Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

(based on our CSCW17’ paper)

Allen Yilun Lin¹, Bowen Yu², Andrew Hall², Brent Hecht¹

1. People, Space, and Algorithms (PSA) Computing Group, Northwestern University

2. GroupLens Research, University of Minnesota, Twin Cities
Article-as-Concept Assumption?
Article-as-Concept Assumption

Concepts

Justin Trudeau
Angela Merkel
Theresa May

Wikipedia article

Justin Trudeau
Angela Merkel
Theresa May

Images from: https://icons8.com/icon/10904/wikipedia
What’s the problem?
Example 1:

Compare multilingual Wikipedias
English wiki v.s. Chinese wiki

English wiki v.s. Chinese wiki

Family information
Family info on **English** Wikipedia:

# of words: 559
% of the article: 3.86%

Trump has five children by three marriages, and has eight grandchildren.[28][29] His first two marriages ended in widely publicized...
Family info on **English** Wikipedia:

# of words: 559
% of the article: 3.86%

Family info on **Chinese** Wikipedia:

# of characters: 578
% of the article: 4.42%

Images from: [https://en.wikipedia.org/wiki/Donald_Trump](https://en.wikipedia.org/wiki/Donald_Trump) | [https://zh.wikipedia.org/wiki/%E5%94%90%E7%B4%8D%C2%B7%E5%B7%9D%E6%99%AE](https://zh.wikipedia.org/wiki/%E5%94%90%E7%B4%8D%C2%B7%E5%B7%9D%E6%99%AE)
Family info on **English** Wikipedia:
# of words: 559
% of the article: 3.86%

Family info on **Chinese** Wikipedia:
# of characters: 578
% of the article: 4.42%
Family info on **English** Wikipedia:
# of words: 559
% of the article: 3.86%

Family info on **Chinese** Wikipedia:
# of characters: 578
% of the article: 4.42%

Images from: [https://en.wikipedia.org/wiki/Donald_Trump](https://en.wikipedia.org/wiki/Donald_Trump) | [https://zh.wikipedia.org/wiki/%E5%94%90%E7%B4%8D%C2%B7%E5%B7%9D%E6%99%AE](https://zh.wikipedia.org/wiki/%E5%94%90%E7%B4%8D%C2%B7%E5%B7%9D%E6%99%AE)
Family info on English Wikipedia:
# of words: 559
% of the article: 3.86%

Family info on Chinese Wikipedia:
# of characters: 578
% of the article: 4.42%
Family info on **English** Wikipedia:
# of words: 559
% of the article: 3.86%

Family info on **Chinese** Wikipedia:
# of characters: 578
% of the article: 4.42%

Family info on **English** Wikipedia:
# of words: 2340
% of the article: 13.91%

Family info on **Chinese** Wikipedia:
# of characters: 578
% of the article: 4.42%

Example 2:

Artificial Intelligence using Wikipedia
Figure from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008.
Figure from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008
Figure inspired from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008
Milne-Witten Semantic Relatedness

Donald Trump

Vladimir Putin

inlinks

outlinks

Figure inspired from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008
Milne-Witten Semantic Relatedness

Figure inspired from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008
Foreign policy

Main article: Foreign policy of Donald Trump

Trump has been described as non-interventionist\(^{[609]}\)\(^{[610]}\) and nationalist.\(^{[611]}\) Trump repeatedly stated that he supports "America First" foreign policy.\(^{[612]}\) He supports increasing United States military defense spending,\(^{[611]}\) but favors decreasing United States spending on NATO and in the Pacific region.\(^{[613]}\) He says America should look inward, stop "nation building", and re-orient its resources toward

Japanese Prime Minister Shinzō Abe and Trump meeting in Trump Tower, Manhattan, on November 17, 2016

Foreign policy of the Donald Trump administration

From Wikipedia, the free encyclopedia

It has been requested that the title of this article be changed to Foreign policy of the Trump administration. Please see the relevant discussion on the discussion page. Do not move the page until the discussion has reached consensus for the change and is closed.

The foreign policy of Donald Trump is Donald Trump's policy regarding dealings with other nations.

Contents [hide]

1 Overview
   1.1 United States Armed Forces and defense spending
   1.2 Diplomacy and U.S. allies
   1.3 Action against terrorists’ families
   1.4 Conducting national security activities in public settings
2 Americas
   2.1 Cuba
   2.2 Mexico
3 Asia
   3.1 Afghanistan
   3.2 China and Taiwan

Images from: https://en.wikipedia.org/wiki/Foreign_policy_of_the_Donald_Trump_administration
Figure inspired from Witten, Ian, and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." AAAI Workshop 2008
Milne-Witten Semantic Relatedness

Donald Trump

Rex Tillerson  G20  Michael Flynn

Donald Trump

ExxonMobil  APEC  Most powerful people

US-Mexican border

Republican

Trump organization

Economy of Russian

2014 Winter Olympics

KGB  Kremlin

inlinks  outlinks

Image source: https://www.christart.com/clipart/image/off-target
Problematizing and Addressing the *Article-as-Concept* Assumption in Wikipedia

Concepts

- Donald Trump

Wikipedia article

- Donald Trump
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

Concepts

Donald Trump

Wikipedia article

Donald Trump
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

- Donald Trump
- Donald Trump
- Family of Donald Trump
- Foreign Policy of Trump Administration

Concepts

Wikipedia article
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

Concepts

Donald Trump

Wikipedia article

Donald Trump

Family of Donald Trump

Foreign Policy of Trump Administration

Parent article

... ...
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

Concepts

Donald Trump

Wikipedia article

Donald Trump

Family of Donald Trump

Foreign Policy of Trump Administration

Parent article

Sub-article
Problematicizing and Addressing the Article-as-Concept Assumption in Wikipedia
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

Concepts

Donald Trump

Wikipedia article

Donald Trump

Family of Donald Trump

Foreign Policy of Trump Administration

Parent article

... ...

Sub-article
Outline

1. Motivation and problem
2. Step 1: Filter sub-article candidates
3. Step 2: Develop training datasets
4. Step 3: Classify sub-article relationship
5. Discussion: audience-author mismatch
6. Summary
Step 1: Filter sub-article candidates
Reduce the # of comparisons?
Step 1: Filter sub-article candidates

Personal life

Early life

Further information: Trump family

Trump was born on June 14, 1946, at the Jamaica Hospital Medical Center in the Jamaica neighborhood of Queens, New York City. He was the fourth of five children born to Frederick Christ "Fred" Trump (1905–1999) and Mary Anne Trump (née MacLeod, 1912–2000). His siblings are Maryanne, Fred Jr., Elizabeth, and Robert. Trump’s older brother Fred Jr. died in 1981 from alcoholism, which Trump says led him to abstain from alcohol and cigarettes.[4]

Family

Main article: Family of Donald Trump

Trump has five children by three marriages, and has eight grandchildren.[32][33] His first two marriages ended in widely publicized divorces.[34]

Step 1: Filter sub-article candidates

Early life

Further information: Trump family

Trump was born on June 14, 1946, at the Jamaica Hospital Medical Center in the Queens neighborhood of New York City. He was the fourth of five children in the family. His parents were "Fred" Trump Sr. (1905–1999) and Mary Anne Trump (née MacLeod). His siblings are Maryanne, Fred Jr., Elizabeth, and Robert. Trump says that his father became an alcoholic, which led him to abstain from alcohol at an early age.

Family

Main article: Family of Donald Trump

Trump has five children by three marriages, and has eight grandchildren. His first two marriages ended in widely publicized divorces. Trump married his first wife, Czech model Ivana Zelníčková, on April 7, 1977, at the Marble Collegiate Church in Manhattan in a ceremony performed by one of America's most famous ministers, the Reverend Norman Vincent Peale. Images from: https://en.wikipedia.org/wiki/Donald_Trump
Step 1: Filter sub-article candidates

Early life

Further information: Trump family

Trump was born on June 14, 1946, at the Jamaica Hospital Medical Center in the neighborhood of Queens, New York City. He was the fourth of five children born to "Fred" Trump (1905–1999) and Mary Anne Trump (née MacLeod). Trump’s older brother Robert was an alcoholic, which Trump says led him to abstain from alcohol and drugs.

Family

Main article: Family of Donald Trump

Trump has five children by three marriages, and has eight grandchildren. His first two marriages ended in widely publicized divorces. Trump married his first wife, Czech model Ivana Zelníčková, on April 7, 1977, at the Marble Collegiate Church in Manhattan in a ceremony performed by one of America’s most famous ministers, the Reverend Norman Vincent Peale.

See also

- List of Presidents of the United States
- List of richest American politicians
- Public image of Donald Trump

Step 1: Filter sub-article candidates

Step 1: Filter sub-article candidates

Donald Trump

45th President of the United States
Incumbent

The Apprentice
Main articles: The Apprentice (U.S. TV series) and The Apprentice (TV series)

Family
Main article: Family of Donald Trump

Palm Beach estate
Main article: Mar-a-Lago

Outline

1. Motivation and problem
2. Step 1: Filter sub-article candidates
3. Step 2: Develop training datasets
4. Step 3: Classify sub-article relationship
5. Discussion: audience-author mismatch
6. Summary
Step 2: Develop training datasets

High-Interest

Donald Trump  Melania Trump

La La Land  Super Bowl Champions
Step 2: Develop training datasets

High-Interest
- Donald Trump
- Melania Trump
- La La Land
- Super Bowl Champions

Random
- Bunbury, Western Australia
- Jean Pasqualini
- List of tallest building in China
Step 2: Develop training datasets

High-Interest:
- Donald Trump
- Melania Trump
- La La Land
- Super Bowl Champions

Random:
- Bunbury, Western Australia
- Jean Pasqualini
- List of tallest building in China

Ad-Hoc:
- Software engineering
- Geostatistics
- Feminism
Step 2: Develop training datasets

High-Interest

- Donald Trump
- Melania Trump
- La La Land
- Super Bowl Champions

Random

- Bunbury, Western Australia
- Jean Pasqualini
- List of tallest building in China

Ad-Hoc

- Software engineering
- Geostatistics
- Feminism
Outline

1. Motivation and problem
2. Step 1: Filter sub-article candidates
3. Step 2: Develop training datasets
4. Step 3: Classify sub-article relationship
5. Discussion: audience-author mismatch
6. Summary
Step 3: Classify sub-article relationship

Relative importance/centrality

\[ E.g. \text{IndegreeRatio} = \]

\[
\frac{\text{Incoming link of parent article}}{\text{Incoming link of sub-article}}\]

Step 3: Classify sub-article relationship

Relative importance/centrality

\[ \text{E.g. IndegreeRatio} = \frac{\text{Incoming link of parent article}}{\text{Incoming link of sub-article}} \]

Implicit Linguistic

\[ \text{E.g. MaxMainTFinSub} = \max\{\text{TF of parent article in sub-article’s summary}\} \]
Step 3: Classify sub-article relationship

Parent article

Donald Trump

From Wikipedia, the free encyclopedia

This article is about the incumbent President of the United States. For other uses, see Donald Trump (disambiguation).

Donald John Trump (born June 14, 1946) is an American businessman, television personality, politician, and the 45th President of the United States. Born and raised in Jamaica, Queens, New York City, Trump received an economics degree from the Wharton School of the University of Pennsylvania in 1968. In 1971, he took charge of his family real

Sub-article candidate

Foreign policy of the Donald Trump administration

From Wikipedia, the free encyclopedia

(Redirected from Foreign policy of Donald Trump)

The foreign policy of Donald Trump is Donald Trump's policy regarding dealings with other nations.

Contents [hide]

1 Overview

Step 3: Classify sub-article relationship

Parent article

Donald Trump

This article is about the incumbent President of the United States. For other uses, see Donald Trump (disambiguation).

Donald John Trump (born June 14, 1946) is an American businessman, television personality, politician, and the 45th President of the United States.

Born and raised in Jamaica, Queens, New York City, Trump received an economics degree from the Wharton School of the University of Pennsylvania in 1968. In 1971, he took charge of his family real


Sub-article candidate

Foreign policy of the Donald Trump administration

From Wikipedia, the free encyclopedia

(Redirected from Foreign policy of Donald Trump)

The foreign policy of Donald Trump is Donald Trump's policy regarding dealings with other nations.

Contents [hide]

1 Overview

{Donald Trump, Foreign policy of Donald Trump}
Step 3: Classify sub-article relationship

Donald Trump
Donald John Trump (born June 14, 1946) is an American businessman, television personality, politician, and the 45th President of the United States. Born and raised in Jamaica, Queens, New York City, Trump received an economics degree from the Wharton School of the University of Pennsylvania in 1968. In 1971, he took charge of his family real

Foreign policy of the Donald Trump administration
The foreign policy of Donald Trump is Donald Trump's policy regarding dealings with other nations.

Contents [hide]
1 Overview


{Donald Trump, Foreign policy of Donald Trump} = 2
Step 3: Classify sub-article relationship

Relative importance/centrality

- **E.g. IndegreeRatio** =
  
  \[ \text{Incoming link of parent article} \div \text{Incoming link of sub-article} \]

Implicit Linguistic

- **E.g. MaxMainTFinSub** =
  
  \[ \text{max\{TF of parent article in sub-article’s summary\}} \]

Explicit Template

- **E.g. MainTemplateRatio** =
  
  \[ \frac{\text{# langs using main template}}{\text{# langs two articles coexists}} \]
Step 3: Classify sub-article relationship

Foreign policy

*Main article: Foreign policy of Donald Trump*

Trump has been described as non-interventionist[^630] and nationalist.[^632] Trump has repeatedly stated that he supports "America First" foreign policy.[^633] He supports increasing United States military defense spending,[^632] but favors decreasing United States spending on NATO and in the Pacific region.[^634] He says America should look toward "self-reliance..." and is against...

Main template used in English and French Wikipedia

{Donald Trump, Foreign policy of Donald Trump}

Step 3: Classify sub-article relationship

Main template used in English and French Wikipedia

Two articles coexist in 5 languages

\{Donald Trump, Foreign policy of Donald Trump\}

Step 3: Classify sub-article relationship

Main template used in English and French Wikipedia

Two articles coexist in 5 languages

\{Donald Trump, Foreign policy of Donald Trump\} = \frac{2}{5}

Classification accuracies across datasets, averaged across languages

Accuracies

- High-Interest
- Ad-Hoc
- Random

Baseline vs Our model
Classification accuracies across datasets, averaged across languages

Classification accuracies across datasets, averaged across languages

High-Interest  Ad-Hoc  Random

Baseline  Our model

Donald Trump  Software Engineering

Classification accuracies across datasets, averaged across languages

- **High-Interest**: Donald Trump
- **Ad-Hoc**: Software Engineering
- **Random**: Bunbury, AU

![Bar chart showing classification accuracies for different datasets.](https://en.wikipedia.org/wiki/Donald_Trump)

![Bar chart showing classification accuracies for different datasets.](https://en.wikipedia.org/wiki/Software_engineering)

![Bar chart showing classification accuracies for different datasets.](https://en.wikipedia.org/wiki/Bunbury,_Western_Australia)
Performance: High-Interest dataset

Classification accuracies on High-interests dataset
by language

Accuracies
Real impact of article-as-concept assumption?
Real impact of article-as-concept problem

<table>
<thead>
<tr>
<th>Impact Metric</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of articles w/ sub-articles</td>
<td>70.8%</td>
</tr>
<tr>
<td>% of page views to articles w/ sub-articles</td>
<td>71.0%</td>
</tr>
<tr>
<td>Avg # of sub-article per article</td>
<td>7.5</td>
</tr>
</tbody>
</table>
Outline

1. Motivation and problem
2. Step 1: Filter sub-article candidates
3. Step 2: Develop training datasets
4. Step 3: Classify sub-article relationship
5. Discuss: audience-author mismatch
6. Summary
Discussion: audience-author mismatch

Audience

Communication
Discussion: audience-author mismatch

Images from: https://www.youtube.com/watch?v=ikkg4NobV_w
Discussion: audience-author mismatch

Audience

Author

Communication

Wikipedia

Images from: https://www.youtube.com/watch?v=ikkg4NobV_w
Images from: Wikimedia Commons
Images from: https://clubrunner.blob.core.windows.net/00000050227/Images/community_icon-300x300.png
Article-as-concept in Wikidata
Family of Donald Trump

From Wikipedia, the free encyclopedia.

This article is about the immediate family of Donald Trump. For the family of his predecessor, see First Family of the United States. For his ancestors from Germany, see Trump family (Germany).

The family of Donald Trump, the President of the United States, is a prominent American family active in real estate, entertainment, business, and politics. Donald Trump's immediate family circle is the First Family of the United States. They are part of the broader Trump family originating from Germany. Donald Trump has children from three different women; Ivana Trump, Maria Maples, and Melania Trump.

Contents

1 Immediate family
   1.1 Ivana Trump
   1.2 Maria Maples
   1.3 Melania Trump
   1.4 Children
      1.4.1 Don Jr., Ivanka, and Eric Trump
      1.4.2 Tiffany Trump
      1.4.3 Barron Trump
   1.5 Grandchildren
2 Ancestry
   2.1 Parents
      2.1.1 Fred Trump
      2.1.2 Mary Anne MacLeod Trump
   2.2 Grandparents
      2.2.1 Frederick Trump
      2.2.2 Elizabeth Christ Trump
   2.3 Relatives
      2.3.1 John G. Trump
      2.3.2 Fred Trump Jr.
      2.3.3 Maryanne Trump Barry
      2.3.4 Robert Trump
   2.4 Genealogical tables

Family of Donald Trump

This article is about the immediate family of Donald Trump. For the family originating from Germany, see the First Family of the United States. They are part of the broader Trump family originating from three different women; Ivana Trump, Maria Maples, and Melania Trump.

<table>
<thead>
<tr>
<th>Contents [hide]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Immediate family</td>
</tr>
<tr>
<td>1.1 Ivana Trump</td>
</tr>
<tr>
<td>1.2 Maria Maples</td>
</tr>
<tr>
<td>1.3 Melania Trump</td>
</tr>
<tr>
<td>1.4 Children</td>
</tr>
<tr>
<td>1.4.1 Don Jr., Ivanka, and Eric Trump</td>
</tr>
<tr>
<td>1.4.2 Tiffany Trump</td>
</tr>
<tr>
<td>1.4.3 Barron Trump</td>
</tr>
<tr>
<td>1.5 Grandchildren</td>
</tr>
<tr>
<td>2 Ancestry</td>
</tr>
<tr>
<td>2.1 Parents</td>
</tr>
<tr>
<td>2.1.1 Fred Trump</td>
</tr>
<tr>
<td>2.1.2 Mary Anne MacLeod Trump</td>
</tr>
<tr>
<td>2.2 Grandparents</td>
</tr>
<tr>
<td>2.2.1 Frederick Trump</td>
</tr>
<tr>
<td>2.2.2 Elizabeth Christl Trump</td>
</tr>
<tr>
<td>2.3 Relatives</td>
</tr>
<tr>
<td>2.3.1 John G. Trump</td>
</tr>
<tr>
<td>2.3.2 Fred Trump Jr.</td>
</tr>
<tr>
<td>2.3.3 Maryanne Trump Barry</td>
</tr>
<tr>
<td>2.3.4 Robert Trump</td>
</tr>
<tr>
<td>2.4 Distant relatives</td>
</tr>
</tbody>
</table>

Donald Trump's immediate family circle is a prominent American family active in real estate, entertainment, business, and politics. The Donald Trump family originated from three different women; Ivana Trump, Maria Maples, and Melania Trump.

家

特朗普有一個兄弟和兩個姊妹。他的姊妹玛丽安娜·特朗普·巴里是聯邦高等法院的法官。他的父親具德國血統，而母親來自蘇格蘭，17歲移民至美國紐約從事僱工，1936年認識特朗普的父親，並於1942年取得美國國籍。

特朗普先後有三任妻子，育有三子兩女。妻子分別為:
facet of (P1269)

topic of which this item is an aspect, item that offers a broader perspective on the same topic

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>facet of</td>
<td>topic of which this item is an aspect, item that offers a broader perspective on the same topic</td>
</tr>
<tr>
<td>Chinese</td>
<td>父話題</td>
<td>此主題的主要方面</td>
</tr>
<tr>
<td>Spanish</td>
<td>faceta de</td>
<td>tema del cual este elemento es un asp</td>
</tr>
<tr>
<td>Traditional Chinese</td>
<td>所屬話題</td>
<td>此主題的主要方面</td>
</tr>
</tbody>
</table>
Limitations and Future Works

● Sub-article relationships that are explicitly encoded by editors
  ○ In future, search for implicit relationship
Limitations and Future Works

- Sub-article relationships that are explicitly encoded by editors
  - In future, search for implicit relationship
- Good on High-Interest articles but struggled on random articles
  - In future, specific modelling on random articles
Limitations and Future Works

● Sub-article relationships that are explicitly encoded by editors
  ○ In future, search for implicit relationship
● Good on High-Interest articles but struggled on random articles
  ○ In future, specific modelling on random articles
● Integrate our system into Wikipedia-based technologies
  ○ E.g. MilneWitten
Summary

● Identified and problematized the article-as-concept assumption
Summary

- Identified and problematized the article-as-concept assumption
- Described the first model to address the problem
Summary

- Identified and problematized the article-as-concept assumption
- Described the first model to address the problem
- Discussed the audience-author mismatch in peer production
Summary

- Identified and problematized the article-as-concept assumption
- Described the first model to address the problem
- Discussed the audience-author mismatch in peer production
- We are releasing the system and the datasets
Problematizing and Addressing the Article-as-Concept Assumption in Wikipedia

Twitter: @bhecht, @cheetah90
Email: allen.lin@eecs.northwestern.edu

Allen Yilun Lin¹, Bowen Yu², Andrew Hall², Brent Hecht¹

1. People, Space, and Algorithms (PSA) Computing Group, Northwestern University
2. GroupLens Research, University of Minnesota, Twin Cities