COMMUNICATION MANAGEMENT FORUM2021

HRZZ
Croatian Science Foundation

SOCRES (IP-CORONA-2020-04)

An exploratory computational analysis of COVID-19 related news in Croatia during 2020

Bilić, P., Furman, I., Yildirim, S., Šnajder, J., Dukić, D., Gjurković, M., Vukojević, I.
Institute for Development and International Relations, Zagreb
Istanbul Bilgi University, Turkey
TakeLab, Faculty of Electrical Engineering, University of Zagreb

COMPUTATIONAL SOCIAL SCIENCE

- Computational methods have found a wide area of application in social sciences and humanities. Some authors argue it has re-conceptualised some of the approaches in sociology towards what is sometimes called digital sociology (e.g. Lupton, 2014; Marres, 2017, Orton-Johnosn, Prior, 2014)
- **The importance of theory and analytical categories in this project:** The PEC allows us to interpret organisational behavior of the media. Successfully combined with computational approaches in recent research (Bilić, Furman, Yildirim, 2018; Furman, Saka, Yildirim, Elbeyi, 2019; Birkinbine and Gomez, 2020)
- **Computational methods used:** Natural Language Processing (NLP), active learning, correspondence analysis, LDA topic modelling

RAW DATASET

- **STEP 1: define which organisations and which news items to retreieve**

- Media organisations selected on the basis of audience reach, regional coverage (Rijeka, Osijek, Split), public ownership, and non-profit ownership. News items selected on the basis of tags and key terms related to the corona virus, civil protection headquartes, scientific advisory, and other scientists commenting on the crisis in the media

- **STEP 2: data retrieval and dataset preparation with computational techniques**

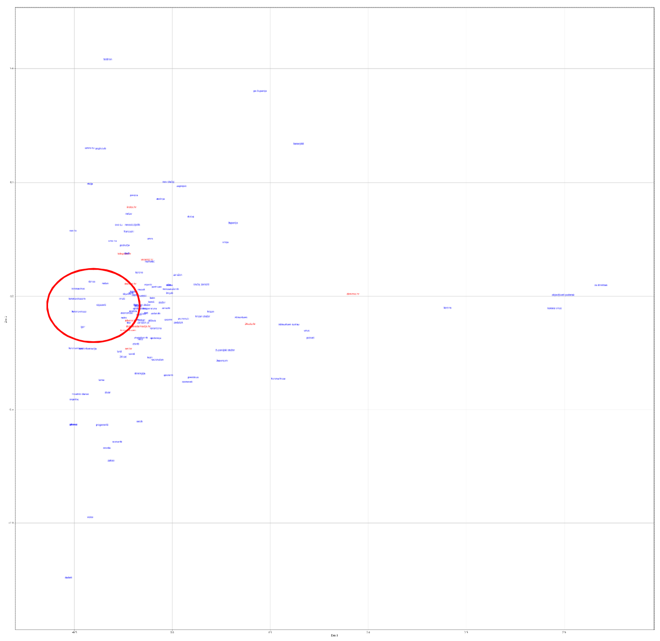
- Textual data crawled from the selected web portals, stored in a database, and then further processed by natural language processing (NLP) tools to extract the relevant pieces of information, ranging from basic lexical cues such as word n-gram counts and collocations to more sophisticated information signals such as named entities and keyphrases. All items published in 2020 were retrieved. Total of 190969 news items.

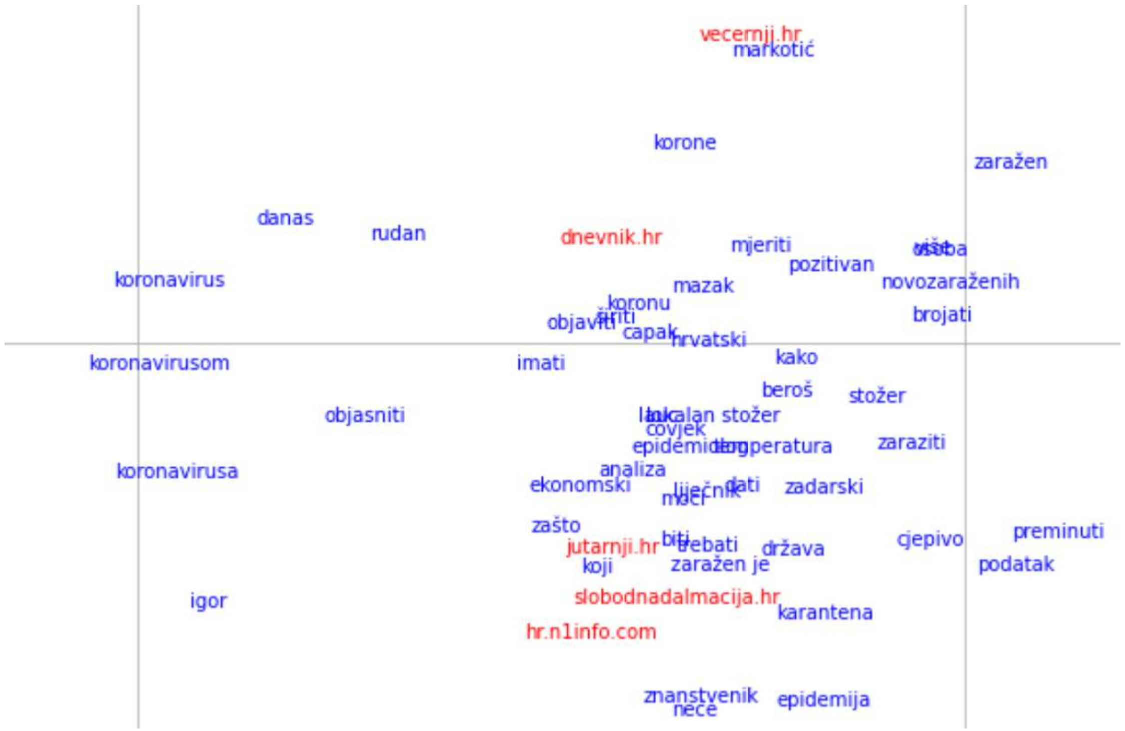
- **STEP 3: find patterns in the preliminary dataset (human interpretative capacity)**

index.hr	Commercial	28278
jutarnji.hr	Commercial	25428
direktno.hr	Commercial	20922
slobodnadalmacija.hr	Commercial, regional	16990
vecernji.hr	Commercial	16385
hr.n1info.com	Commercial	15977
net.hr	Commercial	14740
dnevnik.hr	Commercial	13134
24sata.hr	Commercial	12921
telegram.hr	Commercial	7775
novilist.hr	Commercial, regional	6860
rtl.hr	Commercial	3788
hrt.hr	Public service	3024
dnevno.hr	Commercial	2610
tris.com.hr	Non-profit	654
glas-slavonije.hr	Commercial, regional	441
h-alter.org	Non-profit	363
tportal.hr	Commercial	305
lupiga.com	Non-profit	223
forum.tm	Non-profit	92
crol.hr	Non-profit	59
		190969

CORRESPONDENCE ANALYSIS (top ten)

COMMUNICATION MANAGEMENT FORUM2021

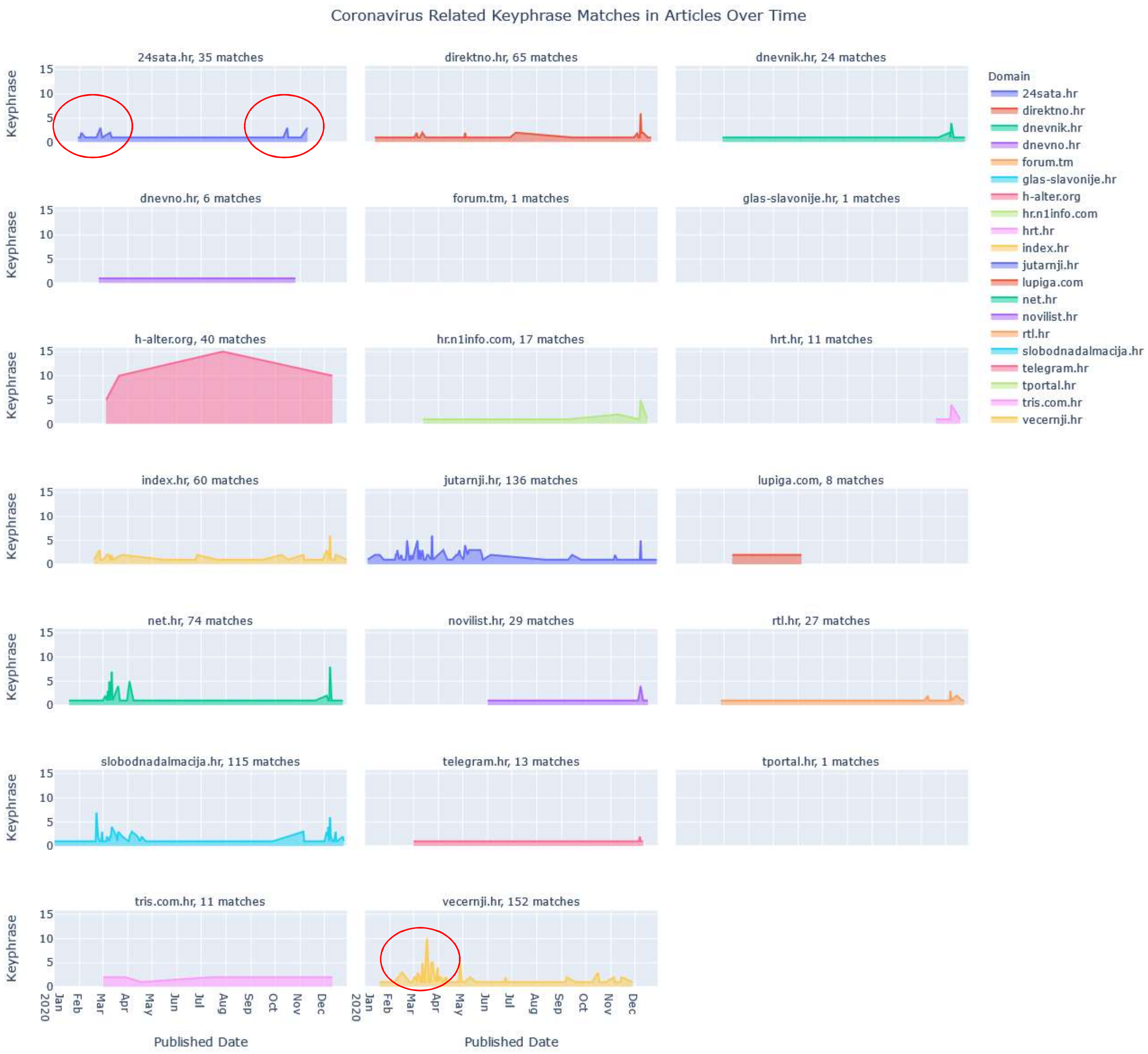




An exploratory computational analysis of COVID-19 related news

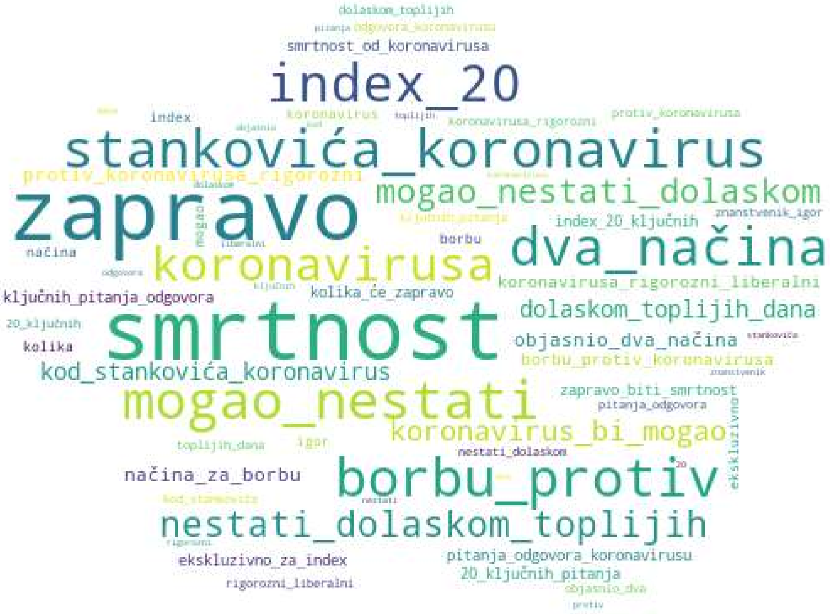
Bilić, Furman, Yildirim, Šnajder, Dukić, Gjurković, Vukojević

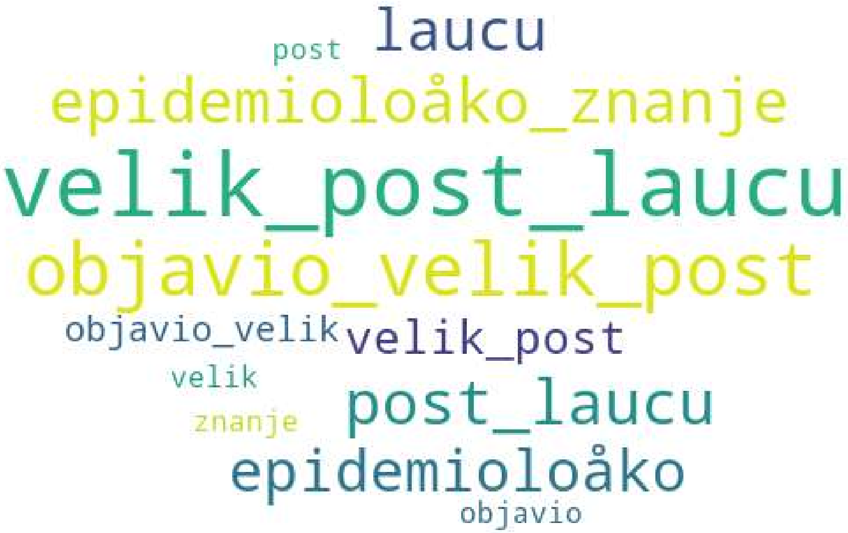
Ministarstvo znanosti i obrazovanja
Edward Bernays University College Communications | Tourism
IDIZ Institut za društvena istraživanja u Zagrebu Institute for Social Research in Zagreb
institutzaturizam institutefortourism



Frequency of the ngram ″Rudan″ in the entire dataset over time

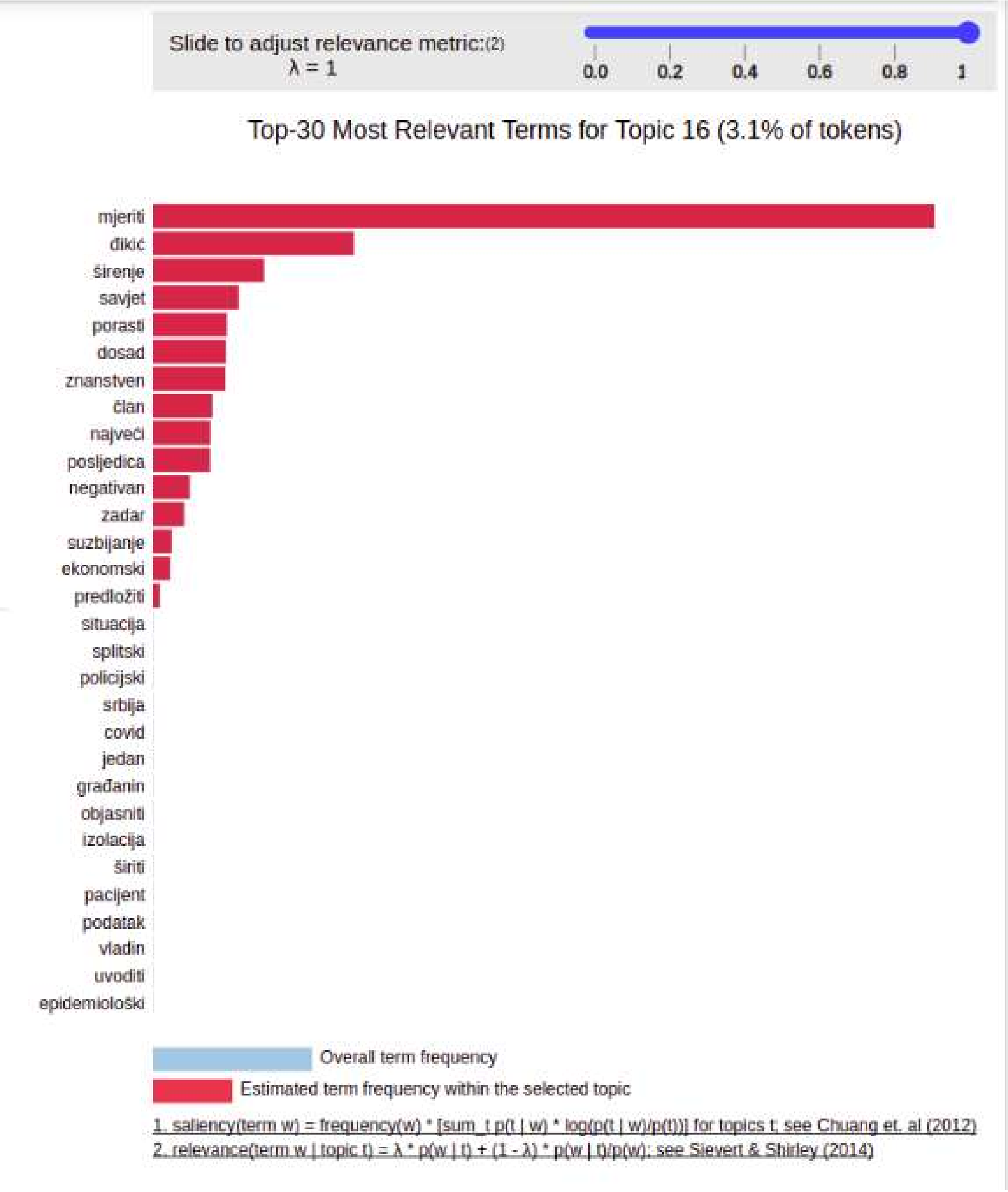
KEYPHRASE ˝Rudan˝ (index.hr March vs December)





RANKED NGRAMS (top ten media)

- Most frequently mentioned names: Beroš (947), Capak (663), Đikić (417), Božinović (408), Markotić (363), Lauc (298), Trump (107), Rudan (87)
- Most frequently mentioned geographical locations: Croatian [hrvatski] (1913), Split [splitski] (420), Europe (412), Italy (389), Serbia (328), Slovenia (290), Swedish (186), etc.
- Wuhan (14) – 42.9% of which relate to jutarnji.hr
- **What are the topics these ngrams are connected to? In other words, what is the context and media frame of these salient ngrams?**



Slide to adjust relevance metric:(2)
λ = 1

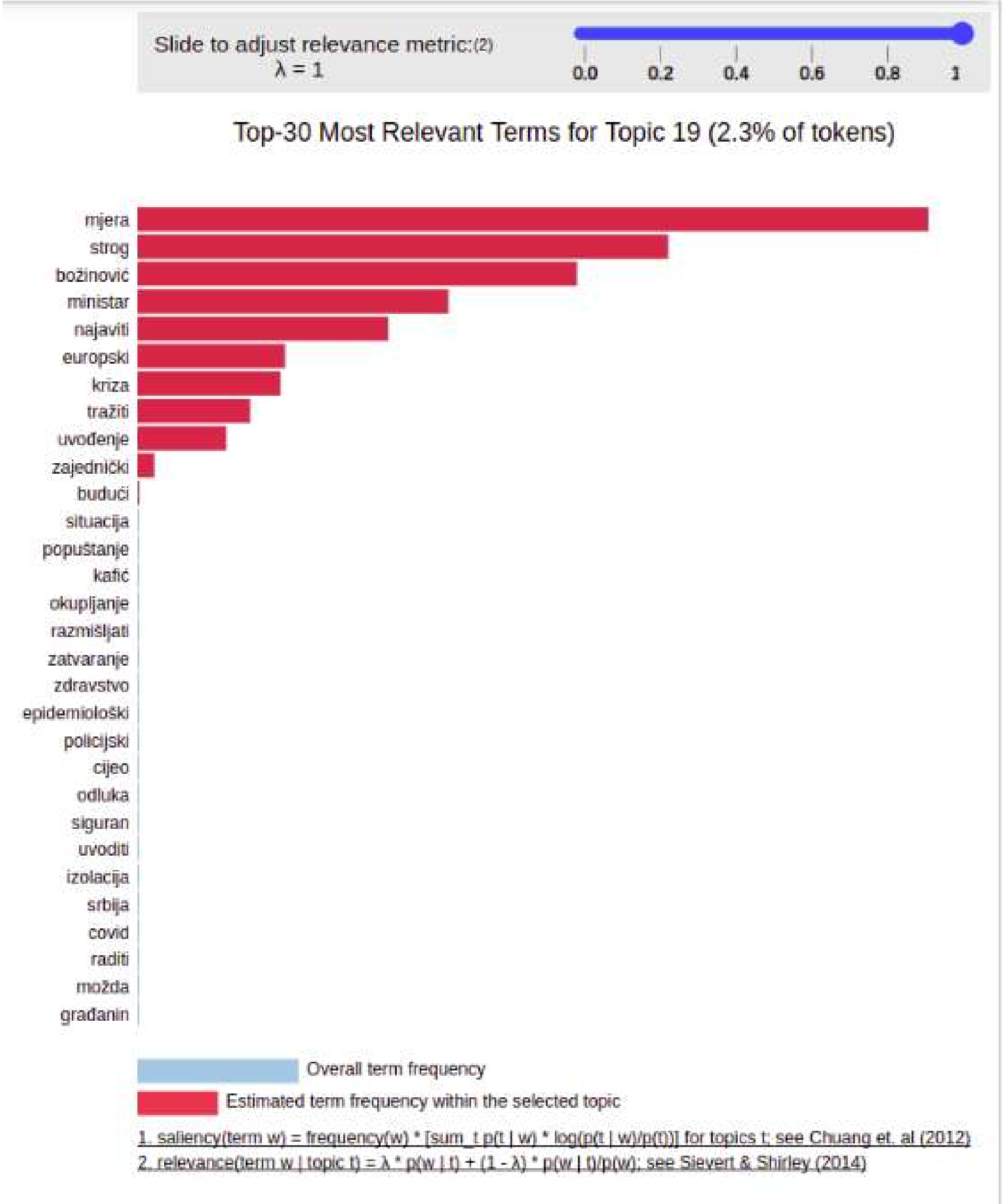
Top-30 Most Relevant Terms for Topic 16 (3.1% of tokens)

🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)



Slide to adjust relevance metric:(2)
λ = 1

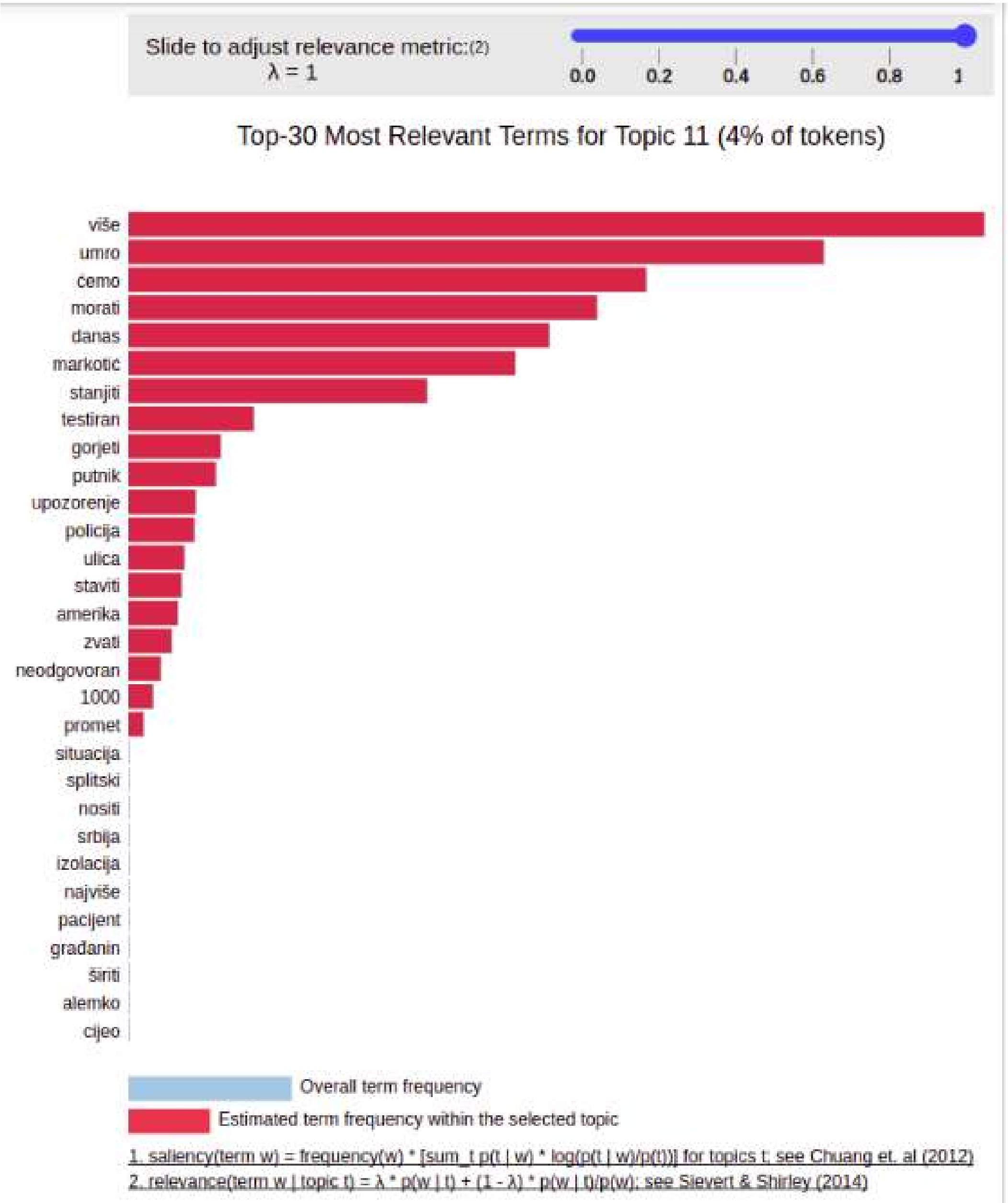
Top-30 Most Relevant Terms for Topic 19 (2.3% of tokens)

🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)



Slide to adjust relevance metric:(2)
λ = 1

Top-30 Most Relevant Terms for Topic 11 (4% of tokens)

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

NEXT STEPS

- **STEP 4:** Cleaning the dataset to remove false positives as well as to annotate for topical specificity (society, politics, economy, healthcare system, scientific communication etc.), and semantic tendency and resilience (coping, adapting, transforming)
- **STEP 5:** Formulate hypotheses and find patterns and differences between commercial, public and non-profit media
- **STEP 6:** Integrate content analysis with audience behavior. Who is engaging with these news items? Do they trust them? Online panel survey being conducted in March 2021

COMMUNICATION
MANAGEMENT
FORUM2021

An exploratory computational analysis of COVID-19 related news

Bilić, Furman, Yildirim, Šnajder, Dukić, Gjurković, Vukojević

THANK YOU FOR YOUR ATTENTION!

The post-pandemic world:
A bad picture or a good opportunity?

Ministarstvo znanosti i obrazovanja

Edward Bernays University College Communications | Tourism

IDIZ Institut za društvena istraživanja u Zagrebu Institute for Social Research in Zagreb

institutzaturizam institutefortourism