

Big Data in Official Statistics: Estimation of job vacancies by using web scraping techniques

Session 27

03.06.2016

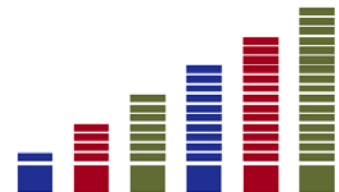
Martina Rengers

Federal Statistical Office, Wiesbaden, Germany

martina.rengers@destatis.de

The ESSnet Big Data at a glance

- in **May 2015** the European Statistical System Committee (ESSC) approves an ESSnet action on Big Data
- a consortium of 22 partners, from Austria, Belgium, Bulgaria, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Spain, United Kingdom, signed on **18 November 2015** a first so-called Framework Partnership Agreement (FPA) with Eurostat
- according to this agreement between the consortium and Eurostat, the project runs **from February 2016 to May 2018**
- the overall objective of the project is to prepare the ESS for integration of big data sources into the production of official statistics



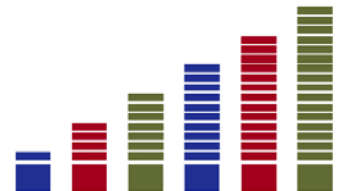
Pilot studies and additional work packages

5 pilot studies as independent work packages (WP)

- web scraping job vacancies
- web scraping enterprise characteristics
- using smart meters for measuring electricity consumption
- using AIS data (AIS – Automatic Identification System to measure real-time ship positions)
- using mobile phone data (geo-localisation)

3 additional work packages dealing with horizontal topics

- using big data sources for early estimates
- using big data sources in multi statistical domains
- general methodological aspects (methodological framework, quality framework, metadata framework, IT infrastructure, skills, partnerships with data providers and scientific community)

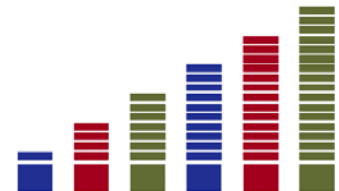


ESSnet BIG DATA: WP-1 job vacancies

The work package focused in a first step the source of job portals. In this case data can be obtained by directly buying from job portal owner(s) or by web scraping.

THE MOST IMPORTANT CHALLENGES

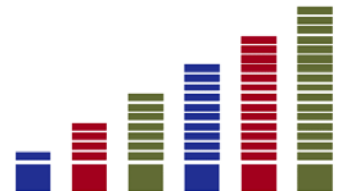
- Legal aspects: Is it legally allowed to do web scraping? Who does the data belong to?
- Internet Security: How can data collection using web scraping be implemented in the IT environment of statistical offices with its high requirements for IT security?
- Assessment of job portals: How many need to be selected and which ones are suitable? How stable is the information that can be obtained from job portals over time?
- Technical implementation: Web scraping or provision of data directly by the portal owners (or third parties)?
- Data quality: Reconciliation with job vacancy statistics from surveys and registers



Assessment of job portals

1. Assessment of job portals **by using web search engines**
 - 1.1 keywords and phrases to find **direct information** about job portals
 - 1.2 keywords and phrases to find **other web sites** which maybe have rankings or assessment analysis of job portals

2. Assessment of job portals **by web scraping** job portals (URLs of job portals are known)
 - 2.1 criteria for assessment of job portals
 - 2.2 criteria to classify and describe job vacancies



Keywords and phrases

First action:

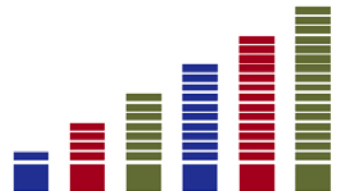
Use Internet search machines such as Google or Bing with the following keywords and phrases

...to find **direct information** about job portals

- *online job portals*
- *ranking of online job portals*
- *assessment of online job portals*

...to find **other web sites** which maybe have rankings or assessment analysis of job portals

- *Competitive Recruiting*
- *Jobcoach*
- *Jobmarketing*
- *HR Reporting / HR recruitment*
- *trade fair for human resources management*

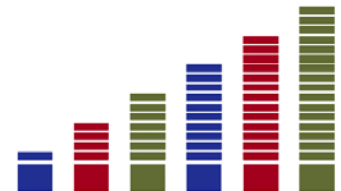


Most import URLs

Result of first action (at least for Germany):

Most important URLs other web sites:

- www.deutschlandsbestejobportale.de
 - ranking list of job portals or job search machines from 2010-2015
 - ICR Studien www.competitiverecruiting.de (Active Sourcing Report 2013, Recruiting report 2014, Recruiting Controlling report 2011, Social Media Recruiting Report 2011, Social Media Recruiting Report 2013)
 - CrossPro Research www.crosspro-research.com
- www.crosswater-job-guide.com
 - www.jobboersen-kompass.de
- www.online-recruiting.net

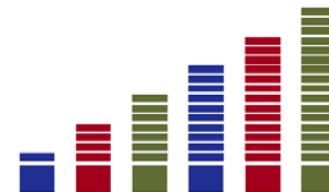


Compilation of ranking lists

No.	Name of job portal	Target group	Number of job vacancies (April 2011)	Alexa popularity ranking (April 2011) *	Number of job vacancies (April 2010)	Employer satisfaction according to Profilo Ranking (March 2011) 7 = very good 1 = absol. not good	User satisfaction according to Crosspro-Research (March 2011) 1 = very good to 4 = absol. not good	Quality of search according to Crosspro-Research (March 2011) 1 = very good to 4 = absol. not good
Nr.	Name der Jobbörse	Zielgruppe	Anzahl der Stellenanzeigen (April 2011)	Reichweite Alexa-Ranking (April 2011)	Anzahl der Stellenanzeigen (April 2010)	Arbeitgeber-Zufriedenheit laut Profilo-Ranking (März 2011) Skala: 7 = sehr gut bis 1 = überhaupt nicht gut	Nutzer-Zufriedenheit laut Crosspro-Research (März 2011) Skala: 1 = sehr gut bis 4 = überhaupt nicht gut	Suchqualität laut Crosspro-Research (März 2011) Skala: 1 = sehr gut bis 4 = überhaupt nicht gut
Allgemeine Jobbörsen								
1	Meinestadt.de	Allgemein	428.813	1.571	265.222		1.87	2.07
2	Arbeitsagentur	Allgemein	373.192	2.255	202.697	4.91	2.13	2.26
3	Jobmonitor	Allgemein	364.001	59.139	124.355			
4	Rekruter.de	Allgemein	279.606	87.295	152.423			
5	Arbeit-Regional	Allgemein	267.749	943.057	288.324			
6	Jobinfo24	Allgemein	263.066	643.503	102.681			
7	Gigajob	Allgemein	195.450	14.072	166.995		1.93	2.14
8	Jobomat.de	Allgemein	94.500	88.972	65.405			
9	Monster Deutschland	Allgemein	68.300	4.915	49.800	4.98	1.96	2.17
10	StepStone	Allgemein	55.282	3.379	36.650	5.62	1.73	1.94

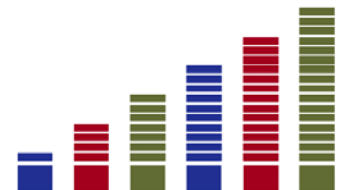
Source: personalmagazin 06 / 11

* This ranking is according to the Internet statistics provider Alexa (www.alexa.com), a subsidiary of Amazon.



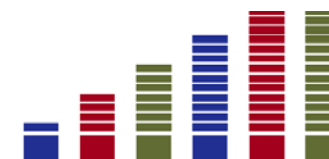
Job portals in Germany (1)

- Number of job portals in 2015 according to ...
 - deutschlandsbestejobportale.de lists more than **1600** job portals
 - jobboersen-im-test.de lists altogether **781** job portals
 - online-recruiting.net lists **99** job portals
- Due to ambiguous rankings the number of job vacancies each job portal indicates was used to make a selection
- altogether **55** general job portals, job search machines and specialized online job portals were selected and examined in more detail



Job portals in Germany (2)

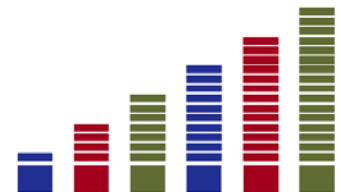
Category	No.	Name	URL	Number of job vacancies in Germany	
				25. Feb 16	13. Mai 16
general job portals	1	Bundesagentur für Arbeit	http://jobboerse.arbeitsagentur.de/vamJB/sta	1,083,929	1,160,595
	2	Gigajob	http://de.gigajob.com/index.html	634,735	682,048
	3	Meine Stadt.de	http://jobs.meinestadt.de/deutschland/stelle	366,524	413,708
	4	Rekruter	http://www.rekruter.de/	277,886	285,664
	5	JobScout24 / Jobs.de	http://www.jobs.de	105,332	113,981
	6	Jobmonitor	http://www.jobmonitor.com/	296,419	107,027
	7	StepStone	https://www.stepstone.de/	59,880	64,665
	8	Jobcluster	https://www.jobcluster.de/	43,741	45,030
	9	JobStairs	https://www.jobstairs.de/	25,900	27,051
	10	Stelleneinzeigen.de	http://www.stellenanzeigen.de/	/	9,029
	11	Süddeutsche Zeitung	http://stellenmarkt.sueddeutsche.de/	8,489	8,512
	12	Kalaydo	http://www.kalaydo.de/jobboerse/	/	6,161
	13	Monster	http://www.monster.de/	/	/
	14	LinkedIn	https://de.linkedin.com/job/	/	/
job search machines	15	CareerNetwork JOBBÖRSE.de	www.jobboerse.de	/	1,500,000



Criteria I

Criteria for assessment of job portals (URLs of job portals are known):

- **kind of job portal**
 - online jobboards (allgemeine Jobportale)
 - specialized online jobboard (spezialisierte Jobportale)
 - Job search engines (Metasuchmaschine)
- **number of job vacancies**



Criteria II

Ten criteria to classify and describe job vacancies:

WHAT ?

- 1 position
- 2 occupation
- 3 education (required education of the candidate)
- 4 type of job (permanent- temporary employment, full time-part time job)

WHEN ?

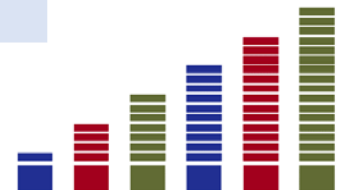
- 5 date of the advertised job vacancy
- 6 data of application deadline
- 7 date to fill a vacancy

WHERE ?

- 8 location of the job

WHO ?

- 9 Direct employer or agency
- 10 Economic activity of the employer (NACE groups)



Many thanks for
your attention!

Dr. Martina Rengers
Federal Statistical Office Germany, Wiesbaden
martina.rengers@destatis.de

