

There is much more you can do
with News than just reading them

Mouna Kacimi
KRDB Summer Seminar
31 July 2020

News are Important, we all know that...

They
influence the
way we think

They
influence the
way we
behave



Why such diversity?

Different sources



Selective reading

Different versions

Increasing amount of information

How can we solve these issues

- Enrich information sources to have a wider view about news events
- Summarize news to have a quick access to information
- Analyze the effect of News on our behavior

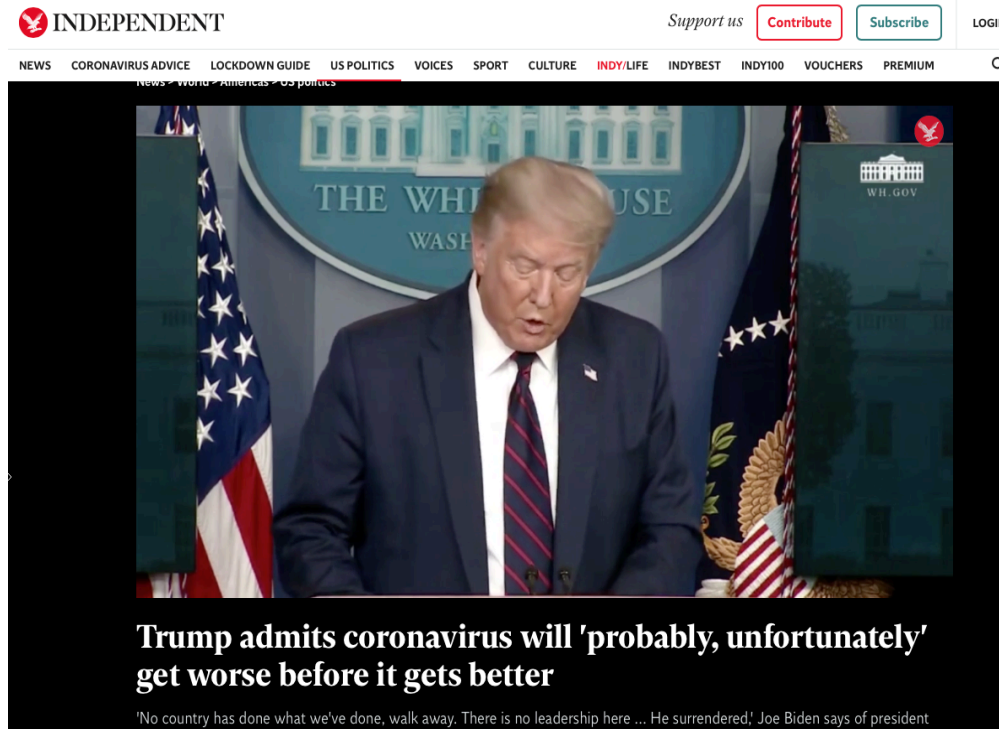
In this talk



News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



*The Independent employs reporters around the world to bring you truly independent journalism. To support us, please consider a **contribution**.*

Donald Trump, after spending months declaring the United States had defeated the **coronavirus** and vowing it soon will "disappear," on Tuesday warned Americans the outbreak that has killed at least 141,000 in the United States is likely to grow more dire.

Published on July 21, 2020

The Article talks about

- Donald Trump warning Americans that coronavirus will get worse
- His previous policy of ignoring the coronavirus problem and pretending that US defeated it.

News Aspects

News
Enrichment

News
Information
Extraction


News Effect
on Stock
Markets

INDEPENDENT

Support us [Contribute](#) [Subscribe](#) [LOGIN](#)

NEWS CORONAVIRUS ADVICE LOCKDOWN GUIDE US POLITICS VOICES SPORT CULTURE **INDY/LIFE** INDYBEST INDY100 VOUCHERS PREMIUM

News > World > Americas > US politics



Trump admits coronavirus will 'probably, unfortunately' get worse before it gets better

'No country has done what we've done, walk away. There is no leadership here ... He surrendered!' Joe Biden says of president

*The Independent employs reporters around the world to bring you truly independent journalism. To support us, please consider a **contribution**.*

Donald Trump, after spending months declaring the United States had defeated the **coronavirus** and vowing it soon will "disappear," on Tuesday warned Americans the outbreak that has killed at least 141,000 in the United States is likely to grow more dire.

Comments

Share your thoughts and debate the big issues

[Learn more](#)

1 Independent Premium comment

75 open comments

Join the discussion

Newest first

Oldest first

Most liked

 **Chainsaw1**

1 week ago

Unbelievably there are people who are still defending Trump. This is for you. Denial is a defense mechanism proposed by Anna Freud which involves a refusal to accept reality, thus blocking external events from awareness. If a situation is just too much to handle,

Impossibility of containing any virus

 **Jellytussle**

1 week ago

Dr. Fauci Statements about COVID-19

As Nives is ignorant of popular culture - doesn't understand references to

Responsibility of China and WHO

Reply

 6  5 

Russian Disinformation

with childish insults. A poor attempt at trolling old man.

 4  0 

Latent Aspects

 0  4 

Problem

Find news articles that match user interests

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

User

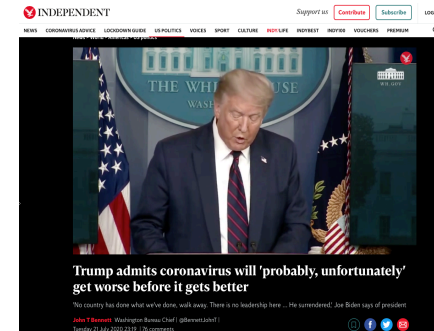


described by
his comments

Latent Aspects

- China & WHO
- Russian disinformation

News Article



described by
its content

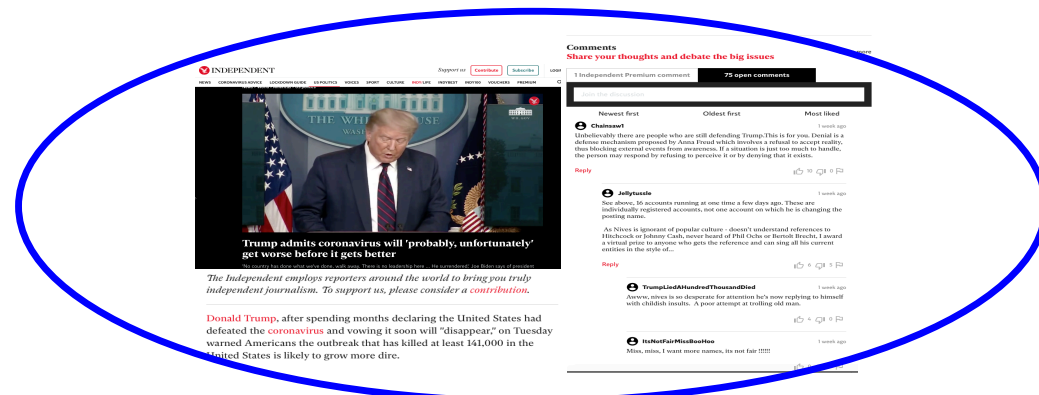
News Aspects

- Donald Trump warning Americans

No matching

Fill the gap between Comments & Articles

- Enrich news articles with comments



Content of a
News Article



Content provided
by each user

Both Articles and Users are described by News Aspects and Latent Aspects

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

Challenges

- How to enrich news articles with comments
 - Use all comments? No because they can be noisy
 - Select a subset of comments, but how?
- How aspects can be modeled and how to extract them from comments

**News
Enrichment**

**News
Information
Extraction**

**News Effect
on Stock
Markets**

News
Enrichment

News
Information
Extraction

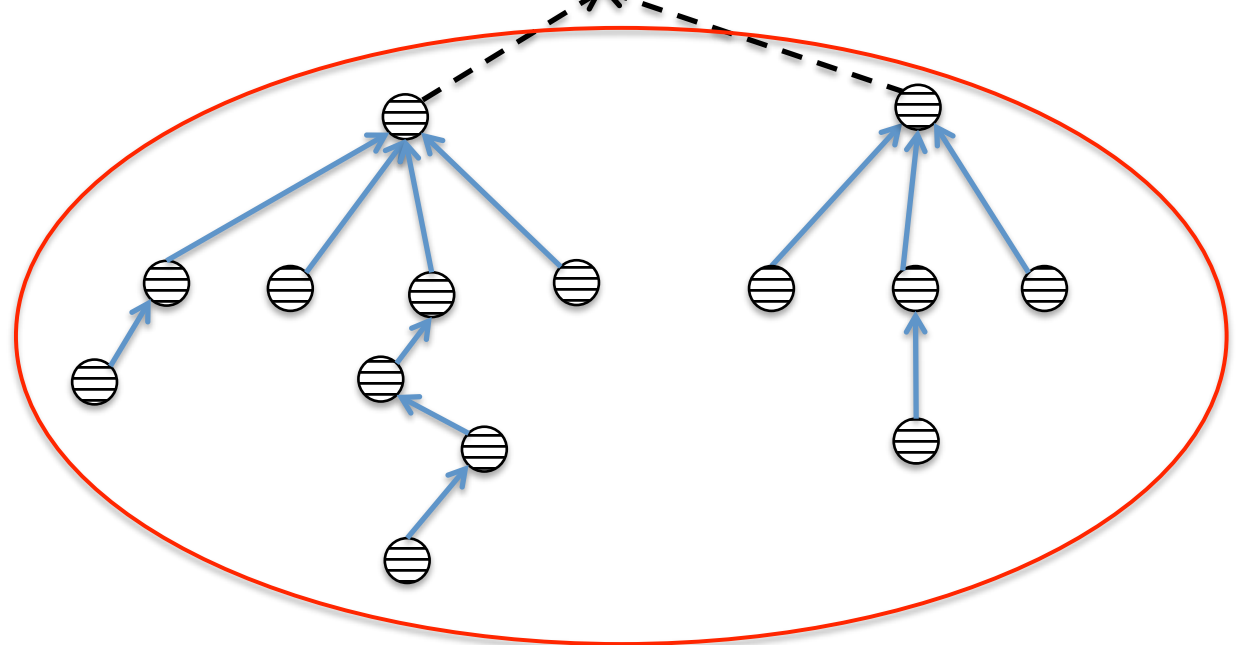
News Effect
on Stock
Markets

Comments Selection

- -> Comment the news article
- Answer Comment



We use Page Rank
Algorithm on the Tree
Structure of Comments



$$h(C) = (1 - m) + m \left(\frac{h(T_1)}{L(T_1)} + \dots + \frac{h(T_n)}{L(T_n)} \right)$$

Comments Diversification

- Comments can be redundant (same idea)



- Two comments can describe the same aspects with 2 different writing style, so they are wrongly detected as diverse
- Use our opinion diversification algorithm proposed in [1]

[1]Mouna Kacimi, Johann Gamper, *Diversifying search results of controversial queries*. CIKM2011: 93-98

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

Aspect Extraction

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

- Extract all words (tfidf)
- Construct phrases from the set of all words
- Phrases are constructed with an incrementing size 1,2,3,...
- Select meaningful phrases based on the correlation of their contained words (PMI)

Example of Generated Aspects

Trump, Clinton, Vote, Supporters, Business, Election, Establishment, President, crazies, Stupid Trump, Good Trump, Crazy Trump, Donald Trump vote, Donald Trump supporters, Donald Trump business, Donald Trump Establishment, Donald Trump president, Donald Trump crazies, Donald Trump trickster, Trump international policy

Dataset Statistics

| | Seeds from CNN | | Seeds from The Telegraph | |
|--------------------|----------------------------|--------------|--------------------------|--------------|
| | #articles | #comments | #articles | #comments |
| <i>CNN</i> | 41, 245 | 12, 056, 789 | 665 | 874, 879 |
| <i>Telegraph</i> | 1, 908 | 1, 257, 645 | 56, 527 | 10, 704, 741 |
| <i>Independent</i> | 1, 412 | 987, 437 | 7, 999 | 1, 608, 665 |
| <i>Al-Jazeera</i> | 801 | 102, 254 | 451 | 62, 835 |
| | Seeds from The INDEPENDENT | | Seeds from Al-Jazeera | |
| | #articles | #comments | #articles | #comments |
| <i>CNN</i> | 528 | 421, 542 | 2, 233 | 1, 652, 875 |
| <i>Telegraph</i> | 23, 272 | 6, 710, 580 | 1, 126 | 894, 710 |
| <i>Independent</i> | 27, 012 | 2, 985, 412 | 394 | 54, 760 |
| <i>Al-Jazeera</i> | 303 | 48, 058 | 9, 313 | 531, 452 |

The **INDEPENDENT**

164 users

The Telegraph

180 users



151 users

ALJAZEERA



150 users

**News
Enrichment**

**News
Information
Extraction**

**News Effect
on Stock
Markets**

Overall Precision of Our Approach

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

| | P@1 | P@3 | P@5 | P@10 | P@20 |
|--------------------|--------------|--------------|--------------|--------------|--------------|
| NoEnrich | 0.424 | 0.494 | 0.481 | 0.513 | 0.540 |
| AllComments | 0.393 | 0.474 | 0.445 | 0.453 | 0.503 |
| Authority | 0.454 | 0.535 | 0.530 | 0.550 | 0.565 |
| Diversity | 0.575 | 0.646 | 0.654 | 0.640 | 0.607 |

[2] Youssef Meguebli, Mouna Kacimi, Bic-Lien Doan, Fabrice Popinau. "Towards better news article recommendation". World Wide Web (2017)

What News Articles Can Tell?

News
Enrichment

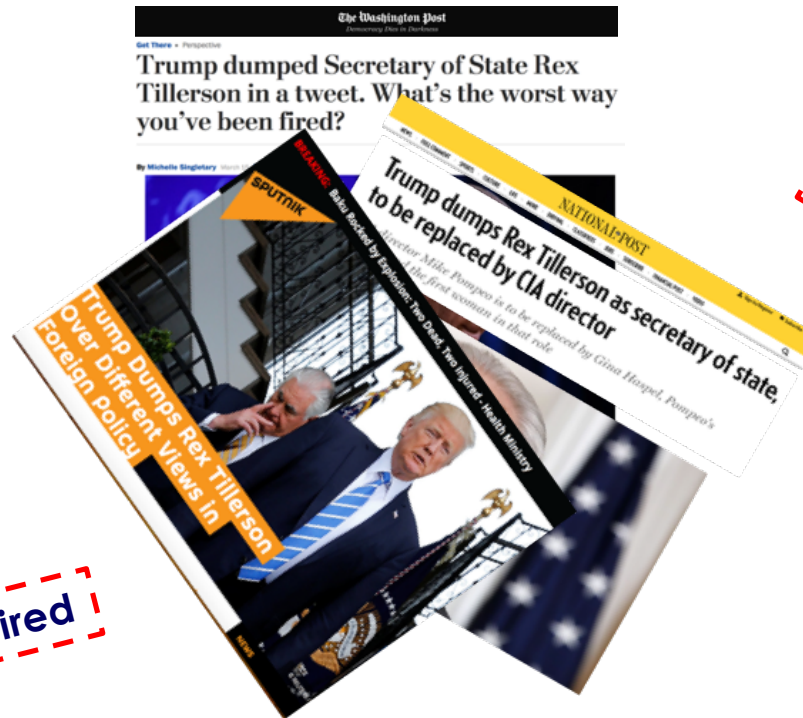
News
Information
Extraction

News Effect
on Stock
Markets

How the firing
was announced

Why Tillerson was fired

What is next?

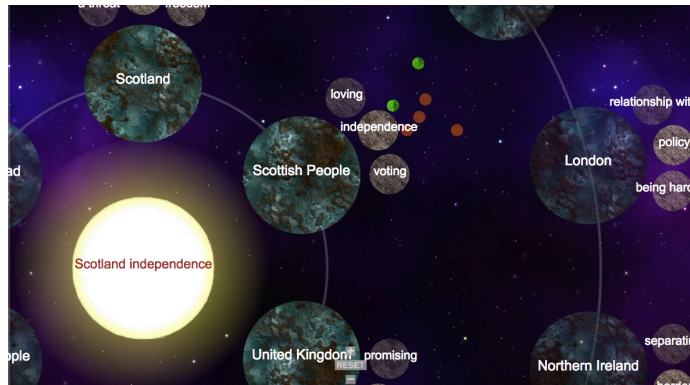


News events are reported by various sources
An event can have several **facets**:
reasons, consequences, time, location...

Why News Event Facets are important?

They enhance

Information Representation



[3] Radityo Eko Prasajo, Mouna Kacimi, Werner Nutt: "Entity and Aspect Extraction for Organizing News Comments". CIKM 2015

[4] Demo <http://orcaestra.inf.unibz.it>

Summarization

- Information aggregation is facilitated
- Fluency can be improved
- Personalized

[5] Radityo Eko Prasajo, Mouna Kacimi, Werner Nutt: "Entity and Aspect Extraction for Organizing News Comments". ESWC2018

[6] Radityo Eko Prasajo, Mouna Kacimi, Werner Nutt: "News: News Event Walker and Summarizer". SIGMOD 2019. (Demo Paper).

News
Enrichment

News
Information
Extraction

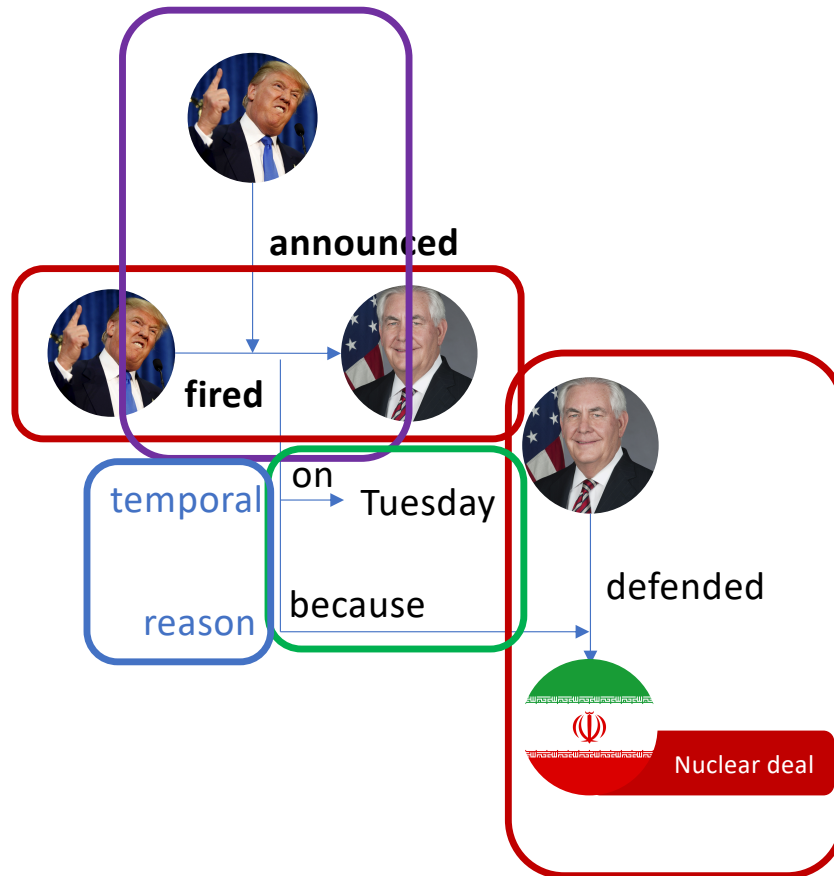
News Effect
on Stock
Markets

Goal: Extract Triples & Facets

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



Triples

Nested Triples

Facets, pointing to

- simple objects
- other triples

Facet labels

President Donald Trump announced that he had fired Secretary of State Rex Tillerson on Tuesday.

Trump fired Mr. Tillerson because he had defended the Iran Nuclear Deal.

Limitations of the State of the Art

- Open Information Extraction Tools (e.g., ClausIE, OLLIE)

- Missing n-tuples,
- coarse-grained argument

Trump announced that he had fired Mr. Tillerson on Tuesday because he had defended the Iran Nuclear Deal.

- Redundancies
- No nested triples
- Do not label any argument

- Semantic Role Labeling (SRL) (e.g., Illinois , SEMAFOR)

- Label some facets

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

StuffIE: Fine-grained Extraction

US President Donald Trump announced that he had fired Secretary of State Rex Tillerson on Tuesday because he had defended the Iran Nuclear Deal.

subject predicate object

1. Donald Trump; announced; _;
 • that; #2; **(Other/Details)** → facet label

2. he<ref#1s>; had fired; Rex Tillerson;
facets { • on; Tuesday; **(Temporal)** } facet labels
 • because; #3; **(Reason)** }

3. He<ref#2o>; defended; the Iran Nuclear Deal;

4. Donald Trump<ref#1s>; <is>; President;

5. Rex Tillerson<ref#1s>; <is>; Secretary of State;

} implicit
triples

News
Enrichment

News
Information
Extraction

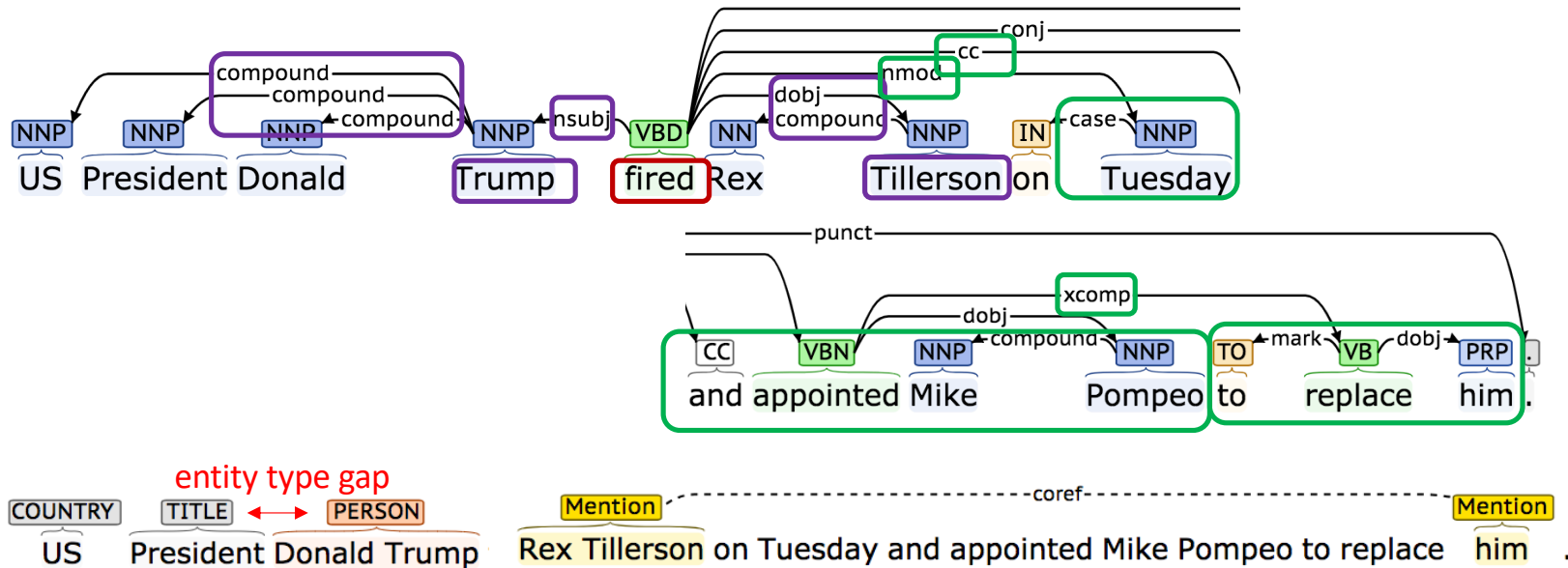
News Effect
on Stock
Markets

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

Triple & Facet Extraction



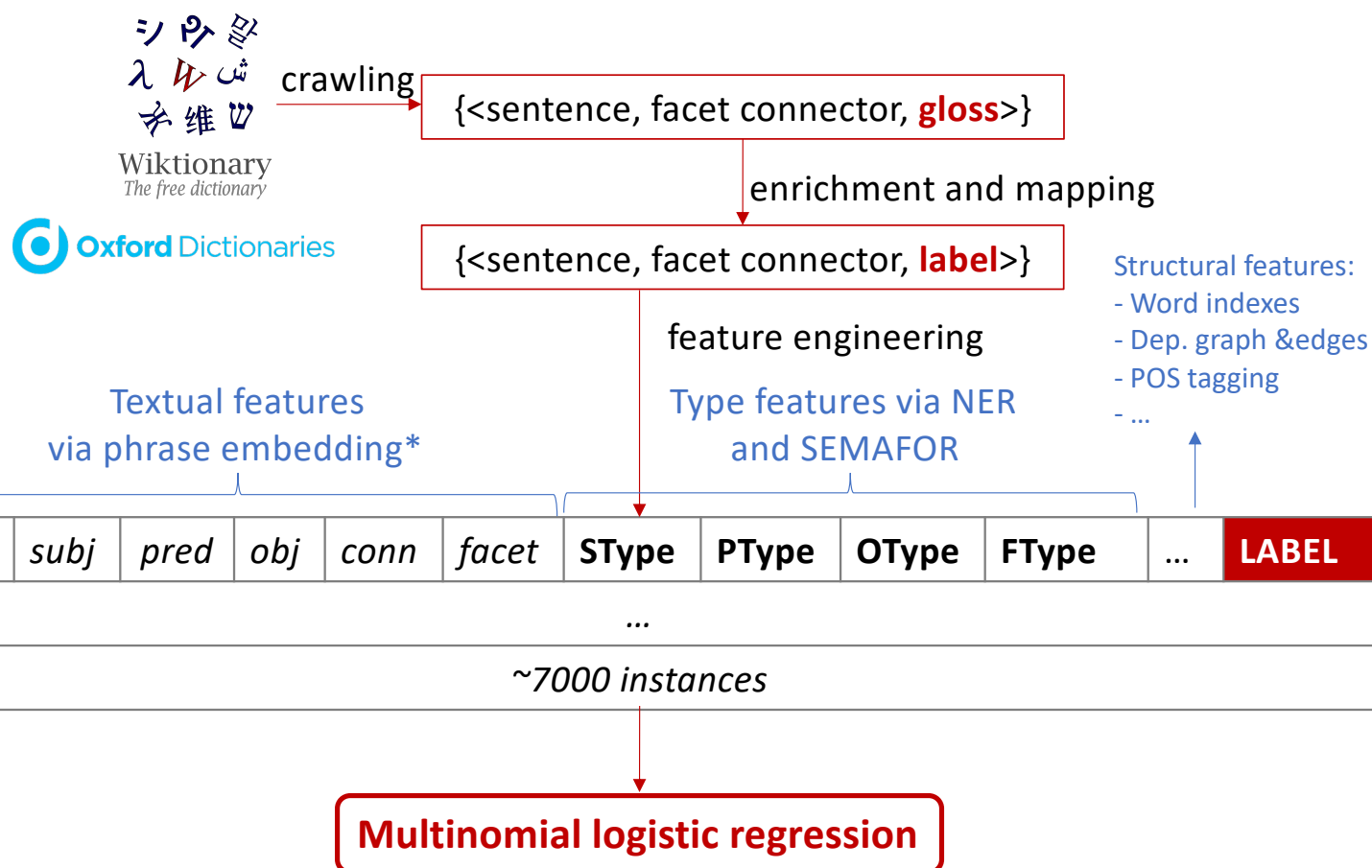
1. Donald Trump; fired; Rex Tillerson;
on; Tuesday;
and appointed; Mike Pompeo;
to replace; him <ref#1o>;
2. Donald Trump<ref#1s>; <is>; US President

Learning Model for Facet Labeling:

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

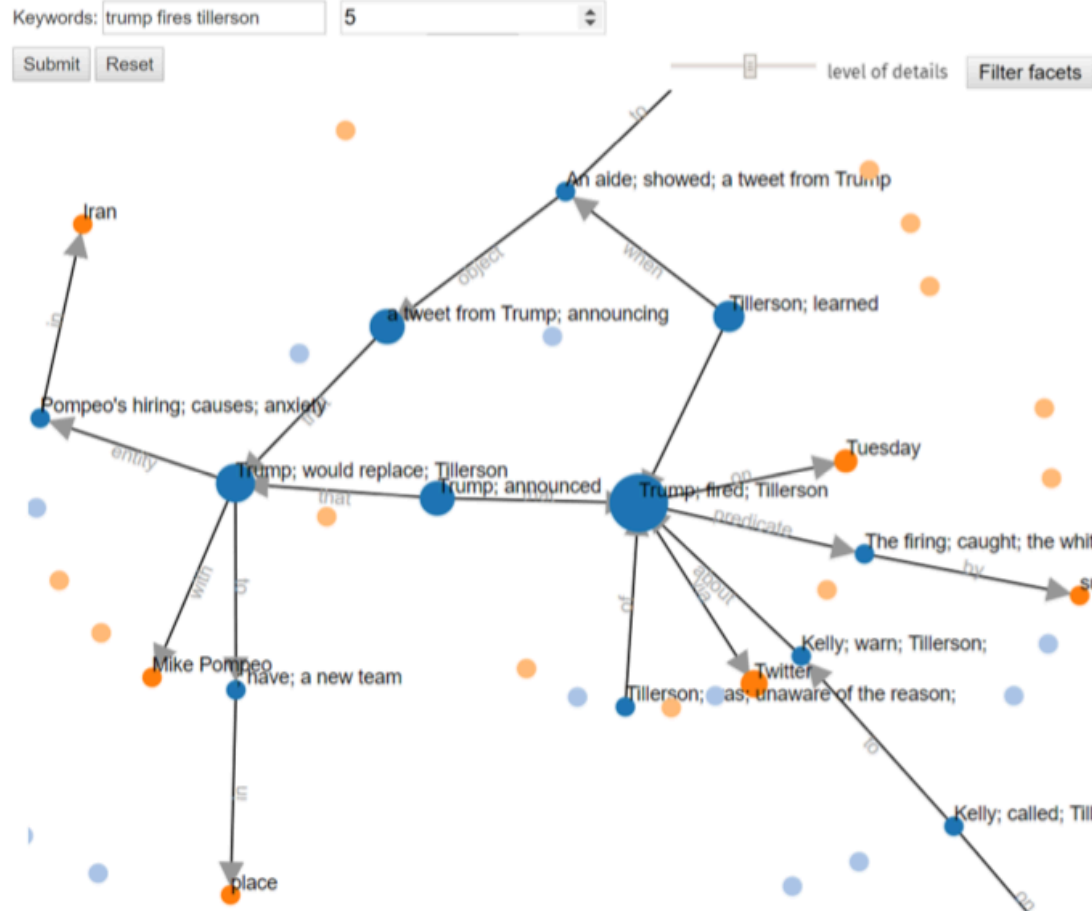


News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

NEWS Event Walker and Summarizer



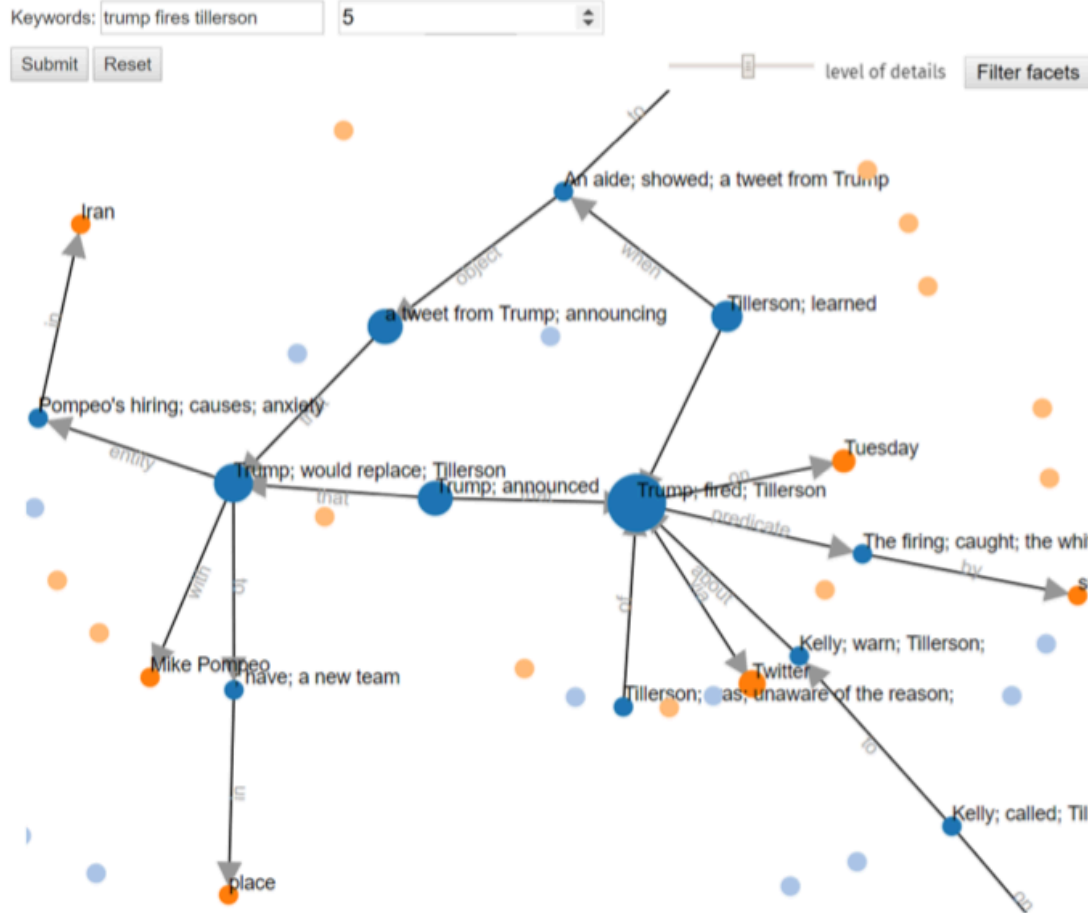
- Compact graph representation
- Entity coreference
- Predicate Coreference

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

NEWS Event Walker and Summarizer



Summary:

US President Donald Trump announced that he had fired Secretary of State Rex Tillerson on Tuesday morning via Twitter and that he would replace him with CIA Director Mike Pompeo to have a new team in place. Mr. Tillerson learned he had been fired when a top aide showed him a tweet from Mr. Trump announcing the change. He was unaware of the reason of his firing. White House chief of staff John F. Kelly called Tillerson on Friday to warn him about the firing.

Summary Generation

- Rank facts
- Specify Depth D and number of facets F
- Proceed with Greedy Selection

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

NEWS Event Walker and Summarizer

Keywords: trump fires tillerson

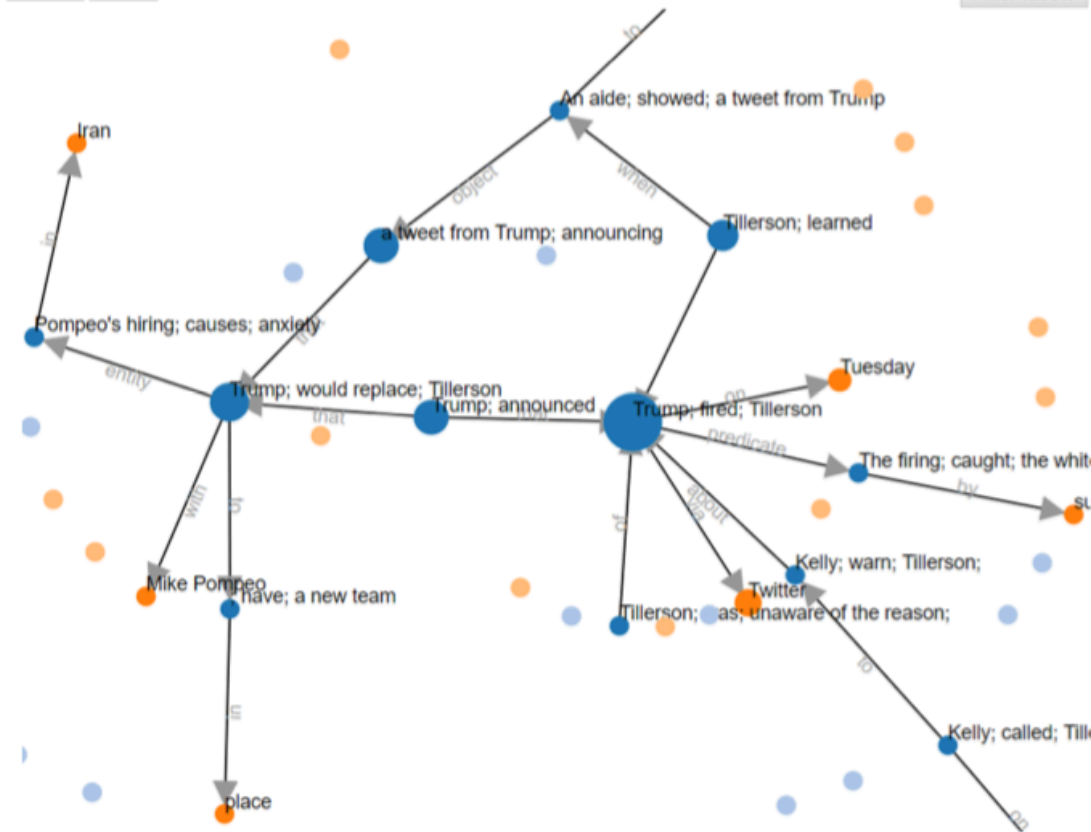
5

Submit

Reset

level of details

Filter facets



Summary:

US President Donald Trump announced that he had fired Secretary of State Rex Tillerson on Tuesday morning via Twitter and that he would replace him with CIA Director Mike Pompeo to have a new team in place. Mr. Tillerson learned he had been fired when a top aide showed him a tweet from Mr. Trump announcing the change. He was unaware of the reason of his firing. White House chief of staff John F. Kelly called Tillerson on Friday to warn him about the firing.

Summary Generation

- Follow the order of facts in the original text
- Group facts with the same predecessor in the same sentence
- Use coreference for reduce redundancy.

Experiments

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

| Method | R-1 | R-2 | R-SU4 | PYRAMID | | LQ |
|--------------|--------------|--------------|--------------|-------------|-------------|----------------|
| | <i>F1</i> | <i>F1</i> | <i>F1</i> | $T_{0.6}$ | $T_{0.65}$ | <i>AvgLL</i> |
| Peng li (RI) | 29.31 | 5.77 | 10.01 | 65.6 | 58.9 | -185.07 |
| Bing et al., | n/a | 11.70 | 14.80 | 90.5 | 79.3 | n/a |
| GRANULA | 43.17 | 12.14 | 16.52 | 90.5 | 79.6 | -144.55 |

TABLE IV

COMPARISON WITH THE BASELINES OVER TAC'11

| Method | R-1 | R-2 | R-SU4 | PYRAMID | | LQ |
|---------|--------------|--------------|--------------------------|-------------|-------------|----------------|
| | <i>F1</i> | <i>F1</i> | <i>F1</i> | $T_{0.6}$ | $T_{0.65}$ | <i>AvgLL</i> |
| 17 | 43.21 | 12.99 | 15.97 | 86.0 | 74.1 | -145.08 |
| 43 | 43.82 | 13.45 | 16.52 | 87.5 | 75.8 | -140.30 |
| 22 | 42.79 | 11.45 | 15.92 | 87.8 | 77.5 | -144.82 |
| GRANULA | 43.17 | 12.14 | 16.52⁷ | 90.5 | 79.6 | -144.55 |

TABLE V

COMPARISON WITH THE TOP-3 EXTRACTIVE METHODS OF TAC'11

How News can affect our trading behavior in stock markets?

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

What type of data we deal with?

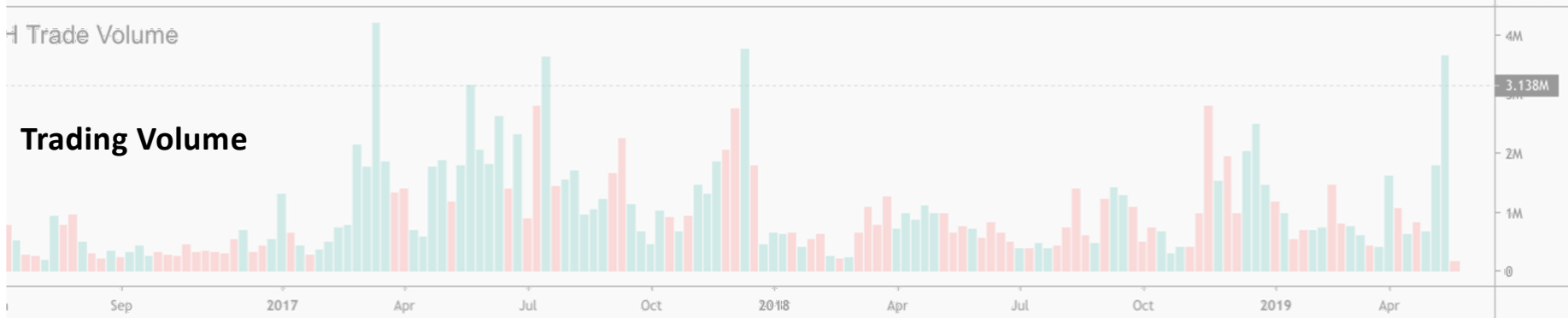
published on TradingView.com, May 20, 2019 15:01:06 UTC

ETHBTC:COINBASE:ETHUSD, 1W 247.28161 ▼ -15.84022 (-5.33%) O:260.47160 H:260.86161 L:239.03105 C:247.28161

Prices



Trade Volume



TradingView

Problem

Text information

Time Series Data

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



How to combine both types of data to predict stock prices?

Challenges

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

News can be many, not all useful for predicting stock prices

What types of News we should look at?
Specific to the stock market of interest
or general news ?

What part of the text we should focus on? Title, keywords, topic, actions, entities, all the text?

Which news have short or long-lasting effects?

Text information



News
Enrichment

News
Information
Extraction

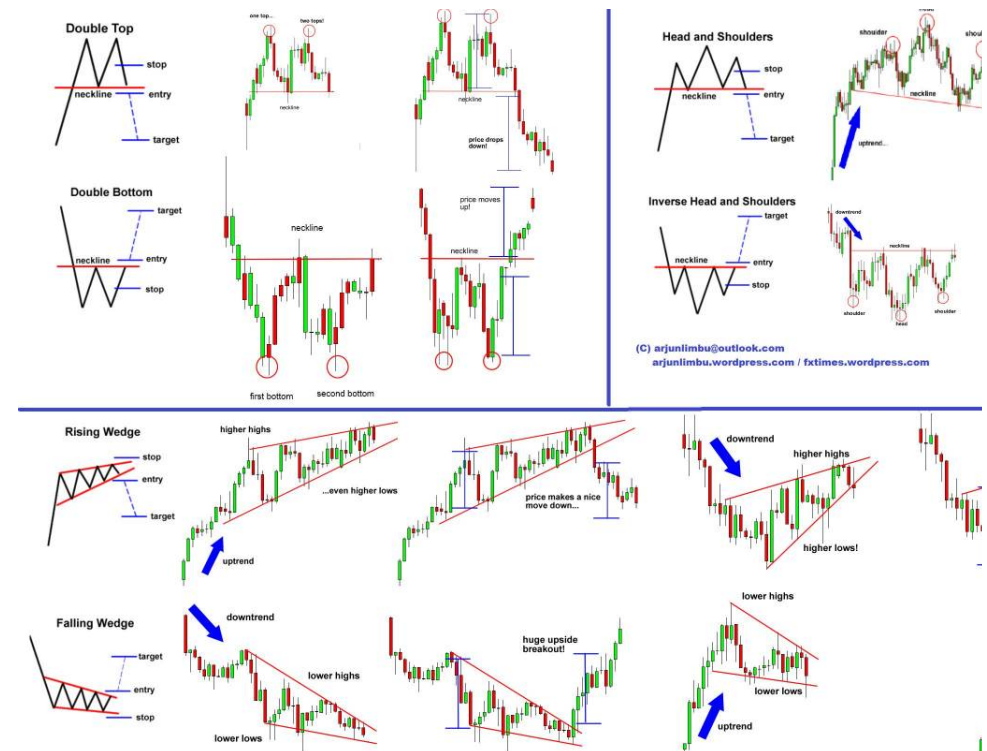
News Effect
on Stock
Markets

Proposal: redefine what a news event

Event: any piece of news that makes price change on a specific stock market



Single prices



Patterns

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

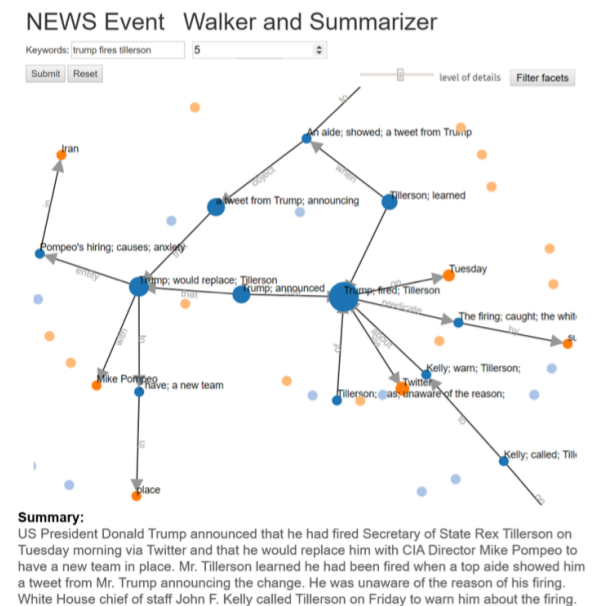
Proposal: exploit our multi-granular summarizer

Summarize a set of news articles about the same event in one single text

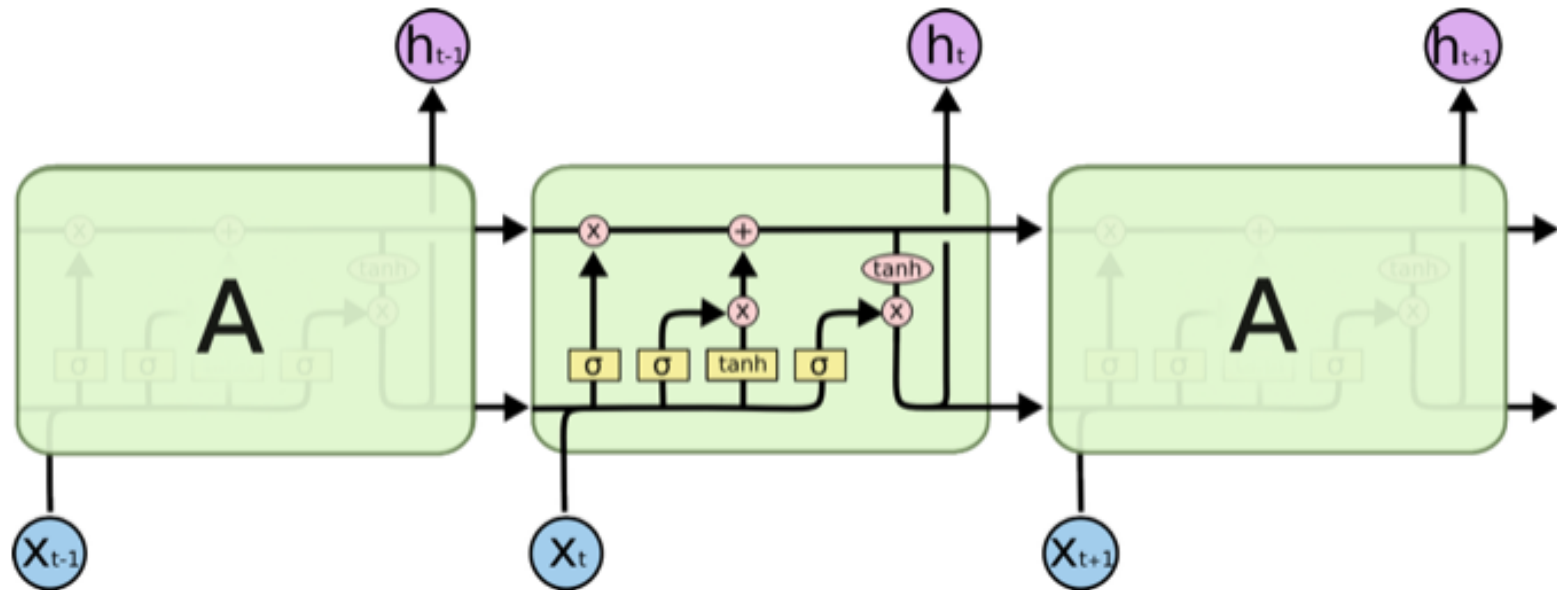
Produce summaries with different granularities ranking from the general topic detailed facts and use content for feature extraction



Stuffie



Proposal: use LSTM networks for stock price prediction



- Long short-term memory RNN Networks
- Typically used for times series data
- Avoid long term dependency problem

In progress...

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

Thank you...

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

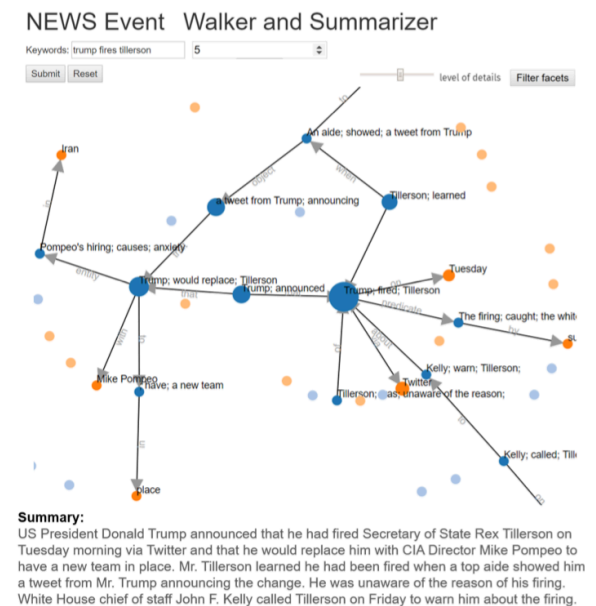
Proposal: exploit our multi-granular summarizer

Summarize a set of news articles about the same event in one single text

Produce summaries with different granularities ranking from the general topic detailed facts and use content for feature extraction



Stuffie



Facet Labeling

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

He; had fired; Secretary of State Rex Tillerson;

- **on; Tuesday;**

→ Temporal?

→ Location?

→ ... ?

Facet labels:
where to
obtain them?

Generating
training data
for
facet labeling

Learning
model for
facet labeling

Facet labels: where to obtain good labels?

Facets are
connected by
prepositions and
conjunctions

| Prep & conj |
|-------------|
| for |
| on |
| because |
| while |
| by |
| ... |

| No | Label |
|-----|-------------|
| 1. | Agent |
| 2. | Beneficiary |
| 3. | Cause |
| 4. | Location |
| ... | |
| 34. | Temporal |
| 35. | Topic |



Illinois SRL



coverage
check

シ ン 局
入 局 局
チ 維 局

Wiktionary
The free dictionary

+additions

Distant Supervision: Building Training Data

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



crawling

Sentence: "The fleet is **on** the American coast."

Facet conn: on

Gloss: "At or near; adjacent to"

Sentence: "He left the house **on** Monday."

Facet conn: on

Gloss: "Some time during the day of"

mapping

Sentence: "The fleet is **on** the American coast."

Facet conn: on

Label: Location

Sentence: "He left the house **on** Monday."

Facet conn: on

Label: Temporal

Enrich The Labels

News
Enrichement

News
Information
Extraction

News Effect
on Stock
Markets

Facet label:
"Temporal"

Synonymous /
related words

WordNet

Context-enriched label:
<"Temporal", "Time",
"Period", "Day", ...>

| No | Label |
|-----|-------------|
| 1. | Agent |
| 2. | Beneficiary |
| 3. | Cause |
| 4. | Location |
| ... | |
| 34. | Temporal |
| 35. | Topic |



| No | Context-enriched label |
|-----|---|
| 1. | <Agent, Instigator, Actor, Initiator,...> |
| 2. | <Beneficiary, Recipient, Receiver, ...> |
| 3. | <Cause, Reason, Grounds, ...> |
| 4. | <Location, Region, Space, ...> |
| ... | |
| 34. | <Temporal, Time, Period, Day, ...> |
| 35. | <Topic, Subject, Theme, ...> |

Map Glosses to Similar Labels!

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets

"Some time
during the day of"

Word
tokenization

<Some, time, during,
the, day, of>

Count overlap

We evaluated the
mapping on a ground
truth of 100 pairs and
reached 93% of precision

| No | Context-enriched label | Overlaps |
|-----|--|----------|
| 1. | <Agent,Instigator,Actor,Initiator,...> | 0 |
| 2. | <Beneficiary,Recipient,Receiver,...> | 0 |
| 3. | <Cause,Reason,Grounds,...> | 0 |
| 4. | <Location,Region,Space,...> | 0 |
| ... | | ... |
| 34. | <Temporal,Time,Period,Day...> | 2 |
| 35. | <Topic,Subject,Theme,...> | 0 |

Overlap tiebreakers

- Dependency edge distance
- Word distance
- Word2vec + WordNet::Similarity averages

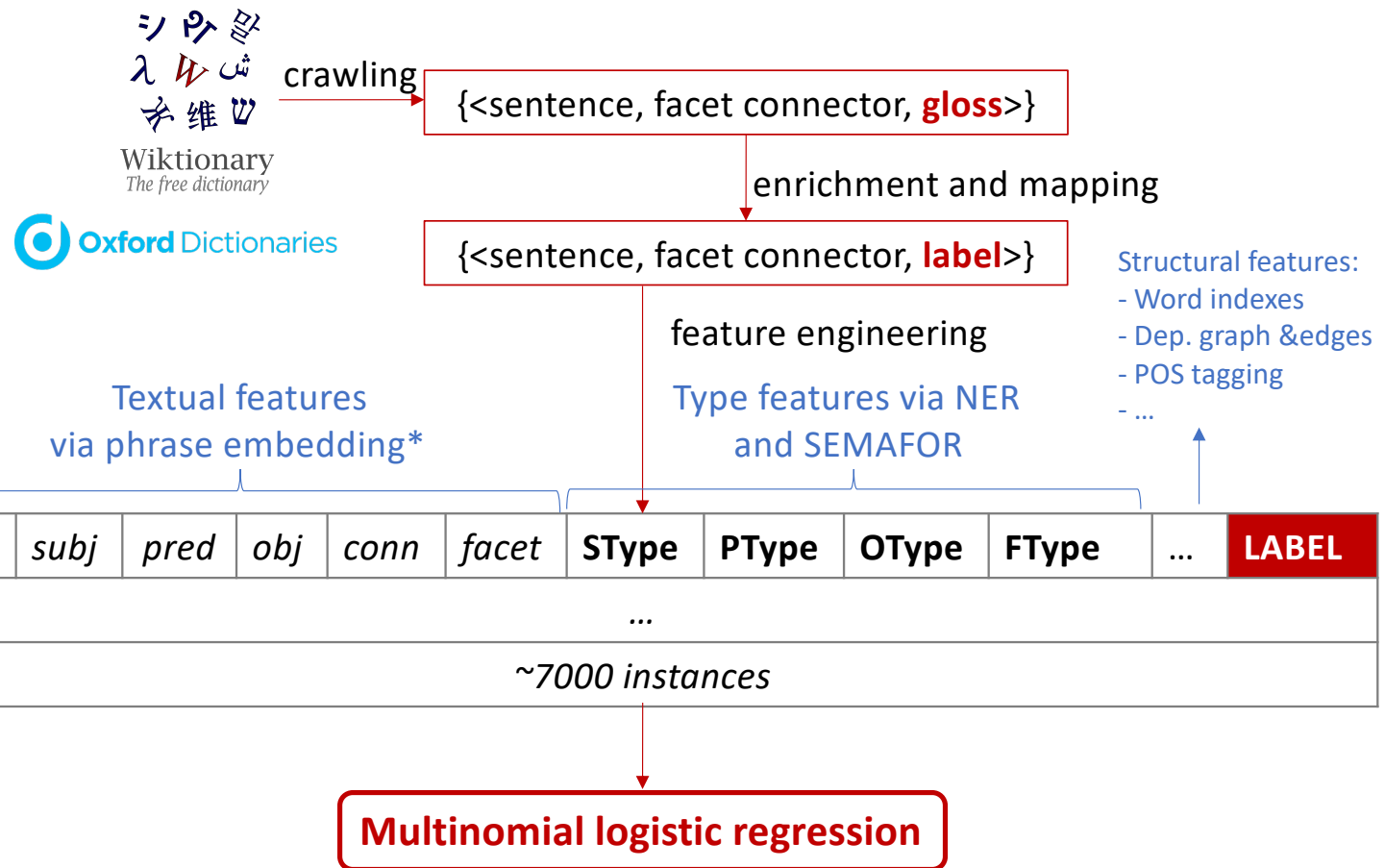
Ranking

Learning Model for Facet Labeling: Summing up

News
Enrichment

News
Information
Extraction

News Effect
on Stock
Markets



Experiments: Questions about StuffIE

- How good is StuffIE's facet labeling model to **predict** its own training data?
 - Which features are the best?
- How does StuffIE's **facet labeling** perform compared to Illinois SRL and SEMAFOR SRL?

Experiments: Questions about StuffIE

- How many triples and facets does StuffIE extract, compared to ClausIE and OLLIE?

Experiments and Results: Triple & Facet Extraction

Dataset: DUC'04 news dataset with ~20,000 sentences

| <i>Method</i> | <i>Triples</i> | <i>Facets</i> | <i>Nested</i> | <i>Miss</i> |
|----------------------|-----------------------|----------------------|----------------------|--------------------|
| ClausIE | 76906 | 105840 | 16381 | 158 |
| OLLIE | 70862 | 69781 | 161 | 642 |
| StuffIE | 83265 | 107210 | 22740 | 0 |

*Miss = missing extraction

Metrics: The number of relations, “facets”, “nested relations”, and missing extractions.

Note: “Facets” for ClauSIE and OLLIE are counted by the number of arguments.

“Nested relations” are counted by redundant extractions.