
ONLINE DATA: The Microscope for Labor Market Studies

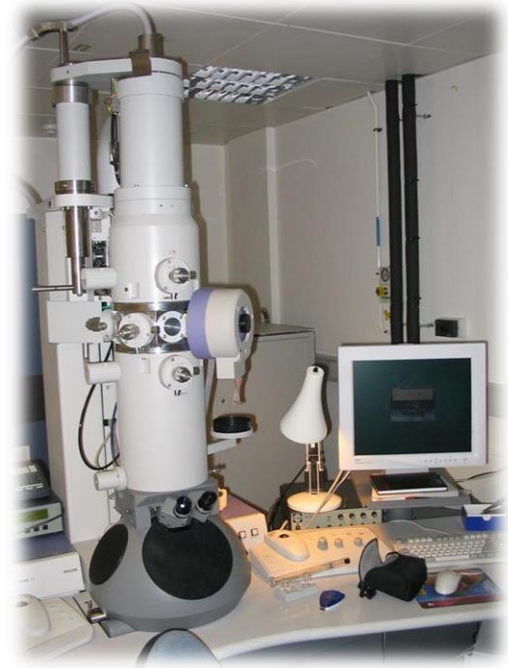
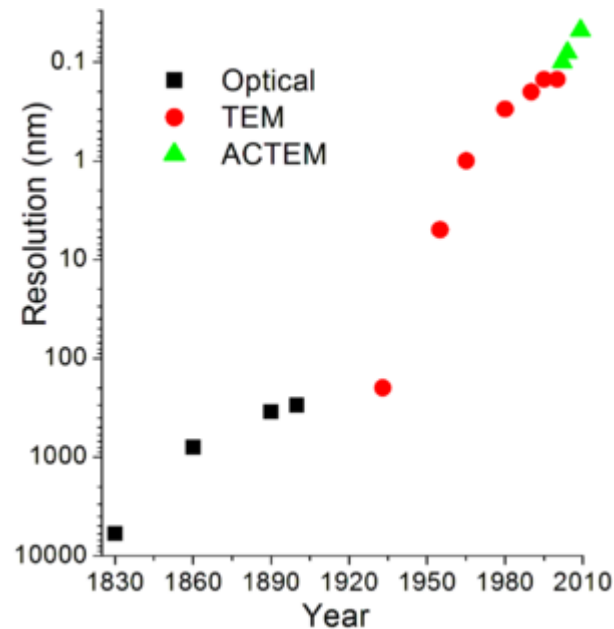
Maxim Bakaev, PhD
(bakaev@corp.nstu.ru)

Novosibirsk State Technical University, Russia

*Human Resource Management conference
University of Economics – Varna, 2021*

Why a microscope?

- Horton J. J., Tambe P. (2015) Labor economists get their microscope: big data and labor market analysis. Big data, 3 (3). P. 130-137.
 - The microscope (XVII век) allowed biology researchers to reach a new level of knowledge, by increasing the level of study detail
 - The labor market data are not Big Data



Traditional statistical sources

- Labor market stats (Russian Committee of Statistics, etc.)
 - ❑ Unavoidable lag, since the data need to come from the origin to the processing body (a state institution), get processed and get published
 - ❑ A certain conservatism in methods: hard to grasp rapid changes in industries, occupation types, positions/professions, etc.)
 - ❑ The typical detail and update frequency: month(s)
 - ❑ The level of detail (aggregation) is not always as desired
 - ❑ The quality of labor market statistical data is not necessarily high: it has low priority for the sources (companies), there is “gray” employment, etc.
- Amount of data transferred daily (online channels): 10^{18}
 - ❑ this is indeed Big Data
- Annual increase in created data: over 50%
 - ❑ of them, 45-90% are never used

Online Data (OD)

- Come from WWW (annual growth of 50%):
 - ❑ Big volumes of electronic data
 - ❑ Is a processable representation of the real world
 - ❑ Is increasingly seen as a source of data by statistical institutions, companies, information services, etc.
 - ❑ Most of the data is unstructured, unless API is provided
 - ❑ Employment of AI methods for collection, structuring and processing of data might be needed
- Not all potential objects of analysis have the “online footprint”
 - ❑ only the ones involved in some kind of “online transactions”
- **The representation is not the reality**

OD in socio-economic research

- The cost/effort is almost independent of the sample size
 - ❑ The initial (fixed) costs might be high
 - ❑ The total costs are not necessarily low
- The increase in data detail and the collection frequency
 - ❑ More attributes of the studied objects, larger samples
 - ❑ Higher personnel skills are needed for the collection and processing
 - ❑ The OD-specific research problems must be specified
- The validity of the results might increase
 - ❑ The responders' subjective factors are removed
 - ❑ The construction of random selection samples is easier
 - ❑ But the interpretation to reality is an important issue

OD employment feasibility checklist

1. A research problem or a customer
 - ❑ What decisions could be made using the analysis outcome?
2. Relevant advantages of OD employment
 - ❑ Frequency/detail? Sample size? Coverage? Costs?
3. The analyzed object's "online footprint" exists
 - ❑ What characteristics need to be collected/monitored?
4. "Concentration" of the OD sources
 - ❑ Collecting from fewer number of sources is easier (+ data sync)
5. Technical aspects of the OD sources
 - ❑ The more reliable and the more structured, the better
6. The legal and ethical considerations
 - ❑ Who is the source of the data and what are their reasons?

Labor market data sources

- LinkedIn – the largest professionals-oriented website in the world
 - ❑ Growth from 300 million users in 2015 to 740 million users in 2021
 - ❑ About 55 million companies registered
 - ❑ Access is blocked in Russia (does not comply with the personal data location requirements)
 - ❑ Has few quantitative data, but allows e.g. monitoring of previous work places, geographic mobility, etc.
- HeadHunter.ru has become the largest platform in Russia
 - ❑ 60 million users, 350 thousand companies
 - ❑ About 700 thousand vacancy ads (daily average)
 - ❑ Covers several Russian-speaking countries (3rd place in the world)
- Other platforms in Russia: zarplata.ru (purchased by HH in 2020), trudvsem.ru, regional websites

HeadHunter – stats.hh.ru

- From 2006 – surveys and occasional “hand-made” reports
 - The Russian employers rating
 - The Russian universities rating (Moscow)
 - “The databank of wages”
 - Some of the reports became non-free
- Now: stats.hh.ru – free aggregated data from the database
 - By regions
 - By industries
 - By periods (limited)
 - “hh.index” – the ratio between active vacancy and resume ads = tension in the labor market
- Absolute values (numbers, wages) are hard to find
- No stats over a year ago

Screenshot: stats.hh.ru

Статистика по России / Сибирский федеральный округ / Новосибирская область

Novosibirsk

Vacancy ads growth

+2% ^

за месяц

+61% ^

за год

+74% ^

январь—август 2021
к январю—августу 2020

Resume ads growth

+3% ^

за месяц

-1% v

за год

+4% ^

январь—август 2021
к январю—августу 2020

Average wages

47 819 ₪ -3% *

средняя предлагаемая

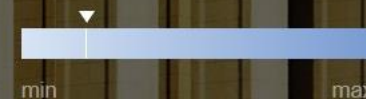
43 984 ₪ -13% *

средняя ожидаемая

* — выше/ниже,
чем в среднем по России

Labor market tension

3 people for 1 vacancy
Moderate competition, in favor of
employees



Labor market monitoring software

- We developed it in 2011, initially for the Novosibirsk City Hall department
 - Has 3 independent modules, launched by schedule or on request:
 - 1. Data collection module:**
 - before 2017: web scraping was used (information collected from web pages using XPath selectors and etc.)
 - from 2017: data collection via API (the sources started to provide it)
 - 2. Data processing and structuring module:**
 - Auto-classification of vacancy and resume ads by industries
 - Duplicates removal = sync between sources and periods
 - 3. Analysis and reporting module**
-

LMM software: details

- Collecting vacancy and resume ads for the Novosibirsk region and some other Siberian regions of Russia
 - From 2020, we started collection for St. Petersburg
 - Contains about 10 million unique records in the database
- The main parameters for the monitoring
 - The numbers of vacancy and resume ads, the ratio between them
 - The wages in vacancy ads (proposed)
 - The wages in resume ads (requested)
- Some challenges and highlights
 - The data is incomplete (disconnects, changes in data sources, etc.)
 - Intelligent classification by industries (+ Other, to be processed)
 - Methodological: e.g. wages “from X Rub” or “up to Y Rub”
 - Identification and labeling of duplicates (not deleted)

Reports: industries by time periods

Новосибирская область, вакансии 2017 - 2019 год

	01.2017 - 06.2017		07.2017 - 12.2017		01.2018 - 06.2018		07.2018 - 12.2018		01.2019 - 06.2019		07.2019 - 12.2019	
	кол-во	ср зп	кол-во	ср зп	кол-во	ср зп	кол-во	ср зп	кол-во	ср зп	кол-во	ср зп
Услуги населению, сервисное обслуживание	7	20079,14	4	40000,00	8	23312,50	8	18750,00	6	15516,67	27	6800,00
Работы, не требующие квалификации	12	16634,67	13	19186,92	15	12490,00	31	18494,74	18	20889,50	36	21911,97
Образование, наука	36	18998,50	1785	15085,60	3536	16629,28	2803	17671,98	2855	18127,50	3290	18899,88
Кадровая служба, управление персоналом	4	20587,50	6	25708,33	12	23980,83	2	33250,00	6	37500,00	10	42806,00
Административная работа, секретариат, АХО	17	23411,76	44	22517,27	124	23248,60	74	23313,16	89	20039,55	140	27213,14
Производство	36	22986,94	6483	17872,65	15703	18197,98	11567	19725,60	13809	19348,93	11993	20425,29
Пищевая промышленность	9	20339,33	1280	15823,02	2549	15871,85	1432	16830,31	1873	17808,25	1602	18764,34
Маркетинг, реклама, PR	1	22000,00	153	16946,55	330	16929,97	201	17981,50	177	19835,07	203	18871,25
Консалтинг, стратегическое развитие, управление	1	20000,00	1427	23187,48	3129	23730,53	2117	25078,80	2371	25144,39	1993	26365,33
Металлургия, металлообработка	7	30642,86	305	24803,20	556	24759,08	346	25952,37	393	26826,10	387	28146,83
Юриспруденция	1	83000,00	463	17936,93	690	19010,25	546	21131,80	743	22473,29	564	21618,93
Продажи, закупки, снабжение, торговля	7	27500,00	1528	16766,28	2564	16900,35	1729	18082,66	2135	18341,53	1986	19057,97
Рабочие специальности	5	14750,00	11	35863,64	20	19140,80	31	37458,19	34	24652,24	23	29833,26
Здравоохранение, спорт, красота, социальное обеспечение	79	31132,18	1701	21204,13	3074	23237,14	1839	27851,43	2537	27961,88	2221	28200,63
Информационные технологии, телекоммуникации, связь	5	46148,00	368	33940,40	656	31055,55	384	29771,23	433	30880,34	408	34523,16
Искусство, культура и развлечения	5	13285,40	239	13127,19	496	15334,94	347	16311,03	384	17961,40	351	18618,85
Машиностроение	1	0,00	17	23462,47	35	23998,03	25	30789,52	29	31518,97	17	25452,94
ЖКХ, эксплуатация	1	11000,00	211	18108,18	533	18230,26	572	18574,31	667	21114,44	583	20460,44
Легкая промышленность	5	20400,00	63	15884,92	100	16742,22	81	19764,30	65	18599,23	94	17532,98
Лесная, деревообрабатывающая, целлюлозно-бумажная промышленность	3	15000,00	207	15502,88	357	16667,83	269	18553,80	283	18466,06	216	18314,01
Безопасность, службы охраны	18	17136,78	18	13602,28	62	16210,03	38	22601,34	30	16332,67	30	26013,33
Государственная служба, некоммерческие организации	4	21562,50	86	21074,45	200	23521,75	143	23513,18	167	24978,99	100	25137,50
Транспорт, автобизнес, логистика, склад, ВЭД	3	24900,00	781	20803,43	1709	21032,45	1319	21832,99	1505	22326,98	1224	24785,43
Сельское хозяйство, экология, ветеринария	1	35000,00	383	13831,08	1019	15214,78	497	17353,75	762	18004,64	563	18707,47
Строительство, ремонт, стройматериалы, недвижимость	10	25750,00	564	23583,95	1112	22834,88	822	26272,90	1008	28523,98	985	34075,82
Финансы, кредит, страхование, пенсионное обеспечение	1	27000,00	473	21702,87	1064	20874,37	806	21000,42	903	21668,05	859	22224,23

Reports: companies

<< вернуться к выбору региона и года

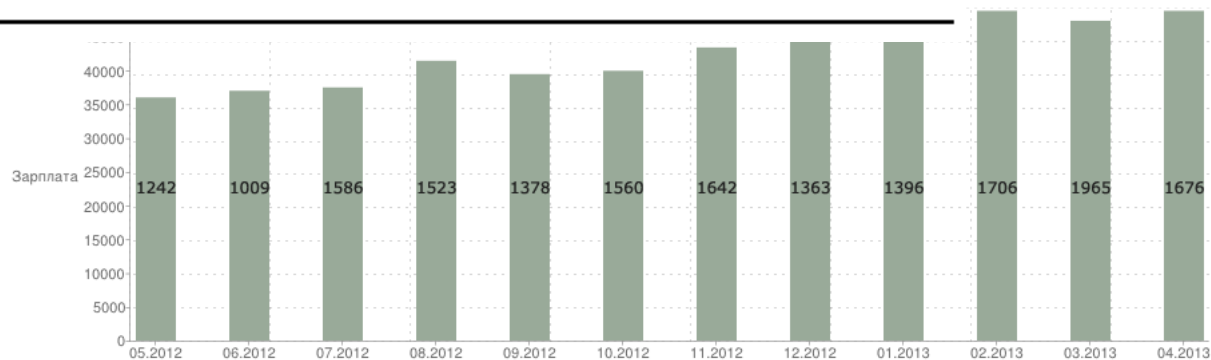
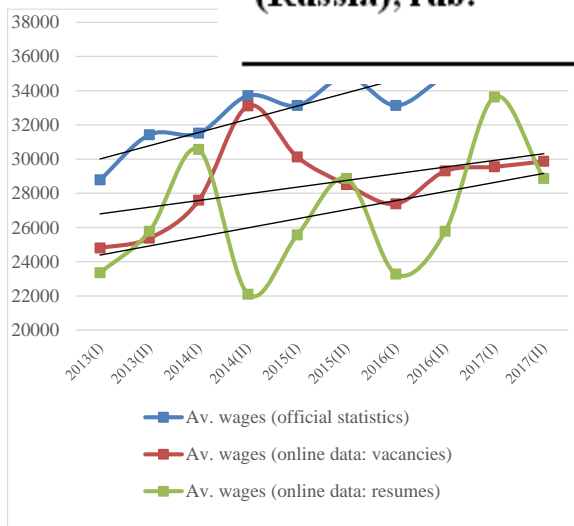
Новосибирская область - 2021 г.

Почта России (176)							Показать
<ul style="list-style-type: none"> Пятерочка (778) АО "РЭС" (382) ООО "Крюгер Хаус" (296) Пункт отбора на военную службу по контракту по Новосибирской области (295) ИП Макеев М.О. (293) People Group, центр подбора и развития персонала (261) Тетрика (238) Сибирская продовольственная компания (224) Магнит, Розничная сеть (220) Ресторанная группа "SunnyDayGroup" (213) ООО Интер-Персонал (211) JSat (210) Билайн (208) Работут (198) Труженик (196) Почта России (176) Сибирская генерирующая Компания (171) ГБУЗ НСО "Доволенская ЦРБ" (168) Главновосибирскстрой (165) ООО "Лента" (153) 							<p>...лен опыт в закупках, бюджетировании. Опытный</p> <p>...разделений на бланочную продукцию, ...ы, почтовые сумки. Контроль использования и оценка В...</p> <p>...работы не менее 2 лет в области продвижения и продаж ... в закупках в...</p> <p>Работа с клиентами (юр.лица B2B, B2G), продажа услуг почтовой связи. Работа с текущими клиентами, поиск потенциальных...</p> <p>https://hh.ru/vacancy/43874645</p> <p>requirement Среднее профессиональное образование. Опыт управления складской деятельностью компании с разветвлённой сетью складов или логистикой в предприятиях интернет торговли - от 3...</p> <p>responsibility Организация эффективной работы смены цеха. Анализ результатов производственной деятельности работников за смену, причин, вызывающих сбои в работе, организация мероприятий по...</p> <p>https://hh.ru/vacancy/43120603</p> <p>requirement Высшее образование (логистика, связь, экономика, математика, статистика). Знание «Excel» – продвинутый уровень: формулы, функции, сводные таблицы. Аналитические способности: склонность к скрупулезной...</p>
1638912	HeadHunter	Начальник смены Фулфилмента	Почта России	2021-04-12 08:10:54	2021-04-14 22:50:22	45000-51750 RUR	

Reports: wages comparisons

The dynamics of the wages, online and official (2017-2020)

	2017	2018	2019	2020	2017-2020
Proposed wages (in vacancy ads), rub.	23,583	25,287 +7.23%	27,450 +8.55%	29,637 +7.96%	+6504 +25.67%
Requested wages (in resume ads), rub.	21,678	22,052 +1.72%	22,068 +0.08%	19,745 -10.53%	-1933 -8.92%
Official wages (Novosibirsk), rub.	32,287	35,686 +10.53%	39,076 +9.50%	41,534 +6.29%	+ 9247 +28.64%
Official wages (Russia), rub.	39,167	43,724 +11.63%	47,867 +9.48%	51,352 +7.28%	+12,185 +31.11%



Highlights in 2020 (Novosibirsk)

- In 2020 (COVID-19 pandemic and lockdowns) we saw rapid drop in **requested** wages (-10,5%), but no notable changes in the **number of vacancies**
 - The online supply in the labor market has **short-term flexibility**, where as the online demand does not
- Correlation analysis by 20 industries:
 - Proposed and requested wages ($r = 0.525$; $p = 0.017$)
 - Numbers of vacancy and resume ads ($r = 0.725$; $p < 0.001$)
 - Tension and wages **are not correlated** ($r = -0.098$)
- Online employment suffers from the “double discount”, probably due to the particular covered positions:
 - Proposed wages – only 71% of the official wages
 - Requested wages – only 78% of the proposed wages (in 2020)

Conclusion

- Online data analysis is a potentially powerful tool for economists and sociologists
 - The *HR Mining* is rapidly developing in the world
 - Google Scholar: “headhunter.ru” – 430, “LinkedIn L.M.” – 168,000
- A microscope sometimes has to be used to hammer nails
 - No customer => we have to invent research problems
 - Local authorities still do not understand or use online sources
- Not everything can be automated, manual work is involved
 - collection: changes in data sources, emergence of new ones
 - structuring: instability in data models (e.g., industries change)
 - analysis: assumptions, data filtering, textual data, etc.
- Online data show us online, not real world
 - How does online reflect the real economy and society?

References

- Bakaev M., Sizova I. *Online Data: The Microscope for Labor Market Studies*. HRM Conference, 2021. – in print.
- Sizova I., Bakaev M. *Labor Demand and Supply Adaptation to the Pandemic-Induced Shock Analysis of Online Recruitment Data in Novosibirsk region of Russia*. Digital Transformation and Global Society (DTGS), 2021. – in print.
- Aletdinova A., Bakaev M. *Intelligent data analysis and predictive models for regional labor markets*. Digital Transformation and Global Society (DTGS), 2019.
- Bakaev M., Avdeenko T. *Prospects and challenges in online data mining: experiences of three-year labour market monitoring project*. Lecture Notes in Computer Science. - 2016. - Vol. 9714 : Data Mining and Big Data. - P. 15-23.