

# Topic Ontologies for Arguments

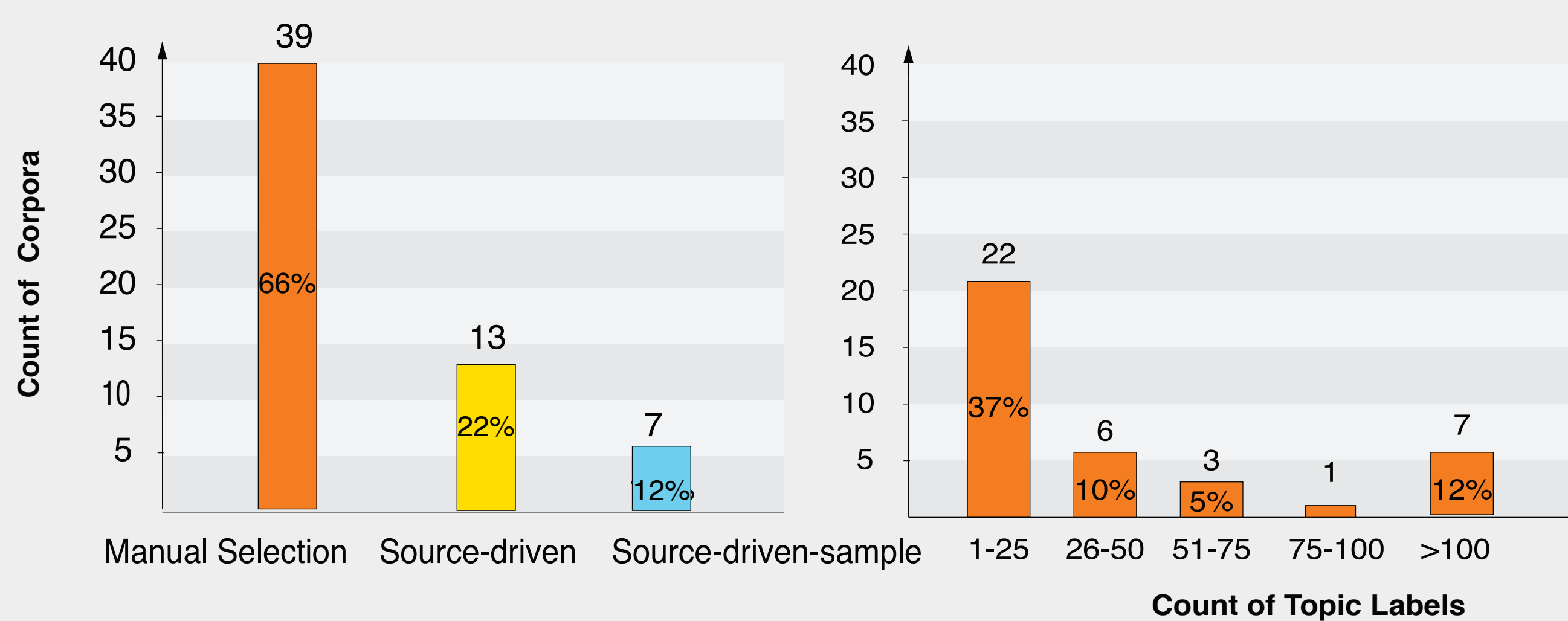
## Motivation

- Can argument mining approaches generalize to new topics?
- To study their generalizability the topic distribution of the test split should be different from the training one.
- Research question: how well do existing argument corpora cover controversial topics?

## Survey of 59 Argument Corpora

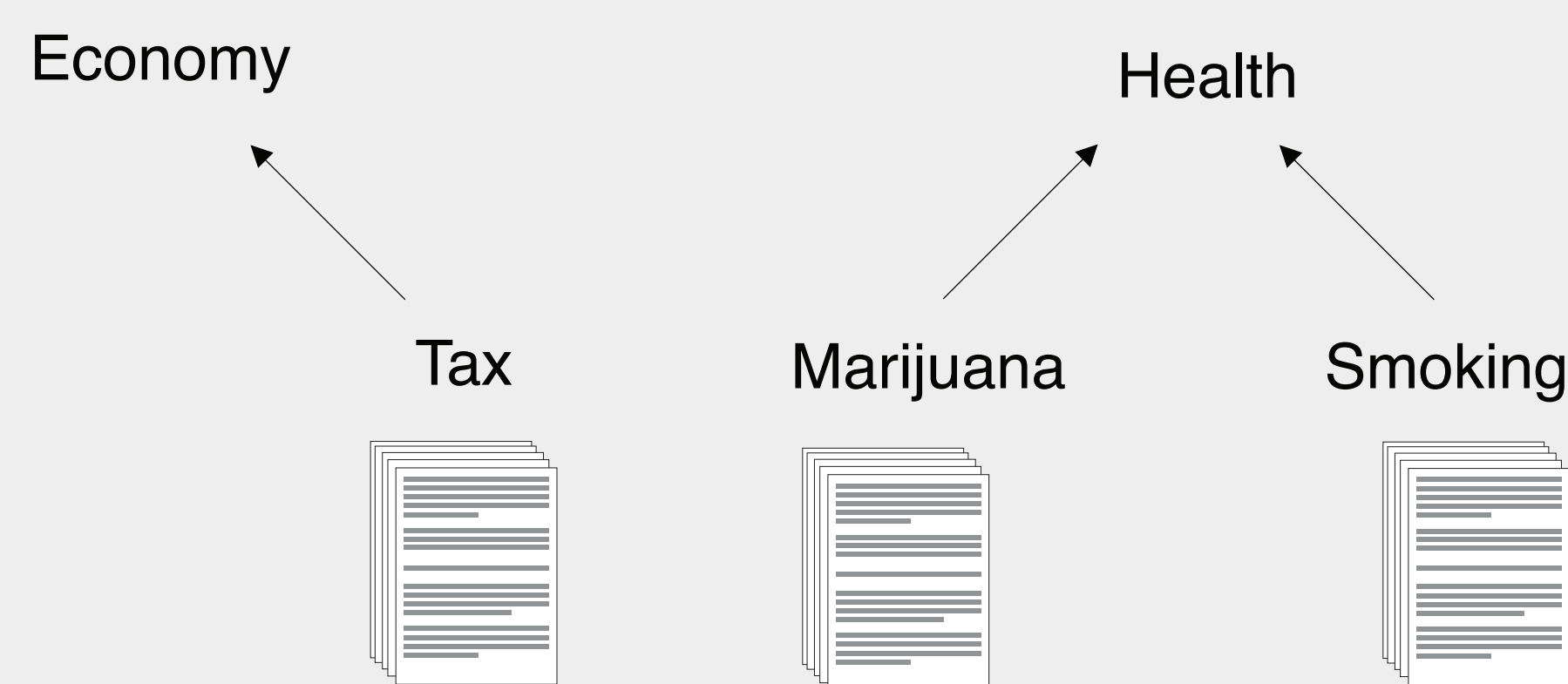
Three ways how researchers approach topics while building corpora:

1. Manual Selection
2. Source-driven
3. Source-driven-sample



## Turstworthy Argument Topic Ontologies

An ontology is an explicit specification of conceptualization: it defines the terms of a specific domain.



We create three topic ontologies starting from

1. World Economic Forum
2. Wikipedia
3. Debatepedia

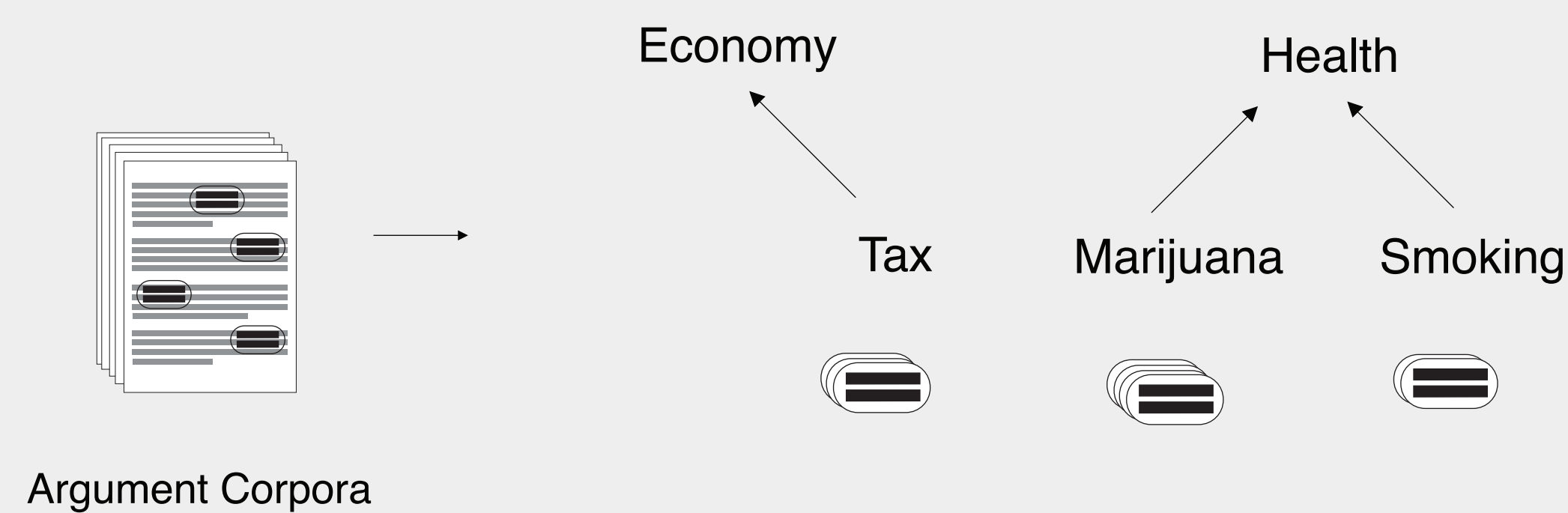


Ontology	Topics	Authors	Documents
World Economic Forum Level-1	137	334	940
World Economic Forum Level-2	822	217	550
Wikipedia Level-1	14	78,014	68
Wikipedia Level-2	748	1,930	1
Debatepedia	89	145	62

## Contributions

- Survey and analysis of 59 argument corpora with regard to their topics.
- Introducing the idea of topic ontologies to model arguments context.
- An effective approach to categorize arguments within a topic ontology.
- Released three argument topic ontologies and the mappings of argument corpora to them.
- **Finding:** Argument corpora cover a small set of topics and the distribution is skewed.

## Mapping Corpora Units to Topic Ontologies



A. Manual mapping of corpora labels to a topic ontology.

1. Normalization (e.g., "Abortion should be legalized" -> abortion)
2. Retrieval: using topic labels as queries and ontology topics as documents.
3. Labeling: is a topic label a subtopic or a synonym of the retrieved ontology topics.

B. Automatic categorization of corpora units

Task: is a corpus unit about an ontology topic?

Direct match checks whether the topic occurs in the corpus unit's text

Semantic indexing calculates the similarity between a corpus unit

Approach	WEF		Wikipedia		Debatepedia
	Level-1	Level-2	Level-1	Level-2	
Direct match	0.29	0.19	0.06	0.40	0.42
Semantic indexing	0.34	0.32	0.43	0.59	0.56
Text2vec_BERT	0.28	0.23	0.47	0.31	0.50

## Topic Coverage and Distribution of Argument Corpora

