Overview of Touché 2022: Argument Retrieval

Touché @ CLEF 2022
Shared Tasks on Argument Retrieval

Alexander Bondarenko
Maik Fröbe
Meriem Beloucif
Shahbaz Syed
Timon Gurcke
Johannes Kiesel
Alexander Panchenko
Chris Biemann
Benno Stein
Henning Wachsmuth
Martin Potthast
Matthias Hagen

[touche.webis.de]

Decision-making processes, be they societal or personal, often come to a point where one side challenges the other with a why-question, thus asking to justify their stance using arguments. Since technologies for argument mining are maturing at a rapid pace, ad-hoc argument retrieval becomes a feasible task in reach. To foster this progress, we invite you to participate in the third Touché lab on argument retrieval featuring three tasks:

○ Argument retrieval for controversial topics, where participants retrieve a gist of arguments from a collection of online debates.

○ Argument retrieval for comparative questions, where participants retrieve argumentative passages from a generic web crawl.

○ Image retrieval for arguments, where participants retrieve images from a focused web crawl that show support or opposition to some stance.

Task 1: Argument Retrieval for Controversial Questions

Should hate speech be banned?

○ Scenario: Users search for an argument gist on controversial topics
○ Task: Retrieve and rank sentences (main claim and premise) that convey key points pertinent to the controversial topic
○ Data: 6.5 mln. pro / con premises and conclusions (sentences)

## Task 2: Argument Retrieval for Comparative Questions

Should I major in Philosophy or Psychology?

○ Scenario: Support users in making informed decisions
○ Task: Retrieve relevant argumentative passages for compared objects and detect their respective stances w.r.t the objects
○ Data: 0.9 million text passages (from web documents)

Task 3: Image Retrieval for Arguments

Should hate speech be banned?

○ Scenario: Users search for images to corroborate their argumentation
○ Task: Retrieve images (from web pages) for each stance (pro/con) that show support for that stance
○ Data: ∼24.000 web images with respective web documents
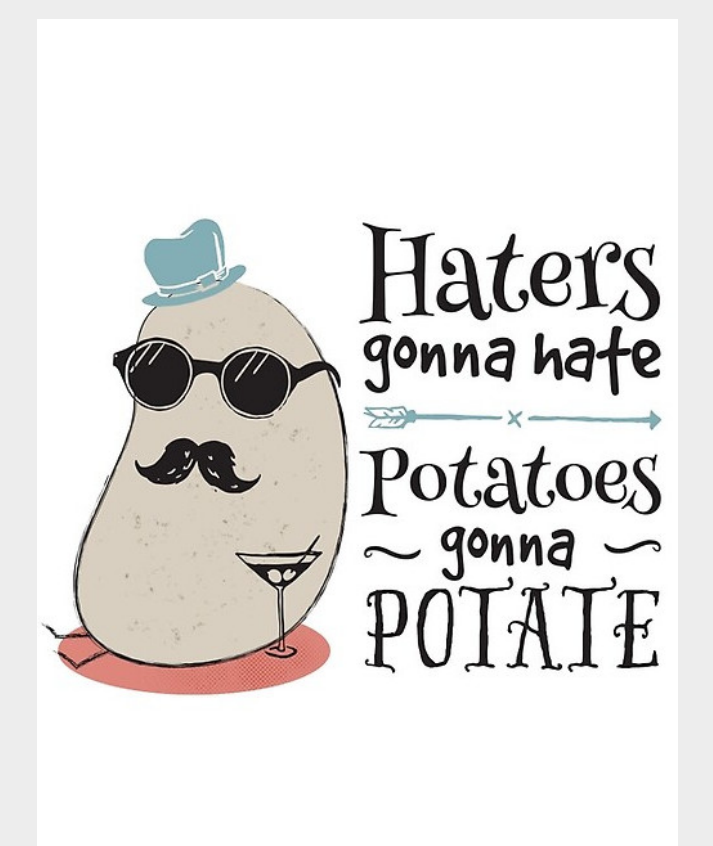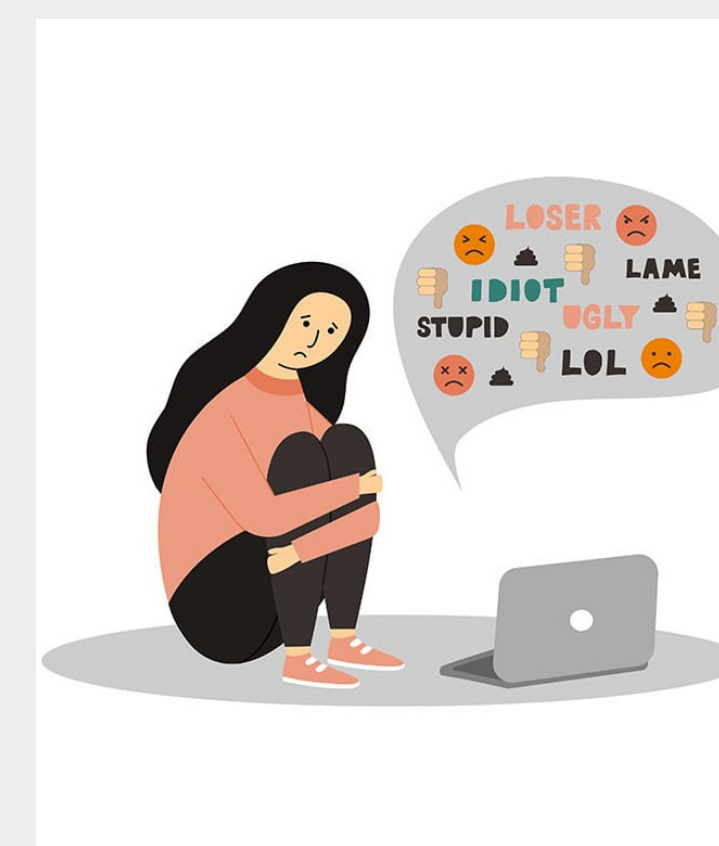
Evaluation: Argument Relevance

Classical (TREC-style) IR relevance judgments

👎 Not relevant   👍 Relevant   🙌 Highly relevant

*Who is a better **pet**, a **cat** or a **dog**?*

🙌 Comparing cats versus dogs as pets

👍 Information about either cats or dogs as pets

👎 Everything else: often ads

Evaluation: Argument Quality

Rhetorical quality: How well written?

👑👑👑 the best !!! Don't even try to argue with me. Yeah, yeah, yeah (Grrrr) I have always had cats, they are sooo cooool, and dogs just suck.

A cat's independent nature generally helps them deal better than dogs with being left alone. Cats also tend to live longer than dogs, which is sometimes a consideration when searching for a lifelong furry companion.

🙌 Proper language, good structure, good grammar, easy to follow

👍 Proper language but broken logic / hard to follow, or vice versa

👎 Profanity, hard to read / follow, grammar issues, or not an argument

Touché 2020 and 2021: Summary

○ Query expansion: Wordnet synonyms / antonyms → GPT-2 generation
○ Document representations using Transformer (e.g., BERT, SBERT)
○ Re-ranking based on argument quality and argument support prediction, on comparativeness features, on premises and claims in documents
○ Using relevance judgments for fine-tuning and parameter optimization
○ Majority improved over baseline in 2021 vs. few in 2020
○ Neural approaches but also BM25

[touche.webis.de]

○ Registration
○ Important dates
○ Tasks and Data
○ Submission
○ Labeled data from previous editions