Towards Analyzing the Bias of News Recommender Systems Using Sentiment and Stance Detection

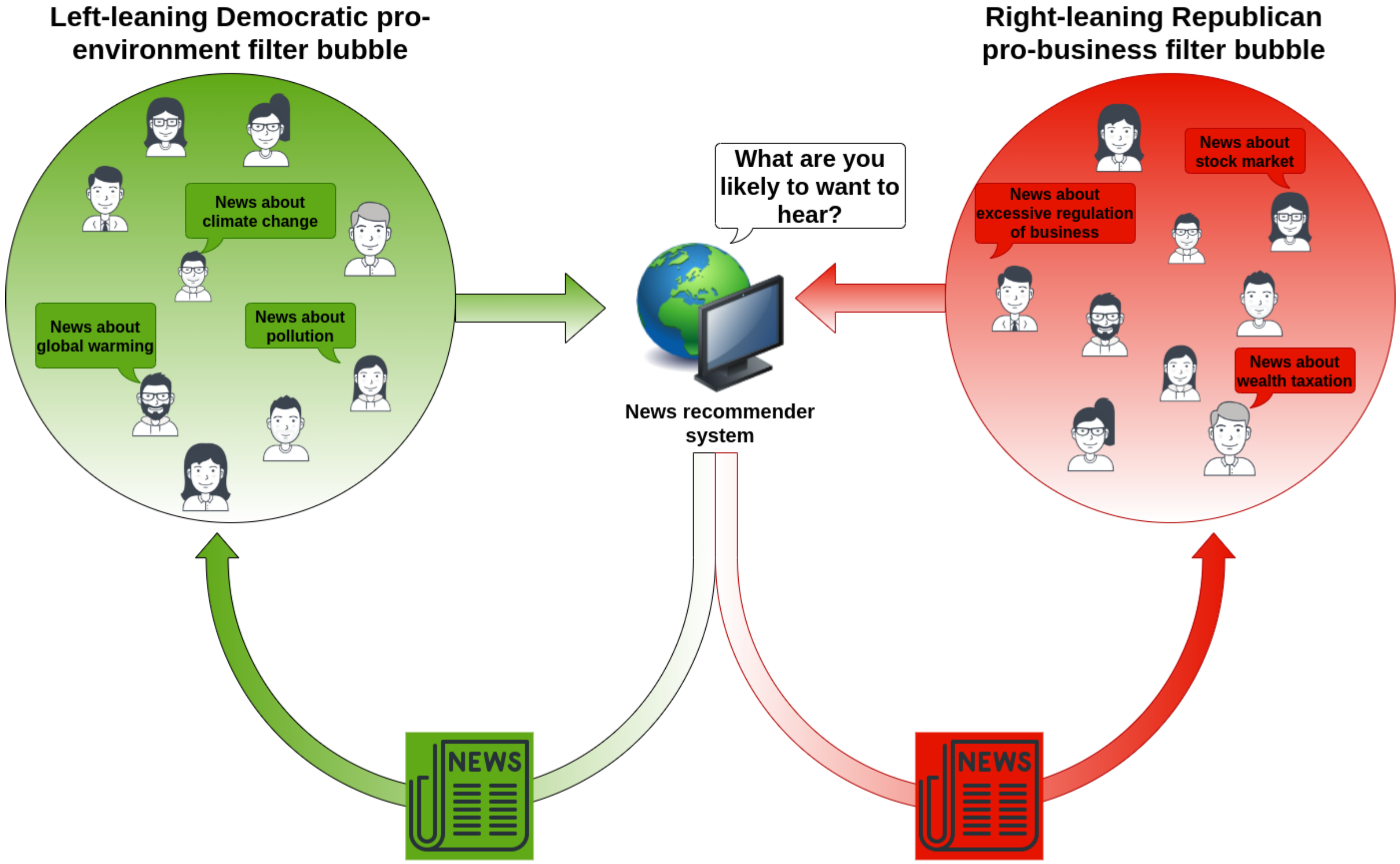
Mehwish Alam, **Andreea Iana**, Alexander Grote, Katharina Ludwig, Philipp Müller, Heiko Paulheim

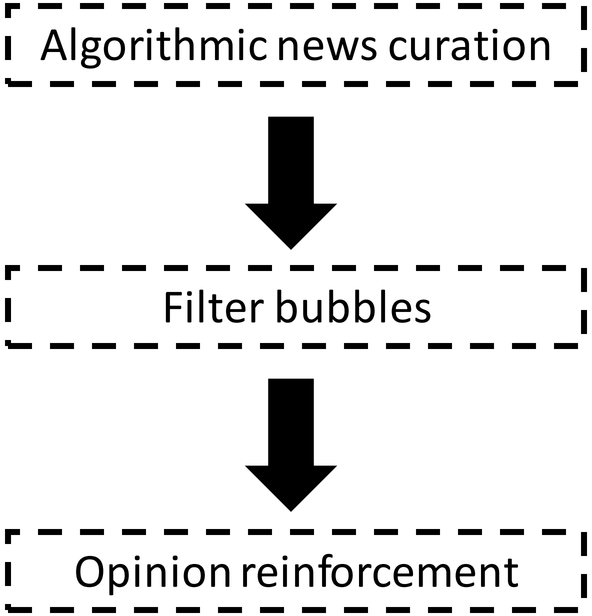
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

UNIVERSITÄT MANNHEIM

KIT
Karlsruher Institut für Technologie

Motivation





Filter bubble: self-reinforcing system state in which only news with a certain standpoint are presented to the user

[1] Pariser, Eli. The filter bubble: How the new personalized web is changing what we read and how we think. Penguin, 2011.

Diversity in News Recommendation

- Source diversity (pluralism of sources)
- Content diversity (pluralism of topics)
- **Viewpoint diversity** (pluralism of stances on a given topic)
  - Sentiment analysis (i.e. determine if text is positive, neutral, negative)
  - Stance detection (i.e. determine an author's viewpoints towards given target issue)

Are different kinds of recommender systems biased towards certain sentiments or stances, and how does this affect diversity of recommendations and users' selective exposure?

[2] Natali Helberger. 2019. On the democratic role of news recommenders. Digital Journalism 7, 8 (2019), 993–1012.
[3] Christian Baden and Nina Springer. 2017. Conceptualizing viewpoint diversity in news discourse. Journalism 18, 2 (2017), 176–194.
[4] Mario Haim, Andreas Graefe, and Hans-Bernd Brosius. 2018. Burst of the filter bubble? Effects of personalization on the diversity of Google News. Digital journalism 6, 3 (2018), 330–343.
[5] Paul S Voakes, Jack Kapfer, David Kurpius, and David Shano-yeon Chern. 1996. Diversity in the news: A conceptual and methodological framework. Journalism& Mass Communication Quarterly 73, 3 (1996), 582–593.

Road Map

- **Corpus Collection**
- Generating a Knowledge Graph of News Articles (GeNeG)
- Sentiment Analysis & Stance Detection
- Evaluating Recommender Systems
- Analyzing Bias in News Recommender Systems

Corpus Collection

Source

- **45** German media **outlets**
- Topic: **refugees & migration** captured using **keywords** (flüchtl*, geflücht*, asyl*, zuwander*, einwander*, immigrant*, immigration*, migration*, migrant*, ausländer, refug*, rapefug*, invasor*)

Inclusion criteria

- Article should contain **at least 2 keywords,** separated by **min. 50 words**
- Article length: **min. 150 words**
- Time frame: published between **01.01.2019 - 20.10.2020**

Exclusion criteria

- Paid articles
- Foreign language articles
- Disclaimers, advertisements, buying options, reader comments, etc.
- Articles consisting only of announcements about publications (e.g. books, movies), TV programs or recommendations of movie or books

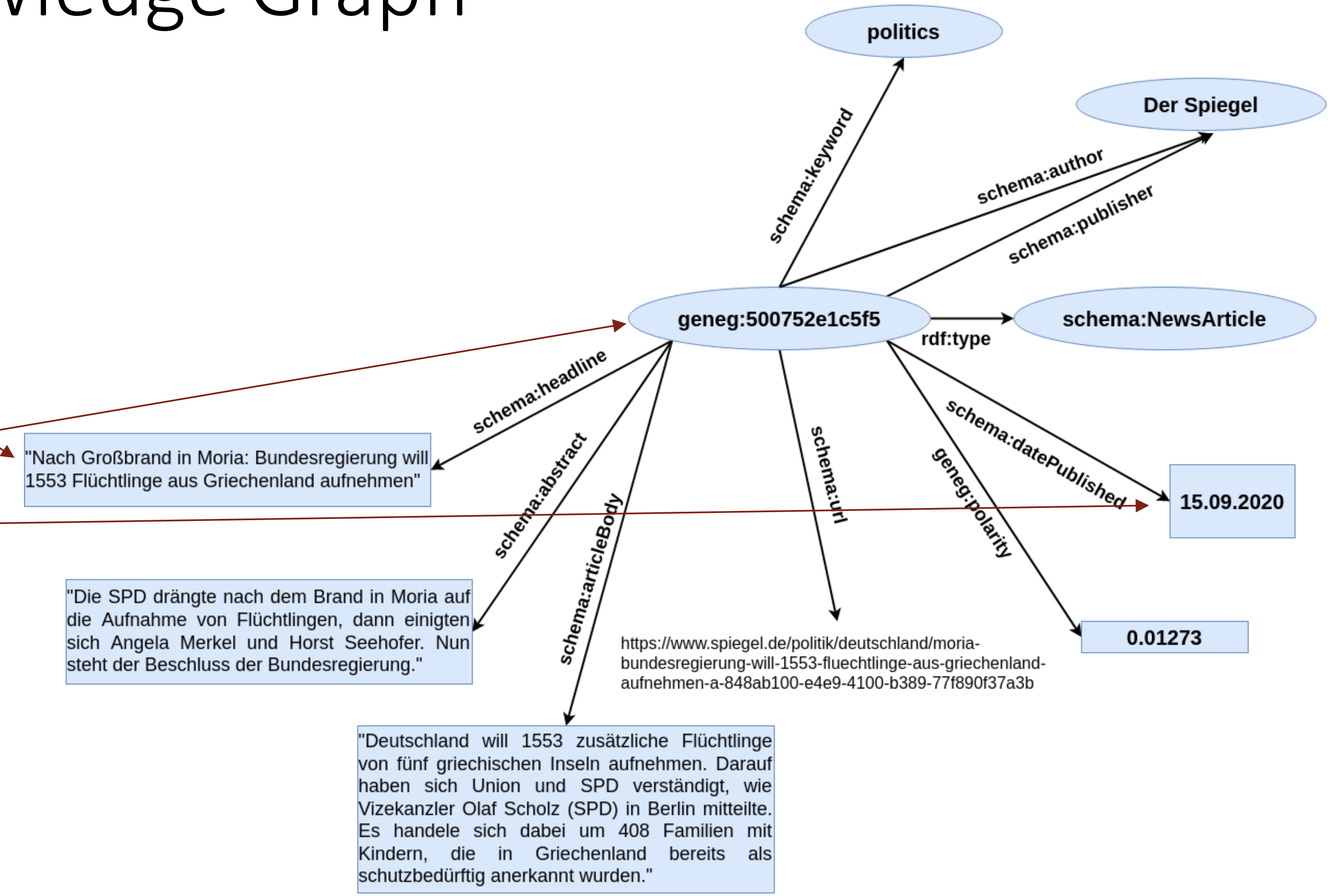
→ **4,557 news articles**

Road Map

- Corpus Collection
- **Generating a Knowledge Graph of News Articles (GeNeG)**
- Sentiment Analysis & Stance Detection
- Evaluating Recommender Systems
- Analyzing Bias in News Recommender Systems

From Article to Knowledge Graph

Metadata information

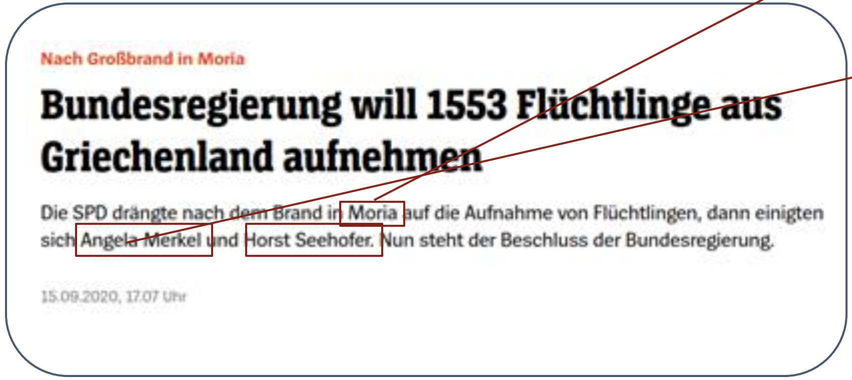


Schemas & Vocabularies

- *Schema.org* -> metadata relations

From Article to Knowledge Graph

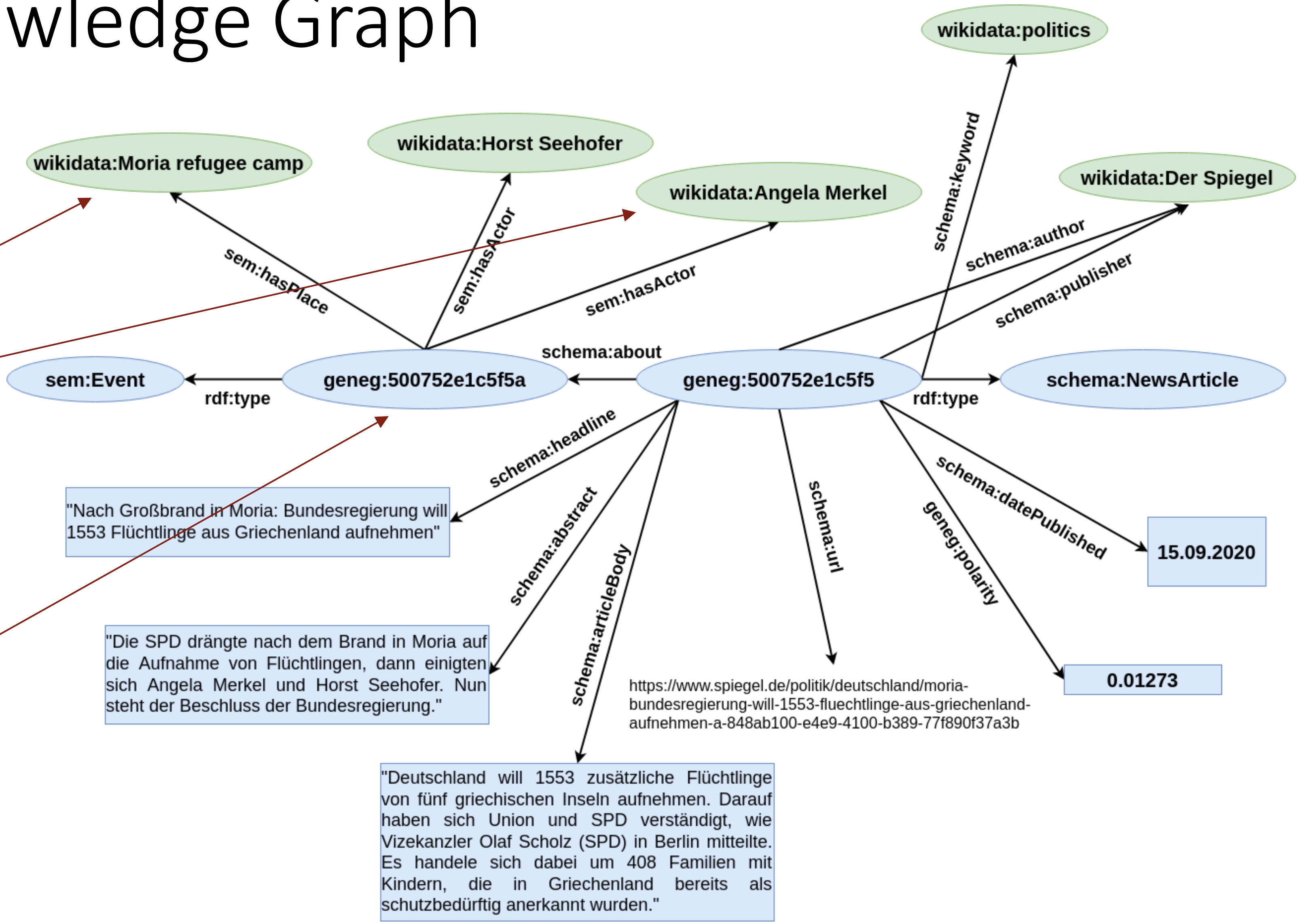
Entities as actors and locations


Nach Großbrand in Moria

Bundesregierung will 1553 Flüchtlinge aus Griechenland aufnehmen

Die SPD drängte nach dem Brand in Moria auf die Aufnahme von Flüchtlingen, dann einigten sich Angela Merkel und Horst Seehofer. Nun steht der Beschluss der Bundesregierung.

15.09.2020, 17.07 Uhr



News discuss events

Schemas & Vocabularies

- *Simple Event Model* -> event modelling

[6] Van Hage, Willem Robert, et al. "Design and use of the Simple Event Model (SEM)." Journal of Web Semantics 9.2 (2011): 128-136.

External Knowledge

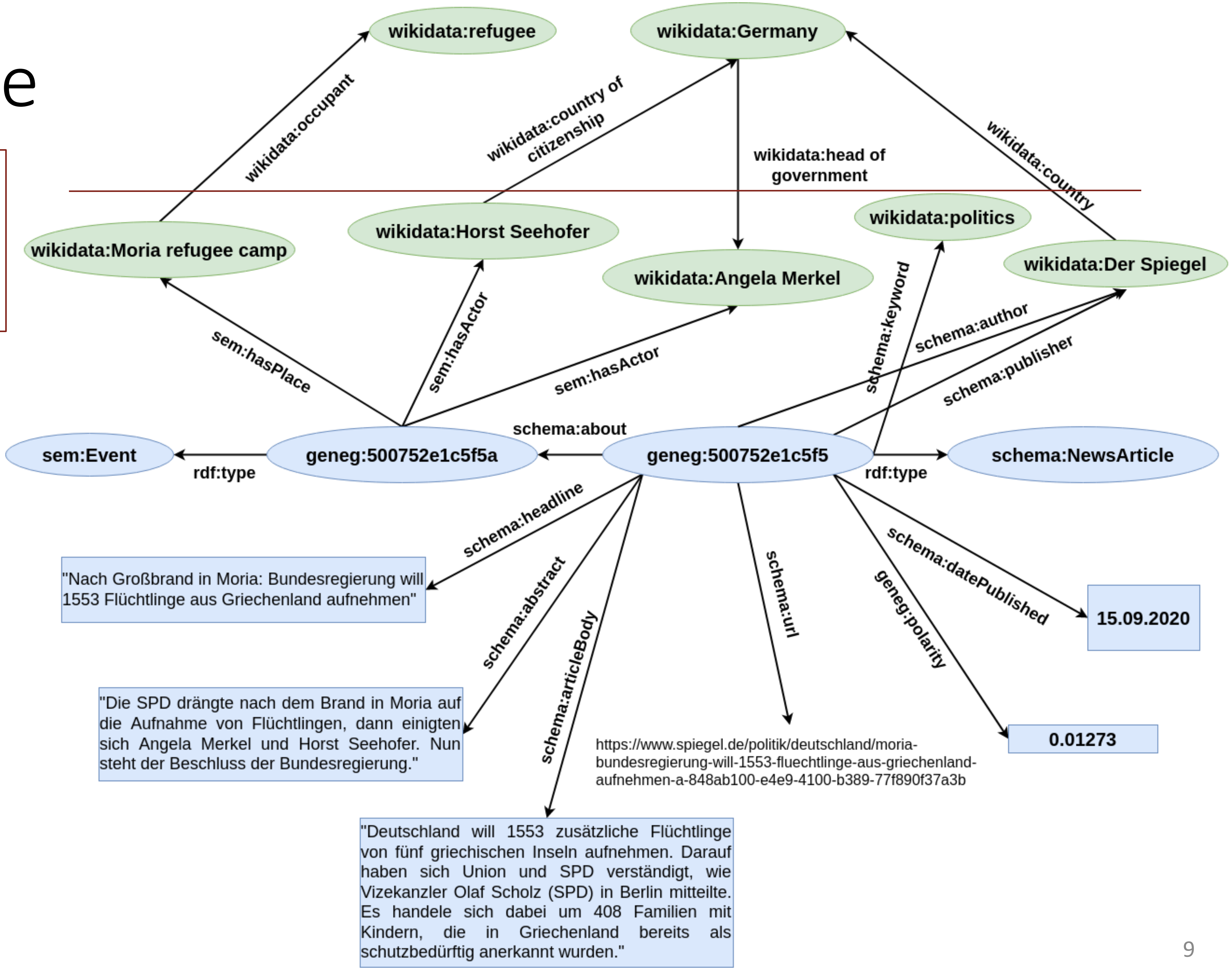
Additional relations and entities can be added from Wikidata

Nach Großbrand in Moria

Bundesregierung will 1553 Flüchtlinge aus Griechenland aufnehmen

Die SPD drängte nach dem Brand in Moria auf die Aufnahme von Flüchtlingen, dann einigten sich Angela Merkel und Horst Seehofer. Nun steht der Beschluss der Bundesregierung.

15.09.2020, 17.07 Uhr



Schemas & Vocabularies

- *Wikidata* -> linkage target for organizations, places and persons

GeNeG

- *Base* graph: textual information + metadata + entities
- *Entities* graph: *base* graph w/o literal nodes + 3-hop entity neighbors from Wikidata
- *Complete* graph: *base* + *entities* graphs

Graph	# Nodes	# Edges	# Properties
Base	54,327	186,584	16
Entities	844,935	6,615,972	1,263
Complete	868,159	6,656,779	1,271

Road Map

- Corpus Collection
- Generating a Knowledge Graph of News Articles (GeNeG)
- **Sentiment Analysis & Stance Detection**
- Evaluating Recommender Systems
- Analyzing Bias in News Recommender Systems

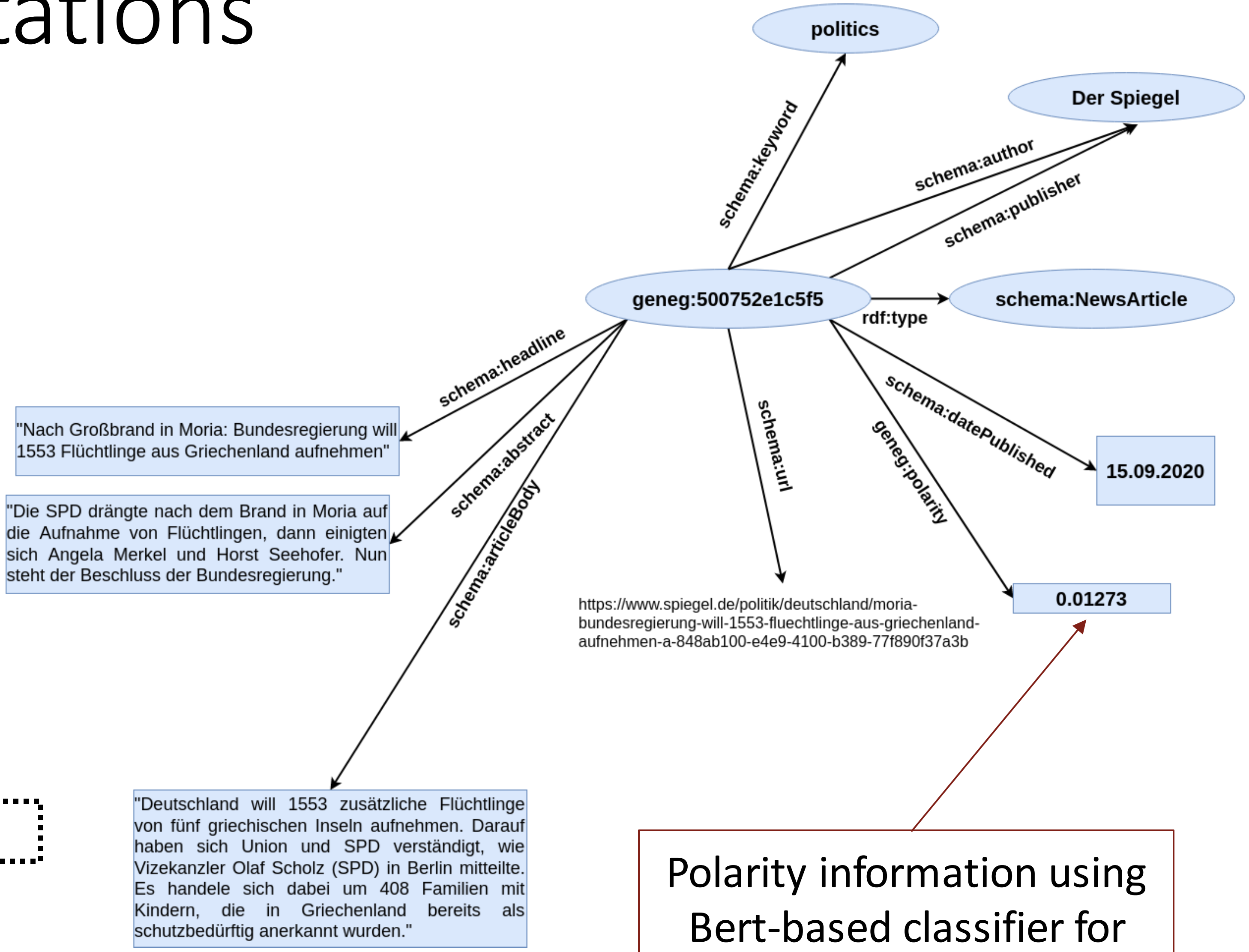
Sentiment Annotations


Nach Großbrand in Moria

Bundesregierung will 1553 Flüchtlinge aus Griechenland aufnehmen

Die SPD drängte nach dem Brand in Moria auf die Aufnahme von Flüchtlingen, dann einigten sich Angela Merkel und Horst Seehofer. Nun steht der Beschluss der Bundesregierung.

15.09.2020, 17.07 Uhr



sentiment score = $\text{probability}_{\text{positive}} - \text{probability}_{\text{negative}}$

Polarity information using Bert-based classifier for German

[7] Guhr, Oliver, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. "Training a broad-coverage German sentiment classification model for dialog systems." In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 1627-1632. 2020.

Stance Detection

- Classify article as being **in favor / against** a given question
- Pre-trained GermanBERT model, fine-tuned on the German subset of the x-stance dataset

Questions	Articles in Favor	Articles Against	Average Score
(Q1) Are you in favor of refugees coming to Germany?	2,165	2,392	-0.050
(Q2) Are you in favor of refugees living in Germany?	2,193	2.364	-0.038
(Q3) Are you in favor of refugees working in Germany?	2,210	2,347	-0.030
(Q4) Should Germany take in refugees?	2,120	2,437	-0.070
(Q5) Should Germany help refugees?	2,192	2,365	-0.038

[8] Branden Chan, Stefan Schweter, and Timo Möller. 2020. German's Next Language Model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6788–6796, 2020
[9] Jannis Vamvas and Rico Sennrich. 2020. X-Stance: A Multilingual Multi-Target Dataset for Stance Detection. CoRR abs/2003.08385 (2020). arXiv:2003.08385

Road Map

- Corpus Collection
- Generating a Knowledge Graph of News Articles (GeNeG)
- Sentiment Analysis & Stance Detection
- **Evaluating Recommender Systems**
- Analyzing Bias in News Recommender Systems

User Data

Goal: measuring the political polarization effect of recommender systems on users

- Article randomly assigned to the participants from one of four recommenders, namely TF-IDF, Word2vec, Transformer, or a random recommendation baseline
- Participants were allowed to choose an article from a preview of six articles, and then to read it
- The user’s choices were included in his or her reading history
- This process was repeated four times

Dataset	Items	Users
Total	3,825	1,417
Training	3,365	1,414
Complete test	1,633	1,174
Random test	316	177

Performance Evaluation

- Evaluation on Click-Through Rate (CTR)
  - Each recommender is applied on every user-article pair from the test set to predict the user's likelihood of clicking the candidate article
- A min-max scaling was applied to the similarity measures generated by the text-based recommenders as an approximation of probability scores

Models	**Complete**			**Random**		
	ACC	AUC	F1	ACC	AUC	F1
TF-IDF	**0.732**	**0.873**	0.647	0.487	0.499	0
Word2Vec	0.514	0.794	**0.674**	0.499	0.474	0.663
Transformer	0.505	0.779	0.671	0.499	0.515	**0.665**
RippleNet	0.553	0.574	0.523	**0.559**	**0.578**	0.531

[8] Wang, Hongwei, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. "Ripplenet: Propagating user preferences on the knowledge graph for recommender systems." In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 417-426. 2018.

Road Map

- Corpus Collection
- Generating a Knowledge Graph of News Articles (GeNeG)
- Sentiment Analysis & Stance Detection
- Evaluating Recommender Systems
- **Analyzing Bias in News Recommender Systems**

Bias Analysis

- Are recommenders prone to stance or sentiment bias?
- Bias score in [-1,+1]
  - -1 represents a user’s or recommender’s tendency for articles with **negative sentiments or stances against** the given topic
  - +1 denotes the opposite situation
- Following biases were computed:
  - Recommender Bias
  - Correlation between Recommender and User Biases

Recommender Sentiment Bias

Do the recommender systems have a tendency to recommend articles with a certain sentiment?

- Average corpus sentiment score: - 0.154
- Users are **significantly** more prone to reading news with **negative sentiments**
- Recommenders likely to suggest news with negative sentiments
- RippleNet appears to be the least prone to sentiment bias

Test set	Avg. user sentiment score	Avg. recommender sentiment score			
		TF-IDF	Word2Vec	Transformers	RippleNet
Complete	-0.171*	-0.162	-0.169*	-0.157	-0.148
Random	-0.169	-0.141	-0.170	-0.160	-0.150

Correlation of Recommender and User Sentiment Bias

How does the recommenders' sentiment bias correlate with the exiting user sentiment bias?

- **No statistically significant difference** between the avg. sentiment score of **text-based recommendations** and the avg. user score
- **Statistically significant difference** between the avg. sentiment score of RippleNet recommendations and the avg. user score (on the complete test set)

- **TF-IDF, Word2vec, Transformer***-based recommendations are **positively correlated** with the user bias
- **RippleNet**-based recommendations are **slightly positively correlated** with the user bias only on the complete test set

* The correlation is significant only on the complete test set

Recommender Stance Bias

Do the recommender systems have a tendency to recommend articles with a certain stance?

- Recommenders show a **tendency** towards news with a **stance against** the topic, for all questions (statistically significant only for Q3)
- RippleNet appears to be the least prone to negative bias

Question	Avg. User Stance Score	Avg Recommender Stance Score (complete/random test sets)			
		TF-IDF	Word2Vec	Transformer	RippleNet
Q1	-0.109 / -0.093	-0.140 / -0.227	-0.165 /-0.219	-0.136 /-0.172	-0.082 /-0.054
Q2	-0.102*/-0.093	-0.132 /-0.220	-0.158 /-0.207	-0.131 /-0.169	-0.074 /-0.038
Q3	-0.092* / -0.081	-0.127* /-0.215	-0.149* /-0.205	-0.116*/-0.164	-0.062* /-0.024
Q4	-0.117 / -0.106	-0.167 /-0.255	-0.178 /-0.268	-0.157 /-0.179	-0.095 /-0.084
Q5	-0.079 / -0.081	-0.130 / -0.237	-0.135 /-0.199	-0.124 /-0.143	-0.060 /-0.055

Correlation of Recommender and User Stance Bias

How does the recommenders' stance bias correlate with the exiting user sentiment bias?

- **Word2vec, TF-IDF****-based recommendations **exacerbate** user's preferences towards news **against** the topic of refugees and migration
- **Word2vec*, TF-IDF*, Transformer**-based recommendations are **statistically significant positively correlated** with the existing user stance bias
- RippleNet-based recommendations show **no statistically significant correlation** with the user stance bias

* The correlation is significant only on the complete test set

** The correlation is significant only on the random test set

Main Findings

- Sentiment and stance annotations can be used to quantify sentiment and stance bias of recommendations generated by different algorithms
- Text-based recommender systems expose amplification of user attitudes with respect to sentiment and stance
- The knowledge-aware recommender appears less prone to both types of biases, at the cost of prediction accuracy

- Future research: how to **balance** between **performance** (prediction accuracy) and **diversity** (of sentiments and stances in recommended news)