

## КЛАСТЕРНИЙ АНАЛІЗ КРАЇН СВІТУ ЗА ЇХ ІНТЕГРАЛЬНИМИ ЗА ОСТАННІ 20 РОКІВ БІБЛІОМЕТРИЧНИМИ ПОКАЗНИКАМИ НАУКОВИХ ПУБЛІКАЦІЙ ТА ЦИТУВАНЬ

*Наведені результати кластерного аналізу 200 країн світу за показниками їх наукової публікаційної діяльності та індикаторів цитування, що представлені в світовій системі Scimago Journal & Country Rank.*

**Ключові слова:** інтелектуальний аналіз, кластерний аналіз, наукометрична база даних, індекси цитування, рейтингове оцінювання, ранжування.

**Вступ, огляд публікацій.** Для надання світовій спільноті цілісної порівняльної картини стану світового наукового середовища та аналізу інтегральної активності, результативності та місця і ролі країн у цих процесах створені та розвиваються світові системи, наприклад, *Scimago Journal & Country Rank (SJR)* [1] та їхні національні складові [2].

В багатьох країнах на основі показників наукометричних баз даних таких як Web of Science, Scopus та Google Scholar і інших визначають рейтинги науковців.

Проте як в їхніх рейтингових таблицях, так і в SJR місця науковців чи країн в основному ранжовані за одним індикатором або послідовно за іншими при однаковості попередніх, наприклад у SJR країни ранжуються за показником загальної кількості наукових публікацій за визначений період років, а для експертного оцінювання результативності діяльності вчених і дослідницьких колективів використовуються індекс Гірша чи йому подібні.

Але, хоч набір таких індикаторів суттєво і розширився, особливо в останні роки, представлені ранжування у вищевказаних *on-line* системах не враховують відносність, взаємозалежність показників, їх впливовість на результуючу оцінку чи місце.

А тому актуальним завданням є здійснення багатоаспектного групування та оцінювання для визначення сукупного рейтингу країн з точки зору їх наукової публікаційної діяльності, визначення місця України та проблем і шляхів їх розв'язання.

Для цього, як було показано в роботах [3-8], доцільно використовувати сучасні інтелектуальні інформаційні технології та

відомі і широко використовувані нейропакети, що дозволяють ефективно здійснювати кластерний аналіз об'єктів із значною кількістю якісних та кількісних ознак і визначати впливовість факторів. Ефективні демонстрації таких підходів та обробка даних, взятих з реальних систем та наукометричних баз, створюють підґрунтя для прийняття оптимальних стратегій стосовно розвитку бібліометричних систем в Україні так і науки в цілому.

Тому **метою** даної роботи є проведення кластерного аналізу 200 країн світу за показниками їх наукової публікаційної діяльності та індикаторів цитування, що представлені в світовій системі SJR за допомогою нейропакетів.

**Об'єктом** дослідження є сукупність таких показників, взятих нами для експерименту з бази SJR за останні 20 років (станом на 16.11.2015) як: загальна кількість (К) документів (Д), К цитованих Д, К цитувань та К само- цитувань, Н-індекс, та додатково розрахованих нами: К Д на 1 тисячу населення, середню кількість цитувань документа та відносні долі цитованих документів та само-цитувань.

**Предметом** дослідження є кластерне групування 200 країн світу і показ переваг застосування інтелектуального аналізу для більш правильного їх ранжування та пояснення причин та особливостей віднесення до рейтингових місць та кластерів з урахуванням впливовості факторів.

В першому експерименті для відібраних 200 країн з вищевказаної бази SJR та запропонованих показників нами було здійснено розбиття множини країн на 7 кластерів, у другому - на 9, але отримані результати не дозволяли виявляти деякі відмінності між підмножинами. Тому було прийнято рішення проводити поділ на 11 кластерів, використовуючи карту Кохонена нейропакету Excel Neural Package [3].

Оскільки всі показники є кількісними, то ці дані після відповідних округлень використовувались безпосередньо без спеціального кодування. Тут ми наводимо лише деякі кінцеві результати та «вікна». На рис. 1 показано фрагмент результуючої таблиці (вікна пакету) з номерами і показниками визначених кластерів та їх місць. У табл. 1 наведені центроїдні характеристики кластерів.

Rank	Country	Documents	Citable documents	Citations	Self-Citations	H index	Population (in thousands)	V_Citations per Document	Number of documents per 1000 population	Fraction Citations	Fraction Self-Citations
1	United States	8 626 193	7 876 234	177 434 935	83 777 658	1 648	321 369	23	27	91	47 A
2	China	3 617 355	3 569 652	19 110 353	10 462 121	495	1 368 660	5	3	99	55 A
3	Germany	2 176 860	2 046 433	35 721 869	9 141 181	887	80 854	17	27	94	26 A
4	Japan	2 074 872	2 008 410	27 040 067	7 619 559	745	126 920	13	16	97	28 A
5	France	1 555 629	1 468 286	24 700 140	5 616 943	811	100 998	17	15	94	22 A
6	Canada	1 227 380	1 134 588	22 162 666	4 136 384	794	35 100	20	35	92	19 A
7	Italy	1 200 448	1 117 013	18 019 464	4 186 908	713	61 855	16	19	93	23 A
8	India	998 544	944 632	6 989 150	2 409 025	383	1 251 696	7	1	96	34 A
9	Spain	952 099	884 670	12 628 097	3 068 362	591	48 146	14	20	93	24 A
10	Australia	890 458	809 027	13 772 961	2 947 945	644	22 751	17	39	91	21 A
11	South Korea	739 229	719 338	7 081 429	1 528 443	424	3 780	10	84	97	22 A
12	Russian Federation	701 029	689 095	4 289 618	1 273 073	390	142 424	6	5	98	30 A
13	Netherlands	681 804	628 678	14 278 721	2 321 446	694	16 948	23	40	92	16 A
14	Brazil	598 234	573 988	5 036 027	1 699 530	379	204 260	9	3	96	34 A
15	Switzerland	493 857	460 824	10 872 269	1 458 098	686	8 122	24	61	93	13 A
16	Taiwan	491 560	477 442	4 790 230	1 075 153	331	23 415	10	21	97	22 A
17	Sweden	460 607	433 674	9 417 604	1 448 940	814	9 802	22	47	94	15 A
18	Poland	431 016	418 917	3 491 958	901 545	371	38 562	8	11	97	26 A
19	Turkey	390 874	368 197	2 938 841	737 423	266	79 414	8	5	94	25 A
20	Belgium	372 093	348 017	6 691 791	948 874	547	11 324	19	33	94	14 A
21	Iran	287 010	278 388	1 504 541	573 656	180	81 824	5	4	97	38 A
22	Israel	272 352	256 036	5 079 652	694 959	496	8 049	20	34	94	14 A
23	Austria	268 472	250 181	4 334 382	583 299	449	8 666	17	31	93	13 A
24	Denmark	263 026	245 115	5 494 671	779 833	518	5 582	22	47	93	14 A

Рис. 1. Результати кластеризації (фрагмент вікна програми)

Центр кластеру	Кластери										
	A1	A3	A2	A4	A10(5)	A7(6)	A8(7)	A11(8)	A6(9)	A5(10)	A9(11)
X <sub>1</sub> - Documents	8 626 193	282 994	1 439 678	1 850	196 627	4 015	2 255	2 307 950	485	4 414	19 066
X <sub>2</sub> - Citable documents	7 876 234	265 435	1 352 490	1 734	189 995	3 729	2 155	2 257 142	405	4 089	18 362
X <sub>3</sub> - Citations	177 434 935	4 760 553	22 005 038	46 847	1 536 625	32 490	30 301	13 049 752	4 273	59 019	133 919
X <sub>4</sub> - Self-Citations	83 777 658	731 220	5 231 040	4 541	389 084	3 499	2 985	6 435 573	272	7 591	21 467
X <sub>5</sub> - H index	1 648	412	739	83,2	218	48,9	50,5	438	26,8	67	82,4
X <sub>6</sub> - Population (in thousands)	321 369	7 035	68 089	1 855	55 384	4 996	8 503	1 310 178	7 967	21 978	29 757
X <sub>7</sub> - V Citations per Document	22,1	17,9	16,4	27,3	7,99	8,76	12,9	6,38	8,67	14,8	6,62
X <sub>8</sub> - Number of documents per 1000 population	26,3	40,1	24,6	2,7	9,19	2,87	1,72	1,72	2,06	1,07	1,68
X <sub>9</sub> - Fraction Citations	91,5	93,2	93,5	93,2	96,5	92,6	95,8	96,6	81,8	92,2	96,5
X <sub>10</sub> - Fraction Self-Citations	47,3	14,7	23,4	11,3	23,5	8,24	7,72	44,5	6,07	8,61	14,5
K-ть елементів в кластері	1	18	7	4	25	42	34	2	6	22	39
оцінка (бали)	11	10	9	8	7(2)	6(5)	5(3)	4	3(6)	2(7)	1
Бали Міся(7_8)	10+10	11+9	8+9	11+6	8+3	7+5	6+4	1+3	4+5	7+1	2+1
Бали Міс (7_8_5)	20+11	20+8	17+10	17+6	11+7	12+3	10+3	4+9	9+1	8+4	3+5

Табл. 1 Характеристика виявлених кластерів (розбиття на 11 кластерів)

Номера кластерів (відповідні стовпчики) у таблиці впорядковані у відповідності до отриманої кількості балів. Це спрощує порівняння кластерів. Аналіз отриманих результатів з цієї таблиці дозволяє зробити наступні висновки. Безперечним лідером за показниками 1-5 (з 10!) є США, до того ж за 6-8 індикаторами вони на 2 місці, а тому лише одні США є у 1 кластері та оцінені нами найвищими 11 балами. До 2 кластеру (A3) ввійшло 18 країн, серед яких: Південна Корея, Нідерланди, Швейцарія, Швеція, Бельгія, Ізраїль, Австрія, Данія, Фінляндія, Греція, Норвегія, Гонг-Конг, Сінгапур, Південна Африка, Нова Зеландія, Ірландія, Ісландія, Монако. Характерним для цієї групи є найвищий у порівнянні з іншими кластерами 8-ий показник (аж 40

документів на тисячу населення!), при середній чисельності населення країн групи, рівній 7млн. чол., 4-місце по 5-ому кумулятивному Н-індексу (419), та 3-тє місце по 7-му, тобто по середній кількості цитувань одного документа (18 раз!). До 3-ої лідируючої групи (A2) входять 7 країн: Німеччина, Японія, Франція, Канада, Італія, Іспанія, Австралія, з середньою чисельністю (70 млн. чол.), що забезпечує їм лідируючі (2-ге чи 3-тє) місця по 1-6 показникам, 3-тє місце по 8-му. Після цих трьох лідируючих груп особливий кластер (A4) складають 4 маленьких країни: Панама, Гамбія, Бермуди, Гвінея-Бісау, що по кількості цитувань одного документа (27 !) займають 1-е місце, хоч по 5-у та 8-у показникам займають середні, 7-е та відповідно 5-е місця. Україна попала у 5-у по рангу групу (A10) з 25 країн, що показані на рис.2.

32	Monaco	1 436	1 322	24 908	2 882	69	31	19	47	92	12 A	3
33					<b>Кластер A10</b>							
34	Russian Federation	701 029	689 095	4 289 618	1 273 073	390	142 424	6	5	98	30 A	10
35	Brazil	598 234	573 988	5 036 027	1 699 530	379	204 260	9	3	96	34 A	10
36	Taiwan	491 560	477 442	4 790 230	1 075 153	331	23 415	10	21	97	22 A	10
37	Poland	431 016	418 917	3 491 958	901 545	371	38 562	8	11	97	26 A	10
38	Turkey	390 874	368 197	2 938 841	737 423	266	79 414	8	5	94	25 A	10
39	Iran	287 010	278 388	1 504 541	573 856	180	81 824	5	4	97	38 A	10
40	Czech Republic	213 209	206 738	1 867 611	440 161	294	10 645	9	20	97	24 A	10
41	Mexico	210 387	201 540	1 962 572	408 914	289	121 737	10	2	96	21 A	10
42	Portugal	189 052	179 134	2 096 242	407 892	297	10 825	12	17	95	19 A	10
43	Malaysia	153 378	148 844	670 387	183 198	165	30 514	5	5	97	27 A	10
44	Argentina	145 416	138 788	1 681 700	354 132	273	43 432	12	3	95	21 A	10
45	Hungary	136 034	130 299	1 660 840	264 809	301	9 898	13	14	96	16 A	10
46	<b>Ukraine</b>	<b>133 650</b>	<b>131 490</b>	<b>635 570</b>	<b>176 428</b>	<b>174</b>	<b>44 429</b>	<b>5</b>	<b>3</b>	<b>98</b>	<b>28 A</b>	<b>10</b>
47	Romania	125 576	122 884	619 956	153 395	167	21 666	5	6	98	25 A	10
48	Egypt	120 493	117 104	818 728	162 544	165	88 487	7	1	97	20 A	10
49	Chile	90 216	86 521	1 014 687	193 534	233	17 508	12	5	96	19 A	10
50	Pakistan	81 612	78 219	425 467	118 262	148	199 086	5	0	96	28 A	10
51	Slovakia	72 847	70 964	561 511	113 733	180	5 445	8	13	97	20 A	10
52	Croatia	72 110	69 675	459 356	97 694	177	4 465	7	16	97	21 A	10
53	Slovenia	64 483	62 170	611 672	115 616	189	1 983	10	33	96	19 A	10
54	Nigeria	53 298	51 223	272 400	61 408	115	181 562	5	0	96	23 A	10
55	Tunisia	51 590	49 230	276 247	60 183	109	11 037	6	5	95	22 A	10
56	Serbia	45 000	43 151	188 381	47 922	100	7 177	4	6	96	25 A	10
57	Lithuania	32 137	31 399	227 339	51 689	133	3 526	7	9	98	23 A	10
58	Estonia	25 458	24 479	313 735	55 005	162	1 265	13	20	96	18 A	10
59					<b>Кластер A9</b>							
60	Thailand	109 832	104 982	976 328	162 255	213	67 976	9	2	96	17 A	9
61	Saudi Arabia	94 460	87 642	547 467	80 252	164	27 752	6	2	96	16 A	9

Рис.2. Фрагмент вікна програми з виділеними в кластер A10 країнами (Україна !).

Але вивчення її показників у порівнянні з середніми по групі показує, що 5-ий (174 !), 7-ий (5 !) та 8-ий (3 !) є нижчими ніж відповідні: (224, 8, 9). Особливий кластер (A11) складають 2 країни: Китай та Індія, для яких характерні значна (за 1 мільярд!) чисельність населення і відповідно значні 1-6 показники, проте за 7 та 8 показниками вони займають передостанні місця. Найбільш відсталу групу A6 складають 6 країн: Афганістан, Сомалі, Ліберія, Андорра, Барбадос, Сент-Китс и Невис.

Аналогічно, можна зробити повний аналіз всіх виділених угруповань. Результати аналізу детальніше будуть висвітлені у

доповіді, а тут, з урахуванням обмежень, ми акцентуємо увагу лише на тому, що отримані центроїдні показники додатково можна обробляти і встановлювати характерні для кожного угруповання країн особливості, в тому числі і на основі відносних змін по кожному з показників. Відмітимо, що загальна кількість документів та сумарні показники цитувань, H-індекс цитування у порівнянні з іншими показниками, що впливають на результуючу оцінку та корелюють з нею, не є достатніми, не завжди розкривають причини та особливості віднесення країн до їх рейтингового місця. Лише розширення списку індикаторів та урахування відносних (7-го та 8-го) дають кращі та точніші результати розбиття об'єктів на кластери.

Для перевірки якості нейромережевої моделі (НММ) кластеризації ми використовували підпрограму Winnet та видозмінені дані шляхом введення для кожного вектора вхідних ознак вихідної змінної, що дорівнювала виставленій кожному кластеру оцінці (у балах!). Відповідно для кожного варіанта розбиття, на 11, 9 чи 7 кластерів, створювалась своя таблиця даних з відповідним діапазоном (1-11, 1-9, 1-7 у балах) вихідної змінної, після чого після процедур навчання визначались прогнозовані відгуки НММ та відповідні відносні похибки оцінювання. Крім того, шляхом введення відхилень від значень ознак вхідних векторів для деяких вибраних країн формувались відповідні вибірки для тестування НММ для кожного з цих варіантів.

Нами завідомо наперед деякі країни, наприклад, Великобританія, були вилучені з навчальних вибірок та введені лише у тестову. Це дозволило більш гарантовано довіряти отриманим результатам. Експерименти показали, що середня похибка оцінювання-прогнозування на створеній вибірці не перевищувала навіть для найгірших ситуацій 10-12% для всіх варіантів розбиття, хоч для деяких екземплярів (країн) вибірки максимальна похибка при недостатній кількості епох навчання могла досягати і 20-30%. А для деяких експериментальних ситуацій, що видно також з рис.3, могла дорівнювати всього 0,02%. Проте для визначення тенденцій і трендів, на наш погляд, при вибраних діапазонах вихідної оцінки такі середні похибки є цілком допустимими, а тому ускладнювати НММ чи суттєво збільшувати час навчання є недоцільним.

Середні похибки оцінювання на тестових вибірках, що також видно з фрагмента-вікна (дивись рис.3) з результатом роботи підпрограми оцінювання, не перевищували декількох відсотків (на рис.3 - 5,85%). Результуючі таблиці з усередненими для кожного

кластера показниками для випадку розбиття вибірки аналізованих країн на 7 кластерів показані в таблиці 2.

Центр кластеру	Кластери						
	A1(1)	A2(2)	A7(3)	A3(4)	A6(5)	A5(6)	A4(7)
X <sub>1</sub> - Documents	6,1e+006	5,69e+005	2,12e+005	4,23e+003	1,62e+004	2,14e+003	779
X <sub>2</sub> - Citable documents	5,71e+006	5,34e+005	2,18e+005	3,95e+003	1,55e+004	1,99e+003	672
X <sub>3</sub> - Citations	9,78e+007	9e+006	1,74e+006	6,41e+004	1,12e+005	2,13e+004	6e+003
X <sub>4</sub> - Self-Citations	4,69e+007	1,85e+006	4,66e+005	7,6e+003	1,77e+004	2,19e+003	453
X <sub>5</sub> - H index	1,07e+003	496	224	79	79,9	43,8	31,6
X <sub>6</sub> - Population (in thousands)	8,48e+005	2,24e+004	1,01e+005	1,21e+004	2,21e+004	7,81e+003	6,8e+003
X <sub>7</sub> - V Citations per Document	13,9	17,5	8,05	17,8	6,95	10,8	8,37
X <sub>8</sub> - Number of documents per 1000 population	14,7	36,1	8,63	2,02	1,76	2	2,68
X <sub>9</sub> - Fraction Citations	95	93,3	96,5	93	96,4	94	84
X <sub>10</sub> - Fraction Self-Citations	51	16,9	23,8	10,1	14,3	7,23	6,62
К-ть елементів в кластері	2	25	26	23	44	71	9
оцінка (бали)	7	6	(5)	(4)	(3)	(2)	(1)

Табл. 2 Характеристика виявлених кластерів (розбиття на 7 кластерів).

201	Somalia	98	83	526	29	14	10 616	6	0	85	6	3	2,6439168	-11,87%
202	India	998 544	944 632	6 989 150	2 409 025	383	1 251 696	7	1	95	34	0	4,3927184	9,82%
203	United Kingdom	2397817	2103145	44011201	10321539	1015	63182	21	38	88	23	0	9,1681039	
204	Japan	2 074 872	2 008 600	27 040 067	7 619 559	745	126 920	13	16	97	28	0	9,1681039	1,87%
205	Canada	1 227 500	1 134 588	22 152 666	4 136 384	794	35 100	20	35	92	19	0	9,1681039	1,87%
206	Switzerland	493 857	460 824	10 872 269	1 458 098	686	8 122	24	61	93	13	0	9,1681039	-8,32%
207	Greece	225 000	209 450	2 702 414	444 424	326	10 776	13	21	93	16	0	11,396655	13,97%
208	Russian Federation	701 029	689 095	4 289 618	1 273 073	390	142 424	6	5	98	30	0	7,3016699	4,31%
209	Ukraine	133 650	131 690	635 570	176 428	174	44 429	5	3	99	28	0	7,3016699	4,31%
210	Slovakia	72 847	71 064	561 511	113 733	180	5 445	8	13	98	20	0	7,3016693	4,31%
211	Thailand	109 832	104 982	976 328	162 255	213	67 976	9	2	96	17	0	1,0050202	0,50%
212	Belarus	28 941	28 543	175 490	32 229	122	9 590	6	3	99	18	0	0,7774829	-22,25%
213	Luxembourg	10 527	9 823	92 832	9 604	105	570	9	18	93	10	0	5,5528683	-7,45%
214	Kenya	21 811	20 250	315 176	49 200	165	45 925	16	0	93	16	0	2,6451601	32,26%
215	Tanzania	10 516	9 866	140 878	21 812	113	51 046	14	0	94	15	0	2,7133774	35,67%
216	Bolivia	3 536	3 384	49 509	4 296	78	10 801	15	0	96	9	0	5,5528677	11,06%
217		3 236	3 084	49 509	4 296	78	10 801	16	0	95	9			
218														0,02%
219														
220														5,85%

Рис.3. Фрагмент вікна програми Winnet з відгуками-оцінками (передостанній стовпчик) та похибками (останній)

На рис.4 показані гістограми законів розподілу досліджуваних показників, що сприяє їх вибору.

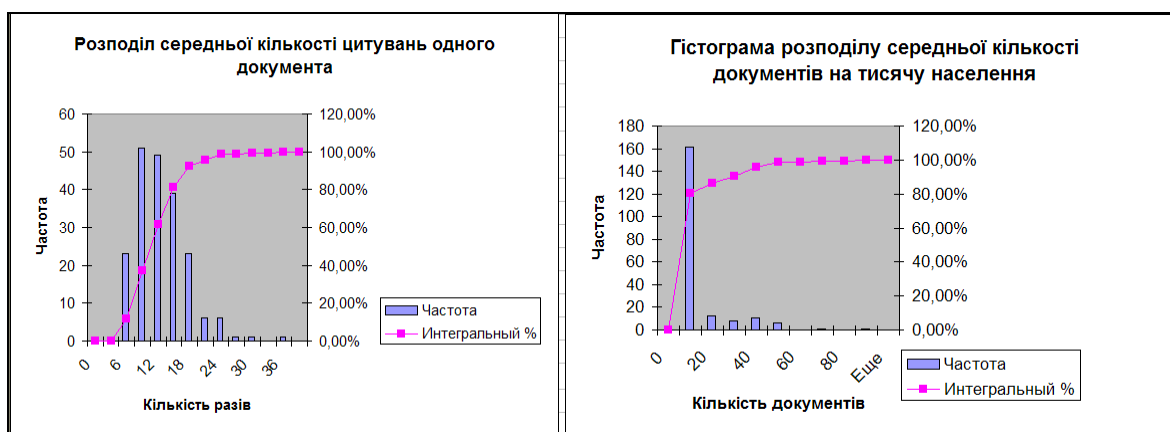


Рис.4. Фрагмент вікон з гістограмами законів розподілу деяких показників

Вигляд емпіричних законів (диференціального та інтегрального) розподілу середньої кількості цитувань одного документа для всієї вибірки країн показує, що може бути прийнята гіпотеза про нормальний закон розподілу цього ( $x-7$ ) показника.

Перевірка цієї гіпотези відомими методами та з використанням для них низки критеріїв підтвердили її істинність. Розрахунки показали, що вибіркоче середнє значення цієї випадкової величини становить 11,15, а середньоквадратичне відхилення дорівнює 5,008.

Аналіз середніх по кожному з виділених кластерів показує, що для кластерів: A1, A3, A2, A4, A8, A5 ці їх середньо-групові значення: 22,1; 17,9; 16,4; 27,3; 12,9; 14,8; більші за 11,15, а для інших 5-ти кластерів менші, при цьому у першу групу кластерів ввійшло всього 86 країн ( $1+18+7+4+34+22=86$ ), а в іншу групу з 5-кластерів ( A10, A7, A11, A6, A9) 114 країн, з Україною в тому числі.

В той же час аналіз показує, що по цьому показнику навіть у кластері A10 є 5 країн, а саме: Естонія, Чилі, Угорщина, Аргентина, Португалія, для яких цей 7-показник вищий навіть за середній по всій вибірці, а не лише за середній по цьому кластеру 7,99, в той же час Україна має значення (5!) цього показника, що менше середньогрупового.

Аналогічний детальний порівняльний аналіз легко зробити і для інших кластерів, як стосовно цього показника, так і інших, використовуючи фрагмент результуючої таблиці середніх значень всіх по всіх кластерах, що показаний на рис.5. Але звертаємо тут увагу на той факт, що перші 6 показників, з урахуванням емпіричних даних статистичної вибірки та їх аналізу мають закони розподілу близькі до гіперболічного закону Ципфа чи йому подібному Бредфорда, які широко використовуються для моделювання та системного вивчення інформаційно-бібліотечних процесів, бібліометричних законів, але тому в той же час є незручними та мало використовуваними для їх застосування в процесах кластерного розбиття методами традиційної статистики.

Наші дослідження показують, що і відносний 8-ий показник, має майже гіперболічний закон розподілу, дивись рис.4 (праворуч), і тому, з урахуванням цих особливостей та різноманітності законів розподілу абсолютних та відносних показників, вибір саме НММ та інтелектуальних технологій для кластеризації є обґрунтованим.

Крім того, як було показано в [8], регресійні моделі по точності та зручності суттєво уступають вибраним інструментам.

Відмітимо, що найбільшу кількість документів на тисячу населення, а саме 84 (перевищує середнє по всіх країнах в  $11,5=84:7,29$  раз !), має Південна Корея, і лише для кластерів A3, A1, A2, A10 цей показник, відповідно середньо-групові: 40, 27, 25, 9, вище за середньо вибіркове значення 7,29. Хоч Україна і попала в кластер A10 цей показник для неї склав всього 3 документа на тисячу, тобто ми суттєво відстаємо. Однією з причин є те, що Україна має невраховані у базах неангломовні публікації. З іншої сторони, навіть такі країни з A10 як: Польща, Чехія (20), Словенія (33!!), Естонія (20!), Португалія, Туреччина та інші мають на порядок кращий цей показник.

212											
213	середні значення вибірки	172 129	162 155	2 454 022	787 138	152	35 639	11	7	94	13
214	вниз середні по кластерах										
215	A1	8 626 193	7 876 234	177 434 935	83 777 658	1 648	321 369	23	27	91	47
216	A11	2 307 950	2 257 142	13 049 752	6 435 573	439	1 310 178	6	2	97	45
217	A2	1 439 678	1 352 490	22 005 038	5 231 040	741	68 089	16	25	93	23
218	A3	282 994	265 435	4 760 553	731 220	419	7 035	18	40	93	15
219	A10	196 627	189 995	1 536 625	389 084	224	55 384	8	9	96	24
220	A9	19 066	18 362	133 919	21 467	87	29 757	7	2	96	15
221	A7	4 015	3 729	32 490	3 499	53	4 996	9	3	93	8
222	A5	4 414	4 089	59 019	7 591	70	21 978	15	1	92	9
223	A8	2 255	2 155	30 301	2 985	54	8 503	13	2	96	8
224	A4	1 850	1 734	46 847	4 541	84	1 855	27	3	93	11
225	A6	493	411	4 273	277	27	7 967	9	2	82	6
226											
227		Documents	Citable documents	Citations	Self-Citations	H index	Population (in thousands)	V_Citations per Document	Number of documents per 1000 population	Fraction Citations	Fraction Self-Citations

Рис. 5. Фрагмент вікна з середньо-груповими для кластерів (груп) значеннями показників

Цікаві висновки можна зробити із аналізу показників: частка цитованих документів (9-ий) та частка само-цитувань (10-ий), які, як показують наші дослідження, мають близькі до нормальних закони їх розподілу. Значення середні по всій вибірці склали: приблизно 94% для частки цитованих з середнім квадратичним відхиленням 3,35% та 13,17% з сігма в 7,84% для частки само-цитувань. Характерним є високі значення середньо-групових 9-ого, а особливо 10-го для лідируючих кластерів (аж 47% для A1, 45% для A11!) показників.

**Висновок.** Проведені дослідження показали, що лише кластерний аналіз на основі розширеного списку відносних індикаторів та використання інтелектуальних технологій групування і нейропакетів дозволяють отримати більш точні оцінки сумарного рейтингу країн стосовно їх науково-публікаційної діяльності.

### Список використаних джерел

1. Scimago Journal & Country Rank (SJR) [Електронний ресурс]. - Режим доступу: <http://www.scimagojr.com/countryrank.php>
2. Бібліометрика Української науки [Електронний ресурс]. - Режим



доступу: [http://nbuviar.gov.ua/bpnu/index.php?page\\_sites=formy](http://nbuviar.gov.ua/bpnu/index.php?page_sites=formy)

3. Круглов, В.В. Искусственные нейронные сети. Теория и практика / В.В. Круглов, В.В. Борисов.– М.: Горячая линия - Телеком, 2002.– 328 с
4. Красиленко, В.Г. Кластерний аналіз науково-дослідницького потенціалу ВНЗ як основи при визначенні їх рейтингів [Текст] / В.Г. Красиленко, Д.В. Нікітович, Л.М. Грущенко // Тринадцята Всеукраїнська наукова Інтернет-конференція. – Тернопіль: Тайп, 2012.– С. 21 – 25.
5. Красиленко, В.Г. Інтелектуальний кластерний аналіз кращих 100 науковців України за показниками наукометричних баз даних / В.Г. Красиленко, Д.В. Нікітович // Наука і навчальний процес: науково-методичний збірник матеріалів НПК ВСЕІ Університету «Україна». – Вінниця, Вінницький соціально-економічний інститут Університету «Україна», 2013. - С. 101 – 104.
6. Красиленко, В.Г. Використання нейропакету Excel Neural Package для кластеризації джерельної бази для супроводу дисципліни «Інтелектуальні інформаційні системи та технології» / В.Г.Красиленко, Р.Ф.Лободзінська, А.В.Карпінська // Комп'ютерні технології: наука і освіта. – Доповіді 5 НПК.- К.: університет «Україна», 2010.-С.116-120.
7. Красиленко, В.Г. Інтелектуальний кластерний аналіз наукометричних показників наукової діяльності кращих науковців планети з нейротематики за даними Google Scholar / В.Г. Красиленко, Д.В. Нікітович // Актуальні питання документознавства та інформаційної діяльності: теорії та іновації: матеріали I МПК. – Одеса: ОНПУ, 2015. - С. 275-280.
8. Красиленко, В.Г. Визначення впливовості факторів на інтегральну рейтингову оцінку кращих науковців світу з нейротематики на основі статистичних даних наукометричної бази Google Scholar / В.Г. Красиленко, Д.В. Нікітович // Актуальні питання документознавства та інформаційної діяльності: теорії та іновації: Матеріали I МПК. – Одеса: ОНПУ, 2015. - С. 333-338.

*Представлены результаты интеллектуального кластерного анализа 200 стран мира на основе их наукометрических показателей в системе Scimago Journal & Country Rank.*

**Ключевые слова:** *интеллектуальный кластерный анализ, наукометрическая база данных, индекс цитирования, рейтинговое оценивание, ранжирование.*

*The article presents the results of cluster analysis of 200 country based on sciento-metric indicators in Scimago Journal & Country Rank.*

**Keywords:** *data mining, cluster analysis, scientometric database.*

УДК 002.6+330

К.С. Корнійчук

## ГАЛУЗЕВЕ ТЛУМАЧЕННЯ СУТНОСТІ ПОНЯТТЯ «ЕЛЕКТРОННИЙ ДОКУМЕНТООБІГ НА ПІДПРИЄМСТВІ» У СУЧАСНИХ ПУБЛІКАЦІЯХ: ТЕРМІНОЛОГІЧНИЙ АНАЛІЗ

*Розглянуто підходи дослідників організації електронного документообігу на підприємствах до визначення його сутності. Охарактеризовано особливості практичного застосування систем електронного документообігу на підприємствах.*

**Ключові слова:** *електронний документообіг, документування, підприємство, автоматизація документообігу.*

Тривалий час понятійно-категоріальний апарат документообігу, зокрема традиційного «паперового», розвивався виключно в межах документознавства. Щодо техніко-технологічної основи електронного документообігу, то ці поняття створювалися та поширювалися в межах технічної інформатики (або інформаційних технологій). Проте нині відбувається широке входження електронного документообігу в різні сфери господарських та суспільних відносин у якості окремої функціональної системи підприємств, установ та організацій. Використання ними специфічних інструментів для роботи з інформацією, а також адаптація спеціальних програмних засобів до специфіки конкретної галузі зумовлюють виникнення вузькогалузевих тлумачень сутності електронного документообігу.

Вивченню електронного документообігу на підприємствах та в організаціях присвячено останні публікації К.В. Безверхого, Л.М. Бойка та О.С. Форманчук, С.В. Домашенка, Л.М. Івашка, Т.А. Лугової та О.Є. Акімова, Л.М. Колечкіної та Р.А. Головащенко, Ж.В. Кудрицької, Г.В. Охріменко, М. Плешакової-Боровинської, Н.О. Пронь, В. Спрінсяна, О.В. Шевченко та інших. Віддаючи належне внеску попередників відзначимо, що у їх працях головне місце відводиться організації та впровадженню систем електронного документообігу, а до