Reading and using CS research

A case study: summarization

Introduction

Hi, I'm Abe.

Student in NLP at UMass: http://slanglab.cs.umass.edu/

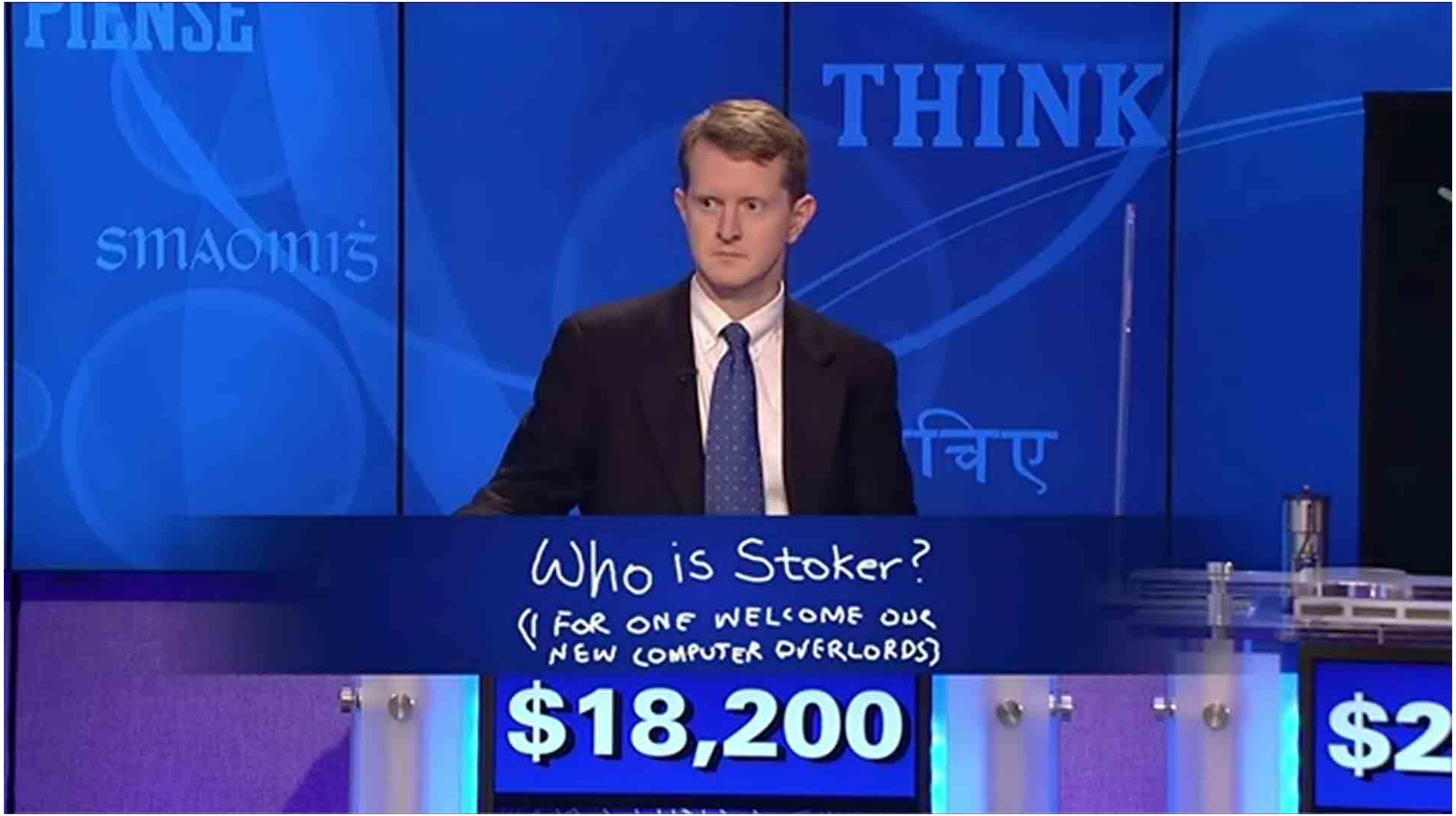
I'm in my 4nd year of the MS/PhD program

Before I became interested in NLP, I worked as a software developer and data journalist

http://www.abehandler.com

I am not an expert researcher! I'm just a grad student sharing what I have learned

IBM Watson


PIENSE
THINK
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$18,200

You can do neat stuff without fancy research methods

So why bother with research?

- CS researcher and you want to do research
- You are a curious person and you want to understand the world around you
- You are on a team that uses research methods

You have a problem that does not have an established solution!

databases vs. summarization

Rookie demo

Find out about a news topic. Build an interactive summary.

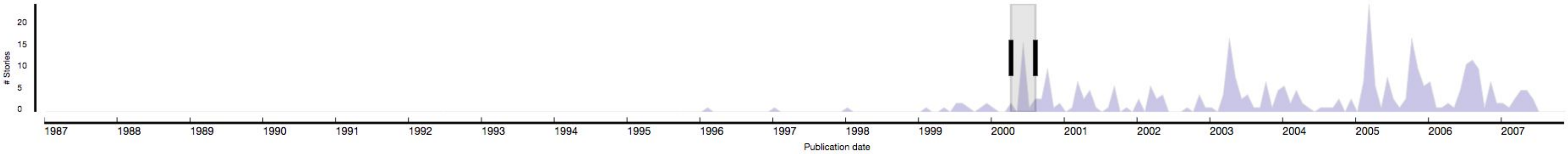
Why is Rookie different?

- Interactive speed
  - Waiting a few seconds for a program is unacceptable
- Criteria for success
  - Task-based. Not supervised learning.
  - What is a "summary"?
- No method does what it needs

Bashar al-Assad

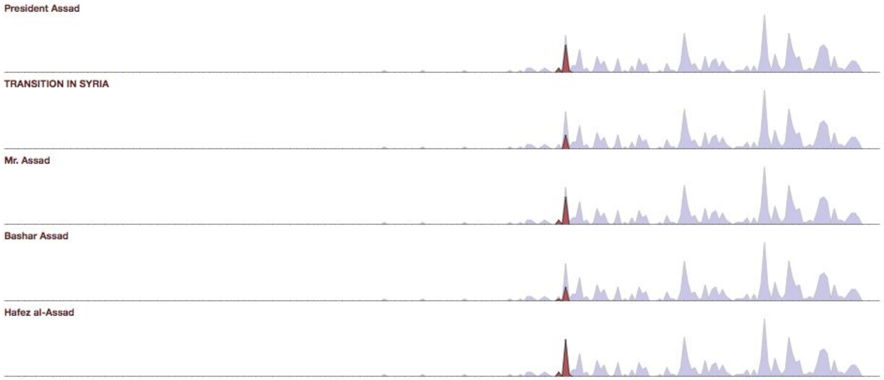
Submit

Found 342 documents for **Bashar al-Assad**



Subjects related to **Bashar al-Assad**

more subjects
page 1 of 28



Summary of 22 documents for **Bashar al-Assad** from **Apr. 2000 — Aug. 2000**X

More →

Jun. 13 2000 | But **Bashar al-Assad**'s apparently smooth -- and so far automatic -- rise to power was challenged by his uncle, Rifaat al-Assad, who has been living in exile in Europe since an unsuccessful attempt to unseat his brother in 1983. ''

Jun. 14 2000 | This was **Bashar al-Assad**'s first real day on the job as the ruler of Syria.

Jun. 14 2000 | The lanky figure of the new leader, **Bashar al-Assad**, the 34-year-old son of Hafez al-Assad, towered over the secretary.

Apr. 18 2000 | The time is not too late and it has not run out,'' said **Bashar al-Assad**, who is considered by many to be the heir apparent to his ailing father and who has been given a considerable role in Syrian affairs. ''

Summarization

Given a collection of one or more documents of length L, produce a short passage that both reads fluidly and includes the "most crucial" information from L.

"summarization"

- Compress a sentence
- Generate a headline
- Compress an article
- Compress lots of articles
- Delete words
- Generate paraphrases
- Query-focused
- Structured
- Domain-specific

What's going on today in the pioneer valley?

Pioneer Valley Volleyball @pvhs_vball · Oct 25
Panthers drop set 1. (25-14)

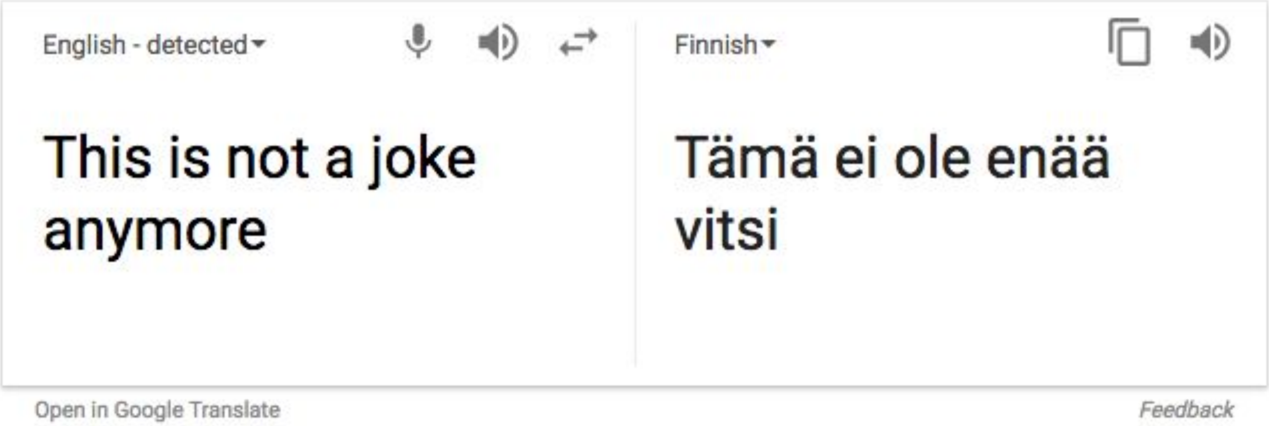
Pioneer Valley Volleyball @pvhs_vball · Oct 25
8-16 Dinuba. PV TIME-OUT

Stockdale HS @stockdalehigh · 18h
Stockdale Football will host **Pioneer Valley** of Santa Maria this Friday in the opening round of the CIF Central Section playoffs. Go Mustangs!

15 49

Scott Foster @AttyScottFoster · 4h
#wmij2018 Good Times working with student #innovators from across the **#PioneerValley**

3


English - detected
Finnish
This is not a joke anymore
Tämä ei ole enää vitsi
Open in Google Translate
Feedback

General research tips

Make a notation sheet

Find a survey paper

Check conferences for recent work

- In NLP: NACCL, ACL, EMNLP, TACL

Learn the "big names" of people who publish in an area

Has this been done? How can you modify existing work?

What is different or unique about your approach?

- Corpus, method, linguistics, math, formalism, application ….

Advice: Big names

Learning the "big names". For summarization:

- Kathleen McKeown
- Sasha Rush
- Regina Barzilay

How to find the "big names"?

- Read a bunch of papers
- https://www.semanticscholar.org
  - Limit to NLP conferences, turn on the author facets. The big names "pop out".

Advice: find a survey paper

When you start a research project, you often don't even know what you are trying to do.

Das and Martin

A Survey on Automatic Text Summarization

- Really old problem (1960s)
- Many formulations
- Lots of disagreement on what you are trying to do
    - Preserve "useful" information
    - Be shorter (compression rate parameter)

Advice: learn the lingo, notation and key concepts

Lingo and notation

- In this class, you try to use consistent names/terms/notation.
- In research, this is often not the case.
- Especially if you want to bridge communities.

Summarization lingo

- Extraction
- Abstraction
- Fusion
- Single vs. multi document
- "Query-focused"
- Sentence compression
- DUC (document understanding conference)
- TREC (text retrieval conference)

Is extractive or abstractive summarization easier?

WASHINGTON — Over the past two decades, Taiwan has slipped from its position atop the list of flash points in the complex relationship between the United States and China. In meetings between President Obama and President Xi Jinping of China, it has typically come up after half a dozen more pressing issues, like trade, cyberattacks and Beijing's aggressive moves in the South China Sea.

Now, though, in a single protocol-shattering phone call with the president of Taiwan, President-elect Donald J. Trump has thrust it back on the table. Not since President Richard M. Nixon met with Mao Zedong in 1972 — when the two issued the Shanghai Communiqué clarifying the status of Taiwan — has an American leader so shaken up the diplomatic status quo on the issue.

"Taiwan is about to become a more prominent feature of the overall U.S.-China relationship," said Jon M. Huntsman, who served as ambassador to China during Mr. Obama's first term. "As a businessman, Donald Trump is used to looking for leverage in any relationship. A President Trump is likely to see Taiwan as a useful leverage point."

Step one: how to find important words?

Some approaches

- word frequency (Luhn 1958)
- word importance (tf-idf, pmi)
- word position: super useful (e.g. scientific abstracts, news...)
- key phrases. "significant" and "we can see that"
- "statistical revolution" (1990s)
- neural networks return (last few years)

Why not just count the most frequent words?

Tf-idf *

- Term frequency = how many times a word type occurs in a document
- Inverse document frequency = 1/(count of documents that contain a word)
- Tf-idf score = tf * idf

* many very similar formulations

Intuition tf-idf

- You have a corpus of 10,000 recent NYT world news articles

- Of 10,000 articles total, 100 contain the word Taiwan
  - document frequency = 100
  - Inverse document frequency = 1/100 = 1/100

- Say 2000 contain the word Obama
  - document frequency = 2000
  - Inverse document frequency = 1/2000

TF and IDF

TF => how many times a term was mentioned

DF => how many documents contain that term

IDF => 1/DF

tf-idf is only one way to find important words

- PMI
- Importance to a "topic" (LDA)
- Proximity to a query word in vector space
- Raw count
- Many more...

Say you searched Google for "Amherst, MA"

What words would have high term frequencies?

What words would have low document frequencies?

WASHINGTON — Over the past two decades, Taiwan has slipped from its position atop the list of flash points in the complex relationship between the United States and China. In meetings between President Obama and President Xi Jinping of China, it has typically come up after half a dozen more pressing issues, like trade, cyberattacks and Beijing's aggressive moves in the South China Sea.

Now, though, in a single protocol-shattering phone call with the president of Taiwan, President-elect Donald J. Trump has thrust it back on the table. Not since President Richard M. Nixon met with Mao Zedong in 1972 — when the two issued the Shanghai Communiqué clarifying the status of Taiwan — has an American leader so shaken up the diplomatic status quo on the issue.

"Taiwan is about to become a more prominent feature of the overall U.S.-China relationship," said Jon M. Huntsman, who served as ambassador to China during Mr. Obama's first term. "As a businessman, Donald Trump is used to looking for leverage in any relationship. A President Trump is likely to see Taiwan as a useful leverage point."

Taiwan = 5 * 1/100 = 5/100

Obama = 2 * 1/2000 = 2/2000

WASHINGTON — Over the past two decades, Taiwan [1] has slipped from its position atop the list of flash points in the complex relationship between the United States and China. [redacted]

Now, though, in a single protocol-shattering phone call with the president of Taiwan, [2] President-elect Donald J. Trump has thrust it back on the table. Not since President Richard M. Nixon met with Mao Zedong in 1972 — when the two issued the Shanghai Communiqué clarifying the status of Taiwan [3] — has an American leader so shaken up the diplomatic status quo on the issue.

[redacted] A President Trump is likely to see Taiwan [4] as a useful leverage point.”

A summary that favors the word Taiwan (tf-idf = 5/100), drops Obama (tf-idf = 2/2000) [redacted] In meetings between President Obama[1] and President Xi Jinping of China, it has typically come up after half a dozen more pressing issues, like trade, cyberattacks and Beijing's aggressive moves in the South China Sea.

[redacted]

"Taiwan is about to become a more prominent feature of the overall U.S.-China relationship," said Jon M. Huntsman, who served as ambassador to China during Mr. Obama[2]'s first term. [redacted]

A summary that favors the word Obama (2/2000) and drops Taiwan tf-idf = 5/100

Which summary is better?

Why?

What other kinds of information would you need besides word importance to make a summary?

How to win DUC 2002

WASHINGTON — Over the past two decades, Taiwan has slipped from its position atop the list of flash points in the complex relationship between the United States and China. In meetings between President Obama and President Xi Jinping of China, it has typically come up afte half a dozen more pressing issues, like trade, cyberattacks and Beijing’ aggressive moves in the South China Sea.

~~Now, though, in a single protocol-shattering phone call with the president of Taiwan, President-elect Donald J. Trump has thrust it bac on the table. Not since President Richard M. Nixon met with Mao Zedong in 1972 — when the two issued the Shanghai Communiqué clarifying the status of Taiwan — has an American leader so shaken up the diplomatic status quo on the issue.~~

~~“Taiwan is about to become a more prominent feature of the overall U.S.-China relationship,” said Jon M. Huntsman, who served as ambassador to China during Mr. Obama’s first term. “As a businessman, Donald Trump is used to looking for leverage in any relationship. A President Trump is likely to see Taiwan as a useful leverage point.”~~

- h/t Chris Kenzie (Columbia PhD student)

Are these examples extractive or abstractive?


LEO

Evaluation:

How do you know a summarization method is good?

When you read a paper: why do you believe the authors?

Human evaluation: How could you test a summarization system on people?

What are the downsides to this method?

Automatic evaluation

ROUGE-N: how many N-grams from gold did you get?

Input: President Donald Trump fired another warning shot Sunday at U.S. companies considering moving their operations out of the country, threatening “retributions or consequences” such as a hefty border tax if they do.

"Gold standard": Donald **Trump** threatened to penalize **companies** who move jobs **out of the country**.

Some system: Trump fired shot at **companies** moving operations **out of the country**.

ROUGE-1 = 6/13

ROUGE-2 = 3/12

Who can find a problem with ROUGE-N ?

Switch from summarization to translation in next examples

Gaming automatic evaluation (precision)

Precision: what percentage of the words in your translation occur in the reference translation?

Input: el gato esta en la estera

precision?

Human: the cat is on the mat

Output: the the the the the the

Precision = total correct / total guesses

Gaming automatic evaluation (precision)

Precision: what percentage of the words in your candidate (machine) translation occur in the reference translation?

Input: el gato esta en la estera
Output: the the the the the the
Human: The cat is on the mat

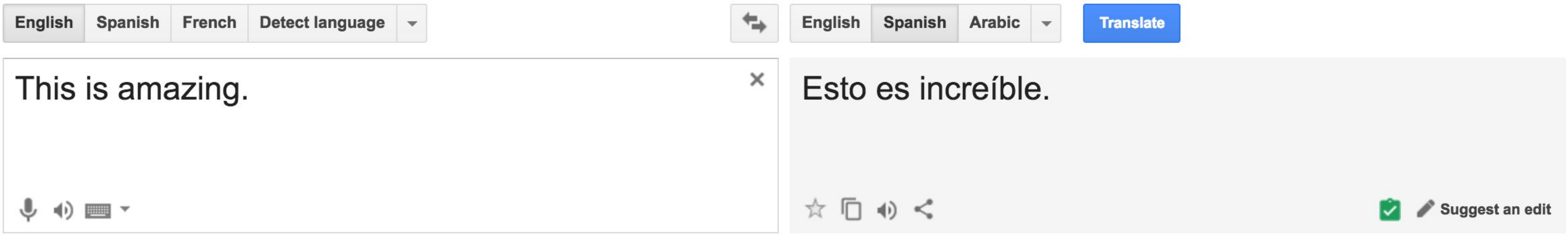
6 out of 6!

Precision = total correct / total guesses

Who can defend ROUGE-N?

- clear
- simple
- interpretable
- fast

Defense of ROUGE-N



	BLEU	Log Perplexity	Decoding time (s)
CPU	31.20	1.4553	1322
GPU	31.20	1.4553	3028
TPU	31.21	1.4626	384

https://arxiv.org/pdf/1609.08144v2.pdf

BLEU scores

We compute the brevity penalty BP,

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \le r \end{cases}.$$

Then,

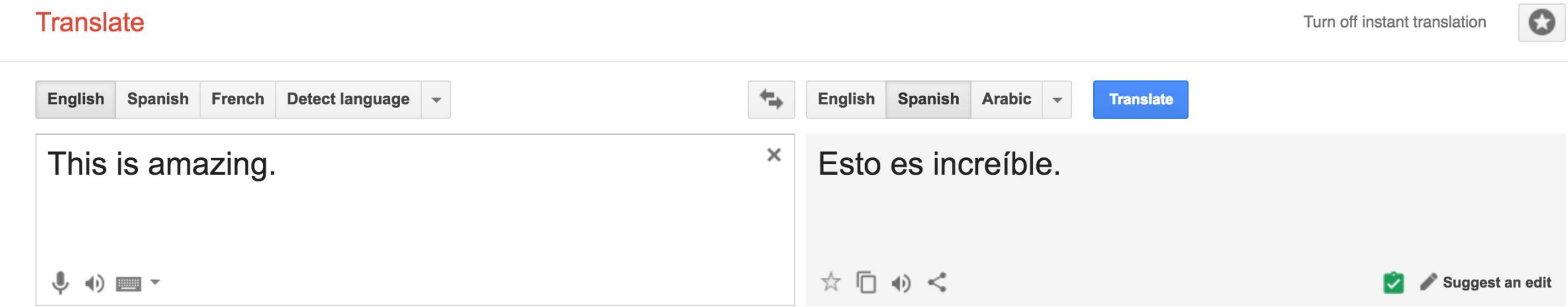
$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \boxed{\log p_n}\right).$$

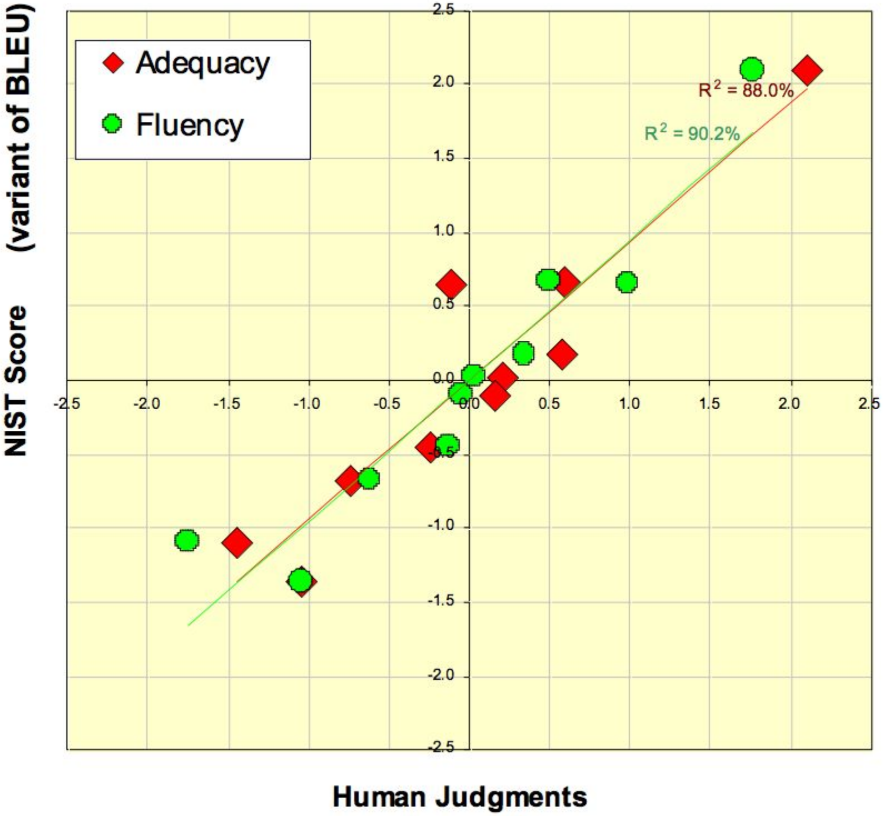
The ranking behavior is more immediately apparent in the log domain,

$$\log \text{BLEU} = \min(1 - \frac{r}{c}, 0) + \sum_{n=1}^{N} w_n \log p_n.$$

http://www.aclweb.org/anthology/P02-1040.pdf

The BLEU score matches human judgement (in some way)





Source :http://web.stanford.edu/class/cs224n/handouts/cs224n-lecture3-MT.pdf

Too much BLEU?

People started optimizing their systems to maximize BLEU score – BLEU scores improved rapidly. The correlation between BLEU and human judgments of quality went way, way down

Coming up with automatic MT evaluations has become its own research field – TER, METEOR, MaxSim, SEPIA, RTE-MT, TERpA

Source :http://web.stanford.edu/class/cs224n/handouts/cs224n-lecture3-MT.pdf

Takeaways

- Without evaluation, you are arguably not doing science
- Auto evaluation is helpful
- Clear, simple, interpretable evaluations are also good
- Human judgement is the ultimate evaluation. If your translation is very choppy or makes no sense, but has a great BLEU score, nobody cares.
- Research communities have set ways of evaluating things. If you submitted a paper on a summarization technique with no ROGUE score, people would not trust your technique.

Oh yeah, machine learning …

Machine learning approaches (via Das)

naive Bayes: Edmunson 1969 and Kupiec 1995

- features (e.g. position, cue words, tf-idf scores)

- classes (class 1 = included, class 2 = not included)

- **HMMs**: Conroy and O'Leary. Sequence model where each sentence gets 1,0 inclusion

- CRFs: Lu Wang. A Sentence Compression Based Framework to Query-Focused Multi-Document Summarization

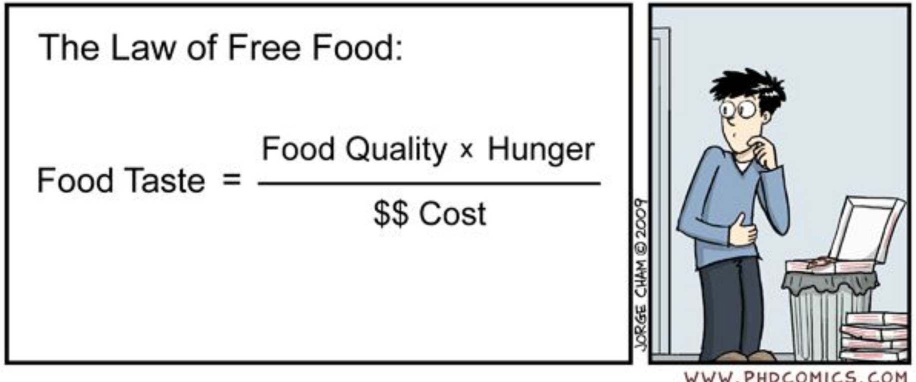
Where do "gold" summaries come from?

Where does this training data come from?

'Don't believe a thing you hear, unless it comes from me': Megyn Kelly responds to reports she is leaving Fox News for CNN by urging fans to ignore rumors - but doesn't rule the move out

- **Megyn Kelly says there is no truth that CNN is reportedly trying to get her to join the network when her Fox News contract expires in July**
- **But she did not rule out the move, and told fans to only believe information that comes from her**
- **Last week, an insider told Drudge Report the network wanted Kelly to anchor either the 8pm or 9pm hour on weeknights**
- **Reports suggested CNN would not match the $20million offer Fox News has already put on the table to get Kelly to re-sign her contract**

http://www.dailymail.co.uk/news/article-3998964/Megyn-Kelly-responds-rumors-leaving-Fox-CNN-saying-don-t-believe-thing.html



OMG papers

Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org

Login

Search or Article ID | All fields

(Help | Advanced search)

Open access to 1,456,318 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance, Statistics, Electrical Engineering and Systems Science, and Economics

Subject search and browse: Physics | Search | Form Interface | Catchup

5 Sept 2018 arXiv looks to the future with move to Cornell CIS
23 Jul 2018: Theoretical Economics and General Economics subject areas added to arXiv
18 Jul 2018: Search interface updated to version 0.4
See cumulative "What's New" pages. Read robots beware before attempting any automated download

Physics

- Astrophysics (**astro-ph** new, recent, search)
  includes: Astrophysics of Galaxies; Cosmology and Nongalactic Astrophysics; Earth and Planetary Astrophysics; High Energy Astrophysical Phenomena; Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- Condensed Matter (**cond-mat** new, recent, search)
  includes: Disordered Systems and Neural Networks; Materials Science; Mesoscale and Nanoscale Physics; Other Condensed Matter; Quantum Gases; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- General Relativity and Quantum Cosmology (**gr-qc** new, recent, search)
- High Energy Physics – Experiment (**hep-ex** new, recent, search)
- High Energy Physics – Lattice (**hep-lat** new, recent, search)
- High Energy Physics – Phenomenology (**hep-ph** new, recent, search)
- High Energy Physics – Theory (**hep-th** new, recent, search)
- Mathematical Physics (**math-ph** new, recent, search)
- Nonlinear Sciences (**nlin** new, recent, search)
  includes: Adaptation and Self-Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons
- Nuclear Experiment (**nucl-ex** new, recent, search)
- Nuclear Theory (**nucl-th** new, recent, search)
- Physics (**physics** new, recent, search)
  includes: Accelerator Physics; Applied Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and Molecular Clusters; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History and Philosophy of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics Education; Physics and Society; Plasma Physics; Popular Physics; Space Physics
- Quantum Physics (**quant-ph** new, recent, search)

arXiv.org

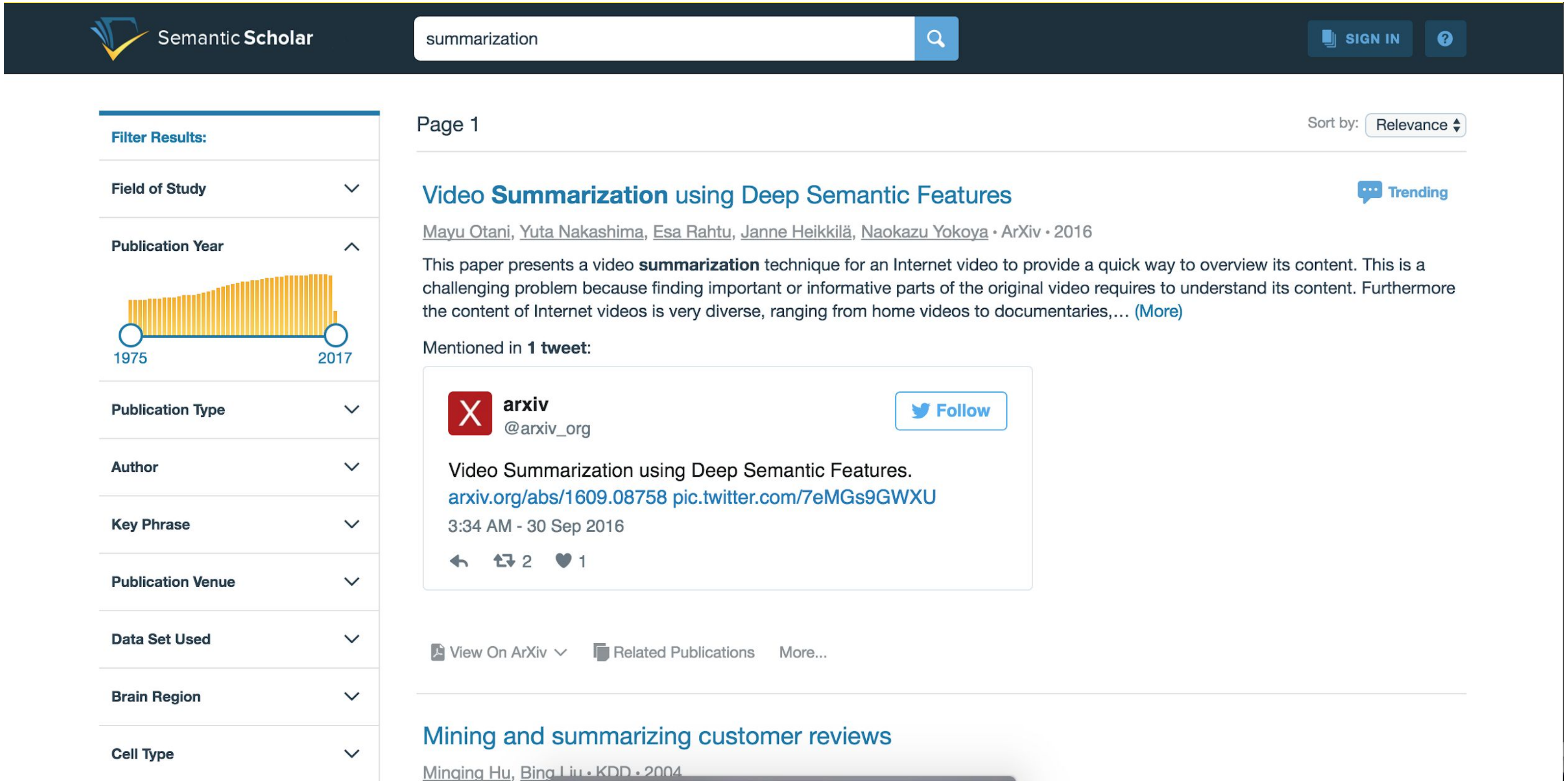
daily feed

Inbox 10
Snoozed
Important
Chats
Sent
Drafts 759
All Mail
Spam 44

\\
arXiv:1810.10566
Date: Wed, 24 Oct 2018 18:18:19 GMT (97kb,D)

Title: Clinical Concept Extraction with Contextual Word Embedding
Authors: Henghui Zhu, Ioannis Ch. Paschalidis, Amir Tahmasebi
Categories: cs.CL
Comments: Submitted to NIPS ML4H workshop
\\
Automatic extraction of clinical concepts is an essential step for turning the unstructured data within a clinical note into structured and actionable information. In this work, we propose a clinical concept extraction model for automatic annotation of clinical problems, treatments, and tests in clinical notes utilizing domain-specific contextual word embedding. A contextual word

semanticscholar.org



heuristics

- The author is someone who you know to do good work
  - Michael Collins, Regina Barzilay, Tommi Jaakkola, Percy Liang, Chris Manning
- The author is in a lab with people you know to do good work
- The paper is very well-written
- The paper has pretty graphics that demonstrate care and attention to detail
- The paper is cited favorably by a source you trust
- The paper won an award at a good conference
- The paper cites work you know is relevant in the field
- The paper is cited a lot
- The paper is on a syllabus from someone you trust
  - http://www.cc.gatech.edu/~jeisenst/

Those heuristics can be totally wrong

Science is a social process

Fast forward to the "state-of-the-art"

(or at least something getting lots of recent attention)

Rush, Chopra, Weston (2015)

A Neural Attention Model for Sentence Summarization

Why is this the state of the art?

Decoder	Model	Cons.	R-1	R-2	R-L
Greedy	ABS+	Abs	26.67	6.72	21.70
Beam	BoW	Abs	22.15	4.60	18.23
Beam	ABS+	Ext	27.89	7.56	22.84
Beam	ABS+	Abs	28.48	$\boxed{8.91}$	23.97

Table 3: ROUGE scores on DUC-2003 development data for various versions of inference. Greedy and Beam are described in Section 4. Ext. is a purely extractive version of the system (Eq. 2)

What is this paper trying to do?

Example One: Person or machine?

Input: a detained iranian-american academic accused of acting against national security has been released from a tehran prison after a hefty bail was posted , a top judiciary official said tuesday

Person or machine? iranian-american academic held in tehran released on bail

Person or machine? detained iranian-american academic released from jail after posting bail

Example One: Person or machine?

Input: a detained iranian-american academic accused of acting against national security has been released from a tehran prison after a hefty bail was posted , a to p judiciary official said tuesday

Real headline: iranian-american academic held in tehran released on bail

Model: detained iranian-american academic released from jail after posting bail

Extractive or abstractive?

Input: a detained iranian-american academic accused of acting against national security has been released from a tehran prison after a hefty bail was posted , a to p judiciary official said tuesday

Real headline: iranian-american academic held in tehran released on bail

Model: detained iranian-american academic released from jail after posting bail

Extractive or abstractive?

Input: a detained iranian-american academic accused of acting against national security has been released from a tehran **prison** after a hefty bail was **posted** , a to p judiciary official said tuesday

Real headline: iranian-american academic held in tehran released on bail

Model: detained iranian-american academic released from **jail** after **posting** bail

- Synonyms, morphology

Example (2)

Input: the white house on thursday warned iran of possible new sanctions after the un nuclear watchdog reported that tehran had begun sensitive nuclear work at a key site in defiance of un resolutions

Real headline: us warns iran of step backward on nuclear issue

Model: iran warns of possible new sanctions on nuclear work

- **Something is deeply wrong with the model's output here.**
- **Can ROUGE-1 detect it? Can ROUGE-2?**

What data do they use?

"The standard sentence summarization evaluation set is associated with the DUC-2003 and DUC-2004 shared tasks"

Science is a social process.

Input and output

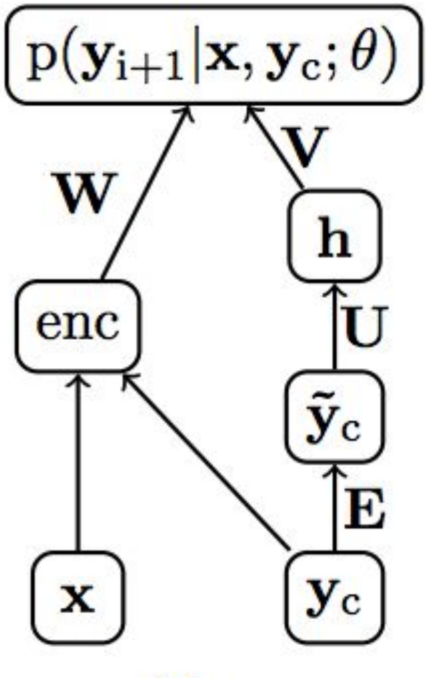
Input: **x**

Output: **y**

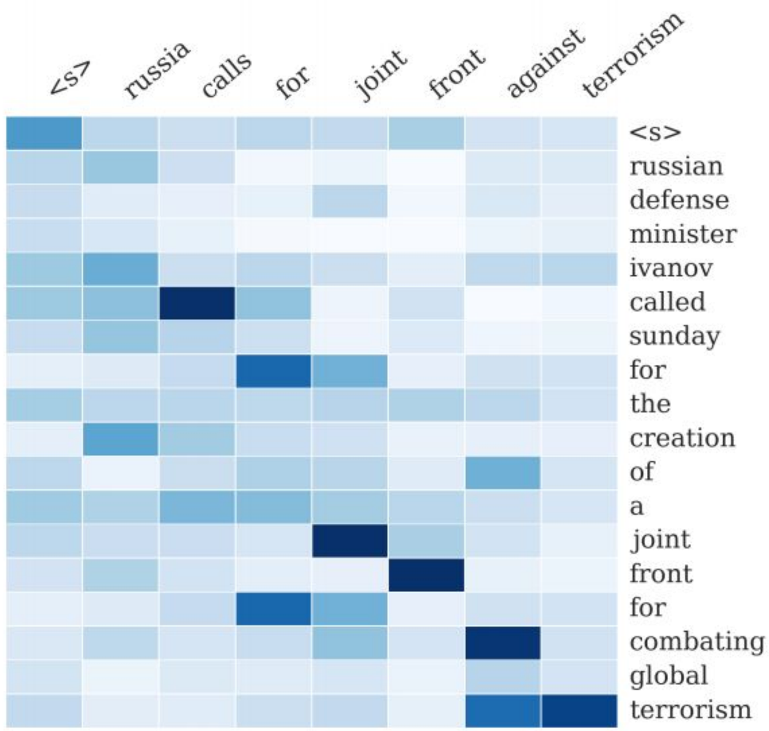
Where **|x|** >> **|y|**

Considers a context of c words

A generative model



(a)

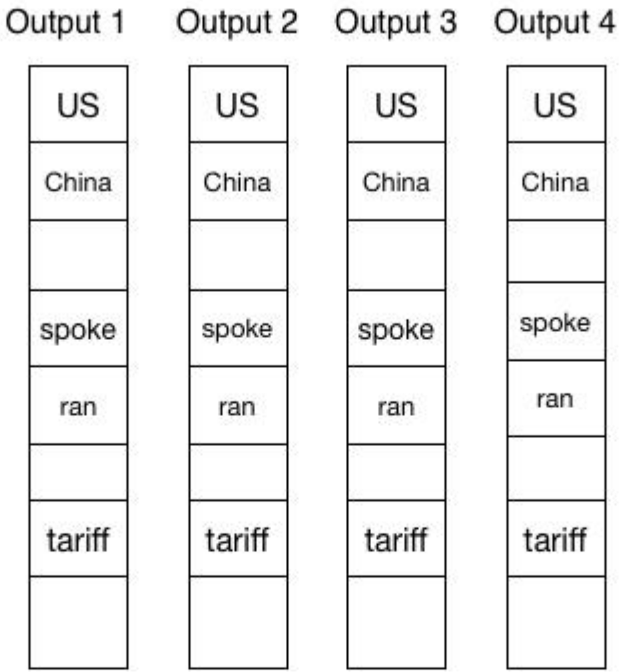


Training loss function

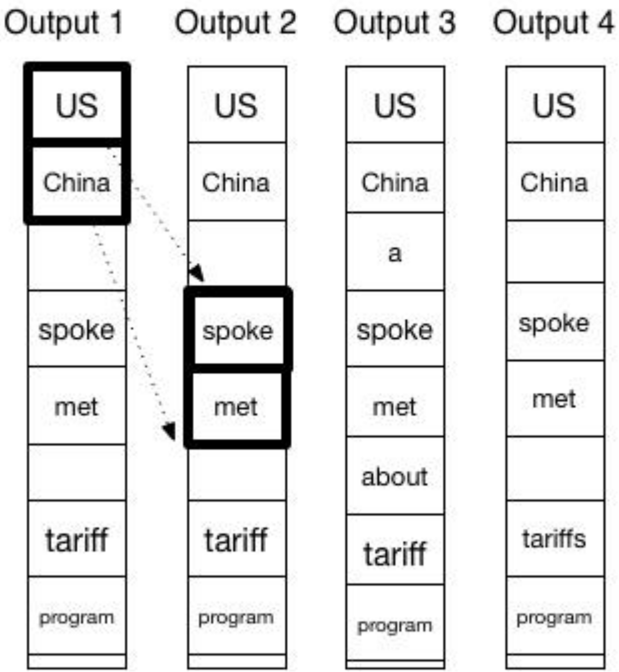
$$
\begin{aligned}
\mathrm{NLL}(\theta) &= -\sum_{j=1}^{J} \log p(\mathbf{y}^{(j)} | \mathbf{x}^{(j)}; \theta), \\
&= -\sum_{j=1}^{J} \sum_{i=1}^{N-1} \log p(\mathbf{y}_{i+1}^{(j)} | \mathbf{x}^{(j)}, \mathbf{y}_{\mathrm{c}}; \theta).
\end{aligned}
$$

J = summary pairs

Beam search decoding



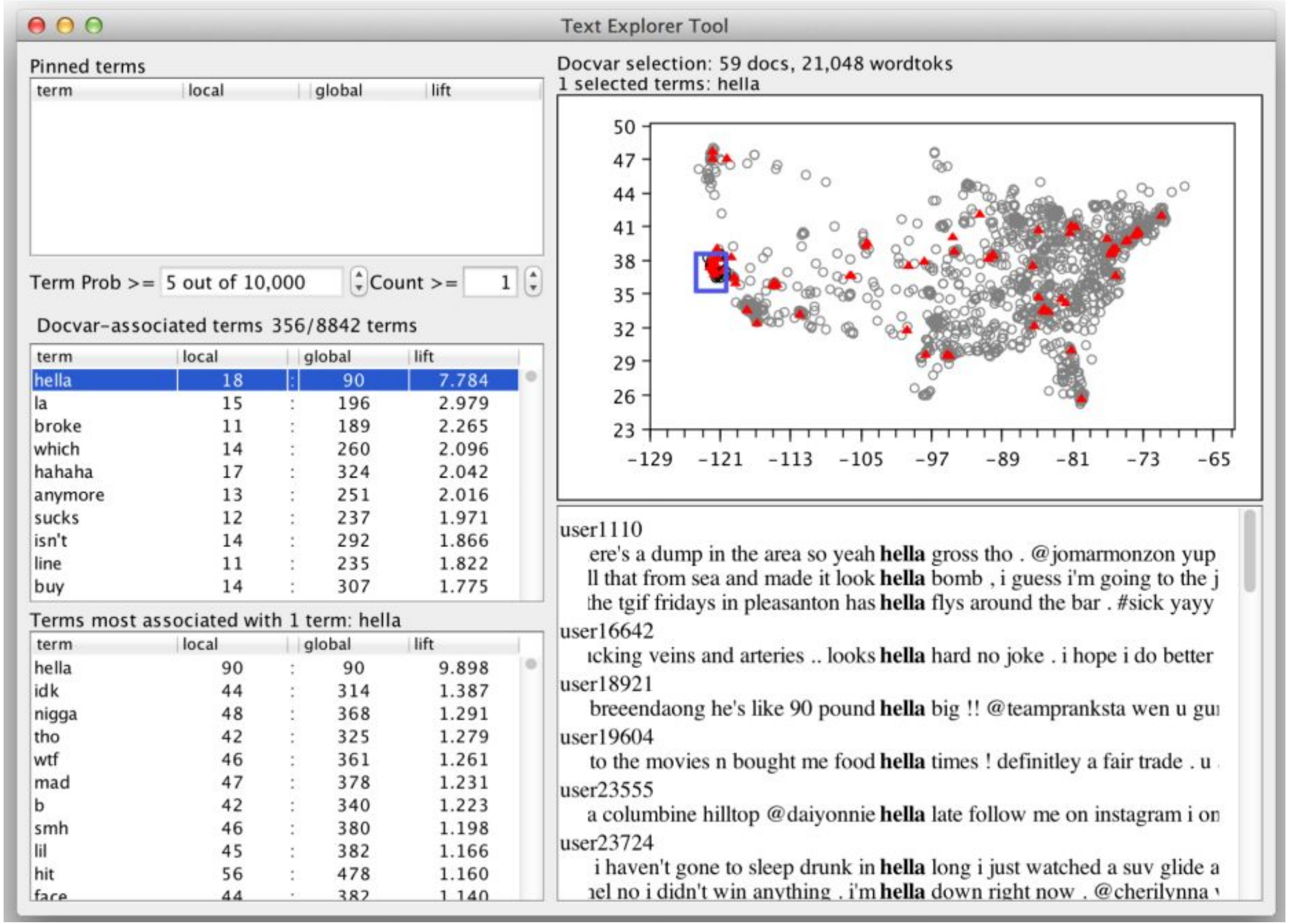
Beam search



Output 1
Output 2
Output 3
Output 4
US
China
spoke
met
tariff
program
US
China
spoke
met
tariff
program
US
China
a
spoke
met
about
tariff
program
US
China
spoke
met
tariffs
program

Practical note: summarization for hackers

- Put sentences in a search engine and query
- Select sentences with query words
- KWIC viewers


Text Explorer Tool
Pinned terms
term local global lift
Term Prob >= 5 out of 10,000
Count >= 1
Docvar-associated terms 356/8842 terms
term local global lift
hella 18 : 90 7.784
la 15 : 196 2.979
broke 11 : 189 2.265
which 14 : 260 2.096
hahaha 17 : 324 2.042
anymore 13 : 251 2.016
sucks 12 : 237 1.971
isn't 14 : 292 1.866
line 11 : 235 1.822
buy 14 : 307 1.775
Terms most associated with 1 term: hella
term local global lift
hella 90 : 90 9.898
idk 44 : 314 1.387
nigga 48 : 368 1.291
tho 42 : 325 1.279
wtf 46 : 361 1.261
mad 47 : 378 1.231
b 42 : 340 1.223
smh 46 : 380 1.198
lil 45 : 382 1.166
hit 56 : 478 1.160
Docvar selection: 59 docs, 21,048 wordtoks
1 selected terms: hella
50
47
44
41
38
35
32
29
26
23
-129
-121
-113
-105
-97
-89
-81
-73
-65
user1110
ere's a dump in the area so yeah **hella** gross tho . @jomarmonzon yup
ll that from sea and made it look **hella** bomb , i guess i'm going to the j
the tgif fridays in pleasanton has **hella** flys around the bar . #sick yayy
user16642
icking veins and arteries .. looks **hella** hard no joke . i hope i do better
user18921
breeendaong he's like 90 pound **hella** big !! @teampranksta wen u gu
user19604
to the movies n bought me food **hella** times ! definitley a fair trade . u
user23555
a columbine hilltop @daiyonnie **hella** late follow me on instagram i on
user23724
i haven't gone to sleep drunk in **hella** long i just watched a suv glide a
el no i didn't win anything . i'm **hella** down right now . @cherilynna

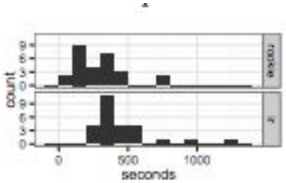
What did they measure?

Does the measurement support their argument?

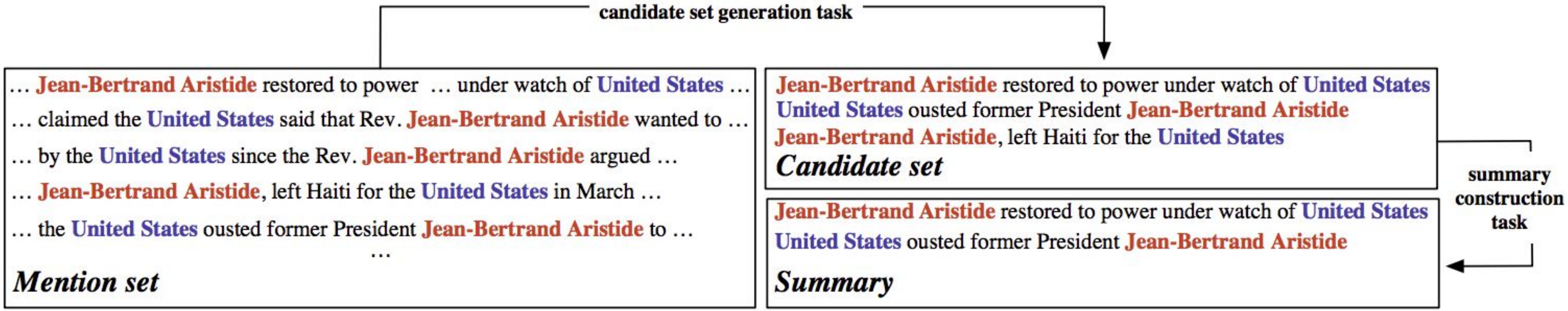
Rookie evaluation

faster; in seconds:

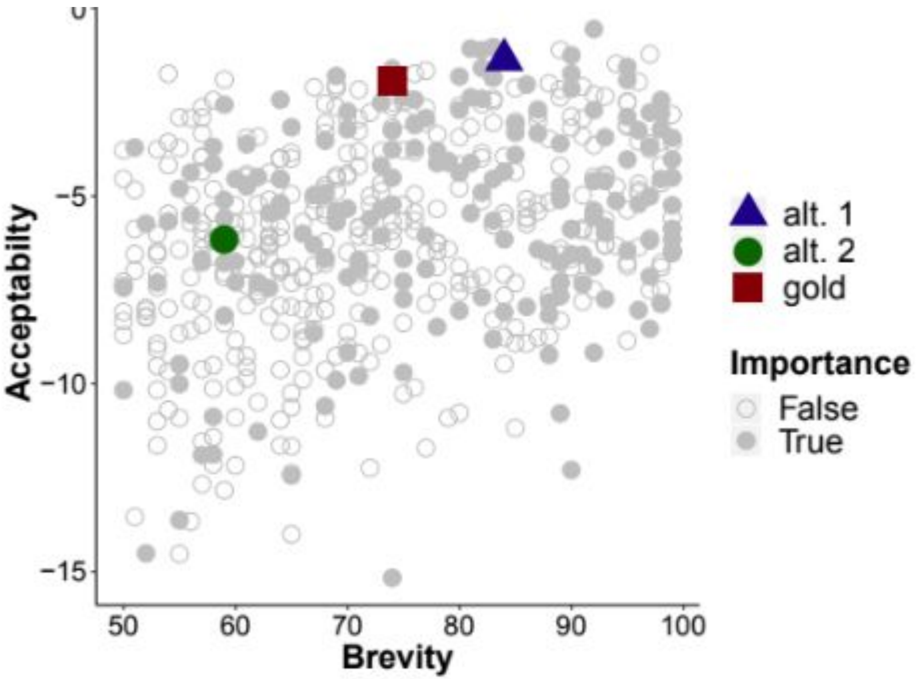
	Mean	Std. Dev.
Rookie	285.1	(169.8)
IR	451.0	(217.5)



"As journalists, it's important to have a large-view grasp of a story before writing about it. This system could be helpful in providing both a snapshot and an ability to then dive deeper into your story"


candidate set generation task
… Jean-Bertrand Aristide restored to power … under watch of United States …
… claimed the United States said that Rev. Jean-Bertrand Aristide wanted to …
… by the United States since the Rev. Jean-Bertrand Aristide argued …
… Jean-Bertrand Aristide, left Haiti for the United States in March …
… the United States ousted former President Jean-Bertrand Aristide to …
…
Mention set
Jean-Bertrand Aristide restored to power under watch of United States
United States ousted former President Jean-Bertrand Aristide
Jean-Bertrand Aristide, left Haiti for the United States
Candidate set
summary construction task
Jean-Bertrand Aristide restored to power under watch of United States
United States ousted former President Jean-Bertrand Aristide
Summary

Sentence. Pakistan launched a search for its missing ambassador to Afghanistan on Tuesday, a day after he disappeared in a Taliban area.
Headline. Pakistan searches for missing ambassador.
"Gold" compression. Pakistan launched a search for its missing ambassador.
Alternate 1. Pakistan launched a search for its missing ambassador to Afghanistan on Tuesday. (A(c) = -1.367, BREVITY = 84 characters max., IMPORTANCE = 1)
Alternate 2. Pakistan launched search Tuesday. (A(c) = -6.144, BREVITY = 59 characters max., IMPORTANCE = 0)



True or False?

Bertrand Aristide fled Haiti?

Bertrand Aristide was a priest?

Bertrand Aristide was President of Haiti?