ПЕРМСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

PERM NATIONAL RESEARCH POLYTECHNIC UNIVERSITY

EVALUATION OF NORMATIVE AND LEGAL SUPPORT OF REGIONAL STRATEGIZING ON THE BASIS OF TEXT MINING

Julia Dubrovskaya, Elena Kozonogova, Maria Rusinova

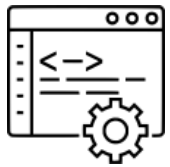
Russia, Perm National Research Polytechnic University,

Economics and Finances Department

The reported study was funded by Russian Foundation for Basic Research according to the research project № 19-010-00449

INTRODUCTION

Without strategizing it is impossible to imagine



solution to operational management tasks

realization of long-term projects

implementation of subsequent control

LITERATURE REVIEW

Strategizing becomes paramount for a wide range of economic entities:

Enterprises

- Anikanova, M.A., Morgunov, A.F.: Criterial evaluation of the possibility of small businesses process automation on public cloud platform. Business informatics no. 3(33), 55–64 (2015).
- Kek-Mandzhieva, Z.V.: Automation of accounting and analytical management system as a necessary component of effective enterprise management. Science and modernity no. 35, 183-191 (2015).
- Nitsenko, V.S., Vysochinskaya, L.N., Marojko, M.N.: Automated forms of accounting as an instrument of increasing intensity information processing. Journal of Economy and Business vol. 4, 120-122 (2016).
- Rabinovich, L.A., Libman, M. L.: The main provisions of the feasibility study of new means of automation of production processes. News of Volgograd State Technical University no. 1, 7-9 (2004).

LITERATURE REVIEW

Government

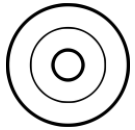
- Lastochkina, M.A. Development of methodology and tools for assessing the degree of modernization in Russia's regions. Problems of development of the territory vol. 4(78), 69-79 (2015).
- Nemets, K.A. Automation of regional management processes and the formation of “Electronic Government”. Scientific and technical statements of the St. Petersburg State Polytechnic University. Economic sciences no. 4(127), 85-89 (2011).
- Nemtsev, A.N., Shtifanov, A.I., Belenko, V.A., Zagorodnyuk, R.A. Nemtsev, S.N., Galtsev, O.V. Designing the automated information system for monitoring of educational departaments and provision the opportunity to use “electronic” services in education sector. Scientific reports of Belgorod State University. Series: Economics. Computer science no. 13-1 (108), 65–76 (2011).
- Robozov, S.A., Erunova, M. G., Maltsev, K.V. Quality assessment of regional and municipal management in an automated information system for monitoring municipalities. Journal of science and technology no. 1, 176-181 (2010).

LITERATURE REVIEW

Financial institutions

- Vasilieva, E.E. Modeling language credit risk assessment of banking activity in the Russian regions, based on fuzzy sets methods. Internet journal "Science of Science" vol. 8, no. 6, 1-18 (2016).
- Ryzhkov, O. Yu., Bobrov, L. K. Complex automation of actuarial work of the insurance company. Vestnik of Astrakhan State Technical University. Series: Management, Computer Sciences and Informatics no.2, 98-108 (2014).
- URL,http://www.standardandpoors.com/prot/ratings/articles/ru/ru?articleType=HTML&assetID=1245322330515, last accessed 2019/08/03.
- Solntsev, O.G., Mamonov, M.E., Pestova, A.A., Magomedova, Z.M. Experience in Developing Early Warning System for Financial Crises and the Forecast of Russia Banking Sector Dynamic in 2012. Journal of the New Economic Association no. 4(12), 41-76 (2011).

THE PURPOSE OF THE STUDY



The research is devoted to the problem of regional strategizing based on benchmarking.

Like benchmarking of companies, benchmarking of the territory is carried out on the basis of continuous monitoring of best practices, the search for new ideas for forming development priorities based on identifying the unique opportunities of the territories.

The tasks of regional benchmarking include determining the leading territory, **analyzing** the main factors of its success and **adapting** the identified benefits to the analyzed region.

THE CONCEPT OF “SMART” BENCHMARKING

The comparison of territories should be carried out on regions with similar institutional conditions.

So a feature of “smart” benchmarking is a preliminary selection of structurally similar territories and further comparison of the analyzed region only with identical regions.

That is why the main task for regional manager who wants to compare regions on a base of “smart” benchmarking tools is an identifying appropriate criteria of regional development.

THE HYPOTHESIS OF THE STUDY

It is obvious that the priorities of the region development and the directions of financing industries in identical regions should be correlated.

The hypothesis of the present work: erroneous self-positioning of the region and incorrect comparison with other territories impedes the correct identification and choice of development priorities.

MAIN STAGES OF THE RESEARCH

Stage 1

Data collection
(data collection is based on the parsing of relevant sites)

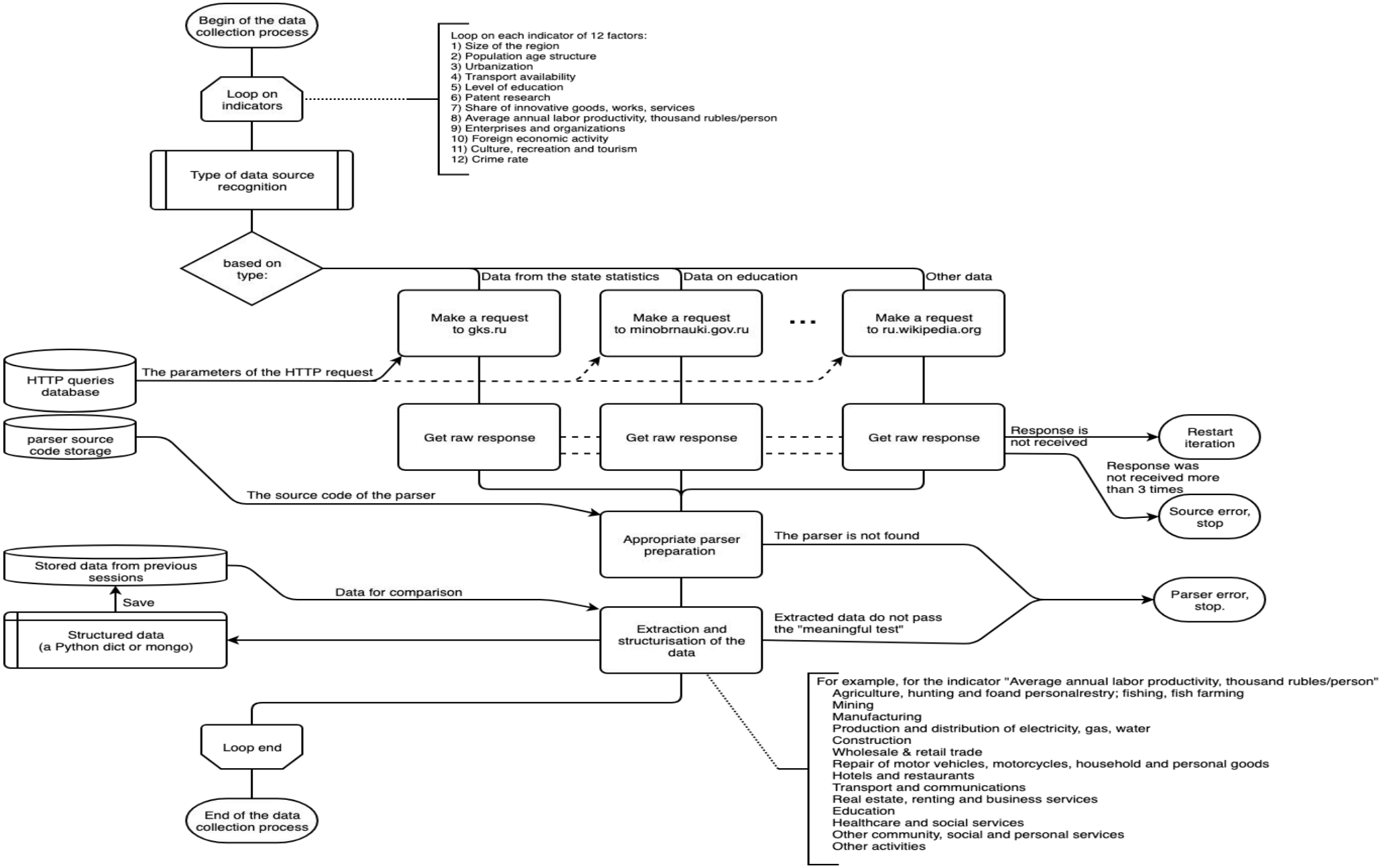
Stage 2

Finding identical regions
(determination of identical regions – by calculating the structural distance indexes)

Stage 3

Evaluation of regional development priorities based on text mining

STAGE 1. GENERAL ALGORITHM FOR PARSING NECESSARY DATA



System of criteria determining regions' similarity*

Criterion	Factor	Statistical indicator		Unit of measurement
Geodemography	Size of the region	x1 x2	- territory size; - population	thousand km^2 thousand people
	Population age structure	x3 x4	- people under working age; - people over working age	% of the total population
	Urbanization	x5 x6	- share of urban population in the total population; - share of rural population in the total population	%
	Transport availability	x7 x8 x9	- number of public buses per 100,000 people; - bus passenger transportation; - density of general-purpose hard-surface highways	units mln people km of roadway per 1000 km^2 of territory
Education	Level of education	x10	- people with above-average education aged 15 and over per 1,000 people who specified their level of education	people
Innovation	Patent research	x11 x12	- provisional patents granted; - utility patents granted	units
	Share of innovative goods, works, services	x13	- share of innovative goods, works, services in the total volume of goods shipped, works performed, service provided	%
Sectoral structure	Distribution of average annual number of employees by economic activity	x14 – x26	- agriculture, hunting and forestry, fisheries and fish farming; mining; manufacturing; production and distribution of electricity, gas, water; constructions; wholesale and retail trade; repair of motor vehicles, motorcycles, household and personal goods; hotels and restaurants; transport and communications; real estate, renting and business services; education; healthcare and social services; other community, social and personal services; other activities	% of the total number of the employed
Investment climate	Enterprises and organizations	x27	- number of enterprises and organizations per 1,000 people	units
Transparency	Foreign economic activity	x28 x29	- exports to far-abroad countries; - exports to CIS countries in current prices	mln U.S. dollars
Social values	Culture and tourism	x30 x31 x32 x33	- number of theatregoers per 1,000 people; - number of visits to museums per 1,000 people; - number of Russian tourists sent by travel agencies on tours in Russia, per 1,000 people; - number of Russian tourists sent by travel agencies to foreign tours, per 1,000 people	people
	Crime rate	x34	- number of reported crimes per 100,000 people	units

STAGE 2. DETERMINATION OF IDENTICAL REGIONS BY CALCULATING THE STRUCTURAL DISTANCE INDEXES

The structural distance index calculated according to formula:

$$(i, i') = \sum_{j=1}^{k} m_j \left(\overline{x_{ij}} - \overline{x_{i'j}}\right)^2 \qquad (1)$$

where $d(i, i')$ – structural distance index of the i-th region;
$\overline{x_{ij}}$ – value of j-index of i-initial region;
$\overline{x_{i'j}}$ – value of j-index of i-“another” region;
m_j – weighting factor.

Weighting factor is calculated according to formula:

$$m_j = \frac{1}{a} / j \qquad (2)$$

were a – number of criteria for comparing regions;
j – number of statistical indicators characterizing the criterion.

The smaller the value of structural distance index, the most similar regions.
The regions with the lowest values of structural distances indices are **identical regions.**

TESTING OF THE PROCEDURE OF "SMART" BENCHMARKING

- We used the statistical data of the Russian regions for 2016 published by the Federal State Statistics Service.
- A complete distance matrix was constructed for the regions of Russia.
- The Republic of Tatarstan was chosen as a region for analysis.
- All regions identical to the Republic of Tatarstan have a number of similar characteristics.
- Firstly, it is a high level of urbanization – the share of urban population in the total population is about 74%.
- Secondly – a highly pronounced specialization in manufacturing, as well as in wholesale and retail trade.

DISTANCE MATRIX OF REGIONS IDENTICAL TO THE REPUBLIC OF TATARSYAN

Region	Republic of Tatarstan	Perm Krai	Krasnoyarsk Krai	Volgograd Oblast	Novosibirsk Oblast	Saratov Oblast	Arkhangelsk Oblast	Primorsky Krai	Kaliningrad Oblast
Republic of Tatarstan		0,112	0,235	0,268	0,293	0,352	0,410	0,415	0,421
Perm Krai	0,112		0,168	0,200	0,369	0,184	0,279	0,220	0,291
Krasnoyarsk Krai	0,235	0,168		0,420	0,285	0,541	0,316	0,184	0,292
Volgograd Oblast	0,268	0,200	0,420		0,256	0,375	0,280	0,382	0,104
Novosibirsk Oblast	0,293	0,369	0,285	0,256		0,511	0,239	0,311	0,235
Saratov Oblast	0,352	0,184	0,541	0,375	0,511		0,280	0,375	0,382
Arkhangelsk Oblast	0,410	0,279	0,316	0,280	0,239	0,280		0,165	0,111
Primorsky Krai	0,415	0,220	0,184	0,382	0,311	0,375	0,165		0,099
Kaliningrad Oblast	0,421	0,291	0,292	0,104	0,235	0,382	0,111	0,099	

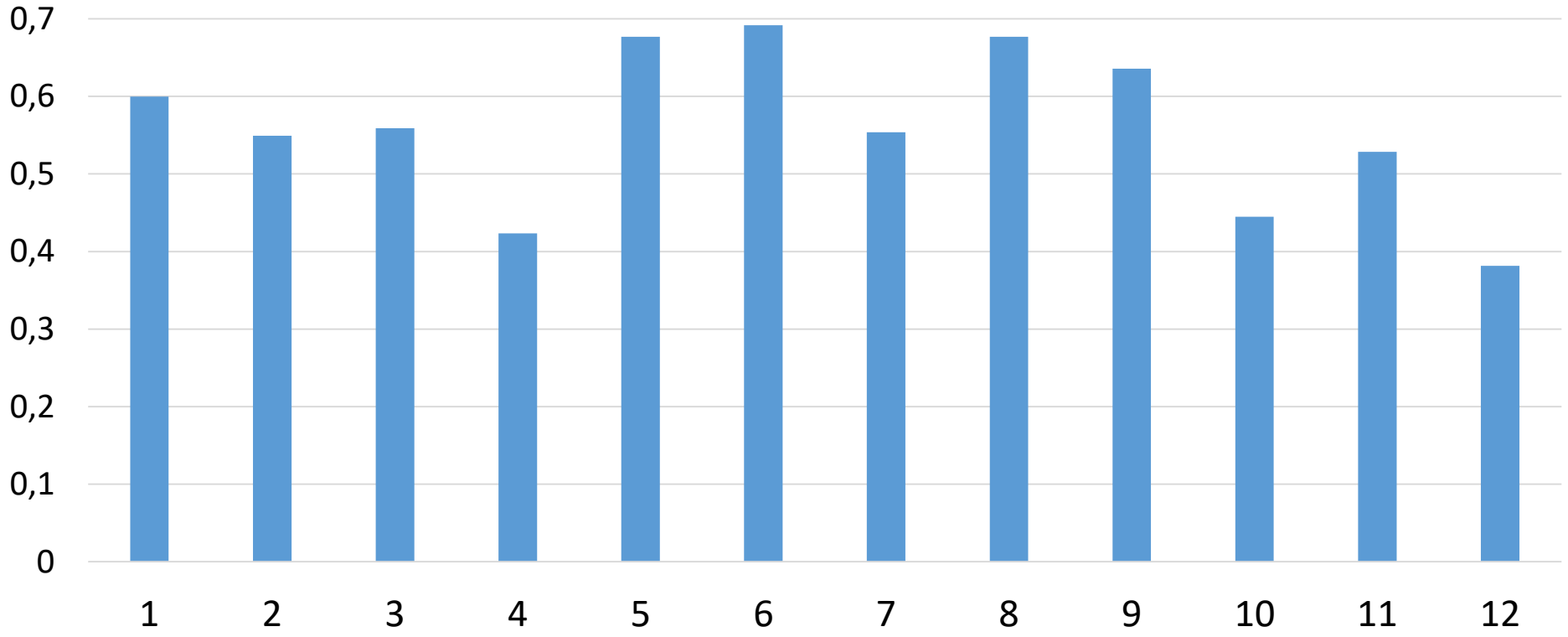
SOFTWARE TOOL “SMART” BENCHMARKING

http://ruclusters.ru/clusters

STAGE 3. EVALUATION OF REGIONAL DEVELOPMENT PRIORITIES BASED ON TEXT MINING

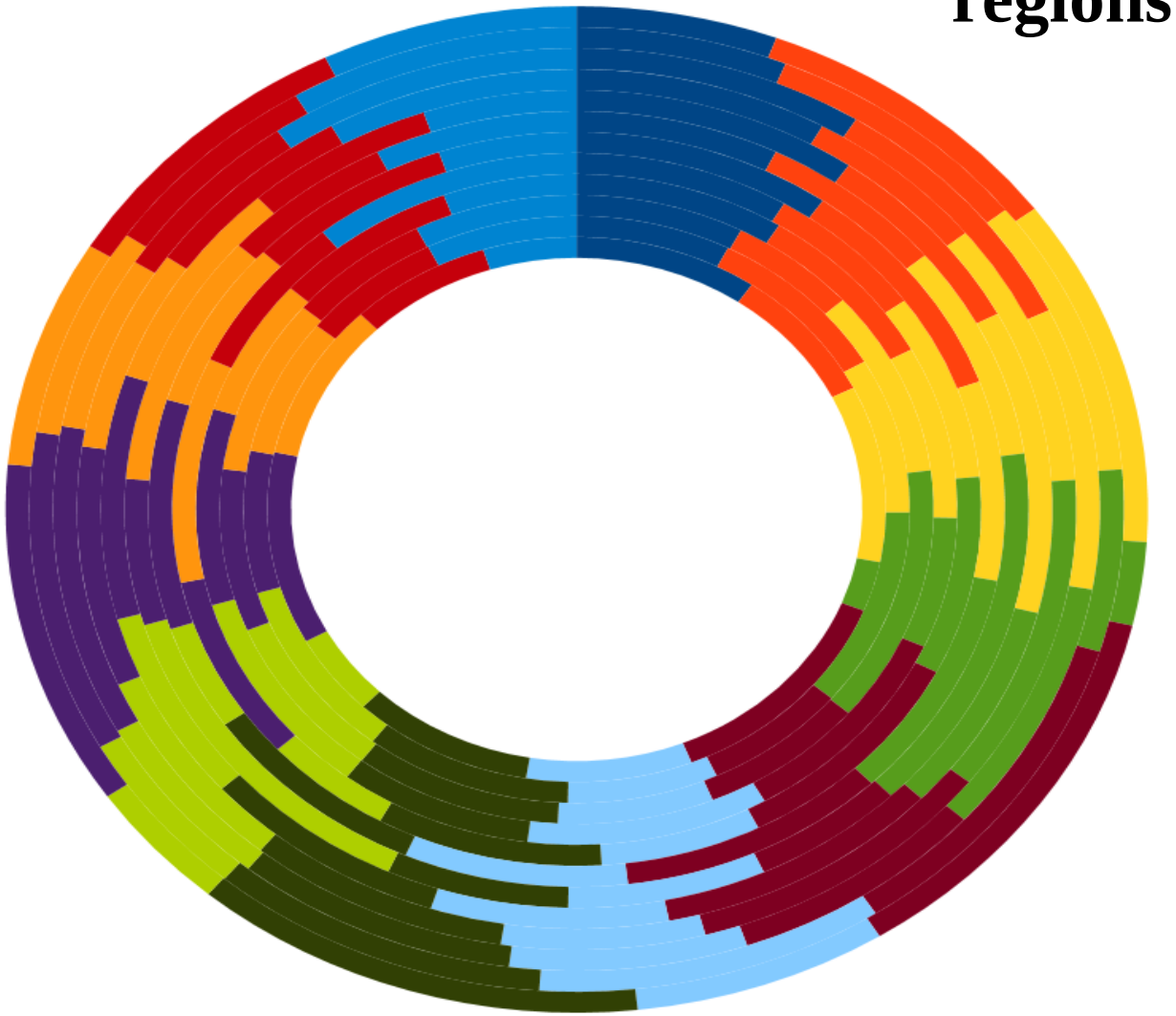
- The Text Mining technique was used to test the hypothesis.
- We evaluated 10 strategies of socio-economic development of the regions, which are identical for the analyzed territory – the Republic of Tatarstan.
- According to the hypothesis of the study, the priorities of the region development and the directions of financing industries in identical regions should be correlated.
- Assessment of selected strategies for the socio-economic development of regions based on text-mining was carried out in 2 consecutive stages using a high-level programming language Python.
- At the first stage, significant concepts were extracted from the strategies.
- At the second stage, a quantitative analysis of the text of the strategies was carried out.

The severity of the main types of economic activity in the list of entities of the Republic of Tatarstan



1 – Agriculture, forestry, hunting, fishing and fish farming; 2 – Mining; 3 – Manufacturing; 4 – Providing electric energy, gas and steam, air conditioning; 5 – Construction; 6 – Transportation and storage; 7 – Activities in the field of information and communication; 8 – Professional, scientific and technical activities; 9 – Public administration and military security; social Security; 10 – Education; 11 – Activities in the field of health and social services; 12 – Activities in the field of culture, sports, leisure and entertainment

The distribution of the severity of the main types of economic activity in the entities list of identical for the Republic of Tatarstan regions



1 – Agriculture, forestry, hunting, fishing and fish farming; 2 – Mining; 3 – Manufacturing; 4 – Providing electric energy, gas and steam, air conditioning; 5 – Construction; 6 – Transportation and storage; 7 – Activities in the field of information and communication; 8 – Professional, scientific and technical activities; 9 – Public administration and military security; social Security; 10 – Education; 11 – Activities in the field of health and social services; 12 – Activities in the field of culture, sports, leisure and entertainment

CONCLUSIONS

The described algorithm for automating the benchmarking procedure is universal and can be applied to any territorial units of all levels, including municipal and regional.

In the future, we plan to develop the text mining technique in relation to Russian regions in order to model the best development option for a specific territory.

Thank you for the attention

for any questions e-mail to:

uliadubrov@mail.ru
elena.kozonogova@gmail.com