



národní
úložiště
šedé
literatury

Využití bibliografických dat a jejich vizualizace pomocí bibliometrických metod

Szarzec, Jakub
2015

Dostupný z <http://www.nusl.cz/ntk/nusl-181070>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 01.10.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Využití bibliografických dat a jejich vizualizace pomocí bibliometrických metod

Jakub Szarzec

Národní technická knihovna

jakub.szarzec@techlib.cz

Úvodem

- ↘ Metody biblometrie a scientometrie v kontextu.
- ↘ Práce s bibliografickými daty.
- ↘ Vizualizace v bibliometrii.
- ↘ Další příklady vizualizace.
- ↘ Teorie + cvičení.

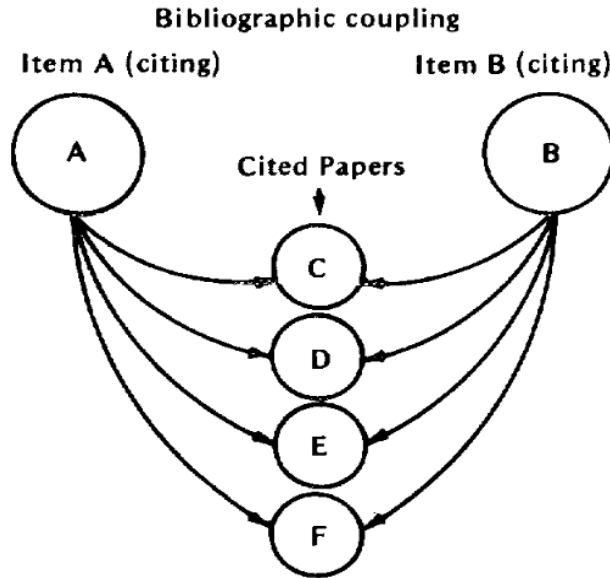
Bibliometrie a scientometrie v kontextu

- ↘ Vědecko-výzkumné metody. Nejsou to pouze tabulky a grafy v Excelu.
- ↘ Bibliometrické metody.
- ↘ Citační analýza je jednou z metod.
- ↘ Bibliografické a citační informace.
- ↘ Analýza vazeb mezi jednotlivými dokumenty.
- ↘ Analýza klíčových slov nebo libovolného textu (abstrakt článku).
- ↘ Interaktivní = Atraktivní. Přehlednost.

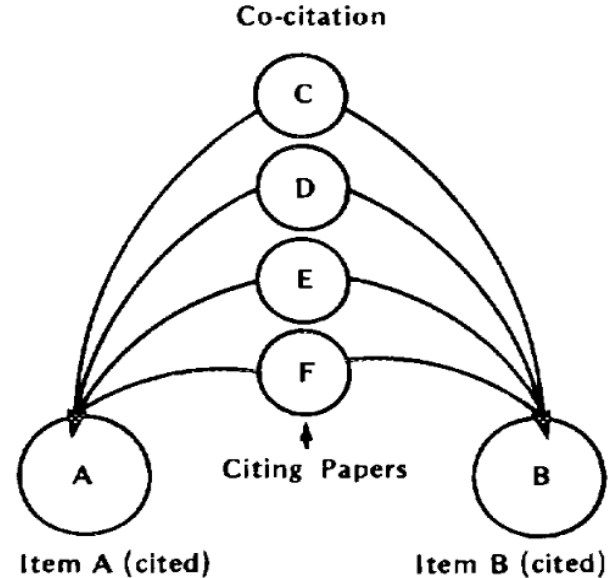
Bibliometrické metody

- ↘ Zakladem je **citační analýza**. Eugene Garfield, 1953. Citační indexace.
- ↘ **Bibliografické spojování** (bibliographic coupling). M. M. Kessler, 1963.
- ↘ **Ko-citační analýza** (co-citation analysis). Henry Small a Irina Marshakova, 1973. B. C. Griffith, 1981.
- ↘ Dvě různé metodologie zkoumající citační vazby mezi dokumenty.
Reference – použitá literatura ve vědecké publikaci.
- ↘ Matematicko-statistické metody.
- ↘ Analýza spoluautorství. Analýza klíčových slov.
- ↘ Jaký je mezi nimi rozdíl?
 - Zkoumají citované (BC) a citující (C-CA) dokumenty.

Bibliografické spojování vs. Ko-citační analýza



Citing papers A and B are related because they cite papers C, D, E, and F.



Papers A and B are associated because they are both cited by papers C, D, E, and F.

Práce s bibliografickými daty

OpenRefine - práce s vybraným datasetem.

Práce s vybraným datasetem.

Práce s daty - cvičení

↘ Dataset = soubor publikací.

↘ Data z RIV

→ VŠCHT FCHI (IČO: 60461373, orjk: **22310**)

↘ Data ze Scopus

→ AF-ID("Vysoka skola chemicko-technologicka v Praze" **60021588**) +
omezení na r. 2014.

↘ Jaký je hlavní rozdíl mezi datasey?

↘ OpenRefine.

OpenRefine

- 2010. Freebase Gridworks / Google Refine.
- Open source **nástroj** sloužící k slouží k získávání, čištění a úpravě dat. Bohatější než Excel.
- Dataset ze Scopus a RIV.
- Využívá **Google Refine Expression Language (GREL)**.
- Použití editace.
- **Vyhledávání** v záznamech a **funkce faset**.
- Základní příkazy jazyka GREL.
- Export.
- <http://openrefine.org>



The screenshot shows a window titled "Authors" with a GREL expression "grel:split(value, ', ')" and a "change" button. Below the expression, it indicates "1831 choices" and "Sort by: name count". The main area displays a list of authors with their names and counts, such as "Pechacek J. 6", "Prech J. 6", "Schreiber I. 6", "Sigler K. 5", "Slavicek P. 6", "Sot P. 6", "Svoboda J. 6", "Tuma J. 6", "Vaclavik J. 6", "Vilhanova B. 6", "Voldrich M. 6", "Bacakova L. 5", "Branyik T. 5", "Cajka T. 5", "Capek J. 5", "Cibulka I. 5", "Coufalova L. 5", "Fojt J. 5", "Jankovsky O. 5", "Jansen J.C. 5", "Januscak J. 5", "Joska L. 5", "Kocourek V. 5", "Kukal J. 5", "Lhotak P. 5", "Macak T. 5", "Maljievskoy A. 5", "Pazout R. 5", "Prajzler V. 5", "Pribyl M. 5", "Prochazka A. 5", "Rezanka P. 5", "Rezanka T. 5", and "Reznickova A. 5".

Author	Count
Pechacek J.	6
Prech J.	6
Schreiber I.	6
Sigler K.	5
Slavicek P.	6
Sot P.	6
Svoboda J.	6
Tuma J.	6
Vaclavik J.	6
Vilhanova B.	6
Voldrich M.	6
Bacakova L.	5
Branyik T.	5
Cajka T.	5
Capek J.	5
Cibulka I.	5
Coufalova L.	5
Fojt J.	5
Jankovsky O.	5
Jansen J.C.	5
Januscak J.	5
Joska L.	5
Kocourek V.	5
Kukal J.	5
Lhotak P.	5
Macak T.	5
Maljievskoy A.	5
Pazout R.	5
Prajzler V.	5
Pribyl M.	5
Prochazka A.	5
Rezanka P.	5
Rezanka T.	5
Reznickova A.	5

Využití OpenRefine – cvičení

- ↘ Čištění a úprava dat.
- ↘ Shluková analýza. Statistika.
- ↘ Výběr a získání potřebných informací.
- ↘ Příprava dat pro další zpracování (import do RIV).

Vizualizace a bibliometrie

Úvod do teorie grafů

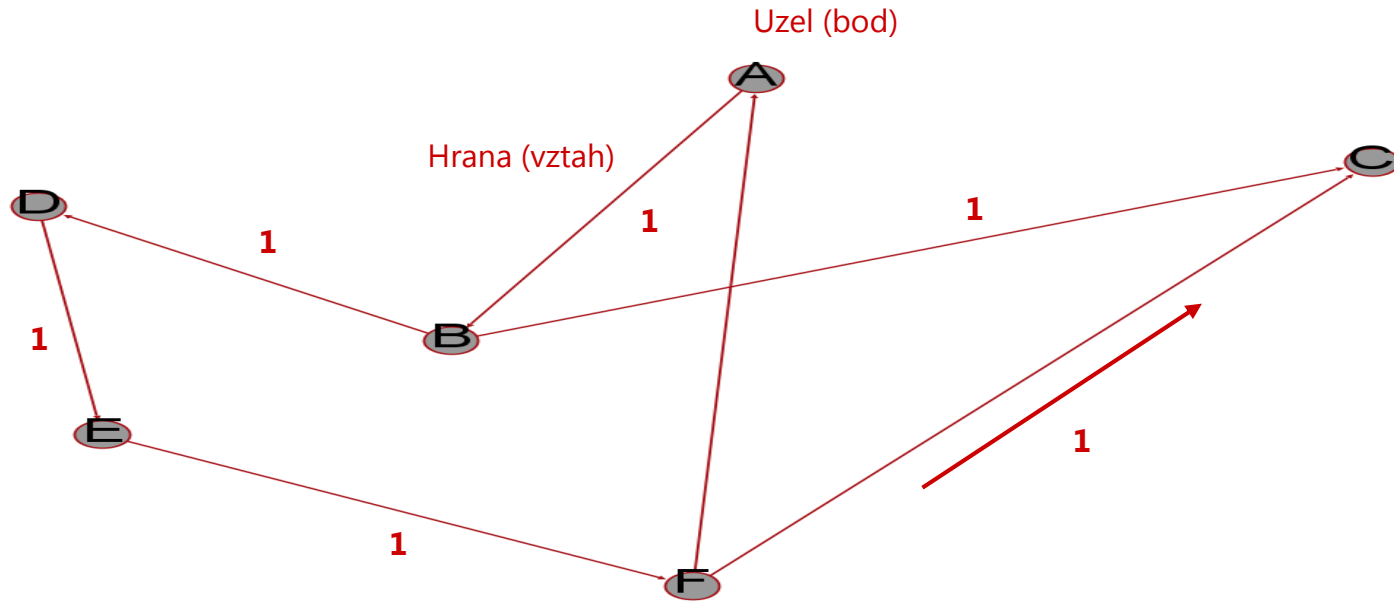
Příklady sítí a jejich vizualizace z pohledu bibliometrie.

Software.

Úvod do teorie grafů

- ↘ Vizualizace podle teorie sítí. Vizualizace BC C-CA
- ↘ **Leonard Euler . Pál Erdős a Alfréd Renyi.**
- ↘ BARABÁSI, Albert-László: **V pavučině sítí.** Paseka. 2009.
- ↘ Využití: matematika, biologie (genetika), lingvistika, informatika, média, informační věda, informatika, atd.
- ↘ **Množina** s vrcholy/uzly a hranami.
- ↘ Grafy mohou být **řízené** nebo **neřízené**.
- ↘ <http://teorie-grafu.cz/>

Příklad sítě (grafu)



Matice sítě (grafu)

Hrany (vztahy) v množině

Uzly (body)
množiny

	A	B	C	D	E	F
A		1	0	0	0	0
B	0		1	1	0	0
C	0	0		0	0	0
D	0	0	0		1	0
E	0	0	0	0		1
F	1	0	1	0	0	

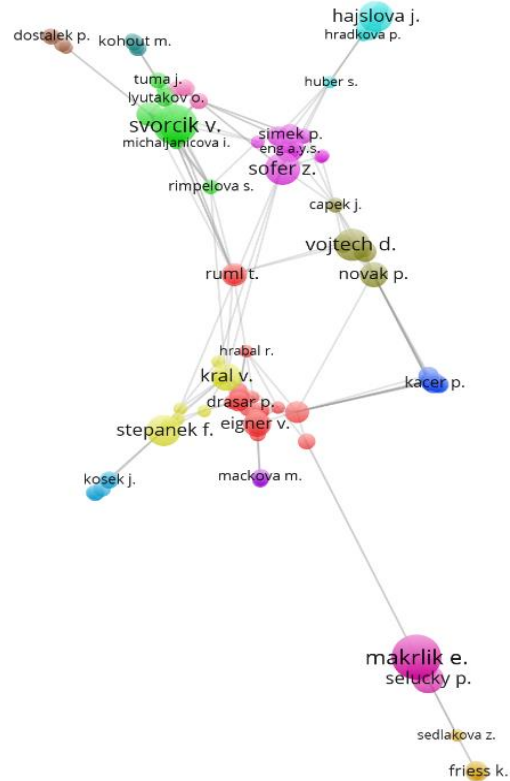
Vizualizace - cvičení

- ↘ **Bibliografické spojování a Ko-citační analýza.**
- ↘ Teorie grafů a sítí.
- ↘ Praktický příklad na staženém datasetu z databáze Scopus (WoS).
- ↘ VOSviewer, ScienceScape a Gephi.

VOSviewer

- ↘ Program pro vytváření vizualizací bibliometrických sítí a map z bibliografických dat.
- ↘ Shluková analýza.
- ↘ Kocitační mapy, spolupráce a klíčová slova.
- ↘ CWTS Leiden University. 2009.
- ↘ Scopus a Web of Science (PubMed).
- ↘ Stažený dataset z databáze Scopus.
- ↘ Export dat.
- ↘ Bez instalace. Pouze stažení.
- ↘ <http://www.vosviewer.com/Home>

brabcova j.



hostasa j.

kukal j.

malijevsky a.

cerveny l.

urban s.

buryan p.

slavicek p.

prochazka a.

sysalova j.

fojt j.

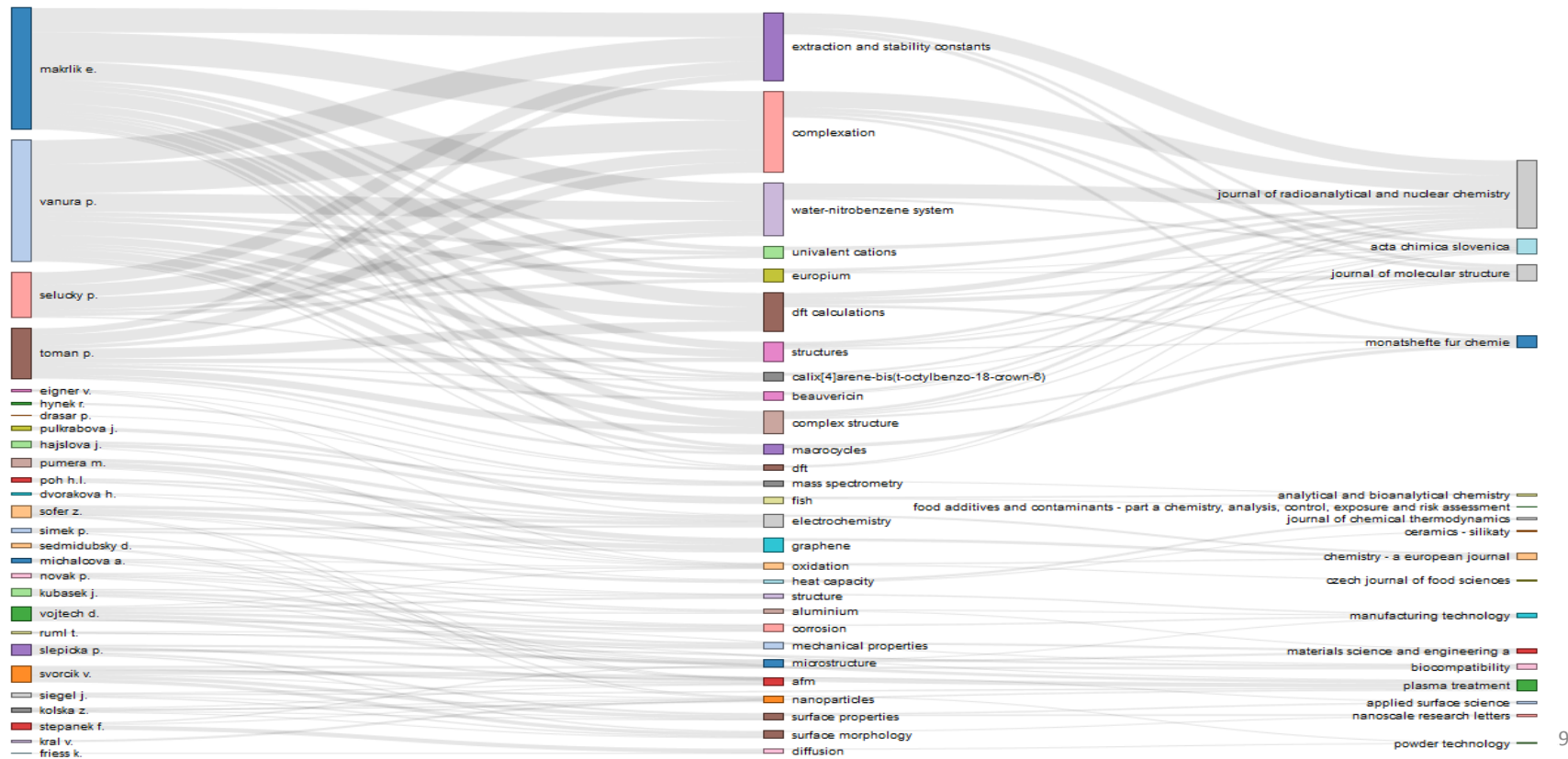
sigler k.

Spoluautoři.

ScienceScape

- ↘ Soubor online nástrojů pro **vytváření vizualizací** bibliometrických sítí a map. Umožňuje zpracování a vizualizaci dat do grafů.
- ↘ Média Lab SciencesPo.
- ↘ OpenSource.
- ↘ Reference, klíčová slova, časopisy.
- ↘ Scopus a Web of Science.
- ↘ Export vizualizace sítě do **Gephi**.
- ↘ Stažený dataset z databáze Scopus.
- ↘ <http://tools.medialab.sciences-po.fr/sciencescape/index.php>

ScienceScape - příklad



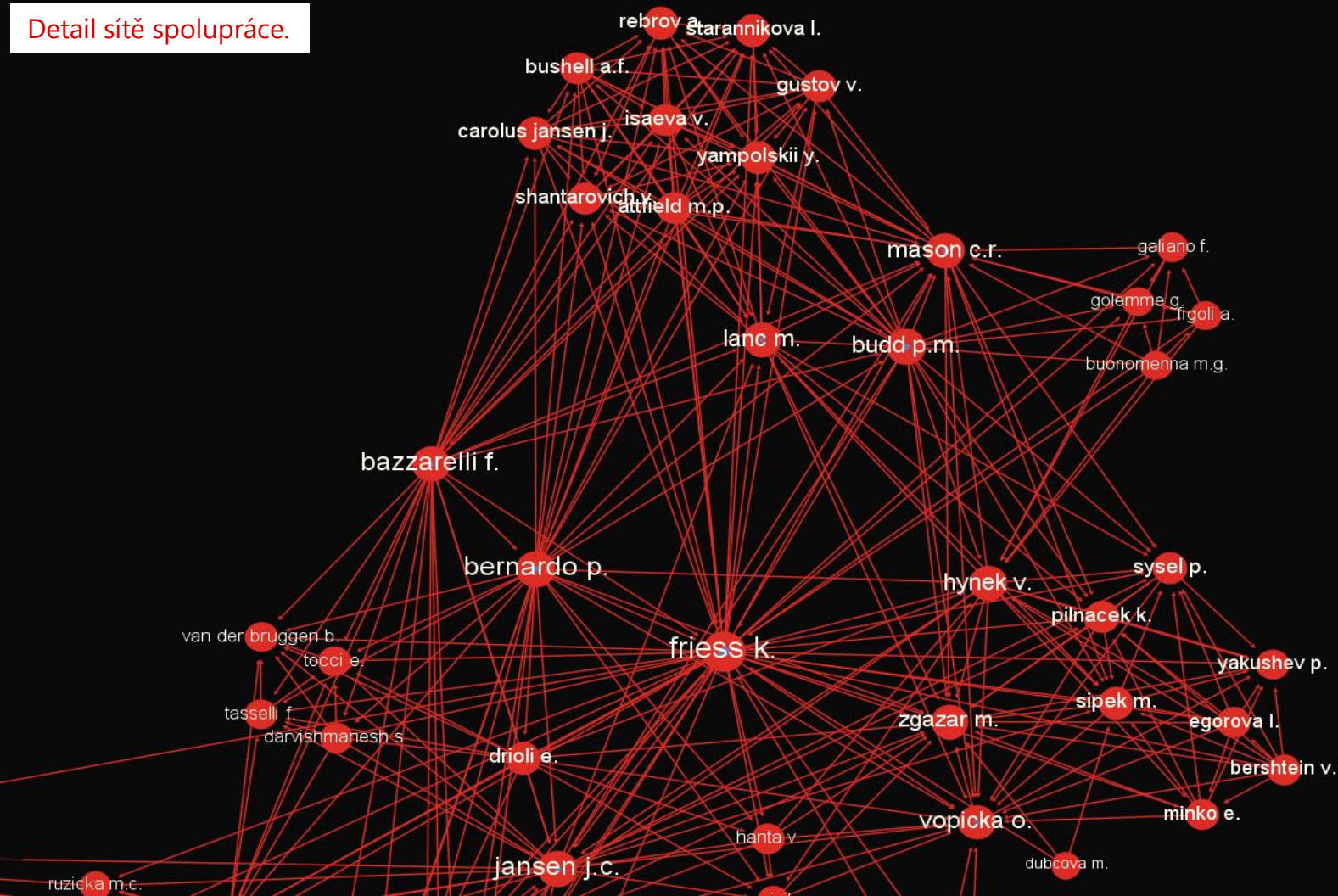
Gephi

- **Vizualizační** nástroj pro vyváření sítí. Vizualní analytika.
- Umožňuje **úpravu a manipulaci s daty**.
- Social network analysis. Statistika a metriky. Shluková analýza. Filtry.
- Grafická úprava.
- Použijeme **export ze ScienceScape** ve formátu GEXF = síť spolupracujících autorů.
- Nutná instalace.
- <http://gephi.github.io/>



Spolupráce mezi jednotlivými autory.

Detail sítě spolupráce.



Další vizualizační programy

↘ Excel template **NodeXL** (formát GraphML)

→ <http://nodexl.codeplex.com/>

↘ **CiteSpace**

→ <http://cluster.cis.drexel.edu/~cchen/citespace/>

↘ **Pajek** (formát NET)

→ <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

↘ **Science of Science (Sci²) Tool**

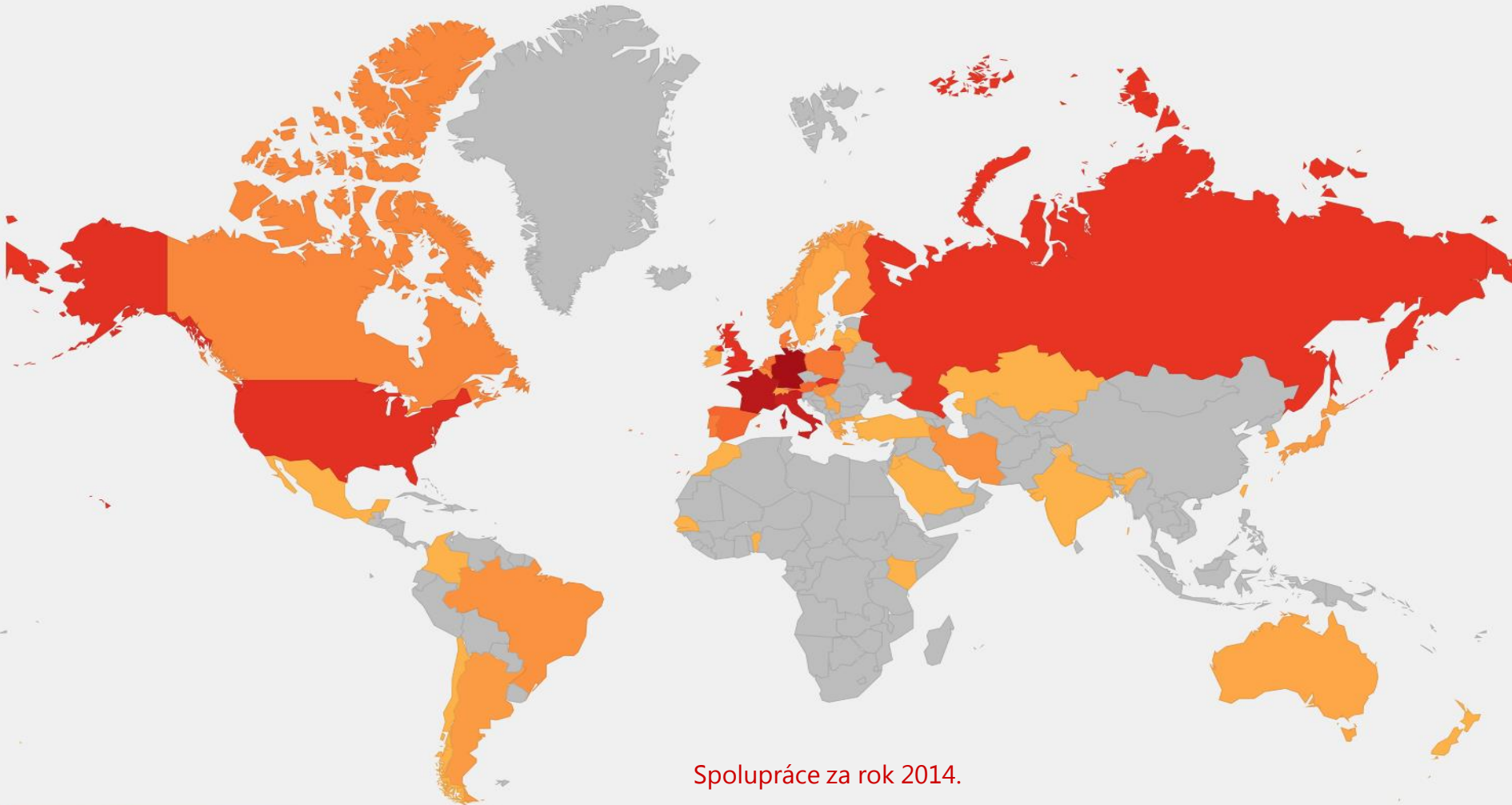
→ <https://sci2.cns.iu.edu/>

Další vizualizace

Příklady.

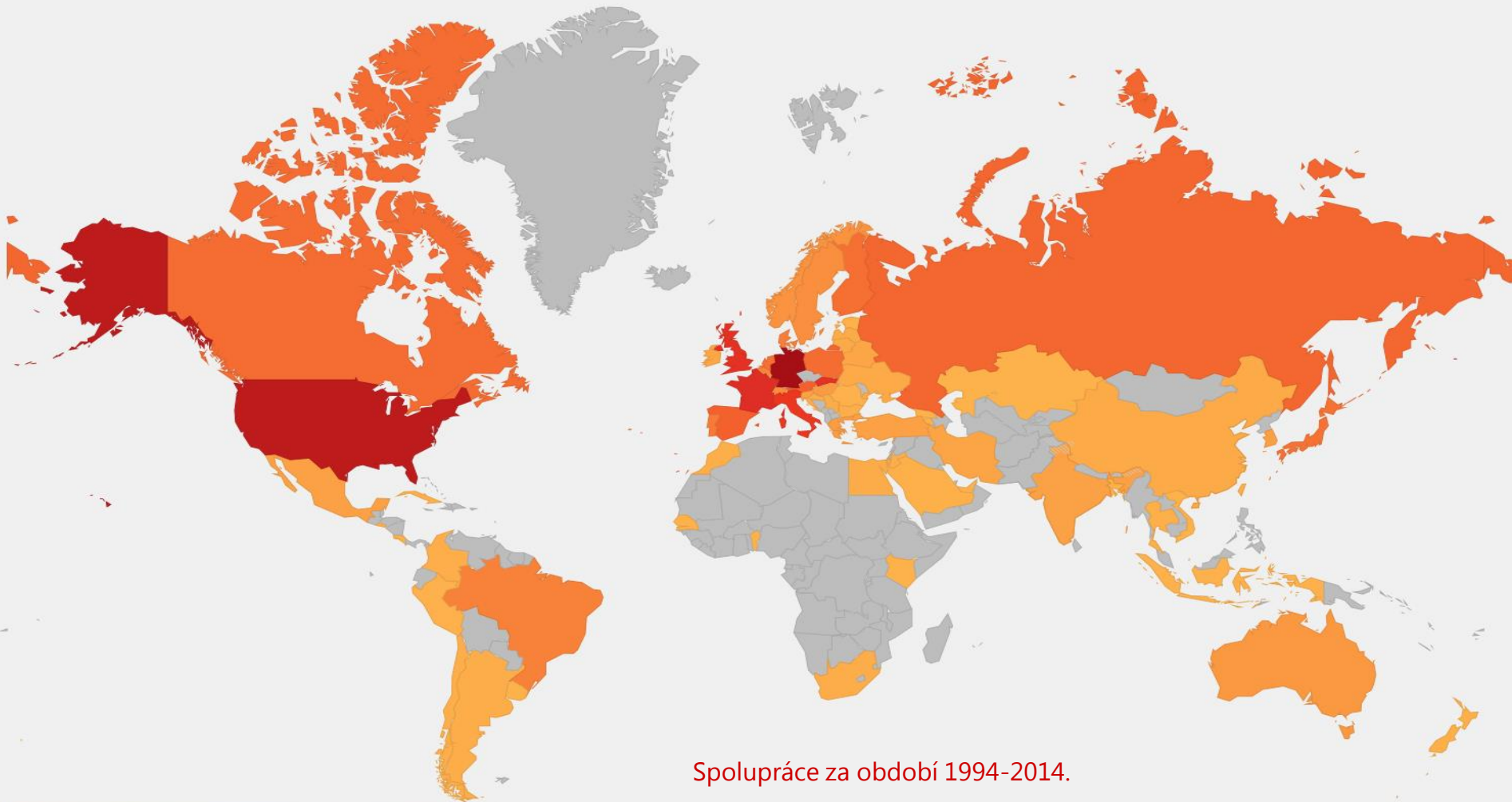
Geomapa zahraniční spolupráce.

Klíčová slova.



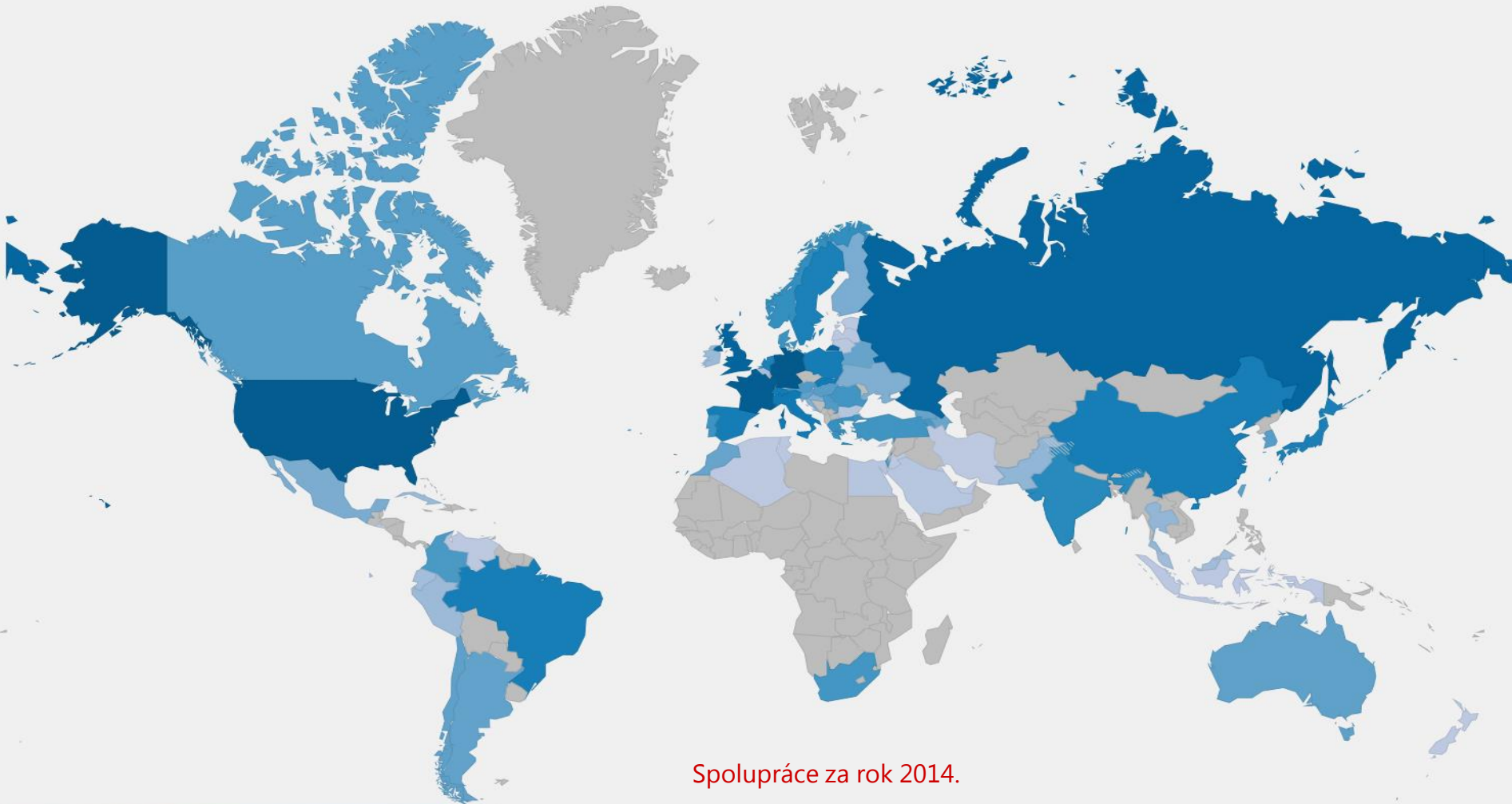
Spolupráce za rok 2014.



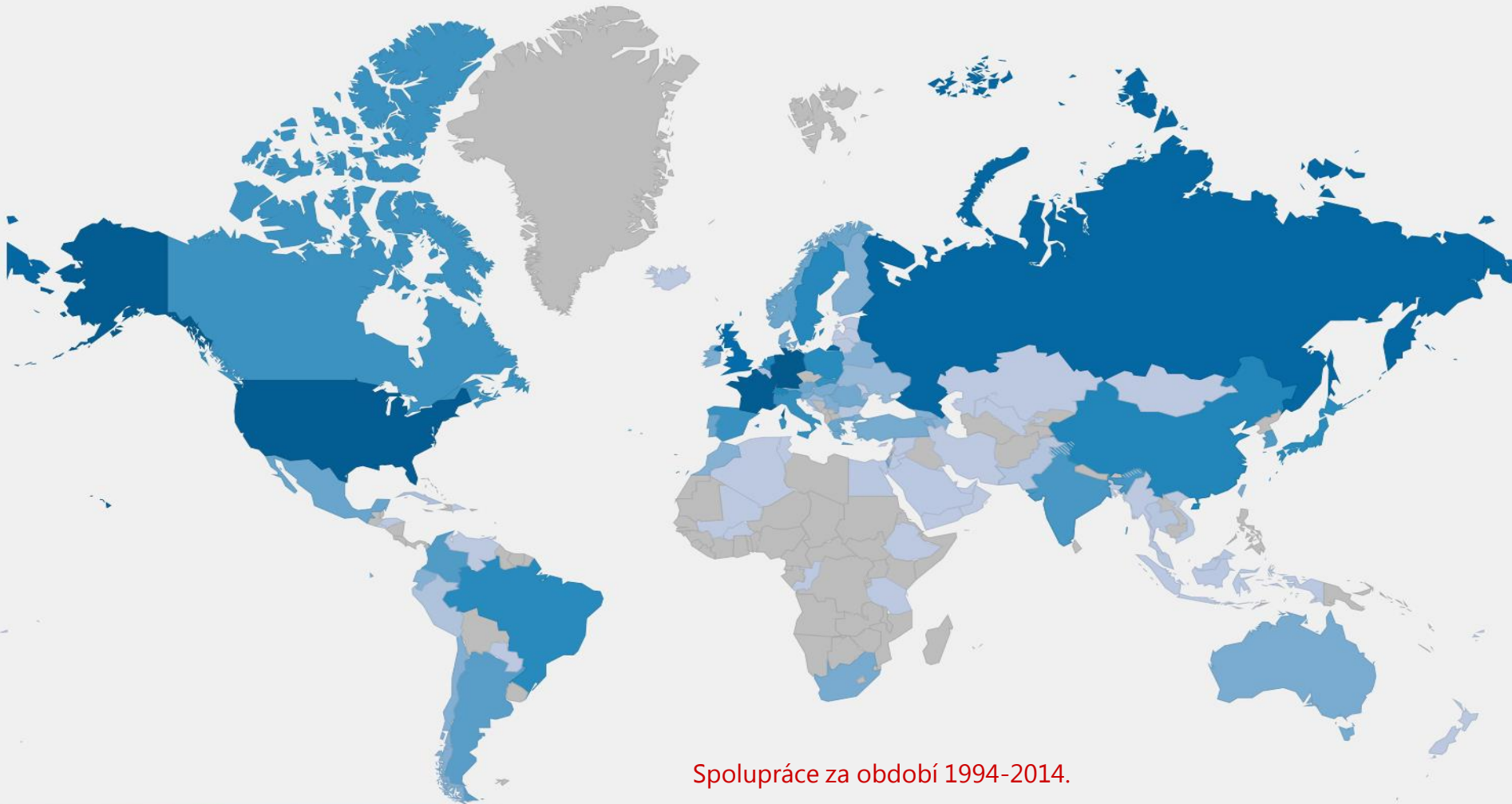


Spolupráce za období 1994-2014.





Spolupráce za rok 2014.



Spolupráce za období 1994-2014.

Search

Alerts

My list

My Scopus

Analyze search results

Export Print E-mail

AF-ID ("Vysoka skola chemicko-technologicka v Praze" 60021588) AND (LIMIT-TO (PUBYEAR , 2014)) [Back to your search results](#)

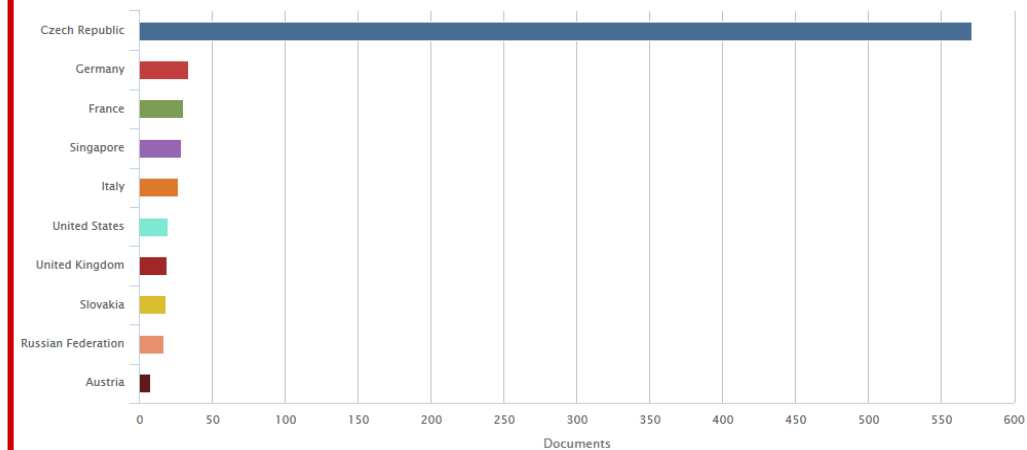
617 document results Choose date range to analyze: 2014 to 2014 Analyze

Year	Source	Author	Affiliation	Country/Territory	Document type	Subject area
------	--------	--------	-------------	-------------------	---------------	--------------

Country/Territory	Documents
<input checked="" type="checkbox"/> Czech Republic	571
<input checked="" type="checkbox"/> Germany	34
<input checked="" type="checkbox"/> France	30
<input checked="" type="checkbox"/> Singapore	29
<input checked="" type="checkbox"/> Italy	27
<input checked="" type="checkbox"/> United States	20
<input checked="" type="checkbox"/> United Kingdom	19
<input checked="" type="checkbox"/> Slovakia	18
<input checked="" type="checkbox"/> Russian Federation	17
<input checked="" type="checkbox"/> Austria	8
<input type="checkbox"/> Spain	8
<input type="checkbox"/> Netherlands	7
<input type="checkbox"/> Denmark	6
<input type="checkbox"/> Portugal	6
<input type="checkbox"/> Poland	6
<input type="checkbox"/> Canada	5
<input type="checkbox"/> Belgium	5
<input type="checkbox"/> Brazil	4
<input type="checkbox"/> Luxembourg	4
<input type="checkbox"/> Iran	4
<input type="checkbox"/> Hungary	4
<input type="checkbox"/> Finland	3

Documents by country/territory

Compare the document counts for up to 15 countries/territories

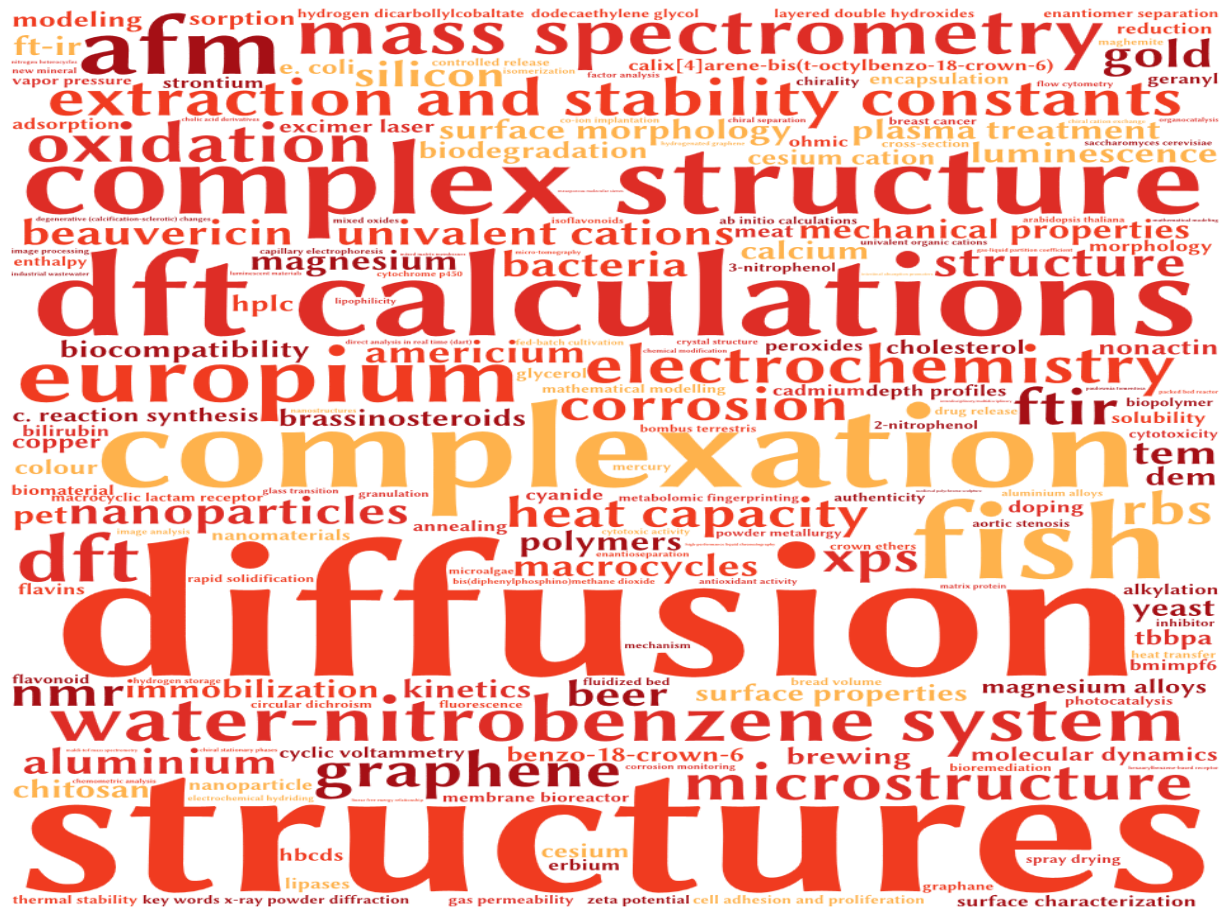


```

1 <html>
2 <head>
3 <script type="text/javascript" src="https://www.google.com/jsapi"></script>
4 <script type="text/javascript">
5   google.load("visualization", "1", {packages:["geochart"]});
6   google.setOnLoadCallback(drawRegionsMap);
7
8   function drawRegionsMap() {
9
10     var data = google.visualization.arrayToDataTable([
11       ['Country', 'Publications'],
12       ['Argentina', 3],
13       ['Australia', 2],
14       ['Austria', 8],
15       ['Belgium', 5],
16       ['Benin', 1],
17       ['Brazil', 4],
18       ['Canada', 5],
19       ['Colombia', 1],
20       ['Denmark', 6],
21       ['Finland', 3],
22       ['France', 30],
23       ['Germany', 34],
24       ['Greece', 2],
25       ['Hungary', 4],
26       ['Chile', 1],
27       ['India', 1],
28       ['Iran', 4],
29       ['Ireland', 2],
30       ['Italy', 27],
31       ['Japan', 3],
32       ['Jordan', 1],
33       ['Kazakhstan', 1],
34       ['Kenya', 1],
35       ['Latvia', 1],
36       ['Lithuania', 2],
37       ['Luxembourg', 4],
38       ['Mexico', 1],
39       ['Morocco', 1],
40       ['Netherlands', 7],
41       ['New Zealand', 1],
42       ['Norway', 3],
43       ['Poland', 6],
44       ['Portugal', 6],
45       ['Russian Federation', 17],
46       ['Saudi Arabia', 1],
47       ['Senegal', 1],
48       ['Serbia', 2],
49       ['Singapore', 29],
50       ['Slovakia', 18],
51       ['South Korea', 2],
52       ['Spain', 8],
53       ['Sweden', 2],
54       ['Switzerland', 3],
55       ['Taiwan', 1],
56       ['Turkey', 1],
57       ['United Kingdom', 19],
58       ['United States', 20]
59     ]);
60
61     var options = {
62       colorAxis: {colors: ['#feb24c', '#f03b2e', '#de2d26', '#a50f15']},
63       backgroundColor: '#f0f0f0',
64       datalessRegioncolor: '#bdbdbd'
65     };
66
67     var chart_div = document.getElementById('regions_div');
68     var chart = new google.visualization.GeoChart(document.getElementById('regions_div'));
69
70     google.visualization.events.addListener(chart, 'ready', function () {
71       chart_div.innerHTML += '';
72       console.log(chart_div.innerHTML);
73     });
74
75     chart.draw(data, options);
76   }
77 </script>
78 </head>
79 <body>
80   <div id="regions_div" style="width: 1500px; height: 1000px;"></div>
81 </body>
82 </html>

```

Stát a počet publikací.



Klíčová slova podle jejich výskytu.

<https://tagul.com/>

Návody a tutoriály k nástrojům

↘Gephi

→<http://gephi.github.io/users/>

↘Google Charts

→<https://developers.google.com/chart/>

↘OpenRefine

→<https://github.com/OpenRefine/OpenRefine/wiki/Screencasts>

→<https://github.com/OpenRefine/OpenRefine/wiki/External-Resources>

↘ScienceScape

→<https://vimeo.com/78916756>

↘VOSviewer

→<http://www.vosviewer.com/Getting-Started>

Závěrem

- ↘ Práce se základními hrubými daty. Podáváme tím přesné a relevantní informace?
 - Jak je dataset čistý?
 - Jsou zobrazené informace relevantní?
 - V jakém jsou kontextu?
 - Velikost datasetu značně ovlivňuje přehlednost sítě/grafu.
 - Statistická spolehlivost.
 - Proces ověření. Pravidlo „Garbage in, Garbage out“ – kontrola dat. Interpretace. Zpětná replikovatelnost výsledků.
- ↘ Využití programovacího jazyka **R** nebo **Python**.
- ↘ Pohodlnější cesta přes nabízené komerční systémy analytických nástrojů.

Děkuji Vám za pozornost...

Dotazy???

Použité zdroje

1. BARABÁSI, Albert-László. *V pavučině síťí*. Vyd. 1. V Praze: Paseka, 2005, 274 s. ISBN 80-7185-751-3
2. ECK, Nees Jan van a Ludo WALTMAN. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* [online]. 2009, vol. 84, is. 2, s. 523-538 [cit. 2015-02-27]. DOI: 10.1007/s11192-009-0146-3. ISSN: 1468-4527. Dostupné z: <http://link.springer.com/article/10.1007%2Fs11192-009-0146-3>
3. GARFIELD, E. Announcing the SCI compact disc edition: CD-ROM gigabyte storage technology, novel software, and bibliographic coupling make desktop research and discovery reality. In: *Essays of an Information Scientist: Science Literacy, Policy, Evaluation, and other Essays*. Vol. 11. Philadelphia, Pa: ISI Press, 1990. ISBN 9780894950841. Dostupné z: <http://www.garfield.library.upenn.edu/essays/v11p160y1988.pdf>
4. KESSLER, M. M. Bibliographic coupling between scientific papers. *American Documentation*. 1963, vol. 14, is. 1, s. 10-25. DOI: 10.1002/asi.5090140103. ISSN: 0096-946X. Dostupné z: <http://onlinelibrary.wiley.com/doi/10.1002/asi.5090140103/abstract>
5. SMALL, Henry. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*. 1973, vol. 24, is. 4, s. 265-269. DOI: 10.1002/asi.4630240406. ISSN: 0002-8231. Dostupné z: <http://onlinelibrary.wiley.com/doi/10.1002/asi.4630240406/abstract>