

D3.2:

Tools to identify populist narratives

WP3 – Narratives Analysis and ICT tools



Lead authors	Róbert Bjarnasson (CF) Iceland
Other contributors	

Due date	31 st of January 2021
Delivery date	30 th of April 2021
Type	Research Results
Dissemination level	Public

Keywords	Populist narratives, CommonCrawl, AI, Open Source
----------	---

Disclaimer

The project PaCE has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 822337.

The opinions express in this document reflect only the author's view and reflects in no way the European Commission's opinions.



Introduction

This deliverable includes two components, a public website called PaCE Dashboard and an Open Source tool to identify populist narratives in the large CommonCrawl dataset using keywords and AI filtering. This document has a brief introduction to both components with links to the actual deliverables, the public website and a GitHub archive with the Open Source software.

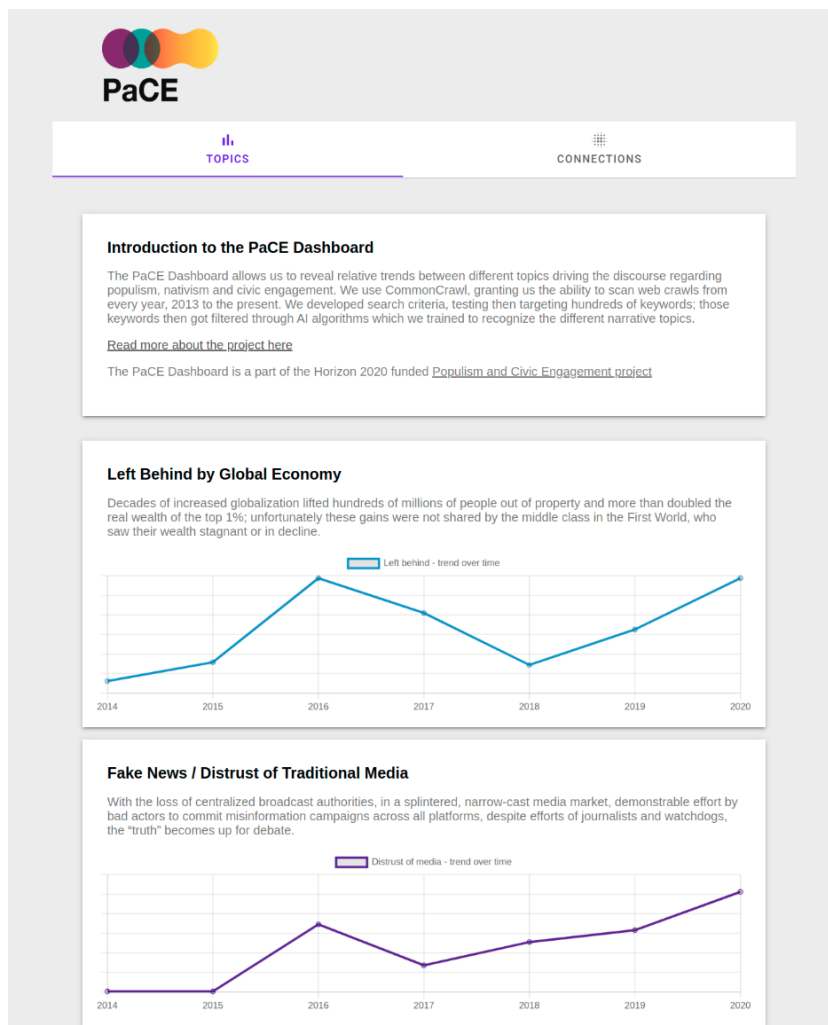


PaCE Dashboard

The PaCE Dashboard allows us to reveal relative trends between different topics driving the discourse regarding populism, nativism and civic engagement. Policy-makers, academics and citizens can use the PaCE Dashboard to see what topics are trending over the past 7 years and see random quotes from each year from each topic to better understand what type of discussion was happening at the time.

We use CommonCrawl, granting us the ability to scan web crawls from every year, 2013 to the present. We developed search criteria, testing then targeting hundreds of keywords; those keywords then got filtered through AI algorithms which we trained to recognize the different narrative topics. The first iteration of the website includes English language topics but we plan to add one or two more languages in the coming months. At the end of 2021 we will scan this year's content and add to the website which we plan to keep open at least throughout 2022.

The public website: <http://pace-dashboard.citizens.is/>



Screenshot of the public website



PaCE CommonCrawl Scanner

The PaCE CommonCrawl Scanner is an Open Source tool to identify populist narratives in the large CommonCrawl dataset using keywords and AI filtering. The tool is not limited to populist narratives, it can be used to search for any topics in any language.

The PaCE CommonCrawl scanner is already being used or will be used in other projects including: The *DEMOS H2020* project to scan for Russian propaganda, the *PopAI H2020* project to scan for topics related to law enforcement and *CODE Europe*, an EEA Funds project, will use it to scan for attitudes to air pollution in nine European countries in 2022.

Open Source code: <https://github.com/CitizensFoundation/pace-commoncrawl-scanner>

rbjarnason Work ...		3 days ago	🕒 301
📁 distanceGraphs	Work	11 days ago	
📁 evaluation	Work on training	2 months ago	
📁 keywords	Work on scan	2 months ago	
📁 processScripts	Work	19 days ago	
📁 scripts	Work on test queries for tud	2 years ago	
📁 src/main/java/org/ct/a...	Work	6 days ago	
📁 training	Work	3 days ago	
📁 utils	Wrok on scan	2 months ago	
📄 .classpath	Initial import after fork	2 years ago	
📄 .factorypath	Work on scan	3 months ago	
📄 .gitignore	Work on scan	4 months ago	
📄 .project	Work on import	5 months ago	
📄 LICENSE.txt	Initial import after fork	2 years ago	
📄 README.md	Work on evaluations	3 months ago	
📄 log4j2.xml	Initial import after fork	2 years ago	
📄 package-lock.json	Work on build	3 months ago	
📄 package.json	Work on build	3 months ago	
📄 pom.xml	Work on scan	3 months ago	
📄 warc.paths	Initial import after fork	2 years ago	

