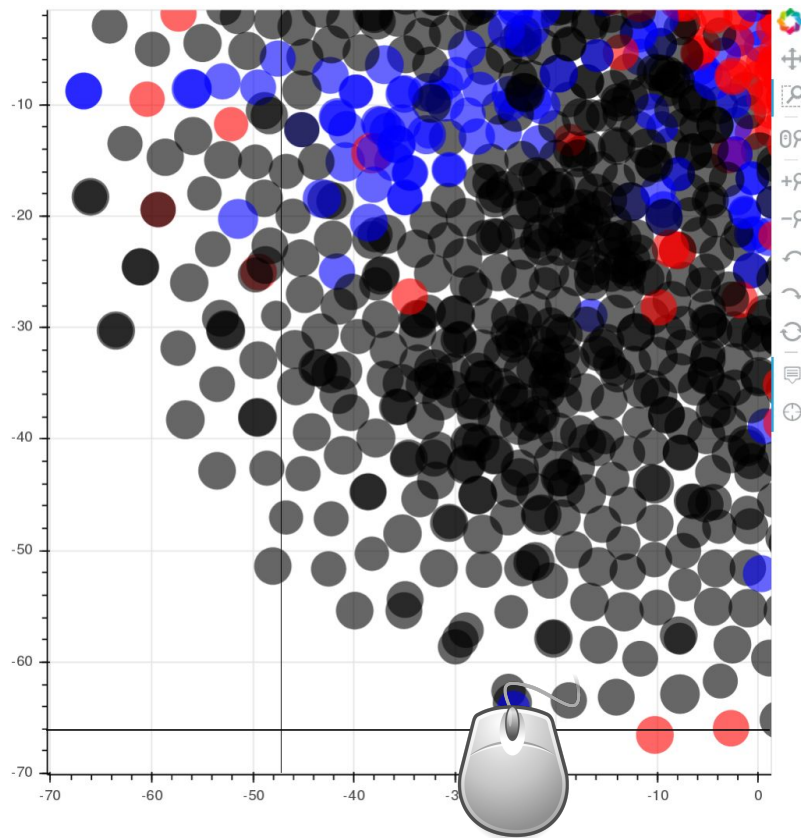
Post-classification - Interactive corpus exploration



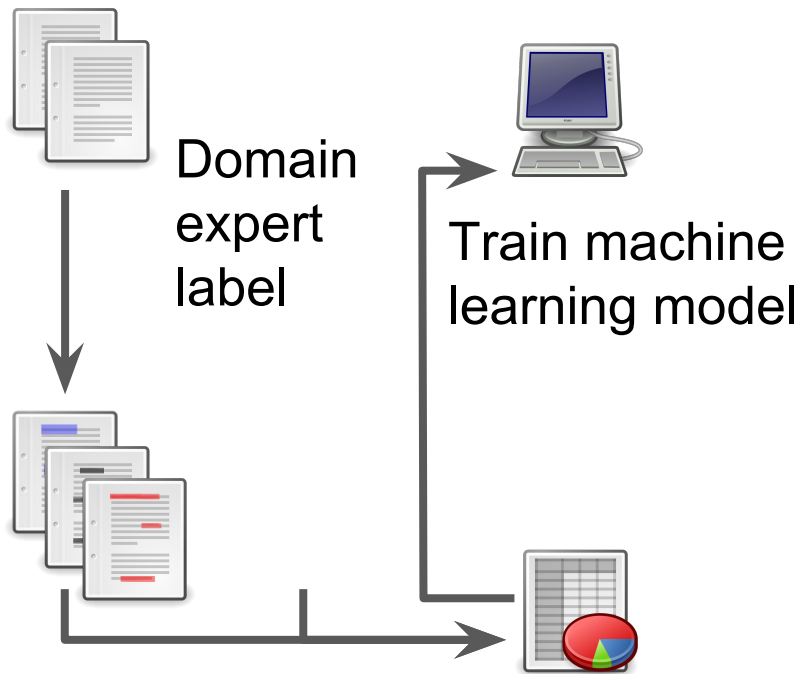
Post-classification - Interactive corpus exploration



Post-classification - Interactive corpus exploration



Classification - Sentence/short text example



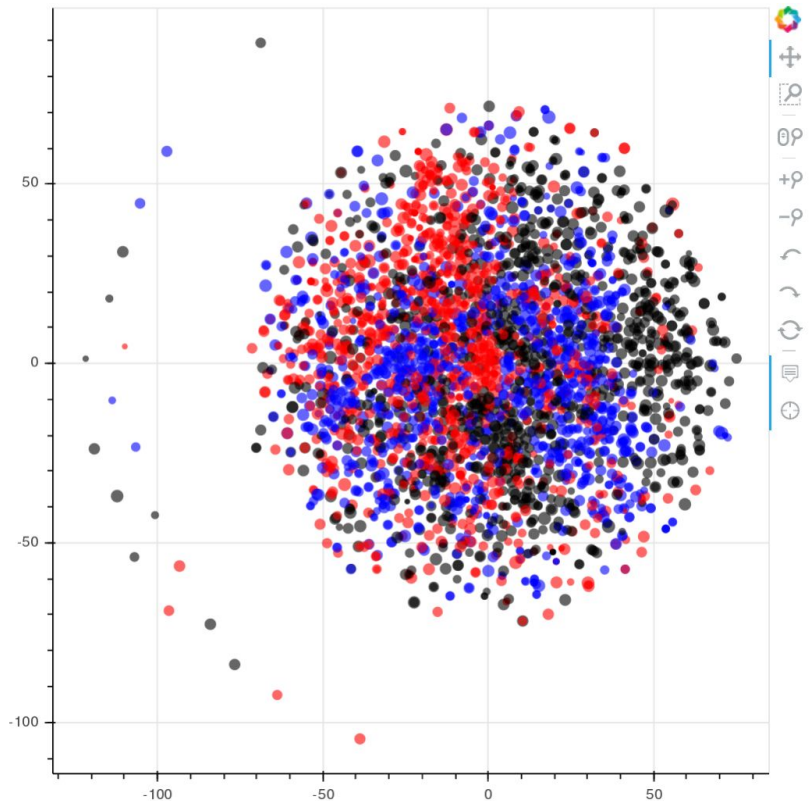
ISEAR corpus (sentences/short text) subset:

- Joy
- Fear
- Sadness

Train machine learning model:

- Data: Sentences/short text
- Labels: Emotion

Post-classification - Sentence/short text exploration



ISEAR corpus (sentences/short text) subset:

- Joy
- Fear
- Sadness

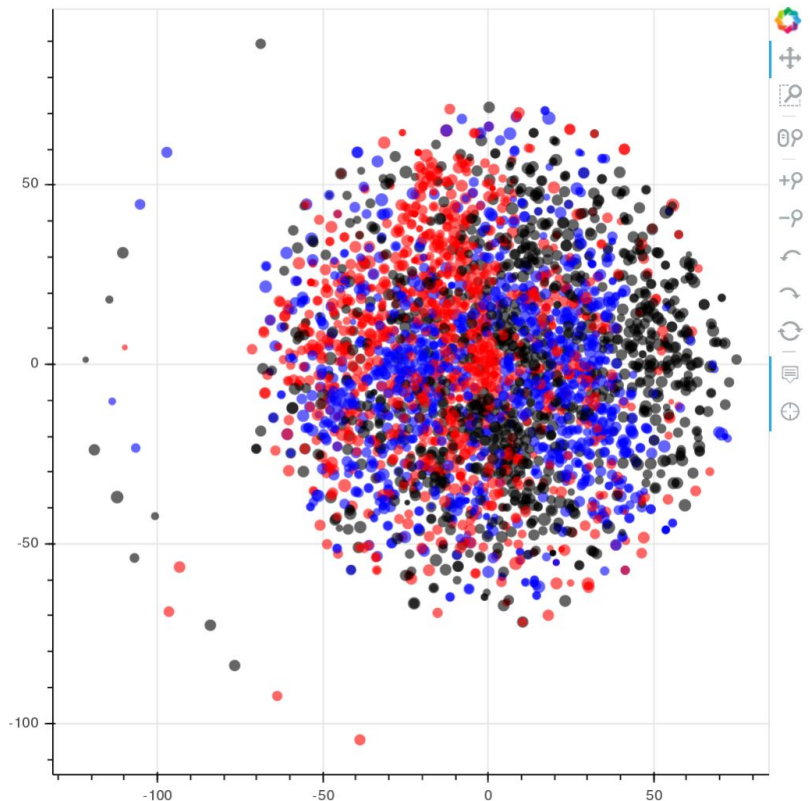
Train machine learning model:

- Data: Sentences/short text
- Labels: Emotion

Reduce the dimensionality:

- t-SNE

Post-classification - Sentence/short text exploration



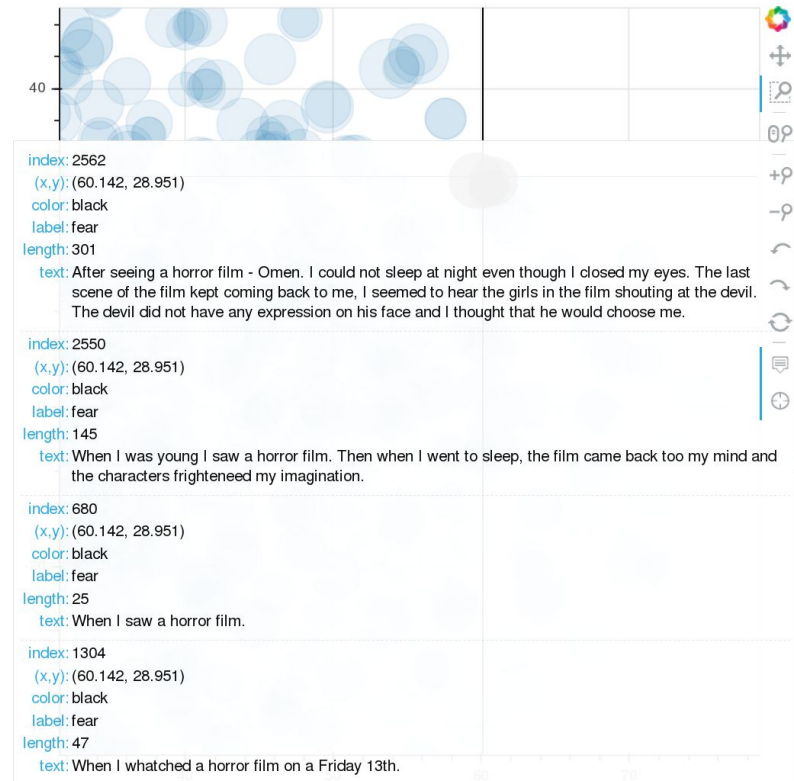
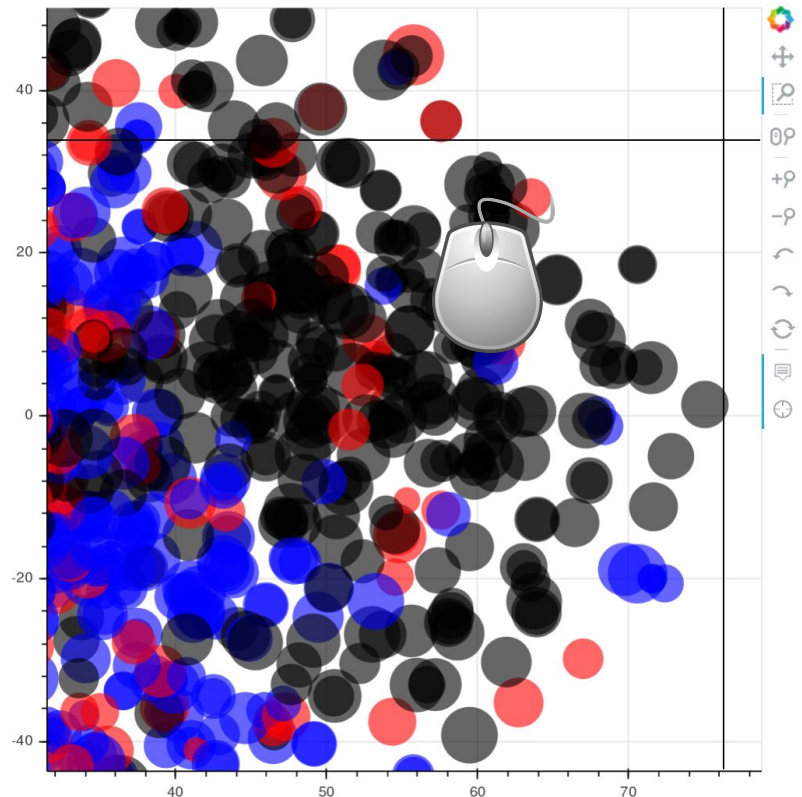
Top predictive features per category:

Joy joy happy passed glad won got
accepted birthday admitted wedding

Fear afraid fear scared night feared dark
threatened frightened friend_love house

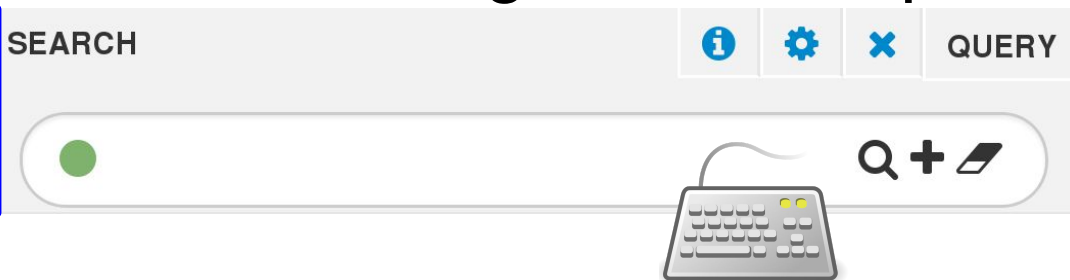
Sadness sad died passed_away
sadness failed death separated left
leave relationship

Post-classification - Sentence/short text exploration



Search and classification - Augmented corpus

Apply classifiers on
project-specific
corpus + search



Apply trained classifiers on corpus

Augment document with classifier outputs

Index augmented corpus in search engine

Explore corpus and shape queries data-driven/interactively

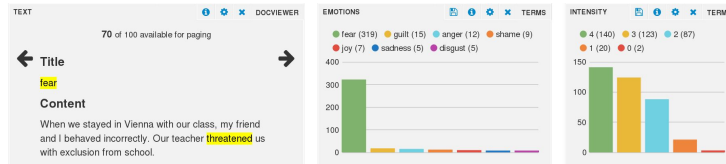
Search and classification - Augmented corpus

Apply classifiers on
project-specific
corpus + search



Application example query:

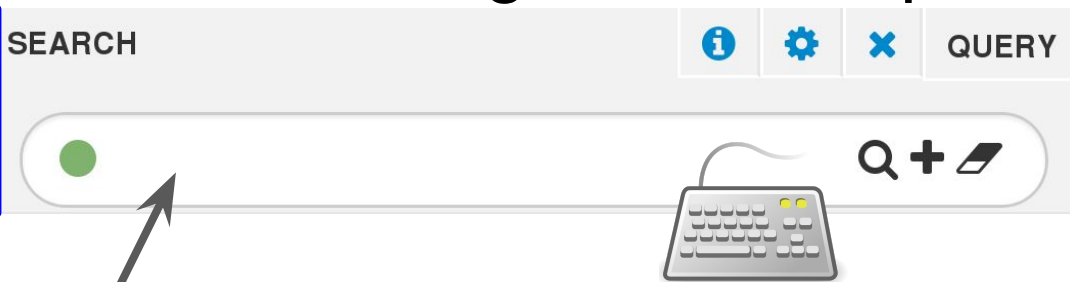
sadness_sentences_count:[2 TO *] **AND** document_type:"A"



Give me all documents with **at least two sad sentences**
AND being of document type **A**

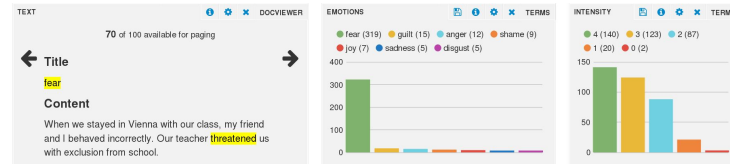
Search and classification - Augmented corpus

Apply classifiers on
project-specific
corpus + search



Application example query:

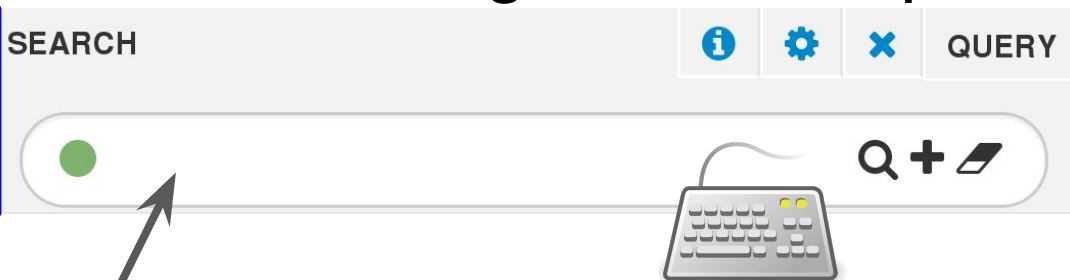
afraid **OR** fear **OR** scared **OR** night **OR** dark **OR** threatened



Give me all documents that contain at least one of my
domain-specific top predictive class features (Fear)

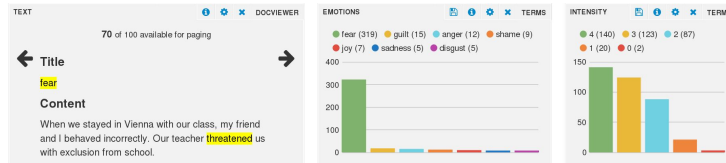
Search and classification - Augmented corpus

Apply classifiers on
project-specific
corpus + search



Application example query:

predicted_label:"class_A" **OR** predicted_label:"class_B"
AND document_date:[2012-01-01 TO 2014-12-31]
AND "some phrase"



Queries involving **date-ranges** are also possible

Summary

Research problems involving **clinical notes** cast as NLP tasks:

NER, information extraction, search, classification, (others ...)

We illustrated **what can be done**, now we would like you to bring us some interesting cases

We can help you with your research

Contact

Thank you for your attention!

We can help you with your research

You are encouraged to get in touch with us now or via email:

Tomasz Oliwa, PhD | toliwa@bsd.uchicago.edu

Brian Furner | bfurner@bsd.uchicago.edu

Center for Research Informatics

Biological Sciences Division

University of Chicago

References 1

- Epic, <http://www.epic.com/>
- Note text modified from example: <http://www.med.unc.edu/medselect/resources/sample-notes/sample-write-up-1> and at the bottom of this page:

Source

Rubin, R. and Strayer, D. Rubin's Pathology. 5th edition. Lippincott Williams and Wilkins, 2008.

- brat rapid annotation tool, <http://brat.nlplab.org/>
- UMLS, <https://www.nlm.nih.gov/research/umls/>
- Apache cTAKES, <http://ctakes.apache.org/>
- Apache Solr, <http://lucene.apache.org/solr/>
- Banana for Solr, <https://github.com/lucidworks/banana>
- Bokeh <https://bokeh.pydata.org/en/latest/>
- scikit-learn and t-SNE: <http://scikit-learn.org> and <https://lvdmaaten.github.io/tsne/>

References 2

Ohsumed dataset obtained from: <http://disi.unitn.it/moschitti/corpora.htm>

See also http://trec.nist.gov/data/t9_filtering.html and <http://trec.nist.gov/data/filtering/README.t9.filtering> for source:

(A) Description of the OHSUMED document collection (files: ohsumed.*)

The OHSUMED test collection is a set of 348,566 references from MEDLINE, the on-line medical information database, consisting of titles and/or abstracts from 270 medical journals over a five-year period (1987-1991). The available fields are title, abstract, MeSH indexing terms, author, source, and publication type. The National Library of Medicine has agreed to make the MEDLINE references in the test database available for experimentation, restricted to the following conditions:

1. The data will not be used in any non-experimental clinical, library, or other setting.
2. Any human users of the data will explicitly be told that the data is incomplete and out-of-date.

The OHSUMED document collection was obtained by William Hersh (hersh@OHSU.EDU) and colleagues for the experiments described in the papers below:

Hersh WR, Buckley C, Leone TJ, Hickam DH, OHSUMED: An interactive retrieval evaluation and new large test collection for research, Proceedings of the 17th Annual ACM SIGIR Conference, 1994, 192-201.

Hersh WR, Hickam DH, Use of a multi-application computer workstation in a clinical setting, Bulletin of the Medical Library Association, 1994, 82: 382-389.

References 3

ISEAR attribution:

Copyright, disclaimer, license, author's website:

<http://www.affective-sciences.org/home/research/materials-and-online-research/research-material/>

License notice on website: All these materials are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.:

<https://creativecommons.org/licenses/by-nc-sa/3.0/>

Title: International Survey On Emotion Antecedents And Reactions (ISEAR)

Short description

Over a period of many years during the 1990s, a large group of psychologists all over the world collected data in the ISEAR project, directed by Klaus R. Scherer and Harald Wallbott. Student respondents, both psychologists and non-psychologists, were asked to report situations in which they had experienced all of 7 major emotions (joy, fear, anger, sadness, disgust, shame, and guilt). In each case, the questions covered the way they had appraised the situation and how they reacted. The final data set thus contained reports on seven emotions each by close to 3000 respondents in 37 countries on all 5 continents.

References

The following publications describe the procedures and report the major patterns of results:

Wallbott, H.G., & Scherer, K. R. (1986). How universal and specific is emotional experience? *Social Science Information*, 24, 763-795.

Matsumoto, D., Kudoh, T., Scherer, K. R., & Wallbott, H.G. (1988). Antecedents of and reactions to emotions in the US and Japan. *Journal of Cross-Cultural Psychology*, 19, 267-286.

Wallbott, H.G., & Scherer, K. R. (1988). Emotion and economic development - Data and speculations concerning the relationships between economic factors and emotional experience. *European Journal of Social Psychology*, 18, 267-273.

Scherer, K. R., & Wallbott, H.G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66, 310-328.

Scherer, K. R. (1997). Profiles of emotion-antecedent appraisal: testing theoretical predictions across cultures. *Cognition and Emotion*, 11, 113-150.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, 73, 902-922.

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin*, 24(7), 769-783.