



WP2 RELATED ACTIVITIES AND SOFTWARE DEMONSTRATIONS

POLAND



AGENDA

- Changes since Sofia meeting
- Testing ISTAT software
- Issues

WP2 SOCIAL MEDIA USE CASE

- Feedback from Olav – thanks!!!
- JSON
- Github
- First phase and second phase
 - Second part of pilot need preparation of dictionaries in particular language

URL RETRIEVAL – TESTING (1/3)

- Correct identification of hyperlinks - 74%

LP	URListat	REGON	IDistat	WWW	nsk
0	http://mapa.targeo.pl/nr-89-ul-szyprow-3-gdansk-4491170/szkola-podstawowa/adres	b'0025556200000'	1	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
1	http://www.szkolnictwo.pl/gimnazjum_gdaŁ.sk.k00609.szkoŁa.podstawowa.nr.89.gimnazjum.nr.49	b'0025556200000'	2	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
2	https://pl-pl.facebook.com/zkpig8gdansk	b'0025556200000'	3	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
3	http://www.gdansk.pl/urzad-miejski/siedem-konkursow-i-ponad-860-uczniowa.53500	b'0025556200000'	4	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
4	http://www.gdansk.pl/urzad-miejski/miedzynarodowy-festiwal-miodych-mundus-cantat-po-raz-pierwszy-wgdansku.a.51421	b'0025556200000'	5	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
5	https://www.facebook.com/zkpig8gdansk/?nr	b'0025556200000'	6	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
6	http://www.schoolius.com/school/635139759905823/zkpig.nr.8	b'0025556200000'	7	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
7	https://gazetabaltycka.pl/promowane/gdansk-konkursy-na-dyrektorow-szkol-beda-jawne	b'0025556200000'	8	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
8	http://na-mapie.info/szkola/szkola-podstawowa-nr-89/	b'0025556200000'	9	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
9	http://www.dziennikbaltycki.pl/wiadomosci/gdansk/a/szkoly-publiczne-w-gdansku-podstawowki-gimnazja-licea-szkoly-srednie-i-zawodowki-lista-adresy.10498095/	b'0025556200000'	10	ZKPIG8.PL	ZESPĄŁ KSZTAŁCENIA PODSTAWOWEGO I GIMNAZJALNEGO NR 8 SZK. P. NR 89 W GDAŁ
10	http://www.eduranking.pl/szkola/specjalny-osrodek-szkolno-wychowawczy-w-tursku-tursko/	b'0019138800000'	1	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
11	http://sosw-tursko.za.pl/cms/news.php	b'0019138800000'	2	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
12	http://www.miastko.pl/pages.php?go=menu&opcja=pokaz&id=55	b'0019138800000'	3	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
13	http://miastko.naszemiasto.pl/tag/sosw-tursko.html	b'0019138800000'	4	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
14	http://www.miastko.pl/pages.php?go=news&opcja=pokaz_news&id=562	b'0019138800000'	5	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
15	http://miastko.naszemiasto.pl/tag/tursko.html	b'0019138800000'	6	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
16	http://www.miastko24.pl/index.php?option=com_content&task=view&id=1672&Itemid=238	b'0019138800000'	7	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
17	http://www.eduranking.pl/szkola/szkola-podstawowa-specjalna-w-tursku-tursko/	b'0019138800000'	8	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
18	http://biblioteka-publiczna.miastko.com.pl/index.php/wydarzenia/274-wyn-konplast.html	b'0019138800000'	9	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
19	http://www.miastko24.pl/index.php?option=com_content&task=view&id=1244&Itemid=1	b'0019138800000'	10	WWW.SOSW-TURSKO.ZA.PL	SPECJALNY OŁŠRODEK SZKOLNO-WYCHOWAWCZY W TURSKU
20	http://www.szkolnictwo.pl/szkoŁa.zawodowa.sl.upsk.b00164.zespÅŁ.szkoŁa.budowlanych	b'0018816200000'	1	ZSBIO.CIV.PL	ZESPĄŁ SZKÅŁ BUDOWLANYCH IM. INŁ»YNIERÅ»W POLSKICH W SŁUPSKU
21	http://ckziu.slupsk.pl/	b'0018816200000'	2	ZSBIO.CIV.PL	ZESPĄŁ SZKÅŁ BUDOWLANYCH IM. INŁ»YNIERÅ»W POLSKICH W SŁUPSKU
22	http://edu.slupsk.eu/placowki/567/kontakt/	b'0018816200000'	3	ZSBIO.CIV.PL	ZESPĄŁ SZKÅŁ BUDOWLANYCH IM. INŁ»YNIERÅ»W POLSKICH W SŁUPSKU
23	http://edu.slupsk.eu/placowki/567/dokumenty/1256.html	b'0018816200000'	4	ZSBIO.CIV.PL	ZESPĄŁ SZKÅŁ BUDOWLANYCH IM. INŁ»YNIERÅ»W POLSKICH W SŁUPSKU
24	http://www.schoolius.com/school/183556891728004/zespÅŁ.szkoŁa.budowlanych.i.ksztaŁ.cenia.ustawic.need.w.sl.upsku	b'0018816200000'	5	ZSBIO.CIV.PL	ZESPĄŁ SZKÅŁ BUDOWLANYCH IM. INŁ»YNIERÅ»W POLSKICH W SŁUPSKU

URL RETRIEVAL – TESTING (2/3)

The most frequent links refer to:

- Social media – ca. 40% units were successfully found on Facebook
- Maps – 49% were found on Targeo, Zumi, Wikimapia
- Governance registers – 37% units were identified in KRS (Krajowy Rejestr Sądowy – National Court Register)
- News in press
- Regional portals
- Trade portals
- Other portals

URL RETRIEVAL – TESTING (2/3)

Our ideas on increasing the accuracy in Polish conditions:

- Add additional attributes
 - REGON (ID)
 - KRS number
 - Postal code

JOB OFFERS WEB SCRAPING

COMMERCIAL

- Sources: portals
- Number of offers:
 - money.pl - 140 thous.
 - gowork.pl - 124 thous.
 - jobs.pl - 46 thous.
 - pracuj.pl - 39 thous.
 - praca.trojmiasto.pl - 9.9 thous.
 - ...

TRADE

- Sources: organizations
- Number of job offers:
 - nabory.kprm.gov.pl - 375 offers in 213 government authorities, 120 cities
 - BIP - depending on units

CASE STUDY - JOB OFFERS

BASIC DATA

- Number of job offers: 132900
- Dates:
22.12.2016-29.05.2017
- Enterprises: 10,3 thous.
- Localizations: 9,5 thous.

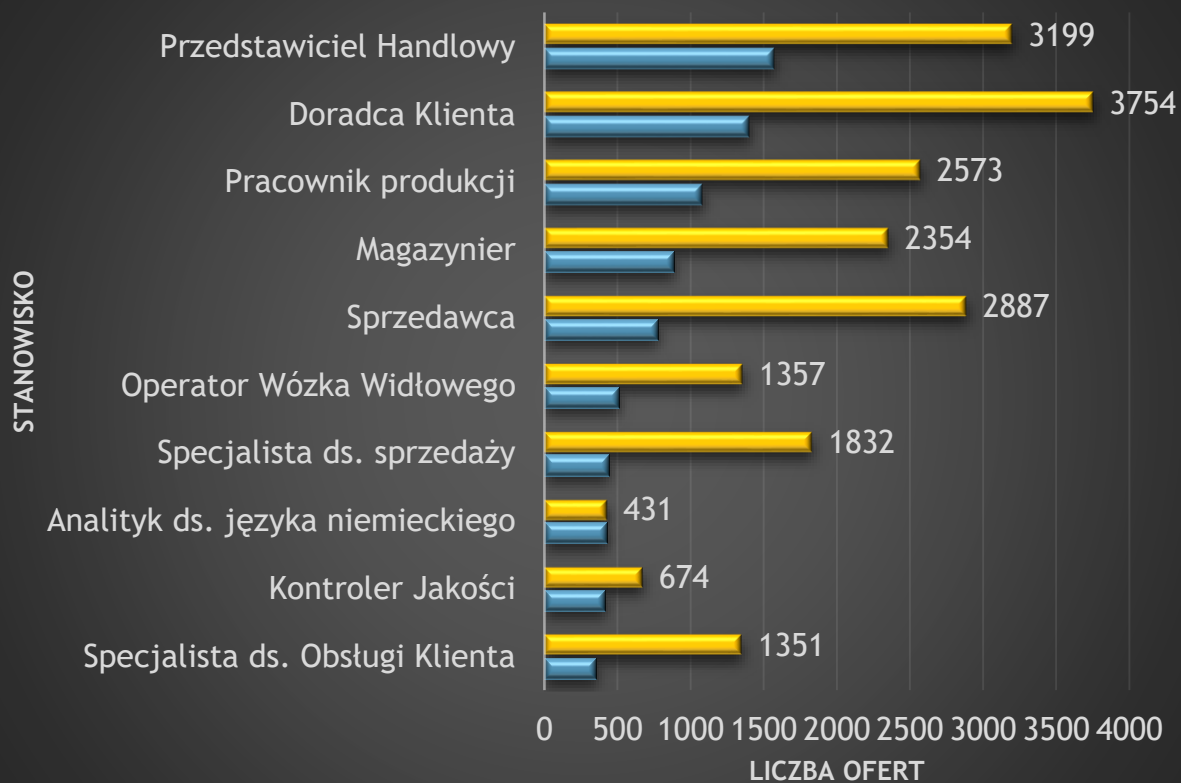
TECHNICAL ISSUES

- Tools:
Apache Spark
Apache Hadoop
- *Python, pyspark*
- **Zasada działania:** Przetwarzanie danych w pamięci operacyjnej i zapisywanie wyników do pliku csv



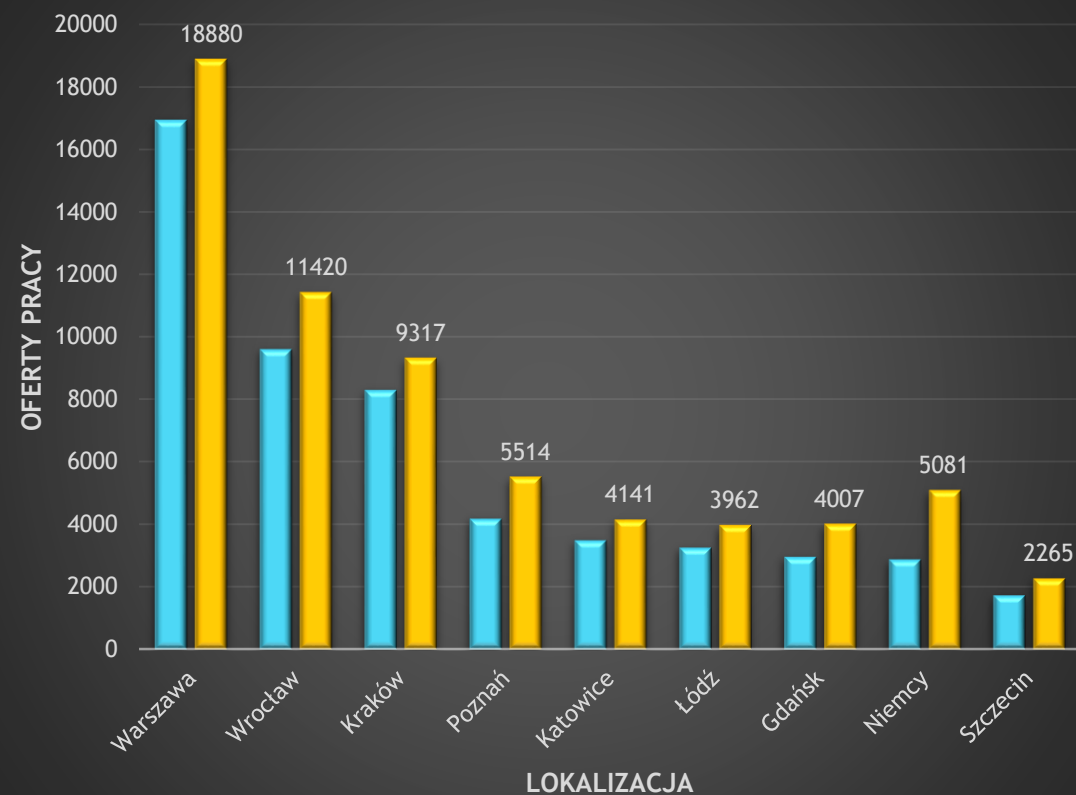
RESULTS - CLASSIFICATION

Oferty pracy według stanowisk



■ Liczba ofert po klasyfikacji
 ■ Liczba ofert przed klasyfikacją

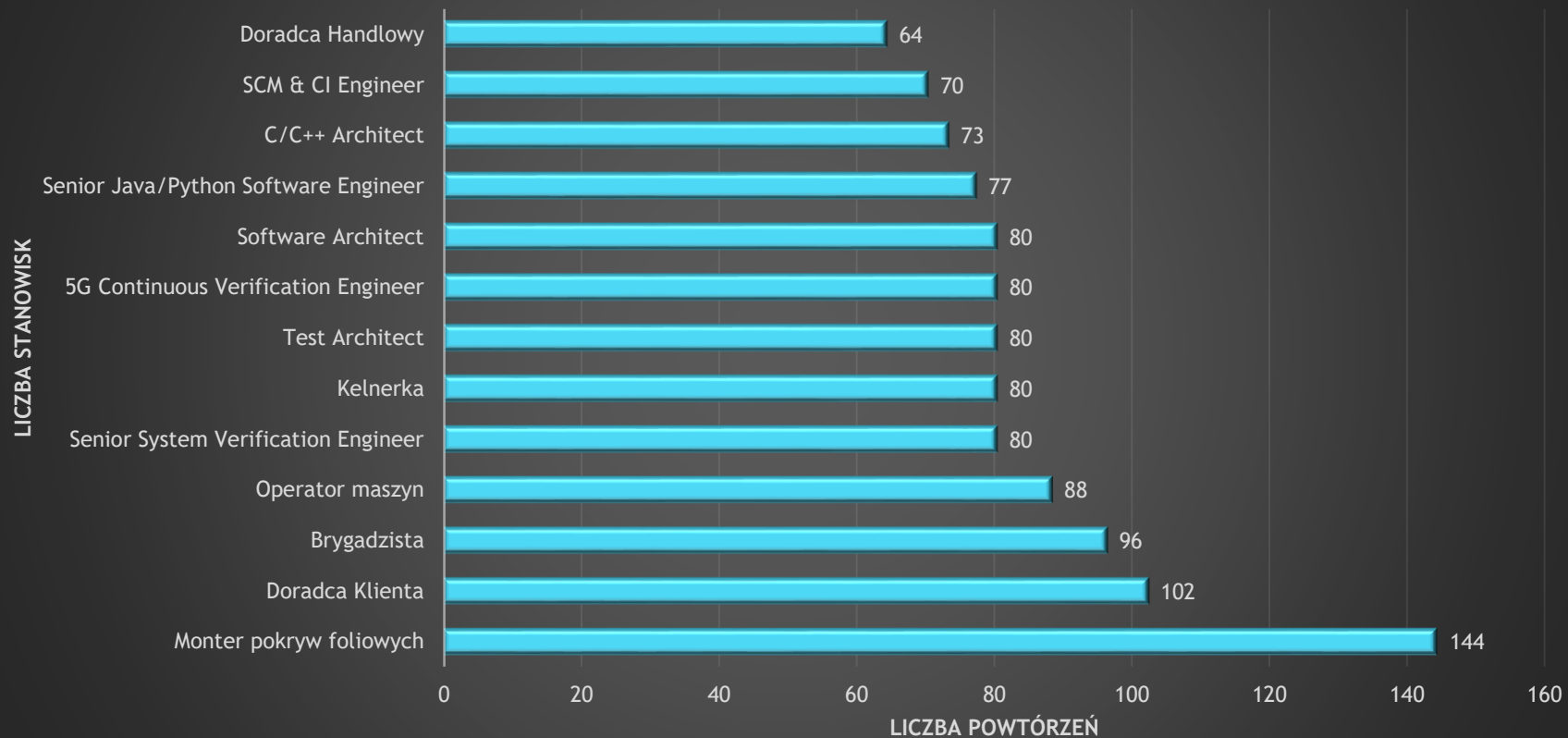
Oferty pracy według lokalizacji



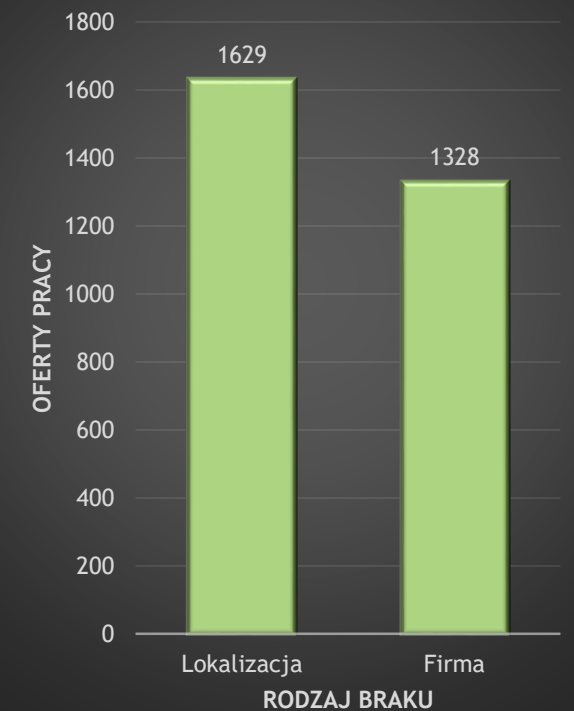
■ Liczba ofert przed klasyfikacją
 ■ Liczba ofert po klasyfikacji

DATA QUALITY - DUPLICATES AND LACKS OF DATA

Oferty pracy powtarzane cyklicznie

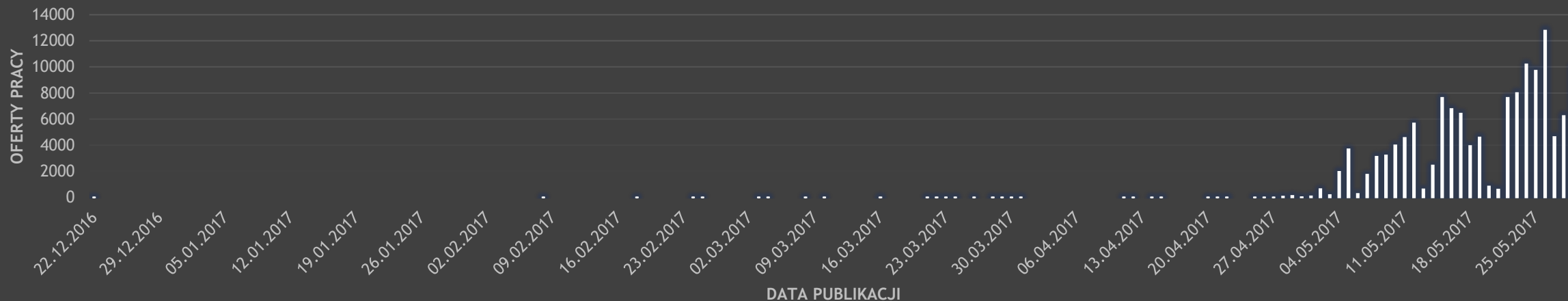


Oferty z brakami lokalizacji i nazwy firmy



TRENDS IN JOB OFFERS

Oferty pracy według daty publikacji w Internecie



Oferty pracy według dnia tygodnia publikacji w Internecie



ISSUES

- Developing Social Media Use case dictionaries for other countries? Need help from other countries.
- Do we have to employ our case on job offers to work for WP2?