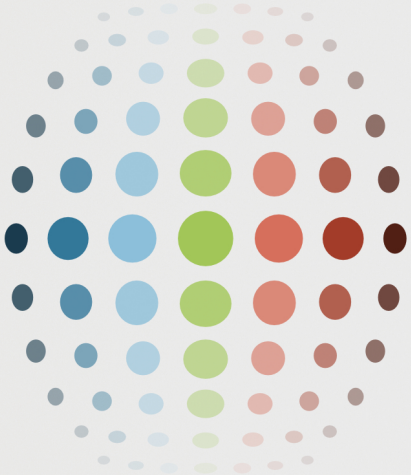
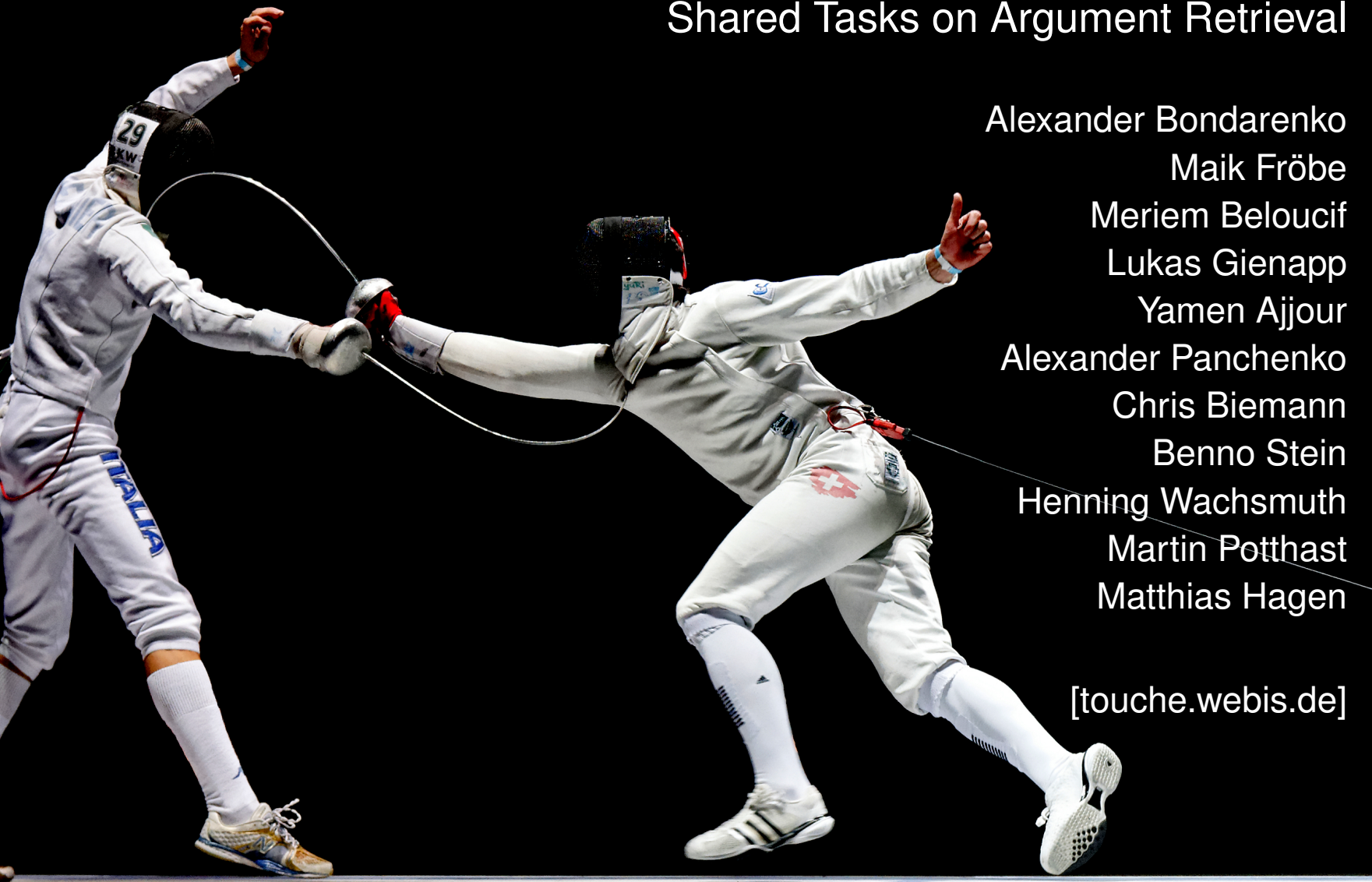


TOUCHÉ 2021

Touché @ CLEF 2021

Shared Tasks on Argument Retrieval



Alexander Bondarenko
Maik Fröbe
Meriem Beloucif
Lukas Gienapp
Yamen Ajjour
Alexander Panchenko
Chris Biemann
Benno Stein
Henning Wachsmuth
Martin Potthast
Matthias Hagen

[touche.webis.de]

A Timeline [Croft 2019]

TOUCHÉ 2021

Document Retrieval

Time

Answer Passage Retrieval

Passages as Features

Sentence Retrieval

Snippet Retrieval

QA Factoid Retrieval

CQA or Non-Factoid QA

Conversational Answer Retrieval

Answer Passage Retrieval Revisited

Response Retrieval/Generation

Question Answering/Machine Comprehension

Complex Answer Retrieval
(Passages as Summaries)

A Timeline [Croft 2019]

TOUCHÉ 2021

Document Retrieval

Time

Answer Passage Retrieval

Passages as Features

Sentence Retrieval

Snippet Retrieval

QA Factoid Retrieval

CQA or Non-Factoid QA

Conversational Answer Retrieval

Answer Passage Retrieval Revisited

Response Retrieval/Generation

Question Answering/Machine Comprehension

Complex Answer Retrieval
(Passages as Summaries)

Argument Retrieval

Touché: Argument Retrieval

Shared Tasks

TOUCHÉ 2021

Task 1: Supporting conversations on controversial topics

- Scenario: Users search for arguments on controversial topics
- Task: Retrieve and rank “strong” pro/con arguments on the topic
- Data: 400,000 “arguments” (short text passages) [args.me]

©touche.webis.de 2021

Touché: Argument Retrieval

Shared Tasks

TOUCHÉ 2021

Task 1: Supporting conversations on controversial topics

- Scenario: Users search for arguments on controversial topics
- Task: Retrieve and rank “strong” pro/con arguments on the topic
- Data: 400,000 “arguments” (short text passages) [args.me]

Task 2: Answering comparative questions with arguments

- Scenario: Users face personal decisions from everyday life
- Task: Retrieve and rank arguments for “Is X better than Y for Z?”
- Data: ClueWeb12 or ChatNoir [chatnoir.eu]

©touche.webis.de 2021

Touché: Argument Retrieval

Shared Tasks

TOUCHÉ 2021

Task 1: Supporting conversations on controversial topics

- Scenario: Users search for arguments on controversial topics
- Task: Retrieve and rank “strong” pro/con arguments on the topic
- Data: 400,000 “arguments” (short text passages) [args.me]

Task 2: Answering comparative questions with arguments

- Scenario: Users face personal decisions from everyday life
- Task: Retrieve and rank arguments for “Is X better than Y for Z?”
- Data: ClueWeb12 or ChatNoir [chatnoir.eu]

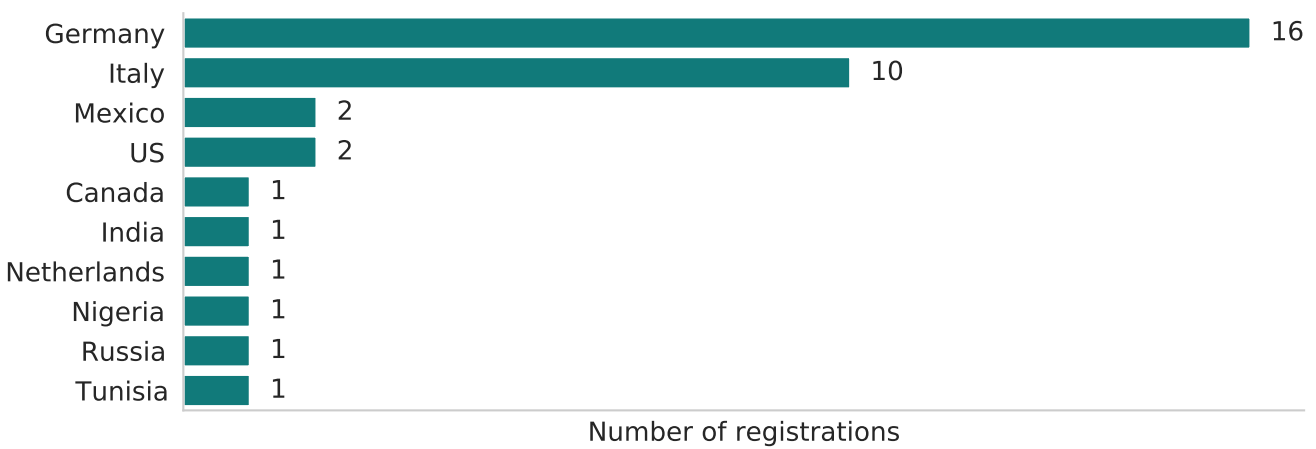
- Run submissions similar to “classical” TREC tracks
- Software submissions via TIRA [tira.io]

©touche.webis.de 2021

Touché: Argument Retrieval

Statistics

- Registrations: 36 teams (vs. 28 teams last year)
- Nicknames: Real or fictional fencers / swordsmen (e.g., Zorro)
- Submissions: 27 participating teams (vs. 17 last year)
- Approaches: 88 valid runs were evaluated (vs. 41 last year)
- Baselines: DirichletLM and BM25F-based ChatNoir [chatnoir.eu]
- Evaluation: 5,787 manual relevance and quality judgments (nDCG@5)



©touche.webis.de 2021

Touché: Argument Retrieval

Workshop Program

September, 24 (times are CEST)	
	Touché: Argument Retrieval Workshop will be held online [free registration]
10:30-10:40	Touché 2021 Welcome *Alexander Bondarenko*
	Session 1: Argument Retrieval for Controversial Questions
10:40-10:50	Overview of Task 1 on Argument Retrieval for Controversial Questions *Lukas Gienapp*
10:50-11:05	Exploring Argument Retrieval for Controversial Questions Using Retrieve and Re-rank Pipelines [paper] *Raunak Agarwal, Andrei Koniaev, Robin Schaefer*
11:05-11:20	Exploring Document Expansion for Argument Retrieval [paper] *Alina Mailach, Denise Arnold, Stefan Eysoldt, Simon Kleine*
11:20-11:35	Team Skeletor at Touché 2021: Argument Retrieval and Visualization for Controversial Questions [paper] *Kevin Ros, Carl Edwards, Heng Ji, ChengXiang Zhai*
11:35-12:00	Lightning talks *Touché Task 1 participants*
12:00-13:00	Lunch break

[touche.webis.de]

©touche.webis.de 2021

Touché: Argument Retrieval

Workshop Program

TOUCHÉ 2021

	Session 2: Argument Retrieval for Comparative Questions
13:00-13:30	**Keynote:** Theory-based Argument Quality for Advanced Argument Retrieval: Opportunities and Challenges *Anne Lauscher*
13:30-13:40	Overview of Task 2 on Argument Retrieval for Comparative Questions *Alexander Bondarenko*
13:40-13:50	DistilBERT-based Argumentation Retrieval for Answering Comparative Questions [paper] *Alaa Alhamzeh, Mohamed Bouhaouel, Előd Egyed-Zsigmond, Jelena Mitrović*
13:50-14:00	Retrieving Comparative Arguments using Ensemble Methods and Neural Information Retrieval [paper] *Viktoriia Chekalina, Alexander Panchenko*
14:00-14:10	Lightning talks *Touché Task 2 participants*
14:10-14:30	Panel discussion and closing remarks

[touche.webis.de]

©touche.webis.de 2021

Touché: Argument Retrieval

Workshop Program

TOUCHÉ 2021

	Session 2: Argument Retrieval for Comparative Questions
13:00-13:30	**Keynote:** Theory-based Argument Quality for Advanced Argument Retrieval: Opportunities and Challenges *Anne Lauscher*
13:30-13:40	Overview of Task 2 on Argument Retrieval for Comparative Questions *Alexander Bondarenko*
13:40-13:50	DistilBERT-based Argumentation Retrieval for Answering Comparative Questions [paper] *Alaa Alhamzeh, Mohamed Bouhaouel, Előd Egyed-Zsigmond, Jelena Mitrović*
13:50-14:00	Retrieving Comparative Arguments using Ensemble Methods and Neural Information Retrieval [paper] *Viktoriia Chekalina, Alexander Panchenko*
14:00-14:10	Lightning talks *Touché Task 2 participants*
14:10-14:30	Panel discussion and closing remarks

[touche.webis.de]

Spoiler: Touché will run again at CLEF 2022

©touche.webis.de 2021

TOUCHÉ 2021

Session 1: Argument Retrieval for Controversial Questions

Moderator: Lukas Gienapp

©touche.webis.de 2021

Touché: Argument Retrieval

Argument and Argumentation

TOUCHÉ 2021

Argument:

- A conclusion (claim) supported by premises (reasons) [Walton et al. 2008]
- Conveys a stance on a controversial topic [Freeley and Steinberg, 2009]

Conclusion *Argumentation will be a key element of conversational agents.*

Premise 1 *Superficial conversation ("gossip") is not enough.*

Premise 2 *Users want to know the "Why" to make informed decisions.*

Argumentation:

- Usage of arguments to achieve persuasion, agreement, . . .
- Decision making and opinion formation processes

©touche.webis.de 2021

Touché: Argument Retrieval

Topics

TOUCHÉ 2021

Example topic for Task 1:

Title	*Should hate speech be penalized more?*
Description	*Given the increasing amount of online hate speech, a user questions the necessity and legitimacy of taking legislative action to punish or inhibit hate speech.*
Narrative	*Highly relevant arguments include those that take a stance in favor of or opposed to stronger legislation and penalization of hate speech and that offer valid reasons for either stance. Relevant arguments talk about [...]*

©touche.webis.de 2021

Touché: Argument Retrieval

Data

TOUCHÉ 2021

Document Collection

- Args.me corpus [Ajjour et al. 2019]
- Argument passages from debate portals: idebate.org, debate.org, . . .
- Download or accessible via the API of args.me search engine [args.me]
- Newest/largest version of the corpus ($\sim 400,000$ structured arguments)

Additional Data

- 50 topics from last year
- relevance judgements from last year

©touche.webis.de 2021

Touché: Argument Retrieval

Statistics

- Submissions: 21 participating teams (up from 13)
- Nicknames: Real or fictional fencers / swordsmen (e.g., Zorro)
- Approaches: 55 valid runs were evaluated (up from 30)
- Baseline: DirichletLM (Lucene Implementation), args.me
- Topics: 50 new topics
- Evaluation: 3,711 manual relevance and quality judgments (nDCG@5)

©touche.webis.de 2021

Touché: Argument Retrieval

Evaluation

TOUCHÉ 2021

Argument retrieval: How good are the results?

- Evaluation w.r.t. argument *relevance* and argument *quality* (new this year)
- Top-5 pooling, removing duplicates
- 3,711 unique arguments (text passages)
- Annotation by eight graduate and undergraduate student volunteers, computer science background
  - Different from last year: crowdsourcing difficult for argument quality
  - Pilot study agreement similar to previous expert studies, follow-up discussion to ensure uniform annotations
- nDCG@5 for relevance and quality

©touche.webis.de 2021

Touché: Argument Retrieval

Evaluation

TOUCHÉ 2021

Relevance: How well do arguments fit the topic?

Spam

Not relevant

Relevant

Highly relevant

Topic: *Should hate speech be penalized more?*

Takes stance for/against stronger legislation on hate speech

About hate speech, but not legal aspects

Argumentative, but not about hate speech

Not an argument

©touche.webis.de 2021

Touché: Argument Retrieval

Evaluation

TOUCHÉ 2021

Rhetorical quality: How well-written is an argument?

> Gender is a social construct cuse we are told when we are first born by a dude what gender but if he didnt tellus that we woudnt have a gender its only cuse he told us that gender that we are that gender.
>
> *(Topic: Is gender a social construct?)*

> Cancel culture gives a voice to disenfranchised or less powerful people. It is a way to acknowledge that you don't have to have the power to change structural inequality. Even if they don't have the power to change all of public sentiment, for many individuals, it is the first time they do have a voice in public discourse.
>
> *(Topic: Is cancel culture good for society?)*

We labeled the quality regardless of relevance:

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

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

Profanity, hard to follow, hard to read, many grammar issues

©touche.webis.de 2021

Touché: Argument Retrieval

Results

(a) Highest **relevance** score per team

Team	**nDCG@5**	
	Rele.	Qual.
Dread Pirate Roberts	0.808	–
Swordsman	**0.756**	–
Elrond*	0.720	0.809
Pippin Took*	0.705	0.798
Robin Hood	0.691	0.756
Asterix*	0.681	0.802
Mercutio	0.678	0.804
⋮	⋮	⋮
Swordsman	**0.626**	**0.796**

(*) different runs (systems) from the same team; baseline **DirichletLM** ranking is shown **in bold**; highest results from 2020 are in gray (no quality).

©touche.webis.de 2021

Touché: Argument Retrieval

Results

(a) Highest **relevance** score per team

Team	**nDCG@5**	
	Rele.	Qual.
Dread Pirate Roberts	0.808	–
Swordsman	**0.756**	–
Elrond*	0.720	0.809
Pippin Took*	0.705	0.798
Robin Hood	0.691	0.756
Asterix*	0.681	0.802
Mercutio	0.678	0.804
⋮	⋮	⋮
Swordsman	**0.626**	**0.796**

(b) Highest **quality** score per team

Team	**nDCG@5**	
	Qual.	Rele.
Heimdall	0.841	0.639
Skeletor	0.827	0.666
Asterix*	0.818	0.663
Elrond*	0.817	0.681
Pippin Took*	0.814	0.683
⋮	⋮	⋮
Swordsman	**0.796**	**0.626**

(*) different runs (systems) from the same team; baseline **DirichletLM** ranking is shown **in bold**; highest results from 2020 are in gray (no quality).

©touche.webis.de 2021

Touché: Argument Retrieval

Summary

- Baseline scores lower, yet majority of teams outperforms vs. few last year
- Quality evaluation shows promising results, improvements w.r.t. baseline
- Two trends among submissions:
  - Deploying “classical” retrieval models with parameter optimization
  - Increased focus on ML for query expansion and assessing quality

©touche.webis.de 2021

Touché: Argument Retrieval

Summary

- Baseline scores lower, yet majority of teams outperforms vs. few last year
- Quality evaluation shows promising results, improvements w.r.t. baseline
- Two trends among submissions:
  - Deploying “classical” retrieval models with parameter optimization
  - Increased focus on ML for query expansion and assessing quality

- All approaches indexed the corpus themselves, no use of `args.me` API
- All approaches used 2020 relev. judgments for training or parameter tuning
- Extended a collection of relevance judgments, additionally argument quality judgments

©touche.webis.de 2021

Touché: Argument Retrieval

Submitted Papers

TOUCHÉ 2021

Team	Paper
Asterix	Smerilli et al.: A Search Engine System for Touché Argument Retrieval task to answer Controversial Questions
Batman	Raimondi et al.: Step approach to information retrieval
Pirate Roberts	Akiki et al.: Learning to Rank Arguments with Feature Selection
G. Ishikawa	Carnelos et al.: Argument Retrieval for Controversial Questions
Heimdall	Gienapp.: Quality-aware Argument Retrieval with Topical Clustering
Hua Mulan	Mailach et al.: Exploring Document Expansion for Argument Retrieval
J.-P. Polnareff	Alecci et al.: Development of an IR System for Argument Search
Macbeth	Agarwal.: Exploring Argument Retrieval for Controversial Questions Using Retrieve and Re-rank Pipelines
Pippin Took	Togni.: Exploring Approaches for Touché Task 1
Shanks	Berno et al.: Shanks Touché Homework
Skeletor	Ros et al.: Argument Retrieval and Visualization
Yeagerists	Green et al.: Exploring BERT Synonyms and Quality Prediction for Argument Retrieval
Baseline	Lucene Implementation of DirichletLM [Zhai & Lafferty 2004] Good results in pilot study [Potthast et al. 2019] args.me [Wachsmuth et al. 2017]

Marked in **bold** are featured talks.

©touche.webis.de 2021

TOUCHÉ 2021

Session 1: Participants' paper presentations

©touche.webis.de 2021

Touché: Argument Retrieval

TOUCHÉ 2021

	Session 2: Argument Retrieval for Comparative Questions
13:00-13:30	**Keynote:** Theory-based Argument Quality for Advanced Argument Retrieval: Opportunities and Challenges *Anne Lauscher*
13:30-13:40	Overview of Task 2 on Argument Retrieval for Comparative Questions *Alexander Bondarenko*
13:40-13:50	DistilBERT-based Argumentation Retrieval for Answering Comparative Questions [paper] *Alaa Alhamzeh, Mohamed Bouhaouel, Előd Egyed-Zsigmond, Jelena Mitrović*
13:50-14:00	Retrieving Comparative Arguments using Ensemble Methods and Neural Information Retrieval [paper] *Viktoriia Chekalina, Alexander Panchenko*
14:00-14:10	Lightning talks *Touché Task 2 participants*
14:10-14:30	Panel discussion and closing remarks

Moderator: Alexander Bondarenko

©touche.webis.de 2021

TOUCHÉ 2021

Keynote:

Theory-based Argument Quality for Advanced Argument Retrieval: Opportunities and Challenges

Anne Lauscher, Bocconi University in Milan

[webpage]

©touche.webis.de 2021

Touché: Argument Retrieval

Shared Task

TOUCHÉ 2021

Task 2: Answering comparative questions with arguments

- Scenario: Users face personal decisions from everyday life
- Goal: Help to come to an informed decision on the comparison
- Task: Retrieve and rank arguments for “Is X better than Y for Z?”
- Data: ClueWeb12 accessible via ChatNoir API [chatnoir.eu]

- Run submissions similar to “classical” TREC tracks
- Software submissions via TIRA [tira.io]

©touche.webis.de 2021

Touché: Argument Retrieval

Statistics

TOUCHÉ 2021

- Registrations: 13 teams, incl. for both tasks (vs. 18 last year)
- Nicknames: Real or fictional fencers / swordsmen (e.g., Katana)
- Submissions: 6 participating teams (vs. 5 last year)
- Approaches: 19 valid runs were evaluated (vs. 11 last year)
- Baseline: BM25F-based ChatNoir [chatnoir.eu]
- Evaluation: 2,076 manual relevance and quality judgments (nDCG@5) (vs. 1,783 last year)

©touche.webis.de 2021

Touché: Argument Retrieval

Topics

TOUCHÉ 2021

Example topic for Task 2:

Title	*Should I major in philosophy or psychology?*
Description	*A soon-to-be high-school graduate finds themself at a crossroad in their life. [. . .]searching for information about the differences and similarities, advantages and disadvantages of majoring in either of them (e.g., with respect to career opportunities).*
Narrative	*Relevant documents will overview one of the two majors in terms of career prospects or developed new skills, or they will provide a list of reasons to major in one or the other. [. . .] Not relevant are study program and university advertisements or general descriptions of the disciplines that do not mention benefits, advantages, or pros/cons.*

©touche.webis.de 2021

Touché: Argument Retrieval

Evaluation

TOUCHÉ 2021

Argument retrieval: How good are web documents with arguments?

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

©touche.webis.de 2021

Touché: Argument Retrieval

Evaluation

TOUCHÉ 2021

Argument retrieval: How good are web documents with arguments?

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.*

We labeled the quality regardless of relevance

🙌	Proper language, good structure, good grammar, easy to follow
👍	Proper language but broken logic / hard to follow, or vice versa
👎	Profanity, hard to follow, hard to read, many grammar issues

©touche.webis.de 2021

Touché: Argument Retrieval

Results

(a) Highest **relevance** score per team

Team	**nDCG@5**	
	Rele.	Qual.
Bilbo Baggins	0.580	–
Puss in Boots	**0.568**	–
Katana*	0.489	0.675
Thor	0.478	0.680
Rayla*	0.473	0.670
⋮	⋮	⋮
Puss in Boots	**0.422**	**0.636**

(*) different runs (systems) from the same team; baseline **ChatNoir** ranking is shown **in bold**; highest results from 2020 are in gray (no quality).

©touche.webis.de 2021

Touché: Argument Retrieval

Results

(a) Highest **relevance** score per team

Team	**nDCG@5**	
	Rele.	Qual.
Bilbo Baggins	0.580	–
Puss in Boots	**0.568**	–
Katana*	0.489	0.675
Thor	0.478	0.680
Rayla*	0.473	0.670
⋮	⋮	⋮
Puss in Boots	**0.422**	**0.636**

(b) Highest **quality** score per team

Team	**nDCG@5**	
	Qual.	Rele.
Bilbo Baggins	–	0.580
Puss in Boots	–	**0.568**
Rayla*	0.688	0.466
Katana*	0.684	0.460
Thor	0.680	0.478
⋮	⋮	⋮
Puss in Boots	**0.636**	**0.422**

(*) different runs (systems) from the same team; baseline **ChatNoir** ranking is shown **in bold**; highest results from 2020 are in gray (no quality).

©touche.webis.de 2021

Touché: Argument Retrieval

Results

(a) Highest **relevance** score per team

Team	nDCG@5	
	Rele.	Qual.
Bilbo Baggins	0.580	–
Puss in Boots	**0.568**	–
Katana*	0.489	0.675
Thor	0.478	0.680
Rayla*	0.473	0.670
⋮	⋮	⋮
Puss in Boots	**0.422**	**0.636**

(b) Highest **quality** score per team

Team	nDCG@5	
	Qual.	Rele.
Bilbo Baggins	–	0.580
Puss in Boots	–	**0.568**
Rayla*	0.688	0.466
Katana*	0.684	0.460
Thor	0.680	0.478
⋮	⋮	⋮
Puss in Boots	**0.636**	**0.422**

(*) different runs (systems) from the same team; baseline **ChatNoir** ranking is shown **in bold**; highest results from 2020 are in gray (no quality).

Team	Represent.	Query processing	(Re-)Ranking features	(Re-)Ranking method
Bilbo Baggins	Bag of words	N. entities, comp. aspects	Credibility, support, ...	Mean sum of scores
Puss in Boots	Bag of words	—	SpamRank	BM25F
Team	**Represent.**	**Query processing**	**(Re-)Ranking features**	**(Re-)Ranking method**
Katana	Bag of words, BERT	—	Comparativeness, relevance, tf·idf	R. Forests, XGBoost, LightGBM, BERT
Rayla	SBERT	Stop words, lemmas, synonyms/antonyms	Relevance, Page-, SpamRank, argument support (DistilBERT)	Weighted features linear combination
Thor	Bag of words	Punctuation, synonyms	Premises & claims with TARGER	Re-index & query w. BM25

©touche.webis.de 2021

Touché: Argument Retrieval

Summary

TOUCHÉ 2021

- All approaches re-ranked ChatNoir results
- All approaches used relevance judgments from 2020
- Majority improved over baseline vs. few last year
- Extended a collection of relevance judgments
- Additionally argument quality judgments
- “Best” so far: query processing and expansion, comparative and argumentative (incl. argument quality) features, neural but also BM25

©touche.webis.de 2021

Touché: Argument Retrieval

Submitted Papers

TOUCHÉ 2021

Team	Paper
Jack Sparrow	Thi Kim Hanh Luu, Jan-Niklas Weder. Argument Retrieval for Comparative Questions Based on Independent Features [paper]
Katana	Viktoriia Chekalina, Alexander Panchenko. Retrieving Comparative Arguments using Ensemble Methods and Neural Information Retrieval [paper]
Mercutio	Daniel Helmrich, Denis Streitmatter, Fionn Fuchs, Maximilian Heykeroth. Touché Task 2: Comparative Argument Retrieval. A document-based search engine for answering comparative questions [paper]
Puss in Boots	Bevendorff, Stein, Hagen, Potthast (ECIR 2018). Elastic ChatNoir: Search Engine for the ClueWeb and the Common Crawl [paper]
Rayla	Alaa Alhamzeh, Mohamed Bouhaouel, Előd Egyed-Zsigmond, Jelena Mitrović. DistilBERT-based Argumentation Retrieval for Answering Comparative Questions [paper]
Thor	Ekaterina Shirshakova, Ahmad Wattar. Thor at Touché 2021: Argument Retrieval for Comparative Questions [paper]
Lab overview	Bondarenko et al. Overview of Touché 2021: Argument Retrieval [paper] [CLEF 2021 Working Notes]

©touche.webis.de 2021

Session 2: Participants' paper presentations

©touche.webis.de 2021

TOUCHÉ 2021

Session 3: Panel discussion and closing remarks

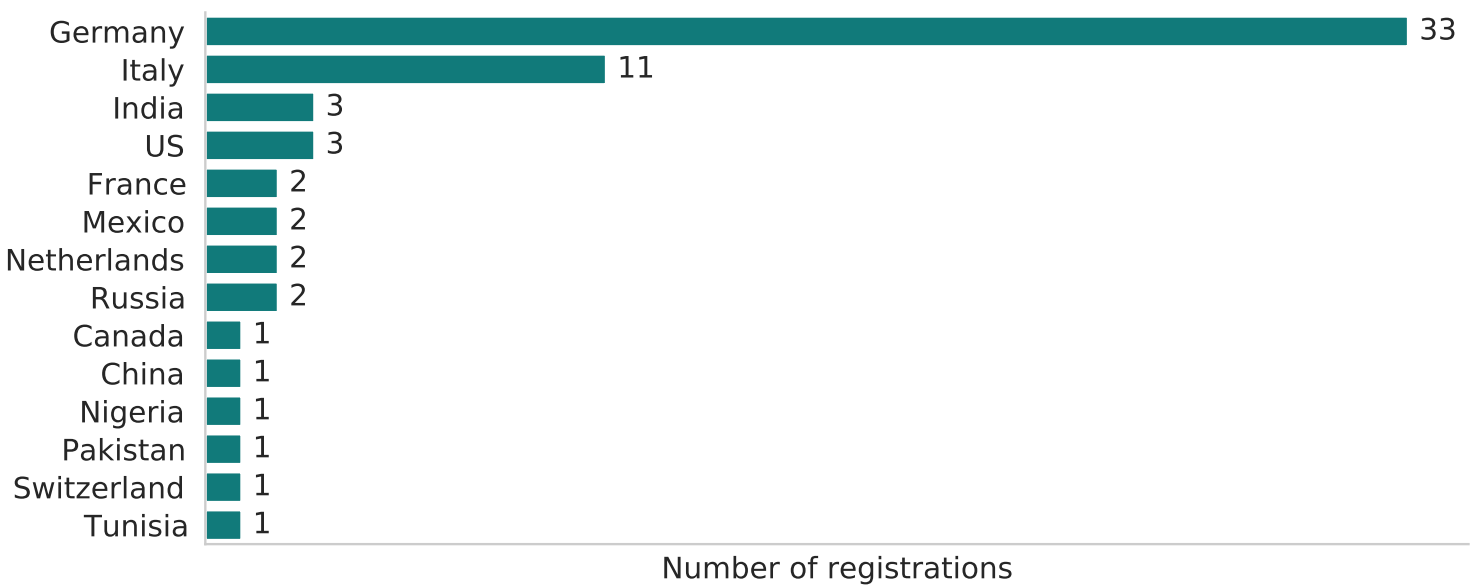
Moderator: Alexander Bondarenko

©touche.webis.de 2021

Touché: Argument Retrieval

Statistics over two years

- Registrations: 64 teams
- Submissions: 44 participating teams
- Approaches: 129 valid runs were evaluated
- Evaluation: 12,832 manual relevance judgments (5,787 additional quality)



©touche.webis.de 2021

Touché: Argument Retrieval

Summary

TOUCHÉ 2021

- Platform for argument retrieval researchers [touche.webis.de]
- Argument relevance / quality corpora / rankings
- Tools for submission and evaluation [tira.io]

- All (almost) used labeled data from 2020
- Majority improved over baselines

©touche.webis.de 2021

Touché: Argument Retrieval

Impact

TOUCHÉ 2021

- 24 participant working notes published in proceedings [CEUR-WS.org]
- Dumani, Schenkel.
  Quality-Aware Ranking of Arguments. CIKM 2020.
- Nilles, Dumani, Schenkel.
  QuARk: A GUI for Quality-Aware Ranking of Arguments. SIGIR 2021.
- Thakur, Reimers, Rücklé, Srivastava, Gurevych.
  BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models. arXiv 2021.
- Cherumanal, Spina, Scholer, Croft.
  Evaluating Fairness in Argument Retrieval. CIKM 2021.

©touche.webis.de 2021

Touché: Argument Retrieval

Outlook 2022

TOUCHÉ 2022

Task 1: Argument Retrieval for Controversial Questions

- Scenario: Users search for argument gist on controversial topics
- Task: Retrieve and rank sentences (main claim and premise) that convey key points pertinent to the controversial topic
- Data: 400,000 “arguments” (short text passages) [args.me]

©touche.webis.de 2021

Touché: Argument Retrieval

Outlook 2022

TOUCHÉ 2022

Task 1: Argument Retrieval for Controversial Questions

- Scenario: Users search for argument gist on controversial topics
- Task: Retrieve and rank sentences (main claim and premise) that convey key points pertinent to the controversial topic
- Data: 400,000 “arguments” (short text passages) [args.me]

Task 2: Argument Retrieval for Comparative Questions

- Scenario: Support users to come to informed decisions
- Task: Retrieve relevant argumentative passages for compared objects and detect their respective stances w.r.t the objects
- Data: > 1 million text passages (from web documents)

©touche.webis.de 2021

Touché: Argument Retrieval

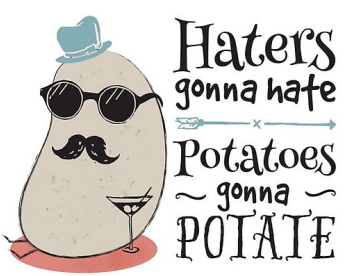
Outlook 2022

TOUCHÉ 2022

Task 3: Image Retrieval for Arguments

- Scenario: Users search for images to corroborate their argumentation
- Task: Retrieve and rank images that can be used to support or attack a given stance
- Data: > 10.000 web images with respective web documents

Should hate speech be banned?



©touche.webis.de 2021

TOUCHÉ 2022

Free discussion

©touche.webis.de 2021

Touché: Argument Retrieval

Related Publications

- Ajjour, Wachsmuth, Kiesel, Potthast, Hagen, Stein. Data Acquisition for Argument Search: The args.me Corpus. Proceedings of KI 2019.
- Bevendorff, Stein, Hagen, Potthas. Elastic ChatNoir: Search Engine for the ClueWeb and the Common Crawl. Proceedings of ECIR 2018.
- Braunstain, Kurland, Carmel, Szpektor, Shtok. Supporting Human Answers for Advice-Seeking Questions in CQA Sites. Proceedings of ECIR 2016.
- Croft. The Relevance of Answers. Keynote at CLEF 2019. https://ciir.cs.umass.edu/downloads/clef2019/CLEF_2019_Croft.pdf
- Freely and Steinberg. Argumentation and Debate: Critical Thinking for Reasoned Decision Making (12th ed.). Boston, MA: Wadsworth Cengage Learning, 2009.
- Potthast, Gienapp, Euchner, Heilenkötter, Weidmann, Wachsmuth, Stein, Hagen. Argument Search: Assessing Argument Relevance. Proceedings of SIGIR 2019.
- Wachsmuth, Naderi, Hou, Bilu, Prabhakaran, Alberdingk Thijm, Hirst, Stein. Computational Argumentation Quality Assessment in Natural Language. Proceedings of EACL 2017.
- Walton, Reed, Macagno. Argumentation Schemes. Cambridge: Cambridge University Press, 2008.
- Zhai,Lafferty. A Study of Smoothing Methods for Language Models Applied to Information Retrieval. ACM TOIS, 22(2), 2004.

©touche.webis.de 2021