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

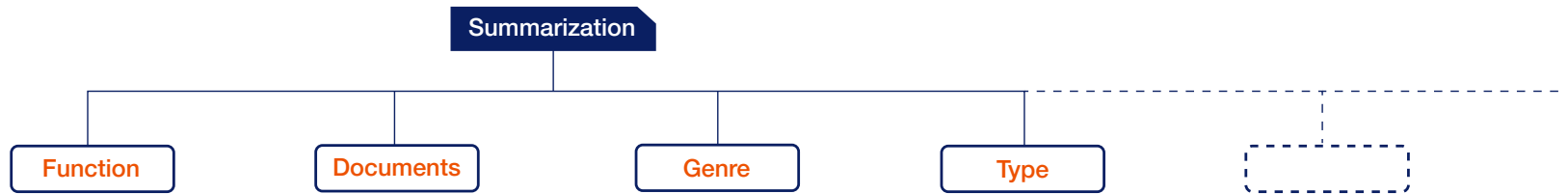- Challenge

- 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   │
                        └──────────────────┘
                                 │
   ┌──────────┬──────────────────┼──────────────────┬ ─ ─ ─ ─ ─ ─ ─ ─
┌──────────┐ ┌───────────┐ ┌──────────┐ ┌──────────┐  ┌ ─ ─ ─ ─ ─ ┐
│ Function │ │ Documents │ │  Genre   │ │   Type   │    ┆         ┆
└──────────┘ └───────────┘ └──────────┘ └──────────┘  └ ─ ─ ─ ─ ─ ┘
```

Summarization

Origins

- *summa* ➜ summary

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

- Polytypic – abstract, precis, epitome, abridgment, review, digest, minutes
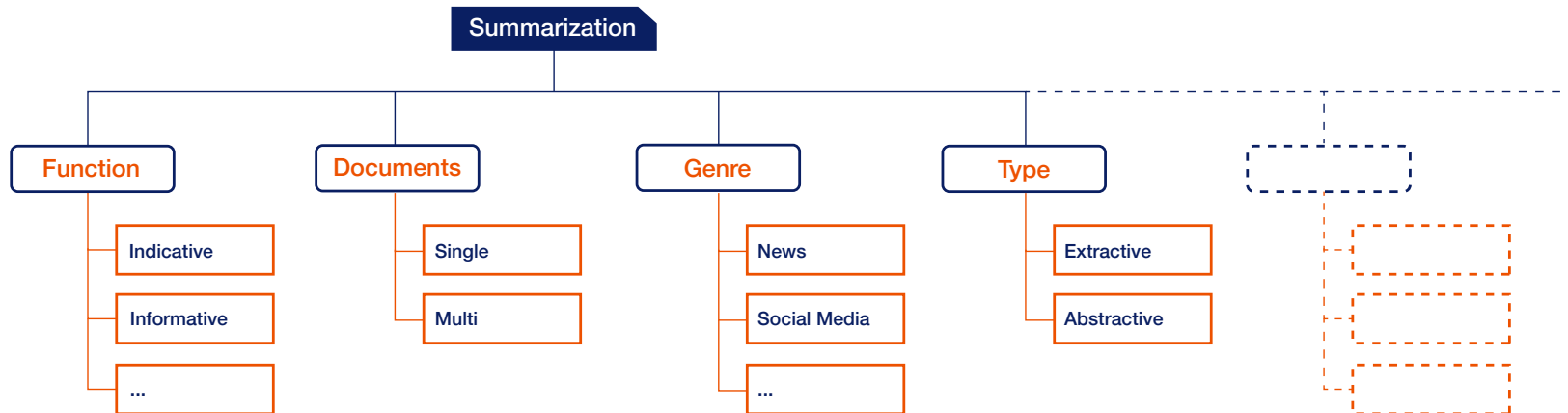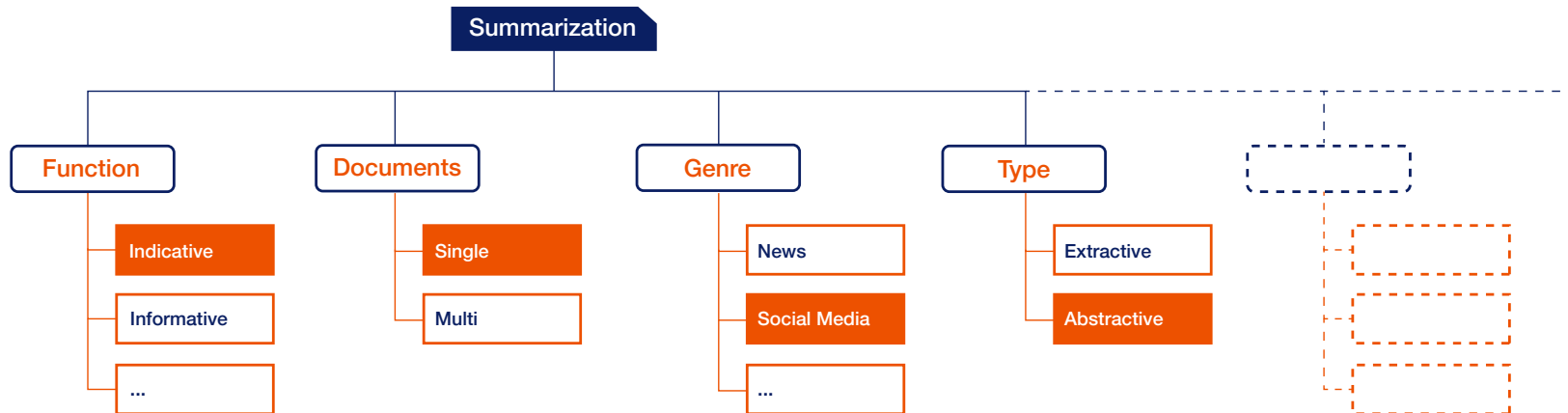
Facets of summarization

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				
Function	**Documents**	**Genre**	**Type**	
Indicative	Single	News	Extractive	
Informative	Multi	Social Media	Abstractive	
...		...		

Webis-TLDR-17

A large-scale summarization corpus for English social media text.

TL;DR – Previously

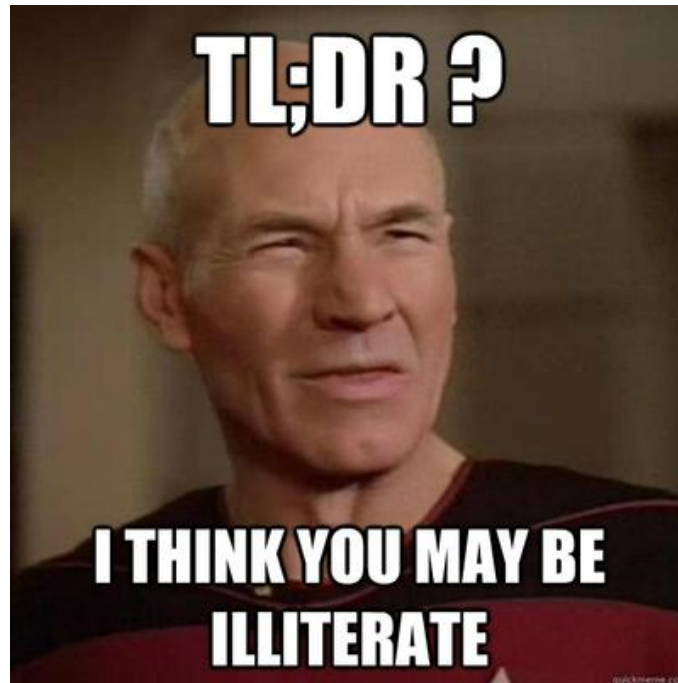Alice: [long impassioned explanation of her views on a particular issue...]

Bob: tl;dr

TL;DR – Previously

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

Bob: tl;dr

Alice:

TL;DR – 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!
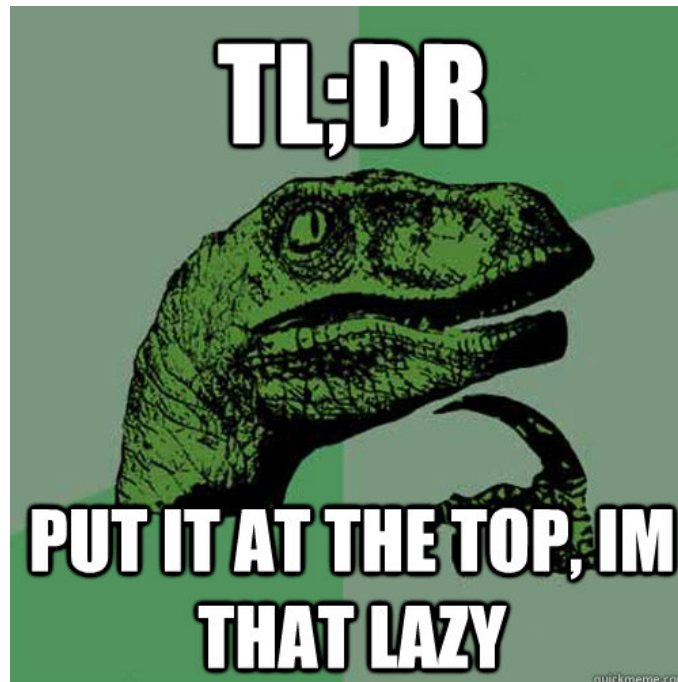
TL;DR – Now

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!

Bob:

TL;DR

PUT IT AT THE TOP, IM THAT LAZY

quickmeme.com

Mining Reddit

askreddit

COMMENTS

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

527

submitted 1 month ago by IfIHadANickle4UsAll

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

Corpus	Genre	Size
Gigaword	News	4 Million
Cornell Newsroom	News	1.3 Million
NYT	News	655,000
CNN-DM	News	312,000
XSum(BBC)	News	226,000
PubMed	Scientific	133,000
arXiv	Scientific	215,000
Webis-TLDR-17	Social Media	3 Million

[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

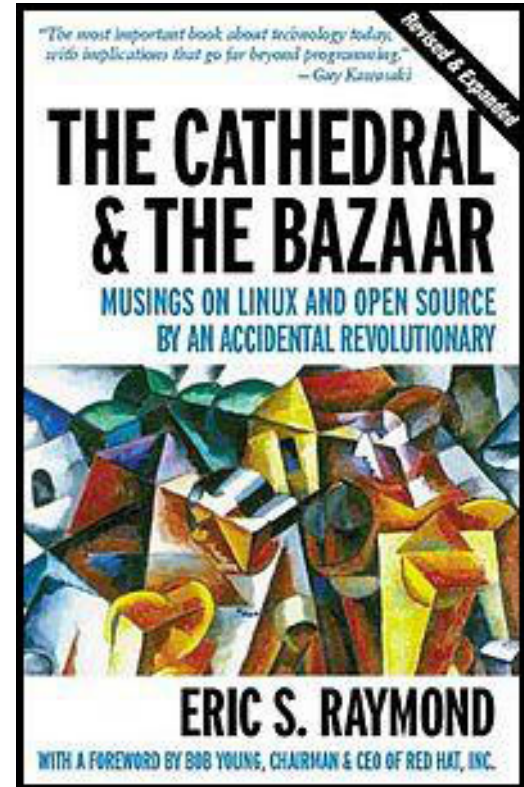# Two styles of writing

News articles:

- Well-defined structure
- Coherent
- Proof read
- Headlines are relatively extractive

User-generated content:

- Casual structure
- Slang
- Poor grammar
- Higher level of abstraction

Image source: Wikpedia

News

(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.

News

(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

News – More extractive
CNN-DM

(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

Social Media

Webis-TLDR-17

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.

Social Media

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.

Social Media– Higher abstraction
Webis-TLDR-17

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.

Social Media– Higher abstraction

Webis-TLDR-17

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.

Social Media– Higher abstraction

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.

Social Media– Slang

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.

Social Media
Webis-TLDR-17

kinda drunk:l 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.

Social Media
Webis-TLDR-17

kinda drunk:l 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

Social Media–Noisy text
Webis-TLDR-17

kinda drunk:l 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

Social Media–Noisy text

Webis-TLDR-17

kinda drunk:l 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

Generated: i told a girl i liked her and she said she was happy[seq2seq+attn]

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

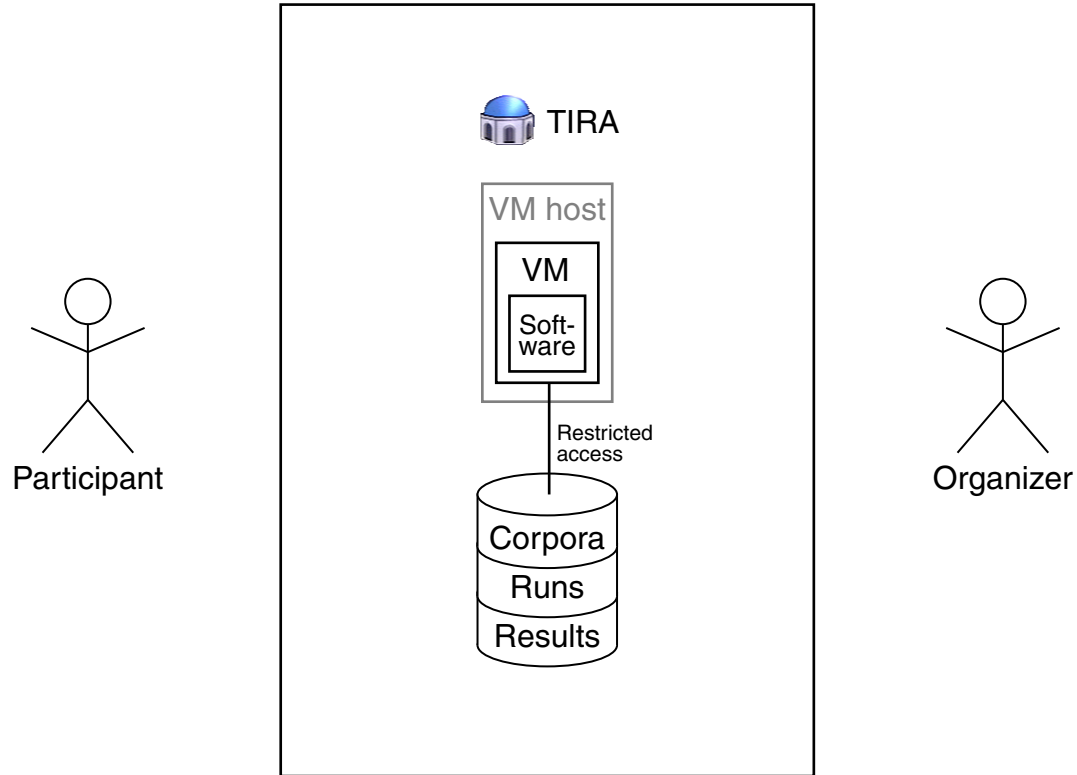- User control

- Blind evaluation

The TIRA way

- Reproducibility ➜ virtualization for safe & diverse software submissions

- User control ➜ UI to self-monitor & control evaluation runs

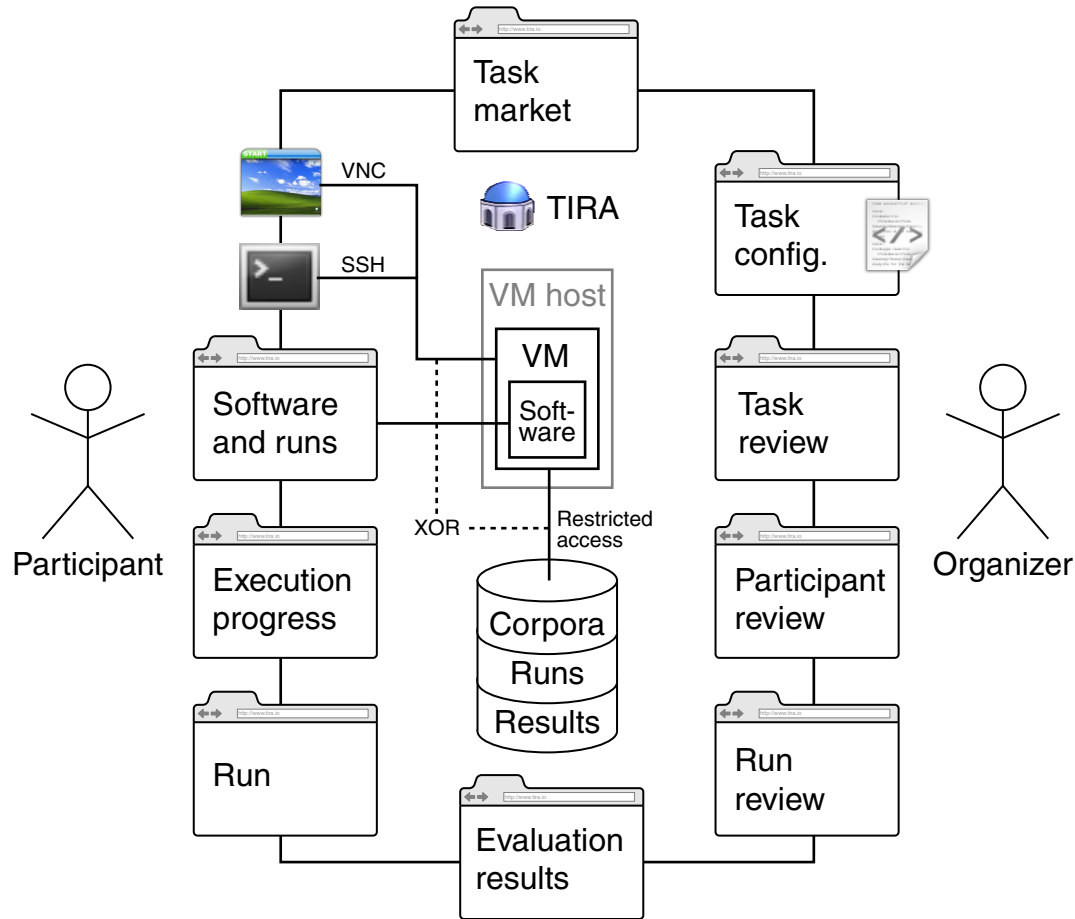- Blind evaluation ➜ sandboxing to avoid test data leakage

The TIRA experiment platform
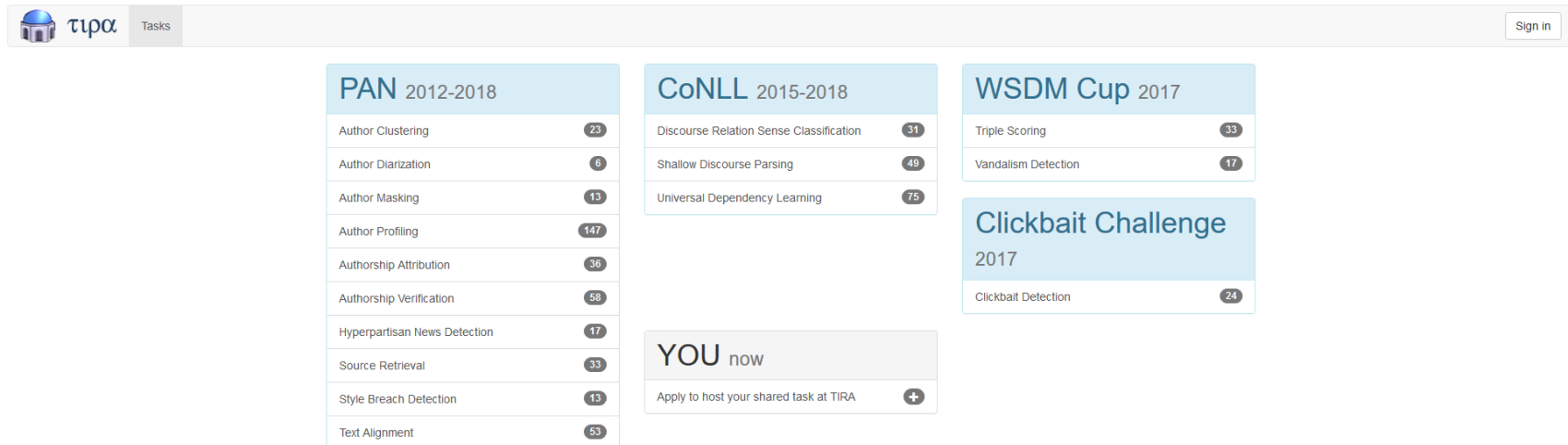
System Architecture: User Interfaces

The TIRA experiment platform

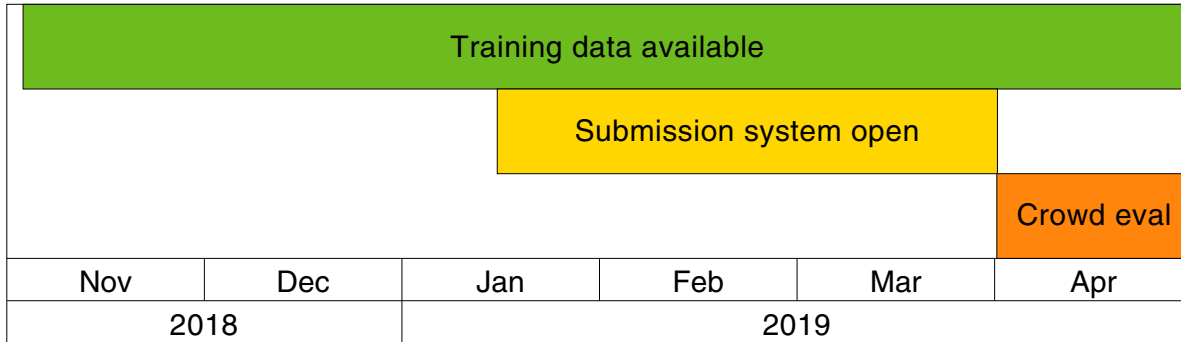## System Architecture: User Interfaces

TIRA adoption

PAN 2012-2018

Author Clustering	23
Author Diarization	6
Author Masking	13
Author Profiling	147
Authorship Attribution	36
Authorship Verification	58
Hyperpartisan News Detection	17
Source Retrieval	33
Style Breach Detection	13
Text Alignment	53

CoNLL 2015-2018

Discourse Relation Sense Classification	31
Shallow Discourse Parsing	49
Universal Dependency Learning	75

YOU now

Apply to host your shared task at TIRA	+

WSDM Cup 2017

Triple Scoring	33
Vandalism Detection	17

Clickbait Challenge 2017

Clickbait Detection	24

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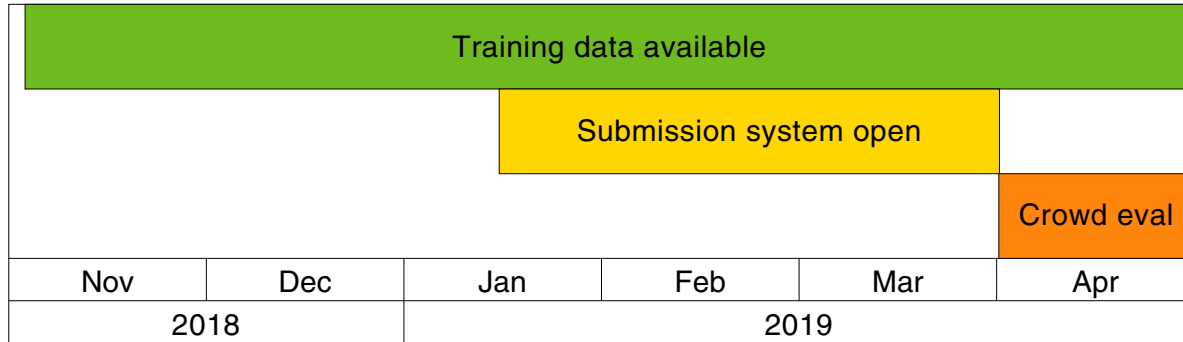- Learn more @ tira.io

Challenge
Timeline

Training data available					
		Submission system open			
					Crowd eval
Nov	Dec	Jan	Feb	Mar	Apr
2018		2019			

Challenge

Timeline

Nov	Dec	Jan	Feb	Mar	Apr

Training data available

Submission system open

Crowd eval

| 2018 | 2019 |

Training

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

Challenge

Timeline

Training data available					
		Submission system open			
					Crowd eval
Nov	Dec	Jan	Feb	Mar	Apr
2018		2019			
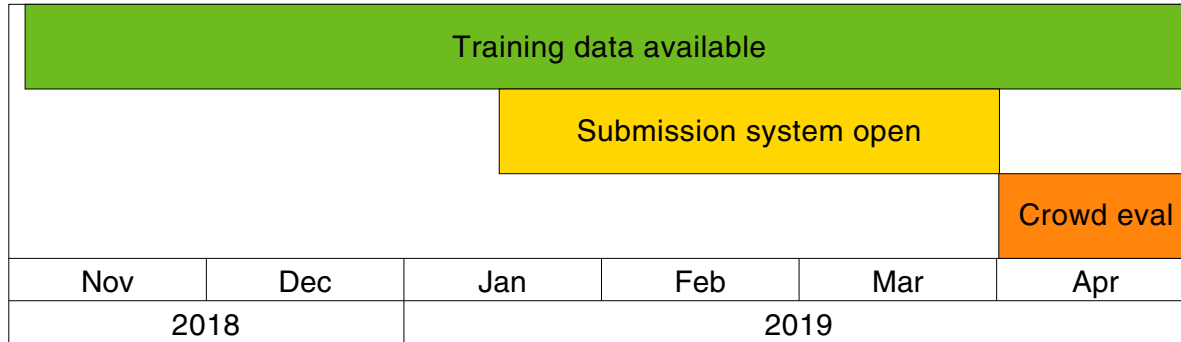
Training

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

Submission

- ❑ TIRA

Challenge

Timeline

Training data available						
			Submission system open			
						Crowd eval
Nov	Dec	Jan	Feb	Mar	Apr
2018		2019			
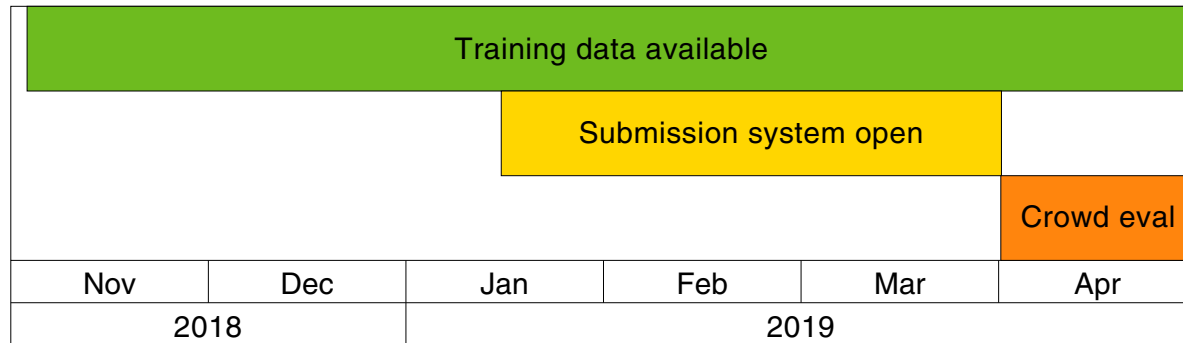
Training

- ❑ 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)

at the wmcp archives , i was able to locate dr morton jones ' thesis , a 20 - page document with ¨ cholera infantum by verina m. harris , columbia , sc " written on the cover page (figure 2) , 4 and a photo of her with 27 of her fellow students (figure 3). 5 morton jones attended wmcp for 4 years , though the school did not require this until 1893 . wmcp was an early adopter among american medical schools of the 4 - year graded curriculum .

Grammaticality: ⓘ
Very Poor ○ ○ ○ ○ ○ Very Good

Non-redundancy: ⓘ
Very Poor ○ ○ ○ ○ ○ Very Good

Referential clarity: ⓘ
Very Poor ○ ○ ○ ○ ○ Very Good

Focus: ⓘ
Very Poor ○ ○ ○ ○ ○ Very Good

Structure and Coherence: ⓘ
Very Poor ○ ○ ○ ○ ○ Very Good

[Liu et.al 2018, Generating Wikpedia 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)