The TL;DR Challenge @ INLG 2018

Shahbaz Syed, Michael Völske, Martin Potthast, Nedim Lipka, Benno Stein, Hinrich Schütze

WEBIS, Germany  Adobe Research, USA  LMU, Germany
Outline

- Summarization
- Webis-TLDR-17 Corpus
- News vs Social Media
- TIRA
- Challenge
- Final thoughts
Summarization

Origins

- *summa* ➔ summary

- One of the oldest tasks performed by humans as part of discourse

- Polytypic – abstract, precis, epitome, abridgment, review, digest, minutes
Summarization

Origins

- *summa* ➜ summary
- One of the oldest tasks performed by humans as part of discourse
- Polytypic – abstract, precis, epitome, abridgment, review, digest, minutes

Facets of summarization
Summarization

Origins

- *summa* ➔ summary
- One of the oldest tasks performed by humans as part of discourse
- Polytypic – abstract, precis, epitome, abridgment, review, digest, minutes

Facets of summarization
Summarization

Origins

- *summa* ➔ summary
- One of the oldest tasks performed by humans as part of discourse
- Polytypic – abstract, precis, epitome, abridgment, review, digest, minutes

Facets of summarization
Webis-TLDR-17

A large-scale summarization corpus for English social media text.
TL;DR – Previously

Alice: [long impassioned explanation of her views on a particular issue...]

Bob: tl;dr
TL;DR – Previously

Alice: [long impassioned explanation of her views on a particular issue...]

Bob: tl;dr

Alice:

TL;DR?

I think you may be illiterate
TL;DR – Now

Alice: [long impassioned explanation of her views on a particular issue...]

Bob: tl;dr ?

Alice: tl;dr - my issue is more important than anything else!
Alice: [long impassioned explanation of her views on a particular issue...]

Bob: tl;dr ?

Alice: tl;dr - my issue is more important than anything else!

Bob:
What are the most ridiculous things you believed to be true as a child? (self.AskReddit)

submitted 1 month ago by lfIHadANickle4UsAll
1000 comments share report

[-] Lostsonofpluto 129 points 1 month ago

My dad spent the better part of his childhood believing his early memories of his family having a pet monkey were just something his kid brain made up. When he asked his parents it turned out they actually did have one and that they gave it to a local zoo before his sister was born.

TL;DR

My dad believed his family didn't have a pet monkey

permalink embed report

load more comments (2 replies)
## Comparison

### Summarization datasets – English

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Genre</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigaword</td>
<td>News</td>
<td>4 Million</td>
</tr>
<tr>
<td>Cornell Newsroom</td>
<td>News</td>
<td>1.3 Million</td>
</tr>
<tr>
<td>NYT</td>
<td>News</td>
<td>655,000</td>
</tr>
<tr>
<td>CNN-DM</td>
<td>News</td>
<td>312,000</td>
</tr>
<tr>
<td>XSum(BBC)</td>
<td>News</td>
<td>226,000</td>
</tr>
<tr>
<td>PubMed</td>
<td>Scientific</td>
<td>133,000</td>
</tr>
<tr>
<td>arXiv</td>
<td>Scientific</td>
<td>215,000</td>
</tr>
<tr>
<td>Webis-TLDR-17</td>
<td>Social Media</td>
<td>3 Million</td>
</tr>
</tbody>
</table>

[Völske et.al 2017, TL;DR Mining Reddit to Learn Automatic Summarization]
News vs Social Media
Two styles of writing

News articles:
- Well-defined structure
- Coherent
- Proof read
- Headlines are relatively extractive

User-generated content:
- Casual structure
- Slang
- Poor grammar
- Higher level of abstraction
Two styles of writing

News articles:
- Well-defined structure
- Coherent
- Proof read
- Headlines are relatively extractive

User-generated content:
- Casual structure
- Slang
- Poor grammar
- Higher level of abstraction

(CNN) – NASA will launch Space Shuttle Endeavour on February 7, which will be the first of five launches this year before the shuttle fleet is retired. Endeavour will blast off from the Kennedy Space Center in Florida on a 13-day mission to the international space station. The mission will include three spacewalks, NASA said. The shuttle will also deliver the final U.S. portion of the space station. This portion will provide more room for crew members. NASA plans to retire its space shuttles Discovery, Endeavour and Atlantis later this year. The space agency has been looking for places, such as museums, to house the shuttles after they are retired. Space Shuttle Discovery will be transferred to the Smithsonian National Air and Space Museum in Washington. The privilege of showing off a shuttle won’t be cheap – about $29 million, NASA said.
(CNN) – NASA will launch Space Shuttle Endeavour on February 7, which will be the first of five launches this year before the shuttle fleet is retired. Endeavour will blast off from the Kennedy Space Center in Florida on a 13-day mission to the international space station. The mission will include three spacewalks, NASA said. The shuttle will also deliver the final U.S. portion of the space station. This portion will provide more room for crew members. NASA plans to retire its space shuttles Discovery, Endeavour and Atlantis later this year. The space agency has been looking for places, such as museums, to house the shuttles after they are retired. Space Shuttle Discovery will be transferred to the Smithsonian National Air and Space Museum in Washington. The privilege of showing off a shuttle won’t be cheap – about $29 million, NASA said.

@highlights
- This will be first of five launches this year before the shuttle fleet is retired
- NASA is scheduled to launch Space Shuttle Endeavour on February 7
- Shuttle will deliver final U.S. portion of the international space station
- NASA has been looking for places to house the shuttles once they are retired
(CNN) – NASA will launch Space Shuttle Endeavour on February 7, which will be the first of five launches this year before the shuttle fleet is retired. Endeavour will blast off from the Kennedy Space Center in Florida on a 13-day mission to the international space station. The mission will include three spacewalks, NASA said. The shuttle will also deliver the final U.S. portion of the space station. This portion will provide more room for crew members. NASA plans to retire its space shuttles Discovery, Endeavour and Atlantis later this year. The space agency has been looking for places, such as museums, to house the shuttles after they are retired. Space Shuttle Discovery will be transferred to the Smithsonian National Air and Space Museum in Washington. The privilege of showing off a shuttle won’t be cheap – about $29 million, NASA said.

@highlights

- This will be first of five launches this year before the shuttle fleet is retired
- NASA is scheduled to launch Space Shuttle Endeavour on February 7
- Shuttle will deliver final U.S. portion of the international space station
- NASA has been looking for places to house the shuttles once they are retired
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.

tl;dr my boyfriend is amazing and bought us an expensive anniversary dinner. Threw it all up, he doesn’t know. Feel horrible guilt and FOMO.
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.

tl;dr my boyfriend is amazing and bought us an expensive anniversary dinner. Threw it all up, he doesn’t know. Feel horrible guilt and FOMO.
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.

tl;dr my boyfriend is amazing and bought us an expensive anniversary dinner. Threw it all up, he doesn’t know. Feel horrible guilt and FOMO.
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.

**tl;dr** my boyfriend is amazing and bought us an expensive anniversary dinner. Threw it all up, he doesn’t know. Feel horrible guilt and FOMO.
I’m so upset at myself. My boyfriend surprised me with an amazing, fancy dinner for our one year anniversary yesterday. I already wasn’t feeling well when he told me we were going to dinner but when I saw what he planned I didn’t have the heart to tell him I wasn’t that hungry. In the end I pushed myself to eat the fixed menu he ordered for us and the bill was over 500, I couldn’t handle it and after dessert I ended up going to the bathroom and throwing it all up. I can’t believe I wasted so much of his money and am so disappointed in myself for not speaking up and simply saying I didn’t feel well. I feel like I’ve wasted the effort he put into planning this. I also feel like I missed out on some amazing food that we would usually never splurge for. He doesn’t know I threw it up and I just told him I loved it because regardless of how I felt health wise I loved that he put in so much effort to make sure I felt special. But I can’t stop stewing in my own feelings. Help.

**tl;dr** my boyfriend is amazing and bought us an expensive anniversary dinner. Threw it all up, he doesn’t know. Feel horrible guilt and FOMO.
kinda drunk: I so i did lie this girl who and i drank a little to much tonight and i kinda
told jher i like her and esked her if she felt the same way. she said that is looking for
a father figure (she has 2 kids) i said i comepletly understand that and now i move
you from the sex list to the awesome fire nd list. and she said that she was happy
wityh how i said that. i hope i didnt screw theis freind ship up casue she really is a n
aeawesome person.
kinda drunk: I so i did lie this girl who and i drank a little to much tonight and i kinda told jher i like her and esked her if she felt the same way. she said that is looking for a father figure (she has 2 kids) i said i comepletly understand that and now i move you from the sex list to the awesome fire nd list. and she said that she was happy wityh how i said that. i hope i didnt screw theis freind ship up casue she really is a n aewesome person.

tl/dr uijksd areyouhumasn is hard to spoelll wihen you are drubnk
kinda drunk:

so i did lie this girl who and i drank a little to much tonight and i kinda told jher i like her and esked her if she felt the same way. she said that is looking for a father figure (she has 2 kids) i said i compeletly understand that and now i move you from the sex list to the awesome fire nd list. and she said that she was happy wityh how i said that. i hope i didnt screw theis freind ship up casue she really is a n aewesome person.

tl/dr uijksd areyouhumasn is hard to spoelll wihen you are drubnk
kinda drunk: I so i did lie this girl who and i drank a little to much tonight and i kinda told jher i like her and asked her if she felt the same way. she said that is looking for a father figure (she has 2 kids) i said i completely understand that and now i move you from the sex list to the awesome fire nd list. and she said that she was happy wityh how i said that. i hope i didnt screw theis freind ship up casue she really is a n aewesome person.

tl/dr uijksd areyouhumasn is hard to spoelll wihen you are drubnk

Generated: i told a girl i liked her and she said she was happy
TIRA

The Incredible Research Assistant
tira.io
Why TIRA?

Evaluation as a service

- Reproducibility
- User control
- Blind evaluation
Why TIRA?

Evaluation as a service

- Reproducibility
- User control
- Blind evaluation

The TIRA way

- Reproducibility  ➔ virtualization for safe & diverse software submissions
- User control  ➔ UI to self-monitor & control evaluation runs
- Blind evaluation  ➔ sandboxing to avoid test data leakage
The TIRA experiment platform

System Architecture: User Interfaces
The TIRA experiment platform

System Architecture: User Interfaces
TIRA adoption

- Battle tested in 12 shared tasks so far
- Supports diverse software stacks
- Robust archiving for high reproducibility
- Learn more @ tira.io
Challenge
Timeline

<table>
<thead>
<tr>
<th></th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2019</td>
</tr>
</tbody>
</table>

- Training data available
- Submission system open
- Crowd eval
Challenge

Timeline

<table>
<thead>
<tr>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td>2019</td>
<td></td>
</tr>
</tbody>
</table>

Training

- Webis-TLDR-17 Corpus (3 Million pairs)
Challenge

Timeline

<table>
<thead>
<tr>
<th></th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Training data available

Submission system open

Crowd eval

Training

- Webis-TLDR-17 Corpus (3 Million pairs)

Submission

- TIRA
Challenge

Timeline

<table>
<thead>
<tr>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2019</td>
</tr>
</tbody>
</table>

Training

- Webis-TLDR-17 Corpus (3 Million pairs)

Submission

- TIRA

Evaluation

- ROUGE for automatic evaluation (content selection)
- Crowdsourcing for qualitative evaluation
Manual Evaluation

- Expensive, subjective yet highly crucial
- Likert scale, Best-to-worst scale
- Asking the right questions (DUC guidelines)

[Liu et.al 2018, Generating Wikipedia by Summarizing Long Sequences]
Final thoughts

- Processing noisy text
- Usability of your model
- Need not be the next state-of-the-art (at least not by ROUGE!)
Final thoughts

- Processing noisy text
- Usability of your model
- Need not be the next state-of-the-art (at least not by ROUGE!)

Visit: tldr.webis.de
Dataset

Mining **Reddit** for ground truths

1. Filter bots

2. Check if the post contains **tl.{0,3}dr**

3. Remove posts with multiple occurrences of **tl.{0,3}dr**

4. Preserve posts with any of the top 25 significant **tl.{0,3}dr** patterns

5. Remove posts where content is shorter than the summary (e.g. edits)