Yamen Ajjour [1,5], Pavel Braslavski [2,3], Alexander Bondarenko [4], Benno Stein [5]

Leipzig Univeristy [1]   HSE University, Moscow [2]   Ural Federal University [3]
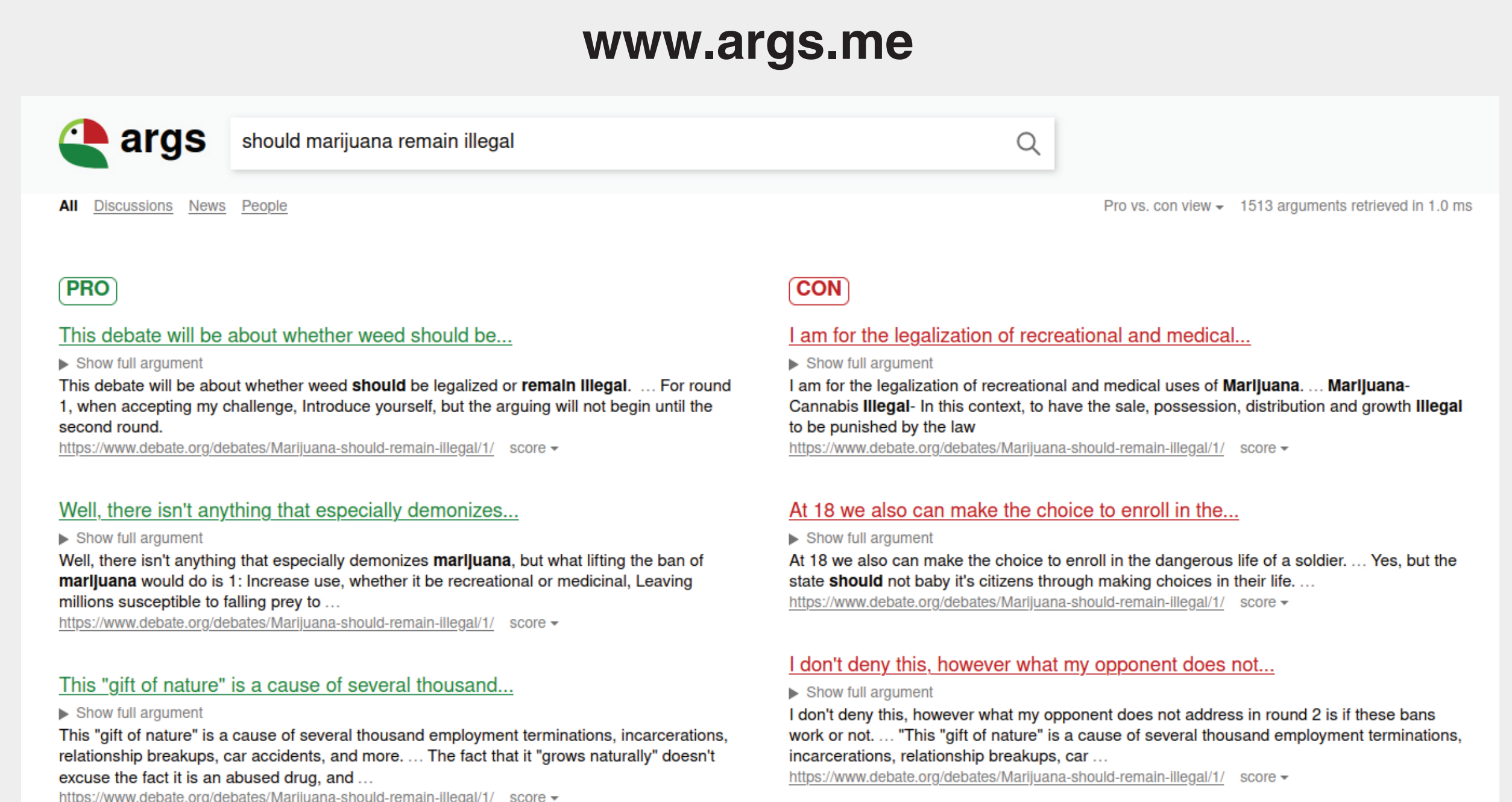Martin Luther Universität Halle-Wittenberg [4]   Bauhaus-Universität Weimar [5]

Identifying Argumentative Questions in Web Search Logs

Motivation

• Argumentative questions look for arguments or opinions for or against a topic.

• Search engines struggle at delivering all perspectives on argumentative questions.

• Argument retrieval systems retrieve pro and con arguments for a given query.

• Research question: how to identify questions that look for arguments in the query logs of a search engine?
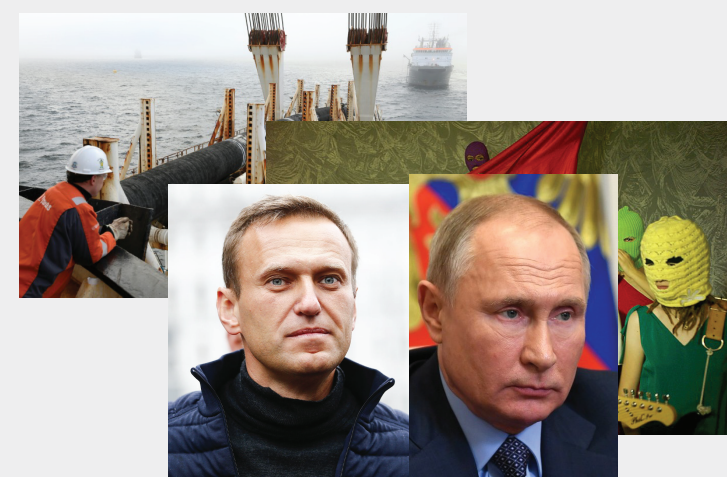
www.args.me

Preparing Russian Question Dataset

1. Filter from Yandex logs 4.5 million Russian questions on 19 controversial topics:

Local Issues
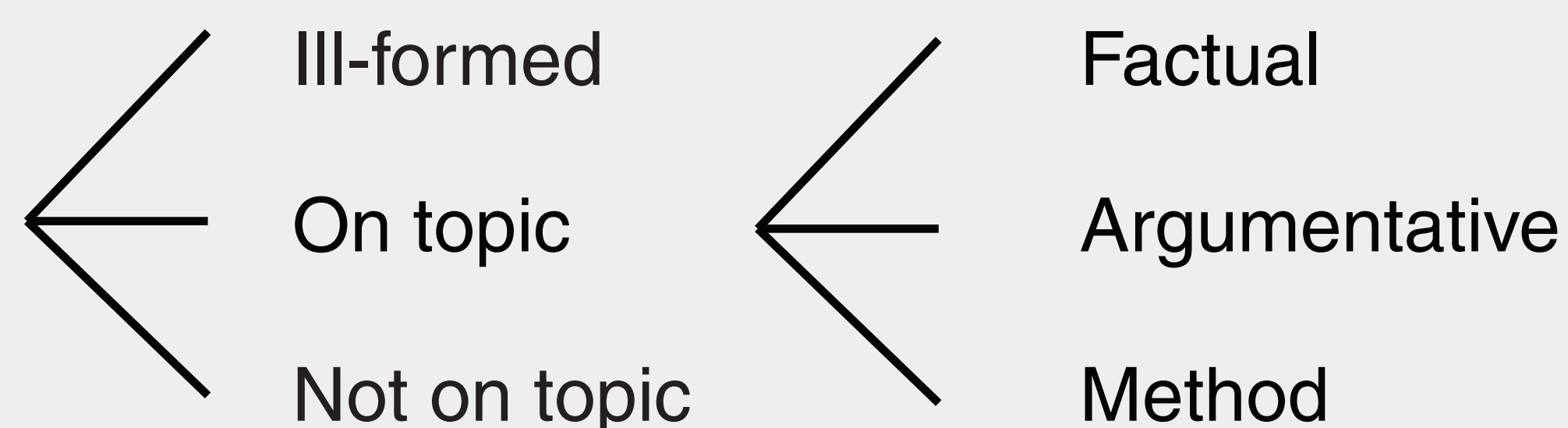
Putin, Navalny, Pussy Riots, Nordstream, ...

Global Issues

Abortion, Death penalty, Gay marriage, ...

2. Sample 54,850 questions and annotate them on a crowdsourcing platform with the annotation scheme:

Ill-formed
On topic — Factual / Argumentative / Method
Not on topic

Factual questions ask about information that most people agree on.

Argumetnative questions seek for arguments or opinions against or for a specific topic.

Method questions seek instructions or a description of a method to reach a goal.

Question Type	Percentage	Count	Example
Factual	64 %	25,332	is marijuana legalized in Belgium?
Argumentative	28 %	10,982	will the president legalize marijuana?
Method	8 %	3,026	how to use medical marijuana?

Contributions

• Effective approach for labeling and identifying argumentative questions in web search logs

• First question dataset on argumentative questions

• Analysis of the structure of argumentative questions

• **Conclusion:** argumentative questions look for reasons and predictions and constitute a substantial share of search questions on controverisal topics.

Quantitative Question Analysis

Comparing argumentative questions with factual and method questions in the dataset using lexical and syntactical patterns.

Characteristic	Factual	Argumentative	Method
Yes/No	7.2 %	**13.8 %**	0.4 %
Predictions	3.8 %	**8.2 %**	0.6 %
Comparisons	3.2 %	**5.7 %**	4.4 %
Personal Pronouns	0.3 %	**3.8 %**	0.4 %
Wh-words	67.0 %	62.0 %	**94.3 %**
why	1.3 %	**20.7 %**	0.0 %
how	7.0 %	7.6 %	**87.5 %**
how much/many	**10.5 %**	2.1 %	0.2 %
how much moeny	**3.4 %**	0.4 %	0.0 %
how many people	**2.4 %**	0.7 %	0.0 %
how much time	**1.3 %**	0.1 %	0.0 %

Examples of argumentative questions for each characteristic:

Characteristic	Question
Yes/No	is marijana legalization possible?
Predictions	will marijuana be legalized in Russia?
Comparisons	should we have partial or full marijuana legalization?
Personal Pronouns	do you think the president will legalize marijuana?
why	why are people in favor of legalizing marijuana?

Question Classification Experiment

Classification of questions on controversial topics into factual, argumentative, and method.

Experimental setting:
- In-topic: 5-fold cross validation
- Cross-topic: train on 18 topics and test on leave-out topic

In-topic results in F1-score

Classifier	Factual	Argumentative	Method	Macro
Random	0.44	0.31	0.13	0.29
Majority	0.78	0.00	0.00	0.26
Rule-based	0.71	0.48	0.62	0.60
Logistic Rgression	0.86	0.67	0.70	0.74
RuBERT	0.90	0.78	0.83	0.84

Cross-topic results in F1-score

Classifier	Factual	Argumentative	Method	Macro
Random	0.43	0.30	0.12	0.28
Majority	0.78	0.00	0.00	0.26
Rule-based	0.69	0.46	0.56	0.57
Logistic Rgression	0.80	0.61	0.52	0.65
RuBERT	0.85	0.74	0.74	0.78