

Document Aboutness via Sophisticated Syntactic and Semantic Features

Marco Ponza, Paolo Ferragina, and Francesco Piccinno

Dipartimento di Informatica
University of Pisa



NLdb 2017

Liège, Belgium

Menu

0. Introduction

- The Document Aboutness Task
- Through a Semantic Representation

1. Entity Salience

- Our Solution
- Three-Stage Framework

2. Experiments

- Competitors & Datasets
- Results

3. Conclusion & Future Work

Introduction

The Document Aboutness Task

Succinct Representation of a

Document's Subject Matter (Bruza, AIR '96)

- ▷ **N-Grams** (Turney, IR '00)
- ▷ **Sentences** (Mihalcea, EMNLP '04)
- ▷ **Terms of Dictionary** (Paranjpe, CIKM '08)
- ▷ **Keywords** (Liu, EMNLP '14)

Goal



Classical
Aboutness
Representation

Limitations

(Hasan, ACL '14)

Introduction

The Document Aboutness Task



Jaguar

Goal



Input Document

- ▷ N-Grams (Turney, IR '00)
- ▷ Sentences (Mihalcea, EMNLP '06)
- ▷ Terms of Dictionary (Paranjape, CIKM '08)
- ▷ Keywords (Liu, EMNLP '14)



or



Jaguar_(felin)

Jaguar_Cars

- ▷ Interpretation
- ▷ Overgeneration
- ▷ Infrequency
- ▷ Redundancy

Limitations

(Hasan, ACL '14)

Introduction

Through a Semantic Representation

- ▷ Interpretation
- ▷ Over-Generation
- ▷ Infrequency
- ▷ Redundancy

Limitations

(Hasan, ACL '14)



*[...] errors could be addressed
using background **knowledge**.*



(Hasan, ACL '14)



Adding **semantics** into the **aboutness** representation

Introduction

Through a Semantic Representation

Semantic aboutness representation

1. **Entity** \in Knowledge Base



▷ Deployment of Entity Annotators

(Cucerzan, EMNLP '07) and many others

Maradona won against *Mexico*



Diego_Maradona



Mexico_National_Football_Team

Introduction

Through a Semantic Representation

Semantic Aboutness Representation

1.

Entity

Knowledge Base



▷ Deployment of Entity Annotators

+

Maradona won against Mexico



Entity Salience Extraction

(Dunietz, EACL '14), (Trani, DocEng '16)

2.

Salience

(Class or Score)

▷ Relevance of an Entity

Introduction

Through a Semantic Representation

Semantic Aboutness Representation

1. Entity \in Knowledge Base



- ▷ Deployment of Entity Annotators



2. Saliency (Class or Score)

- ▷ Relevance of an Entity

Entity Salience

Our Solution



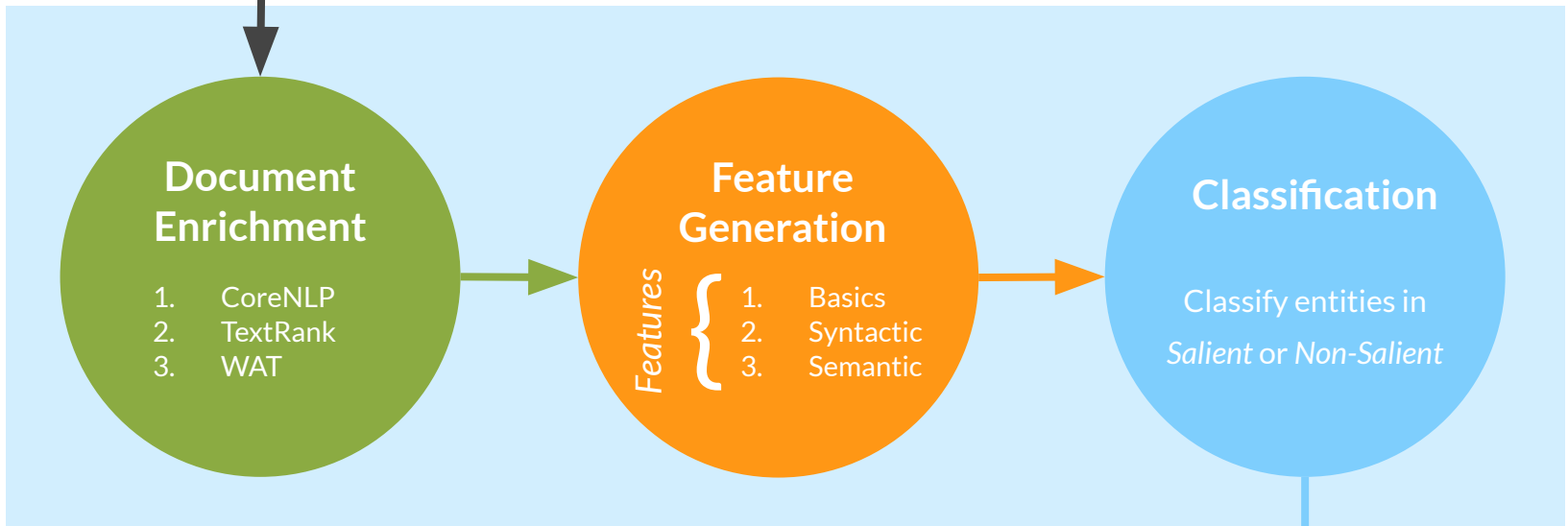
- ▷ **Three-Stage Framework** for **Entity Salience Extraction**
- ▷ **In-Depth Feature Engineering:**
 - **Syntactic:**
 - Sentence Ranking
 - Dependency Trees
 - **Semantic:**
 - Entity Annotations
 - Relatedness Graph
- ▷ **Improves** current solutions
 - Up to +9.8%
- ▷ The **first publicly** available **API**

Entity Saliency

General Structure

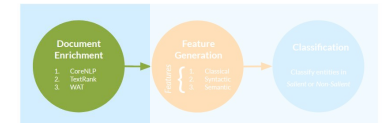


Input Document



Salient Entities





Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

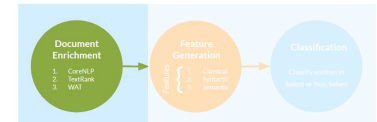
At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

1. Document Enrichment

▶ CoreNLP (Manning, ACL '14)



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

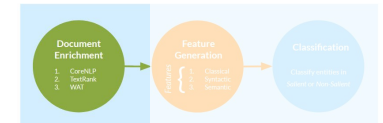
- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

Module

Sentence Splitting



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

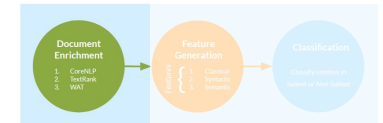
1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

Module

Sentence Splitting

Tokenization



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

Part-of-Speech:

1 Do nt look for presidential announcements from Senators Barack Obama .

1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

Module

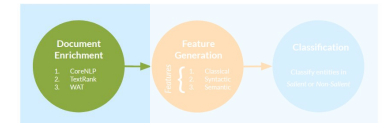
Sentence Splitting

Tokenization

POS-Tagging

Images via

<http://corenlp.run>



Entity Salience

Three-Stage Framework

The New York Times

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

Part-of-Speech:

1 Do nt look for presidential announcements from Senators Barack Obama .

Named Entity Recognition:

1 Do nt look for presidential announcements from Senators Barack Obama .

Images via
<http://corenlp.run>

1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

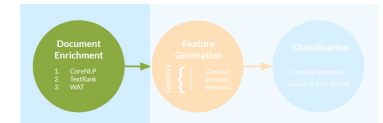
Module

Sentence Splitting

Tokenization

POS-Tagging

Named Entity Recognition



Entity Salience

Three-Stage Framework

The New York Times

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

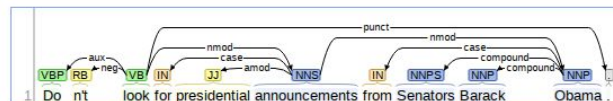
Part-of-Speech:

VB VP RB IN VB IN JJ NNS IN NNPS NNP NNP
1 Do nt look for presidential announcements from Senators Barack Obama .

Named Entity Recognition:

PERSON
1 Do nt look for presidential announcements from Senators Barack Obama .

Basic Dependencies:



1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

Module

Sentence Splitting

Tokenization

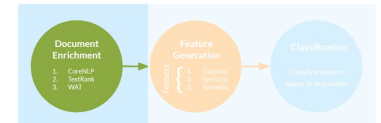
POS-Tagging

Named Entity Recognition

Dependency Parsing

Images via

<http://corenlp.run>



Entity Salience

Three-Stage Framework

The New York Times

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

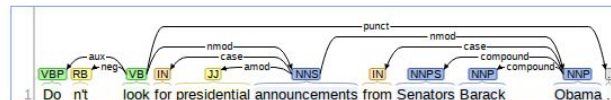
Part-of-Speech:

1 Do nt look for presidential announcements from Senators Barack Obama .

Named Entity Recognition:

1 Do nt look for presidential announcements from Senators Barack Obama .

Basic Dependencies:



1. Document Enrichment

▷ CoreNLP (Manning, ACL '14)

Module

Sentence Splitting

Tokenization

POS-Tagging

Named Entity Recognition

Dependency Parsing

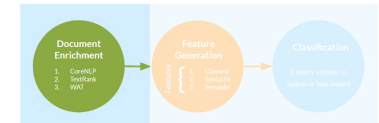
Coreference

Images via

<http://corenlp.run>

Entity Salience

Three-Stage Framework



The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

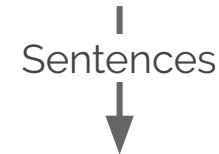
At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS

1. Document Enrichment

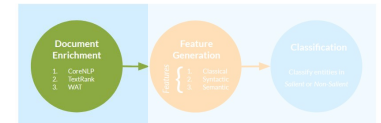
▷ CoreNLP (Manning, ACL '14)



▷ TextRank (Mihalcea, EMNLP '04)

- Graph-Based Summarizer
 - Nodes = Sentences
 - Weights = Normalized Token Overlap
- Sentence Ranking via PageRank





Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



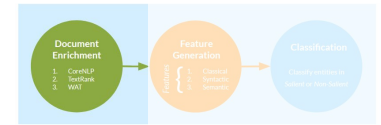
1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



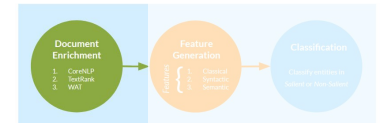
1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

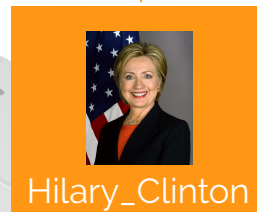
By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to wake up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



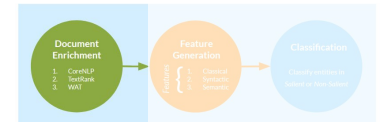
1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

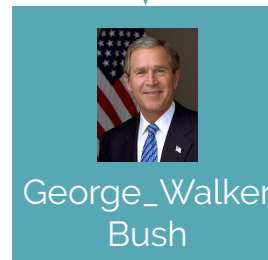
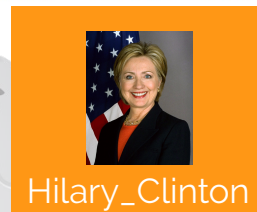
By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



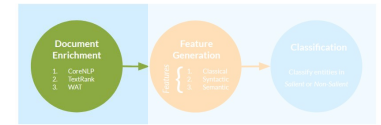
1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

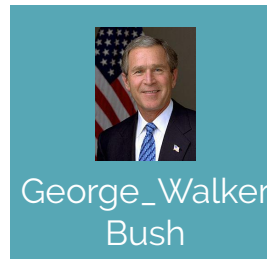
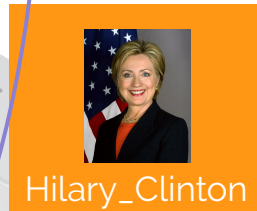
By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



1. Document Enrichment

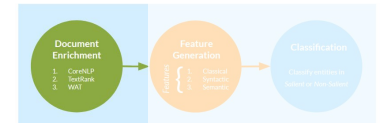
▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns

↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities



Entity Salience

Three-Stage Framework

The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

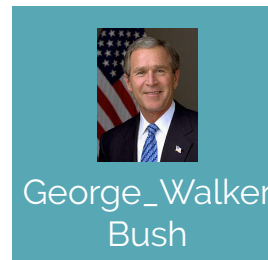
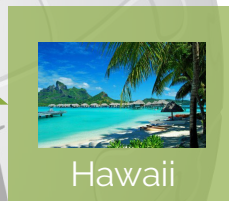
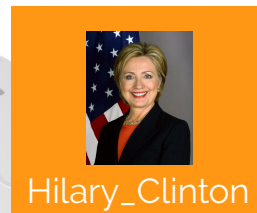
By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns

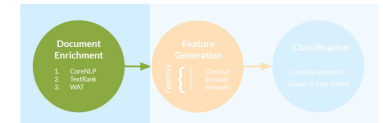
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities

Entity Salience

Three-Stage Framework



The New York Times

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICAL ACTION; Decisions on the Horizon

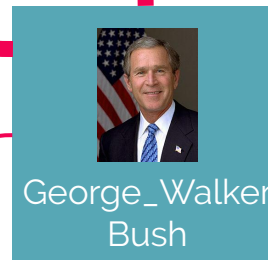
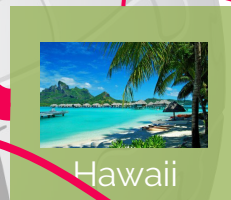
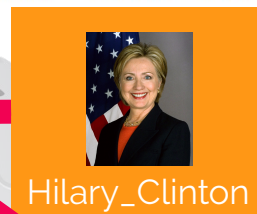
By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

Don't look for presidential announcements from Senators **Barack Obama** and **Hillary Rodham Clinton** anytime soon, but stay tuned.

At least that is the word from their associates. **Mr. Obama**, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after **President Bush's** State of the Union address on Jan. 23. As he walked out of the **Capitol** on a recent afternoon, **Mr. Obama** only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, **Mr. Obama** said he intended to announce his decision after returning from a holiday vacation in **Hawaii** where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

- FACEBOOK
- TWITTER
- GOOGLE+
- EMAIL
- SHARE
- PRINT
- REPRINTS



1. Document Enrichment

▷ **CoreNLP** (Manning, ACL '14)

Named Entities + Proper/Common Nouns
↓

▷ **WAT** (Piccinno, SIGIR '14)

- Annotates them with Wikipedia Entities
- Relatedness Graph
 - Nodes = Entities
 - **Weights** = Wikipedia **Jaccard In-Links**

Entity Salience Three-Stage Framework



Barack Obama

From Wikipedia, the free encyclopedia

For other uses, see **Barack (disambiguation)** and **Obama (disambiguation)**.

Barack Hussein Obama II (US: /ˈbɑːrɑːk huːˈseɪn oʊˈbɑːmɑː/ *ə-RAHK-hoo-SAYN oh-BAH-mɑː*^[1]) born August 4, 1961) is an American politician who served as the 44th President of the United States from 2009 to 2017. He is the first African American to have served as president. He previously served in the U.S. Senate representing Illinois from 2005 to 2008, and in the Illinois State Senate from 1997 to 2004.

Obama was born in Honolulu, Hawaii, two years after the territory was admitted to the Union as the 50th state. Raised largely in Hawaii, Obama also spent one year of his childhood in Washington State and four years in Indonesia. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago. In 1998 Obama enrolled in Harvard Law School, where he was the first black president of the *Harvard Law Review*. After graduation, he became a civil rights attorney and professor, teaching constitutional law at the University of Chicago Law School from 1992 to 2004. Obama represented the 13th District for three terms in the Illinois Senate from 1997 to 2004, when he ran for the U.S. Senate. Obama received national attention in 2004, with his unexpected March primary win, his well-received July Democratic National Convention keynote address, and his landslide November election to the Senate. In 2008, Obama was nominated for president, a year after his campaign began, and after a close primary campaign against Hillary Clinton. He was elected over Republican John McCain, and was inaugurated on January 20, 2009. Nine months later, Obama was named the 2009 Nobel Peace Prize laureate.

During his first two years in office, Obama signed many landmark bills. Main reforms were the Patient Protection and Affordable Care Act (often referred to as "Obamacare"), the Dodd–Frank Wall Street Reform and Consumer Protection Act, and the Don't Ask, Don't Tell Repeal Act of 2010. The American Recovery and Reinvestment Act of 2009 and Tax Relief



Barack Obama

44th President of the United States

In office
January 20, 2009 – January 20, 2017

Vice President Joe Biden

Preceded by George W. Bush

Succeeded by Donald Trump

United States Senator from Illinois

In office
January 3, 2005 – November 16, 2008

Preceded by Peter Fitzgerald

Succeeded by Roland Burris

Member of the Illinois Senate from the 13th district

President of the United States

From Wikipedia, the free encyclopedia

"**PTSD**" redirects here. For the post-traumatic stress disorder, see PTSD (disambiguation).
For other uses, see **President of the United States (disambiguation)**. For the use and list of Presidents of the United States.

The **President of the United States** (informally referred to as "**PTSD**") is the head of state and head of government of the United States. The president directs the executive branch of the federal government and is the commander-in-chief of the United States Armed Forces.

The president is considered to be one of the world's most powerful political figures, as the holder of the only contemporary global superpower. PTSD's role includes being the commander-in-chief of the world's most powerful military with the second largest nuclear arsenal and leading PTSD's foreign relations with the largest economy by nominal GDP. The office is both a significant trust and both power both domestically and abroad.

Article II of the U.S. Constitution vests the executive power in the president. The president is elected for a four-year term by the Electoral College.



Bill Clinton

From Wikipedia, the free encyclopedia

For other uses, see **William Clinton (disambiguation)**.

William Jefferson Clinton (born **William Jefferson Blaine Clinton**; August 19, 1946) is an American politician who served as the 42nd President of the United States from 1993 to 2001. He was previously the 46th Governor of Arkansas from 1979 to 1983, and the 40th governor from 1985 to 1991. Clinton was the first president to have been born after 1929. A member of the Democratic Party, Clinton was ideologically a New Democrat and was one of the presidents to succeed "The New" political philosophy.

Clinton was born in Hopewell, Arkansas as the son of a farmer of German-American descent who was a member of Kappa Kappa Psi and Phi Kappa Phi Honor Society, the son of a Baptist. Following his graduation from the University of Central Arkansas, he married Hillary Rodham Clinton, who served as United States Secretary of State from 2009 to 2013 and U.S. senator from New York from 2005 to 2008, and was the Democratic nominee for Vice President in 2008. He and Hillary Clinton both earned their law degrees from Yale Law School, where they met and began dating. An governor of Arkansas, Clinton was elected to the White House in 1992, after a narrow victory over George H. W. Bush, when they were both largely doing. An governor of Arkansas, Clinton was elected to the White House in 1992, after a narrow victory over George H. W. Bush, when they were both largely doing.



Bill Clinton

42nd President of the United States

In office
January 20, 1993 – January 20, 2001

Preceded by George H. W. Bush

Succeeded by George W. Bush

Illinois

From Wikipedia, the free encyclopedia

This state is about the U.S. state of Illinois. For other uses, see **Illinois (disambiguation)**.

Illinois (IPA: /ɪˈllɪnoʊ/) is a state in the Midwestern region of the United States, occupying a substantial portion of the central United States. It is the 14th most populous state and the 25th largest by area. It is bordered by Kentucky to the east, Indiana to the east, Missouri to the west, and Wisconsin to the north. It is bordered by the Mississippi River to the west and the Gulf of Mexico to the south. Illinois has a total population of approximately 12.8 million, with Chicago being the largest city. It is the only state to be a member of the United States' Great Lakes region. The state is a member of the United States' Great Lakes region. The state is a member of the United States' Great Lakes region. The state is a member of the United States' Great Lakes region.



Illinois

State of Illinois

Illinois is a state in the Midwestern region of the United States, occupying a substantial portion of the central United States. It is the 14th most populous state and the 25th largest by area. It is bordered by Kentucky to the east, Indiana to the east, Missouri to the west, and Wisconsin to the north. It is bordered by the Mississippi River to the west and the Gulf of Mexico to the south. Illinois has a total population of approximately 12.8 million, with Chicago being the largest city. It is the only state to be a member of the United States' Great Lakes region. The state is a member of the United States' Great Lakes region. The state is a member of the United States' Great Lakes region.

Hillary Clinton

From Wikipedia, the free encyclopedia

Hillary Diane Rodham Clinton (/ˈhɪlɑːriaɪ daɪˈroʊdəm ˈklɪntɒn/; born October 26, 1947) is an American politician who was the 67th United States Secretary of State from 2009 to 2013, U.S. Senator from New York from 2001 to 2009, First Lady of the United States from 1993 to 2001, and the Democratic Party's nominee for President of the United States in the 2016 election.

Born in Chicago, Illinois, and raised in the Chicago suburb of Park Ridge, Clinton graduated from Wellesley College in 1969, and earned a J.D. from Yale Law School in 1973. After serving as a congressional legal counsel, she moved to Arkansas and married Bill Clinton in 1975. In 1977, she co-founded Arkansas Advocates for Children and Families. She was appointed the first female chair of the Legal Services Corporation in 1978 and became the first female partner at Rose Law Firm the following year. As First Lady of Arkansas, she led a task force whose recommendations helped reform Arkansas's public schools.

As First Lady of the United States, Clinton was an advocate for gender equality and healthcare reform. Her relationship with her husband came under public speculation during the Lewinsky scandal, forcing her to issue a statement reaffirming her commitment to the marriage. Clinton was elected in 2000 as the first female senator from New York. She was re-elected to the Senate in 2006. Running for president in 2008, she won far more delegates than any previous female candidate, but lost the Democratic nomination to Barack Obama.^[1]

As Secretary of State in the Obama administration from 2009 to 2013, Clinton responded to the Arab Spring, during which she advocated the U.S. military intervention in Libya. She helped organize a diplomatic isolation and international sanctions regime against Iran, in an effort to force curtailment of that country's nuclear program; this would eventually lead to the



Hillary Clinton

67th United States Secretary of State

In office
January 21, 2009 – February 1, 2013

President Barack Obama

Deputy James Steinberg

Preceded by Condoleezza Rice

Succeeded by John Kerry

United States Senator from New York

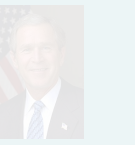
In office
January 3, 2001 – January 21, 2009

Preceded by Daniel Patrick Moynihan

Succeeded by Kirsten Gillibrand

First Lady of the United States

George Walker Bush



George Walker Bush

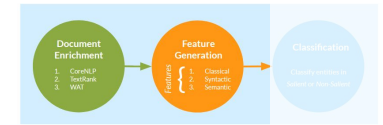
WAT (Piccinno, SIGIR '14)
Annotates them with
Wikipedia Entities

- Relatedness Graph
- Nodes = Entities
- Weights = Wikipedia Jaccard In-Links

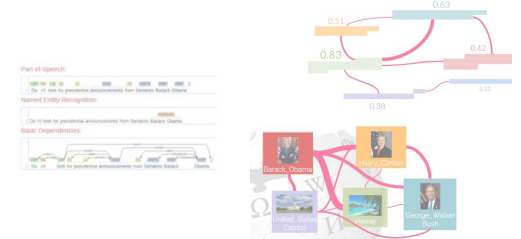


Entity Salience

Three-Stage Framework



Tokens, POS Tags, Dependency Relations, Coreference Chains, Sentence Ranks, Wikipedia Entities and their Relatedness



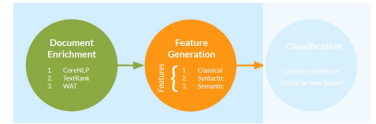
Standard Entity Features	
$ef(e, d), idf(e), ef-idf(e, d)$	Entity frequency (number of times WAT annotates e in d), inverse document frequency for e and their product.
$position-stats_{\{s,t\}}(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of sentence- (resp. token-) positions of e in d .
$mention-title(e, d)$	Presence of a <i>mention</i> of e in the title of d .
$entity-title(e, d)$	Presence of e in the title of d .
$is-upper(e, d)$	True if one of the mentions of e appear in d in uppercase, false otherwise.
Cmu-Google's Features	
$1st-loc(e, d)$	Index of the sentence in which the first mention of e appears in d , modeled as a binary indicator.
$head-count(e, d)$	Frequency of head word of entity e in the document d , modeled as a binary indicator.
$mentions(e, d)$	Sum between entity frequency and co-referenced frequency of e in d , modeled as a binary indicator.
$headline(e, d)$	POS tag of each word of e that appears in at least one mention and also in the headline of d , normalized via feature hashing.
$head-lex(e, d)$	Lower-cased head word of the first mention of e in d , normalized via feature hashing.
$google-centrality(e, d)$	PageRank score of e on the entity graph generated from d , where weights are the co-occurrence probability of two entities, computed on the training set.
Novel Features Introduced	
$spread_{\{s,t\}}(e, d)$	Difference between the maximum and minimum sentence- (resp. token-) positions of e in d .
$bucketed-freq_{\{s,t\}}(e, d)$	Vector of bucketed frequencies through sentence- (resp. token-) positions of e in d .
$textrank-stats(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of TEXTRANK scores of sentences where e appears in d .
$dep-freq(e, d)$	Frequency of e in d when it appears as dependent of the dependency relation dep .
$dep-bucketed-freq_{\{s,t\}}(e, d)$	Vector of bucketed frequencies through sentence- (resp. token-) positions of e in d where only the mentions where e appears as dependent of a dependency relation dep are considered.
$dep-position-stats_{\{s,t\}}(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of sentence- (resp. token-) positions of e in d , where only the mentions where e appears as dependent of a dependency relation dep are considered.
$dep-textrank-stats(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of TEXTRANK scores where only the sentences where e appears as dependent of a dependency relation dep are taken into account.
$rel-stats(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of the relatedness scores between e and all other entities annotated in d .
$rel-bucketed-stats_{\{s,t\}}(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of the relatedness scores between e and all other entities present in d , bucketed over document positions (both at sentence- and token-level).
$comm-stats(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of the <i>commonness</i> values of e in d (see WAT).
$p-stats(e, d)$	Minimum, maximum, arithmetic mean, median, standard deviation and harmonic mean of the p -score values of e in d (see WAT).
$rel-centrality(e, d)$	Degree, PageRank, Betweenness, Katz, HITS, Closeness and Harmonic [2] scores of e computed on the entity graph of d . Details in text.

2. Feature Generation

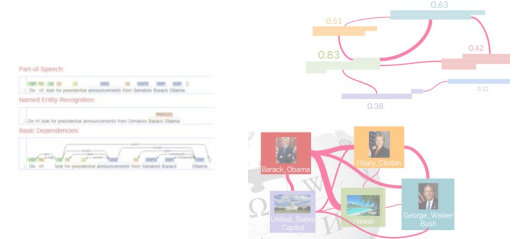
- ▷ Standard Entity Features
 - Frequency
 - Positions
 - ...
- ▷ CMU-Google Features
 - POS-Tags, Coreference Freq.
 - PageRank on a graph whose weights are based on co-occ.
 - ...
- ▷ Syntactic Features
 - Statistics on Sentence Ranks
 - Frequency/Positions of Dependency Relations
 - ...
- ▷ Semantic Features
 - Statistics on annotations (coherence, commonness)
 - Graph Centralities on Relatedness Graph
 - Relatedness over Positions
 - ...

Entity Salience

Three-Stage Framework

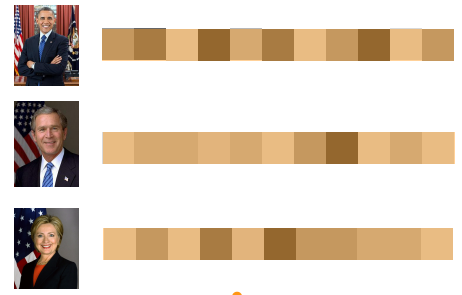


1. Document Enrichment



Tokens, POS Tags, Dependency Relations, Coreference Chains, Sentence Ranks, Wikipedia Entities and their Relatedness

2. Feature Generation



Entity Feature Vectors

3. Salience Classification

dmlc
XGBoost
(Chen, SIGKDD '16)

Salient Entities



Experiments

Experiments

Competitors & Datasets

- ▷ **CMU-Google System** (Dunietz, EACL '14)

Re-implemented

- Proprietary modules substituted with open-source tools



- ▷ **SEL** (Trani, DocEng '16)

Limitations

- No comparison with CMU-Google System
- Benchmark on small dataset
- Not publicly available

Dataset (365 news, 4747 entities)

Experiments

Results

Dataset

New York Times

System	Micro		
	P	R	F1
CMU-Google (Dunietz, EACL '14)	60.5	63.5	61.5
CMU-Google-ours	58.8	62.6	60.7
SWAT	62.2	63.0	62.6

Experiments

Results

Dataset

New York Times

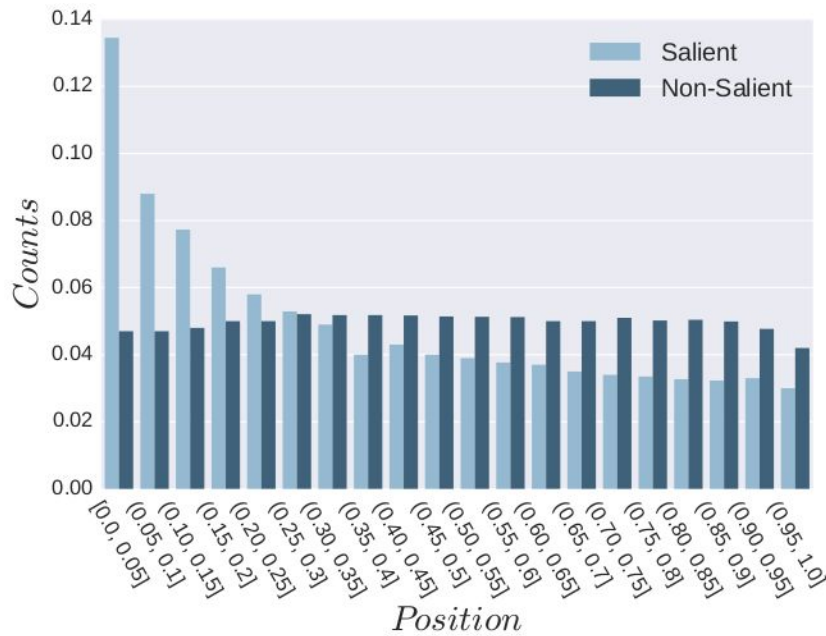
System	Micro		
	P	R	F1
CMU-Google (Dunietz, EACL '14)	60.5	63.5	61.5
CMU-Google-ours	58.8	62.6	60.7
SWAT	62.2	63.0	62.6

▷ Up to +1.9 % on Micro-F1

Experiments

Results

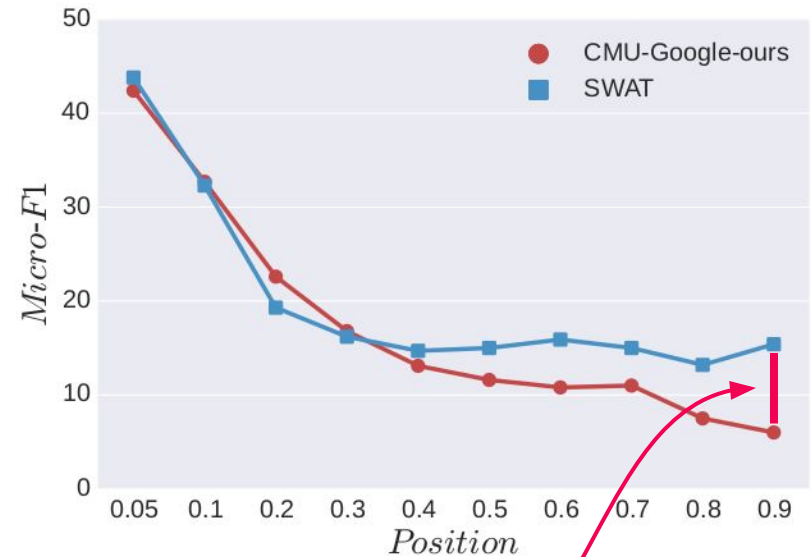
Independence from position of salient entities



Dataset

New York Times

Goal



+9%

Experiments

Results

Dataset

Wikinews

System	Macro		
	P	R	F1
CMU-Google-ours	35.7	48.6	38.0
SEL (Trani, DocEng '16)	42.0	51.0	43.0
SWAT	47.3	55.2	47.8

Experiments

Results

Dataset

Wikinews

System	Macro		
	P	R	F1
CMU-Google-ours	35.7	48.6	38.0
SEL (Trani, DocEng '16)	42.0	51.0	43.0
SWAT	47.3	55.2	47.8

▷ Up to +9.8% on Macro-F1

Conclusion & Future Work

Conclusion & Future Work

- ▷ We proved that the deployment of **semantic knowledge** eliminates the limitations described by (Hasan, ACL '14)

- ▷ Interpretation
- ▷ Over-Generation
- ▷ Infrequency
- ▷ Redundancy

Limitations
(Hasan, ACL '14)

- ▷ We aim to
 - Improve SWAT
 - Deployment of WAT 2.0 & Deep Learning
 - Ranking of Entities
 - Other datasets (not only news)
 - Deploy our system for **other applications**
 - Entity Annotation of Query
 - Document Similarity
 - ...
 - *Please, suggest your favourite one!*

Thanks!

Any questions?