Automatic Text Summarization and Applications

March 1, 2014
Yasemin Altun
Google Research, Zurich

Team Members:
Enrique Alfonseca
Massimiliano Ciaramita
Katja Filippova
Evangelos Kanoulas
Srini Narayanan
Daniele Pighin
Ioannis Tsochantaridis
Outline

- Big picture of automatic text summarization
- Text summarization applications
  - Entity timelines
  - Storylines of events
  - Sentence compression
  - Event understanding
  - Summarization of user-generated content
“I took a speed-reading course and read *War and Peace* in twenty minutes. *It involves Russia.*”

- Create a concise representation that retains most relevant information.

“events surrounding the French invasion of Russia, and the impact of the Napoleonic era on Tsarist society, as seen through the eyes of five Russian aristocratic families”
Text Summarization

- Create a concise representation that retains most relevant information.
- Relevant information with respect to particular aspect:

“events surrounding the French invasion of Russia, and the impact of the Napoleonic era on Tsarist society, as seen through the eyes of five Russian aristocratic families”

**War:** “events surrounding the *French invasion of Russia* during Napoleonic wars”

**Social context:** “story of five *aristocratic families* during and the entanglement of their lives with the history of 1805 to 1813”
Summarize the Web

- Web: Enormous amount of information; Huge diversity (Style, quality, language)
- How to summarize? Content
- What for? Aspect
Summarize the Web

- Search engines organizes information for accessibility and usefulness
- Match keywords to queries (words)
- Richer meaning: Refer to entities, objects (people, places, things) in real world.
Google Knowledge Graph

- Searching for entities, instantly provide information relevant about the entity
- Provide connections: Relations between entities
- An ever growing database of structured knowledge
  - 500 million entities
  - 3.5 billion defining attributes and connections

Marie Curie

Physicist

Marie Skłodowska-Curie was a Polish and naturalized-French physicist and chemist who conducted pioneering research on radioactivity.

Wikipedia

Born: November 7, 1867, Warsaw, Poland
Died: July 4, 1934, Passy, Haute-Savoie, France
Spouse: Pierre Curie (m. 1895–1906)
Discovered: Radium, Polonium
Children: Irène Joliot-Curie, Ève Curie
Education: University of Paris (1903), University of Paris (1894), University of Paris (1891–1893), Flying University

People also search for

Pierre Curie
Antoine Henri Becquerel
Albert Einstein
Irène Joliot-Curie
Louis Pasteur
Google Knowledge Graph

- Find the right thing

Bush (band) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Bush_(band)
Bush are a British rock band formed in London in 1992 shortly after vocalist/guitarist Gavin Rossdale and guitarist Nigel Pulsford met. It was not long before they ...
Discography - Sixteen Stone - Razorblade Suitcase - The Sea of Memories

George W. Bush - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/George_W._Bush
George Walker Bush (born July 6, 1946) is an American politician and businessman who served as the 43rd President of the United States of America from 2001 ...
Laura Bush - Barbara Pierce Bush - Jenna Bush Hager - George Bush

News for bush
George W. Bush the painter about to show his art in gallery for first time
South China Morning Post - 3 minutes ago
George W. Bush, the painter, will soon have his first gallery showing. The former US president's artwork - a source of fascination for fans and ...
Bush (band) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Bush_(band) ▼
Bush are a British rock band formed in London in 1992 shortly after vocalist/guitarist
Gavin Rossdale and guitarist Nigel Pulsford met. It was not long before they ...
Discography - Sixteen Stone - Razorblade Suitcase - The Sea of Memories

BUSH Official Website
bushofficial.com/ ▼
Bush Live at Stubb's Now Available in North America for Blu-Ray & DVD; video ...
2013, a DVD and Blu-ray package from British alternative rock band BUSH.
Video - News - Contact - Shush

Bush - Chemicals Between Us - YouTube
www.youtube.com/watch?v=TrnTFcfe6w8 ▼
Feb 10, 2010 - Uploaded by setzerlock
Led by guitarist/vocalist Gavin Rossdale, Bush became the first post-Nirvana British band to hit it big in ...

The Official Sam Bush Website
www.sambush.com/ ▼
This is Sam Bush's official website featuring the fan centric members area "The Bush League". You can find tour dates, press information, latest news, Sam Bush ...

Bush Bands Bash
bushbandsbash.net.au/ ▼
Thank you to everyone that made the 2013 Bush Bands Bash such a success! There were some 3000 people at the event and all the bands put on an amazing ...
Moving forward

- Entity-centric knowledge base is one side of the effort to establish a "semantic Web"

- Manual annotation requires large collaborative effort (slow, lack of comprehensiveness,...)

- Smaller machines to access information

- Entity-centric summarization of the content, with special emphasis of *news-worthy events*
Entity-centric summarization

- Identify most relevant information with respect to a given entity.
  - Annotate the content with references to entities
- Multi-document summarization to leverage redundancy
- Methodology
  - Extractive (selecting sentences from the documents)
  - Abstractive (generate the sentences)
- Beyond academic definition: For machine or user consumption
Outline

- Big picture of automatic text summarization
- Text summarization applications
  - Entity timelines
  - Storylines of events
  - Sentence compression
  - Event understanding
  - Summarization of user-generated content
Entity Timelines

Given an entity and a time period, provide a summary of the most memorable events involving this entity.

- Instantly provide concise information about what happened to/about/around an entity within a particular time.

Joint work with Evangelos Kanoulas and Enrique Alfonseca
Google Knowledge Graph

Find the right thing
Bush (band) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Bush_(band)
Bush are a British rock band formed in London in 1992 shortly after vocalist/guitarist Gavin Rossdale and guitarist Nigel Pulsford met. It was not long before they ...

Discography - Sixteen Stone - Razorblade Suitcase - The Sea of Memories

BUSH Official Website
bushofficial.com/
Bush Live at Stubb's Now Available in North America for Blu-Ray & DVD; video ... 2013, a DVD and Blu-ray package from British alternative rock band BUSH.
Video - News - Contact - Shush

Bush - Chemicals Between Us - YouTube
www.youtube.com/watch?v=TrmTFCFe6w8
Feb 10, 2010 - Uploaded by setzerlock
Led by guitarist/vocalist Gavin Rossdale, Bush became the first post-Nirvana British band to hit it big in ...

The Official Sam Bush Website
www.sambush.com/
This is Sam Bush's official website featuring the fan centric members area "The Bush League". You can find tour dates, press information, latest news, Sam Bush ...

Bush Bands Bash
bushbandsbash.net.au/
Thank you to everyone that made the 2013 Bush Bands Bash such a success! There were some 3000 people at the event and all the bands put on an amazing ...
Prince's **band** release live footage of Shepherds Bush Empire gig
NME.com - Feb 13, 2014
Prince's **band** release live footage of Shepherds Bush Empire gig ...
Prince's new **band** 3DEYEGIRL have released footage from their last ...

Watch: Prince Perform Live At Shepherd's Bush Empire
Clash Magazine - Feb 13, 2014
Prince unveils 'Chaos And Disorder' video from Shepherds Bush gig ...

Gigwise - Feb 13, 2014
all 13 news sources »

Concert review: Kristian **Bush** puts his soul into solo show
Access Atlanta (blog) - Feb 2, 2014
Chances are, fans attending Kristian **Bush**'s first plugged-in, full-**band** solo concert at The Buckhead Theatre Saturday night weren't expecting a ...

**WCBE Presents Yonder Mountain String **Band w/ Sam **Bush** @ The ...**
WCBE 90.5 FM - Feb 1, 2014
Bending bluegrass, rock and countless other influences that the **band** cites, Yonder has pioneered a sound of their own. They're recently ...

**PTO Ladies Night Out Fundraiser to Feature Randolph H.S. **Musical ...**
TheAlternativePress.com - 2 hours ago
The featured performers for the night include juniors Jesse **Bush** and ... **Bush** had been part of the **band** "Flashback," cartaking in multiple ...
Entity Timelines

News Collections:
- Similar news published from various sources within few days.

Assumption:
- News in the same collection describe one main event.

Methodology:
For a given entity,
- Identify and rank news collections in order of importance
  - Prominence of collection, prominence of entity in collection
- Concisely summarize each news cluster
  - Select sentence representative of collection & focus on entity
Entity Timeline

Showing news from January 1st, 2008

Top 10 events

- November 4, 2011
  - Gavin Rossdale: Bush should have been huge in UK

- October 9, 2011
  - Gavin Rossdale agonised over Bush reunion

- September 21, 2011
  - Gavin Rossdale said sons made their mark on his music with Bush

- August 2, 2011
  - Bush is back: interview with Gavin Rossdale

- July 11, 2011
  - Bush to release first studio album in 10 years, 'The Sea of Memories'

- June 22, 2010
  - Bush announces reunion; new song, album, and tour
Entity Timeline

Showing news from January 1st, 2008

July 11, 2011

Bush to release first studio album in 10 years, 'The Sea of Memories'

Gavin Rossdale is set to officially bring Bush back to life in September with the band's first album since it disbanded in 2002.

According to The Hollywood Reporter, Bush has announced that they will release The Sea of Memories on Sept. 13.

"When making music, you have a choice to repeat what you've done or move on," said frontman Rossdale.

June 22, 2010

Bush announces reunion; new song, album, and tour
Outline

- Big picture of automatic text summarization
- Text summarization applications
  - Entity timelines
  - Storylines of events
  - Sentence compression
  - Event understanding
  - Summarization of user-generated content
Storylines of events

*Identify and summarize events that leads to the event of interest*

- Provide background information for an event
- Provide structured timelines for a given entity

Joint work with Evangelos Kanoulas
Storylines of events

News Collections:

- **Similar** news published from various sources within few days.

Assumption:

- News in the same collection describe one **main event**.

Methodology:

- Define a graph of events (news collections) using similarity
- Identify heaviest path ending in a given event
- Summarize the events on this path with respect to prominent entities in the target event.
Storylines of events

Syrian rebel factions clash as UN inspectors secure sites
South China Morning Post - 18 minutes ago

Aid Barred From Gassed, Starved Syrian Town
Nowruz - 20 minutes ago

Weapons experts begin Syria mission
Express.co.uk - 20 minutes ago

Chemical experts start Syria mission amid UN calls for access
Reuter - 21 minutes ago

Chemical experts start Syria mission amid UN calls for access
GlobalPost - 1 hour ago

Tsarnaev lawyers allege mistreatment
The Keene Sentinel - 8 minutes ago

Tsarnaev lawyers want prison restrictions lifted as violate his right to free speech
The Voice of Russia - 2 hours ago

Dzhokhar Tsarnaev disposed of key bombing evidence, Justice Department says
Washington Times - 2 hours ago

Dzhokhar Tsarnaev disposed of Boston bombing evidence, memo says
Los Angeles Times - 4 hours ago

Tsarnaev's Lawyers Complain About Harsh Conditions
WGBH NEWS - 4 hours ago

UNHCR favours supporting countries hosting Syrian refugees
Nowtrack India - Oct 1, 2013

Syria’s neighbours demand greater international assistance for refugees
A&D Online - Sep 30, 2013

U.S. Is Arming Syrian Rebels, But Refugees Who’ve Aided Them Are ...
ProPublica - Sep 30, 2013

Syria conflict: Neighbours plead for international help
BBC News - Sep 30, 2013

Syrians affected by civil war share their stories
gulfnews.com - Sep 30, 2013
Storylines of events

- July 29, 2013
  At least nine detained in Turkey over Gezi protests

- June 19, 2013
  Turkish police unleash water cannon, tear gas on demonstrators

- June 17, 2013
  Defying gov't, turkish protesters to stay in park

- June 15, 2013
  Defiant turkey stands'still and silent' as govt contemplates the use of army

- June 10, 2013
  Erdogan's threat: turkey's protests will be over in 24 hours
Outline

- Big picture of automatic text summarization
- Text summarization applications
  - Entity timelines
  - Storylines of events
  - Sentence compression
  - Event understanding
  - Summarization of user-generated content
Sentence Compression

*Given a sentence, generate a shorter one while preserving the essential information.*

- Move beyond Extractive summarization onward to Abstractive summarization
- Challenging research problem

Joint work with Katja Filippova
Sentence Compression

Assumption:
● Title of news article is a compression of the first sentence.

Methodology:
● Limited publicly available training data:
  ○ Automatically extract noisy label data: Project title to first sentence
    Releasing the data soon!
● Use of syntactic information to ensure grammaticality
● Use semantic/lexical information to preserve the essential content
First sentence:
*Three men were arrested Wednesday night and Dayton police said their arrests are in connection to a west Dayton bank robbery.*

Original Headline:
*3 men arrested in connection with Bank robbery*

Projectioned Headline:
*Three men were arrested are in connection to a bank robbery*
Outline

● Big picture of automatic text summarization
● Text summarization applications
  ○ Entity timelines
  ○ Storylines of events
  ○ Sentence compression
  ○ Event understanding
  ○ Summarization of user-generated content
Event Understanding

In some domains (e.g. celebrity news, sports) headlines may be subjective, to-the-point descriptions:

LaLa Vasquez has married Carmelo Anthony!!

Ciara rocks a sexy Versace Spring 2010 mini to LaLa Vasquez and Carmelo Anthony's wedding (photos)
Event Understanding

Learn how events are referred to in text and represent event mentioning text in predicate argument structure.

- Limited training data: Unsupervised learning
- No restriction in domain of events.
- Abstractive summarization: Generate headlines for news.
- Semi-automatic updates the Knowledge Graph

Daniele Pighin, Katja Filippova, Enrique Alfonseca
Event Understanding

News Collections:
● **Similar** news published from various sources within few days.

Assumption:
● News in the same collection describe one **main event**.
● Titles and first sentences involving same set of entities describe the same event. Patterns between entities refer to the same event.

Methodology: For a given set of news collections,
● Extract compressed patterns between entities from news collections
● Clustering patterns that refer to the same event.
Patterns and news collections

Anna Faris weds actor fiance in Bali
``The House Bunny'' star Anna Faris has married her fiancé, actor Chris Pratt.
'House Bunny' star Anna Faris weds actor fiance Chris Pratt in Bali, Indonesia
Chris Pratt and Anna Faris are seen during the Malibu...
Anna Faris ties the knot with actor beau
Anna Faris marries Chris Pratt in private ceremony
Anna Faris weds actor fiance in Bali ...

Jessica Alba marries Cash Warren
The 27-year-old actress quietly wed producer Cash Warren on Monday, her publicist, Brad Cafarelli, said Tuesday in an e-mail to The Associated Press.
Jessica Alba ties the knot with Cash Warren
Jessica Alba Marries Boyfriend Cash Warren
Jessica Alba Quietly Marries Fiance
Jessica Alba and Cash Warren a hot family

[actor] weds [actor]
[actor] has married [actor]
[actor] and [actor] are seen
[actor] ties the knot with [actor]
[actor] marries [actor]
[person] weds [person]
...

[actor] marries [actor]
[actor] ties the knot with [actor]
[person] weds [person]
[person] marries [person]
[celebrity] weds [actor]
[celebrity] and [actor] a hot family
...
Headline examples

**Actress Alyssa Milano to wed Hollywood agent**
Publicists for Milano said Tuesday that the actress accepted a marriage proposal on Dec. 18 from David Bugliari, an agent at Creative Artists Agency.

**Alyssa Milano is engaged to David Bugliari**
Alyssa Milano is set to marry David Bugliari
Alyssa Milano, David Bugliari engaged
Alyssa Milano to marry David Bugliari
Alyssa Milano engaged to David Bugliari

**Alyssa Milano To Marry Hollywood Agent**
The information you provide will be used only to send the requested e-mail and will not be used to send any other e-mail communications.

**Actress Alyssa Milano to wed Hollywood agent**
LOS ANGELES --

**PHOTO: ALYSSA MILANO ENGAGED TO MARRY AGENT**
Bugliari proposed to the former Charmed star just before Christmas, according to Us magazine.com.

**Alyssa Milano Ready To Walk The Aisle Again!**
Los Angeles 08/01/2009- Looks like the Charmed actress has floored someone again with her magical charm.

**Actress Alyssa Milano to wed Hollywood agent**
AP Breaking News Video Publicists for Milano said Tuesday that the actress accepted a marriage proposal on Dec. 18 from David Bugliari, an agent at Creative Artists Agency.
Outline

- Big picture of automatic text summarization
- Text summarization applications
  - Entity timelines
  - Storylines of events
  - Sentence compression
  - Event understanding
  - Summarization of user-generated content
Summarization of User Generated Content

*Identify main aspects and summarize for machine or user consumption.*

- Unedited user content: eg. YouTube comments
  - Noisy, unstructured, ungrammatical
  - Huge volume
- Use a centrality summarization model on the comments for a video
- Video categorization to provide the user a better search experience

Katja Filippova and Ioannis Tsochantaridis
YouTube comments summarization

 Uploaded on Apr 12, 2011
 Play the DUDE PERFECT GAME here!
 iPhone - http://bit.ly/DPGameiPhone

 Tags of awesomeness

 "dude perfect game" "dude perfect game android" "dude perfect game iphone" "dude perfect game ipod" "dude perfect game ipad"
 "dude perfect game trailer" "basketball game" "iPhone game"
 "android game" "march madness" NBA "Jeremy Lin" instantly
 "summer camp" "backyard edition" "trick shots" "trick shots" "hunger games" "hunger games trailer" "basketball trick shots" "addicting games" "game trailer" "movie trailer"

The iPhone is a line of mobile phones designed and developed by Apple Inc. It was introduced on September 28, 2007, and later released on October 25, 2007. The iPhone was the first smartphone to feature a touch screen, which allows users to interact with the device using gestures and touch input. The iPhone also features a built-in camera and a variety of applications, including email, web browser, and social networking. The iPhone is widely regarded as a technological innovation and has become a popular device among consumers.

Basketball is a team sport played on a rectangular court with two goals at opposite ends. The objective of the game is to score points by shooting a ball through the opposing team's basket. The game is played with a group of players on each team, and the game is divided into four quarters. The team with the most points at the end of the game wins. Basketball is a widely popular sport and is played at various levels, from amateur to professional.
Where to go?

- Entity-centric summarization
  - Extractive/Abstractive
  - User/Machine consumption

- Improve performance: coverage and accuracy

- More structure: E.g. Combining with storyline generation with event understanding to reason about events in aggregation

- More on User Generated Content: E.g. Linking to news stories/events