

# Messy Problem

Currently impossible to **quickly** show the public a **visual summary** of all the **results** Sida has achieved.



Vastly different types of results in EN/SV

# Messy Problem

Currently impossible to **quickly** show the public a **visual summary** of all the **results** Sida has achieved.



Vastly different types of results in EN/SV



Hidden in 100 000 files with different formats

# Messy Problem

Currently impossible to **quickly** show the public a **visual summary** of all the **results** Sida has achieved.



Vastly different types of results in EN/SV



Hidden in 100 000 files with different formats



100 000 files scattered across openaid.se



Problem Focus

How can we visually summarize *qualitative* results hidden in *pdf* files?



Problem Focus

How can we visually summarize *qualitative* results hidden in *pdf* files?

Idea for Change

Weighted wordcloud Pretty Smart Solution



www.resi.se



Header catches

www.openaid.se

## Why a wordcloud?

#### Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data

CA DePaolo, K Wilkinson - TechTrends, 2014 - Springer

☆ 💯 Cited by 31 Related articles All 8 versions

## Assessing the Use of Key Words in Student Responses

One possible way in which word clouds can be used in assessment is to gauge the use of key words in student responses to essay or short answer questions. For example, in a statistics course, an instructor wished to assess if students understand



# "Word clouds have emerged as a straightforward and visually appealing visualization method for text."

<u>F Heimerl, S Lohmann, S Lange...</u> - System Sciences (HICSS ..., 2014 - ieeexplore.ieee.org

☆ 切り Cited by 83 Related articles All 5 versions



## Solution: How?



#### Locate pdf files



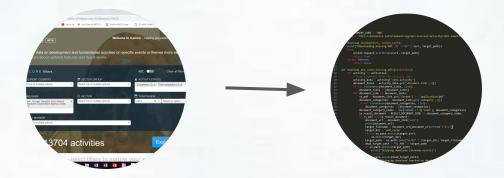
- 1) Get IATI-identifiers
- 2) Make API calls
- 3) Filter activities
- 4) Obtain URLs
- 5) Determine language

### Solution: How?



#### Locate pdf files

#### **Automate with code**



- 1) Get IATI-identifiers
- 2) Make API calls
- 3) Filter activities
- 4) Obtain URLs
- 5) Determine language

- 1) Run NLP locally
- 2) Stem words
- 3) Remove stop words
- 4) Filter pre-defined keywords
- 4) Weigh keywords (k-NN)

## Solution: How?



#### Locate pdf files

#### **Automate with code**

#### **Qualitatively visualise**







- 1) Get IATI-identifiers
- 2) Make API calls
- 3) Filter activities
- 4) Obtain URLs
- 5) Determine language

- 1) Run NLP locally
- 2) Stem words
- 3) Remove stop words
- 4) Filter pre-defined keywords
- 4) Weigh keywords (k-NN)

- 1) Fetch JSON data
- 2) Generate word cloud

