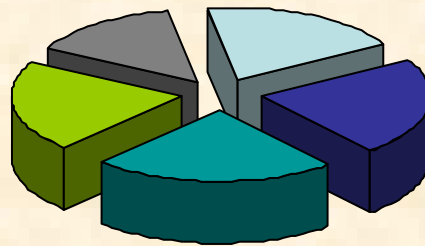


Treść Wikipedii

Porównania międzyjęzykowe

1. Przedmiot i zakres badań

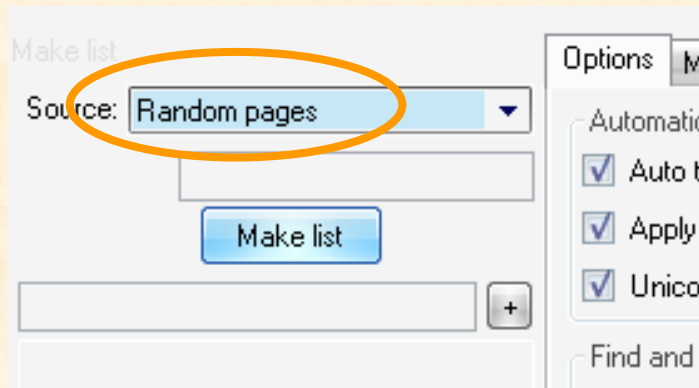
- Jimbo Wales:
- “Wyobraź sobie, że każda osoba na Ziemi mogłaby dzielić się wolnym i pełnym dostępem do ludzkiej wiedzy.”
- **Kwestia:** Jaka jest struktura ludzkiej wiedzy?
- Sprobujemy zanalizować to na Wikipedia



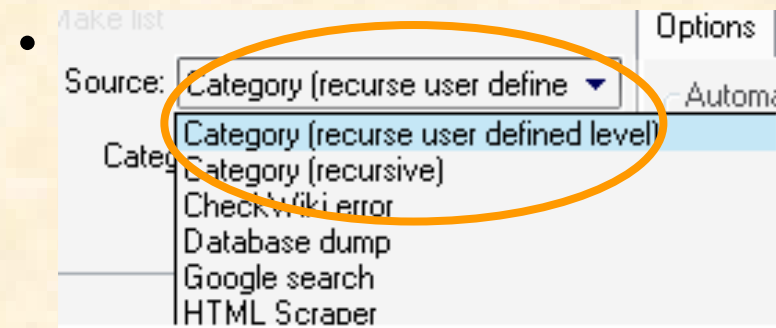
dzedziny wiedzy

2. Metodologia

- Wybór losowy
- Np. Przez AWB



- Analiza rekurencyjna kategorii



- Lub ScanCat

CatScan V2.0β by Magnus Manske [\[Керівництво\]](#) [\[Переклад інтерфейсу\]](#)
Мова інтерфейсу: [\[CS\]](#) [\[DA\]](#) [\[DE\]](#) [\[EL\]](#) [\[EN\]](#) [\[ES\]](#) [\[FI\]](#) [\[FR\]](#)

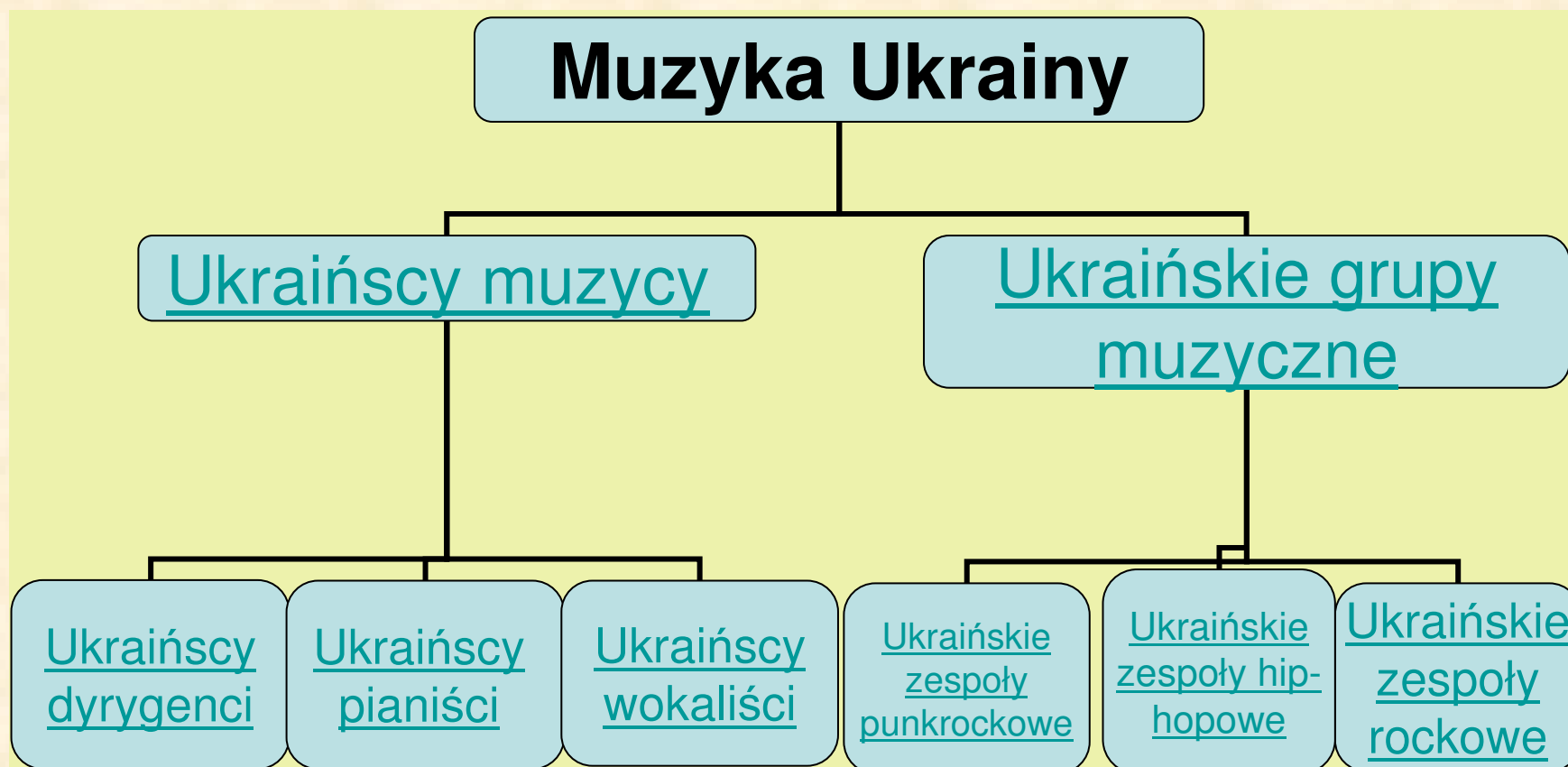
Мова	uk	
Проект	wikipedia	
Глибина	0	(глибина 0 = не шукати в підкатегоріях)
Категорії	Польща	

2.1 Dobór losowy

Próbka i Błąd	En	PI	Ru	Uk	Cs
100 art. 1%	35K	7,5K	6.5K	2,5K	1,9K
250 art. 0,4%	14K	3K	2,6K	1K	760
1000 art. 0,1%	3,5K	750	650	250	190

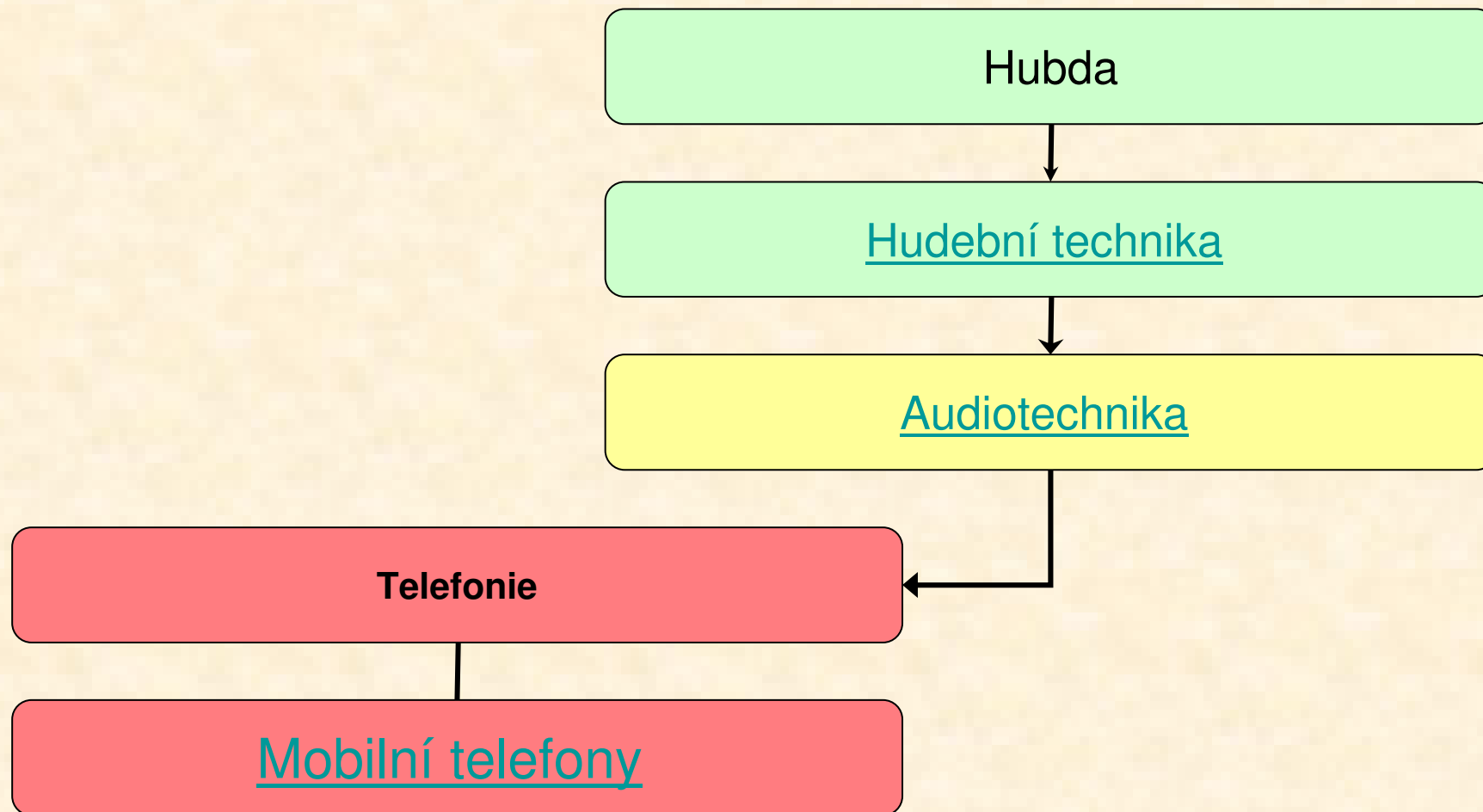
2.2 Analiza rekurencyjna – kategoryzacja tranzytywna, podzbiory wg

treści
polska Wikipedia, [Kategoria:Muzyka Ukrainy](#)



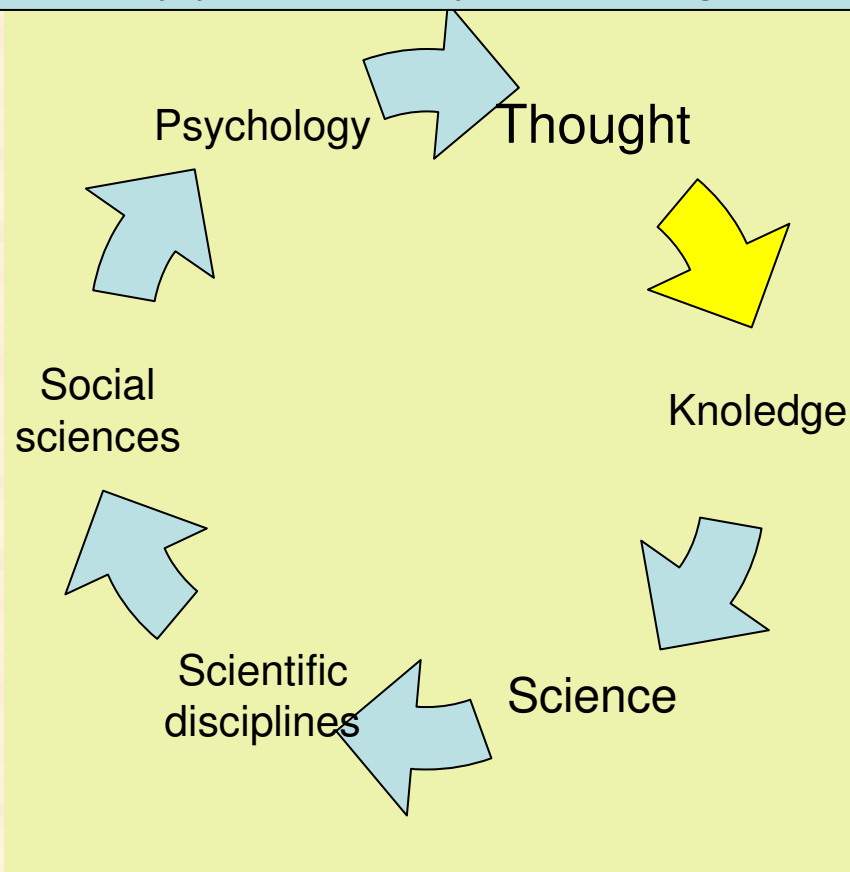
2.3 Analiza rekurencyjna – kategoryzacja tranzytywna – podzbiory nie wg treści

Wikipedia czeska, Kategoria:Hudba

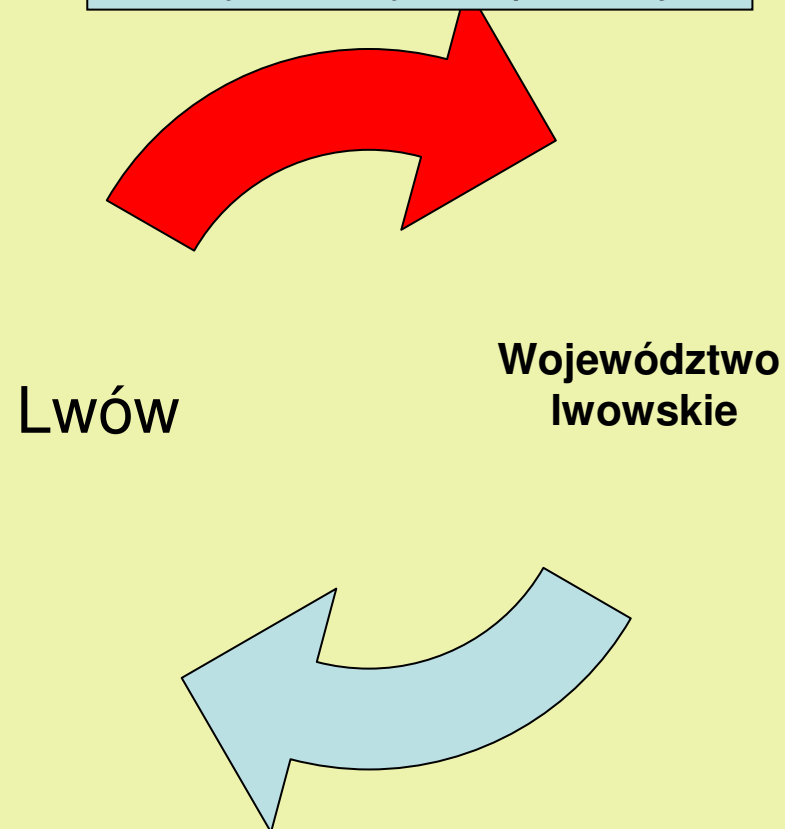


2.4 Analiza rekurencyjna – kategoryzacja nietranzytywna

Nietranzytywność klasyczna w angielskiej



Kiedyś tak było w polskiej...

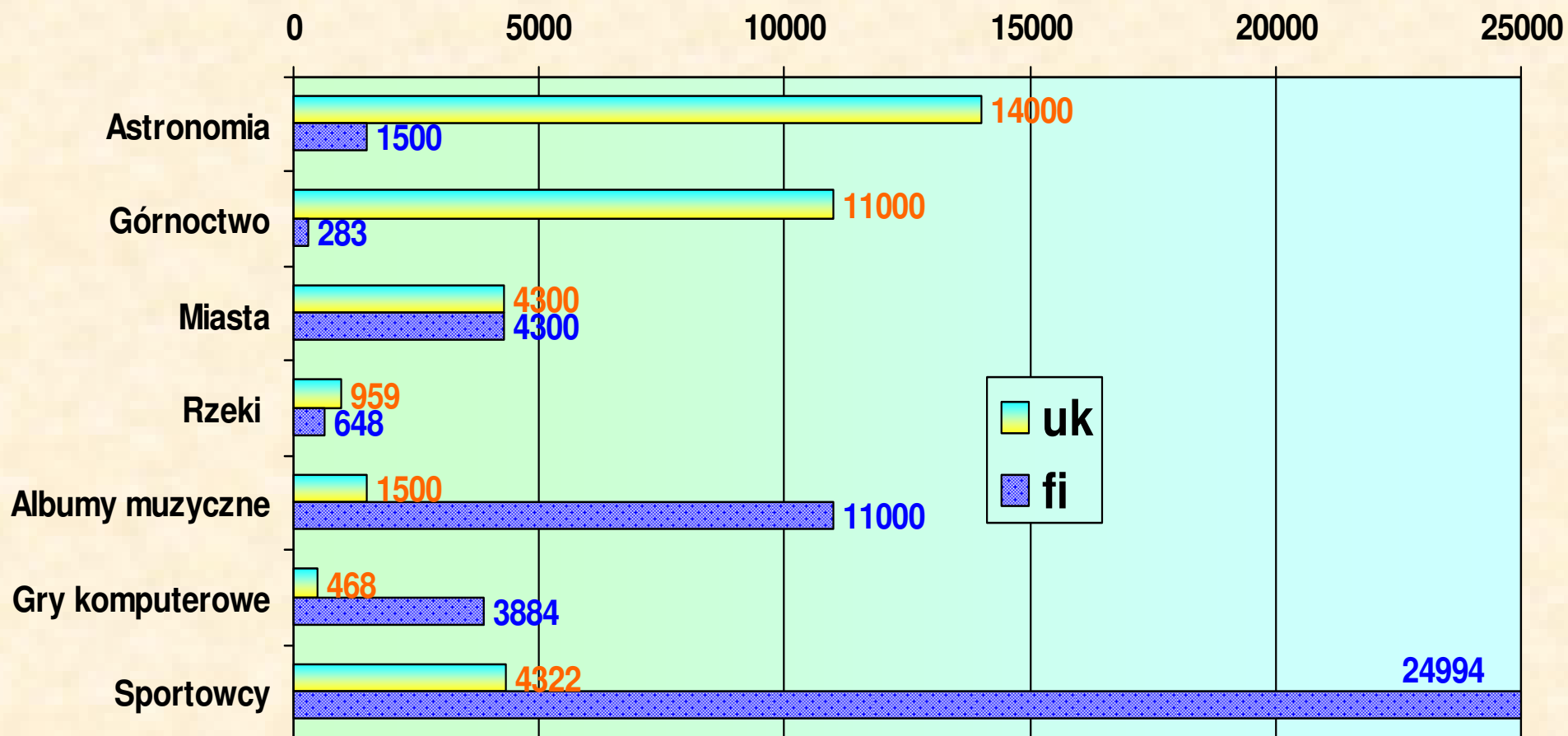


2.5 podsumowanie

- Wybór losowy – dla obszarów szerszych
- Analiza rekurencyjna – dla obszarów węższych

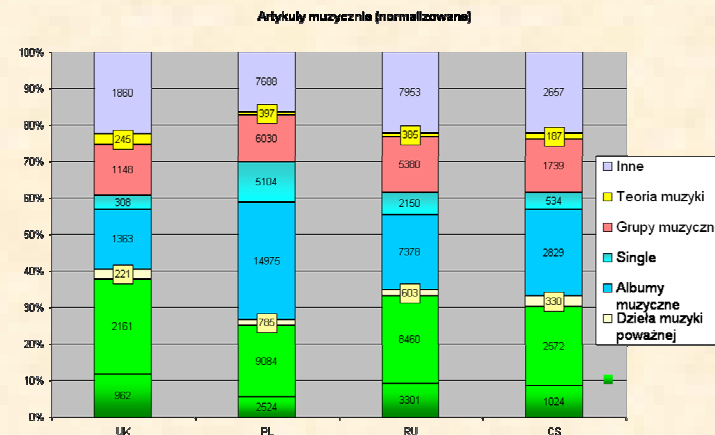
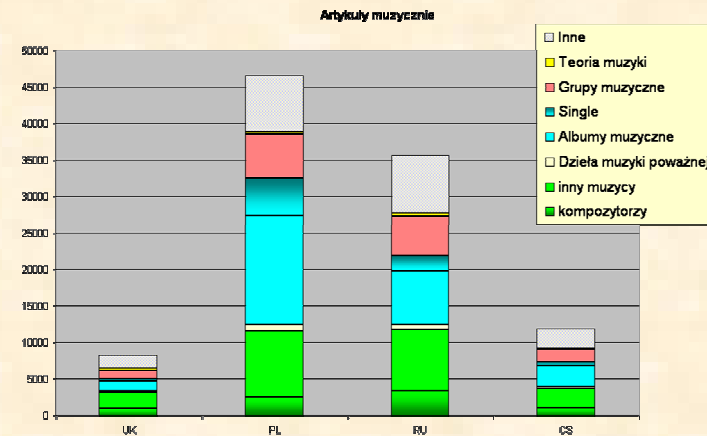
3.1. Badania specjalnie – Ukr vs Fin

Porównanie ukraińskiej i fińskiej Wikipedii w momencie wyprzedzania (~255K art.)



3.2. Badania specjalne – Muzyka

- Struktura artykułów o tematyce muzycznej:
 - Wykres słupkowy absolutny
 - Wykres słupkowy znormalizowany
 - Zielony - muzycy
 - Niebieski – albumy i single
 - Czerwony – grupy
 - Żółty – teoria muzyki
- ❖ UK (Українська)
- ❖ PL (Polska)
- ❖ RU (Русский)
- ❖ CS (Česky)



4.1. Badania ogólne

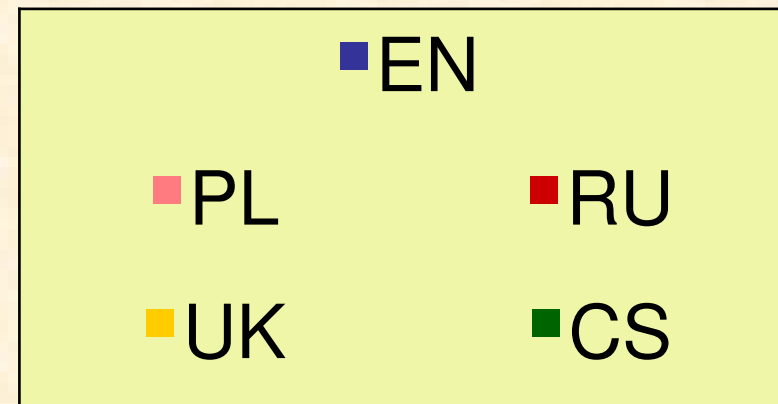
1. Robimy losowe próby

2. Analizujemy:

- Wg tematyki
- Wg typu

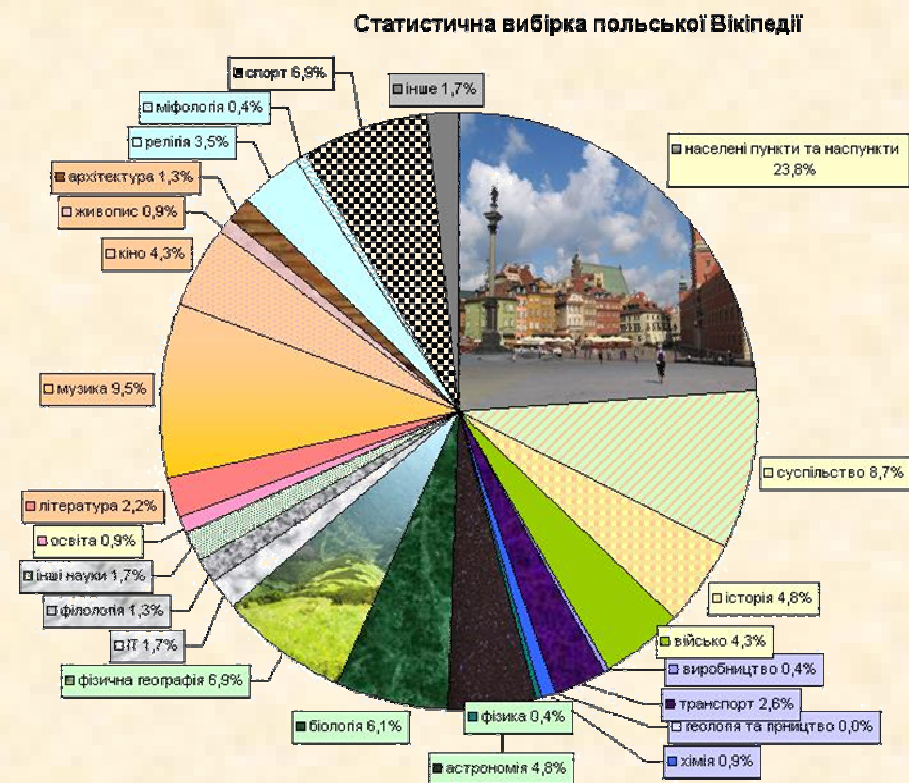
3. Robimy diagramy

1	стаття	тематика	тип предмета	повторюв	повторюв	зведена таблиця			зведена таблиця		
2	Pałac Michny z Vácinoва	архітектура	об'єкт	22	3	тематика	знайдено	частота	тематика	знайдено	частота
3	Vtřaz	архітектура	поняття	51	3	соц. географія	27	населені	14.4%	поняття	51
4	Alfred Mosher Butts	архітектура	людина	41	3	суспільство	10	суспільств	5.3%	об'єкт	22
5	Sojuz TMA-17	астрономія	модель	8	4	історія	14	історія	7.4%	людина	41
6	Messier 67	астрономія	об'єкт	22	4	військо	2	військо	1.1%	організація	9
7	Německé středisko pro letectv	астрономія	організація	9	4	промисловість	2	виробницт	1.1%	населеній пу	19
8	Kap (měsíc)	астрономія	об'єкт	22	4	транспорт і зв'яз	6	транспорт	3.2%	адміністратив	7
9	Cardale Dabington	біологія	людина	41	13	геологія та гірн	2	геологія та гірн	1.1%	людя	6
10	Centrální dogma molekulární b	біологія	поняття	51	13	археологія	0	археологія	0.0%	персонаж	2
11	Kosatcec bahenní	біологія	поняття	51	13	фізика	3	фізика	1.6%	дисамбіг	9
12	Dynaktin	біологія	поняття	51	13	хімія	1	хімія	0.5%	модель	8
13	Homo antecessor	біологія	поняття	51	13	астрономія	4	астрономія	2.1%	список	6
14	Koryši	біологія	поняття	51	13	біологія	13	біологія	6.9%	огляд	1
15	Mitozom	біологія	поняття	51	13	медицина	3	медицина	1.6%	споруда	0
16	Památný strom	біологія	поняття	51	13	фіз. географія	14	фізична ге	7.4%	твор	11
17	Tegumentum	біологія	поняття	51	13	математика	5	математика	2.7%	видання	1
18	Dvouděložné	біологія	поняття	51	13	IT	10	IT	5.3%	інше	5
19	Sběrač rosný	біологія	поняття	51	13	філологія	2	філологія	1.1%		
20	Lewisuchus	біологія	поняття	51	13	освіта	1	освіта	0.5%		
21	Korandův smrk	біологія	об'єкт	22	13	література	8	література	4.3%		
22	Canadian War Museum	військо	об'єкт	22	2	музика	11	музика	5.9%		
23	Zahar Prilepin	військо	людина	41	2	кіно	8	кіно і ТВ	4.3%		
24	Hydrodynamická vertikální zon	геологія та гірницт	поняття	51	2	театр	1	театр	0.5%		
25	Richard Lydekker	геологія та гірницт	людина	41	2	журналіст	2	журналіст	1.1%		
26	1696	дата	список	6	3	архітектура	3	архітектура	1.6%		
27	1995	дата	список	6	3	грашки	3	грашки	1.6%		
28	Ochotnýj tjad	дисамбіг	дисамбіг	9	9	релігія	7	релігія	3.7%		
29	Horčícka	дисамбіг	дисамбіг	9	9	міфологія	0	міфологія	0.0%	усього (без д	200
30	Oběhání Brna	дисамбіг	дисамбіг	9	9	спорт	17	спорт	9.0%		
31	DeSoto County	дисамбіг	дисамбіг	9	9	інше	9	інше	4.8%		
32	Letka	дисамбіг	дисамбіг	9	9	усього (без д	186		100.00%		
33	П	дисамбіг	дисамбіг	9	9	дисамбіг	9				
34	Imogta	дисамбіг	дисамбіг	9	9	інше	3				

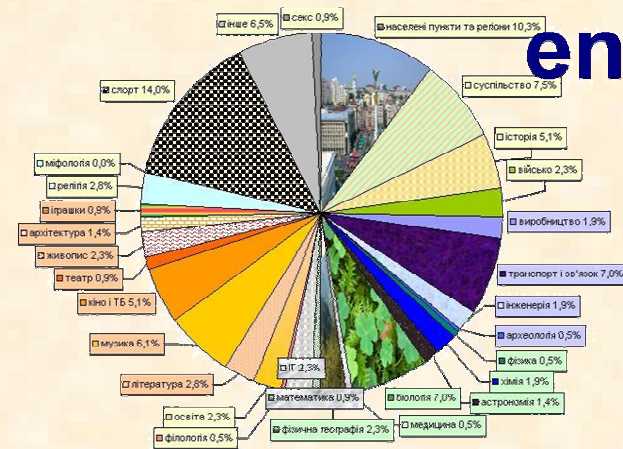
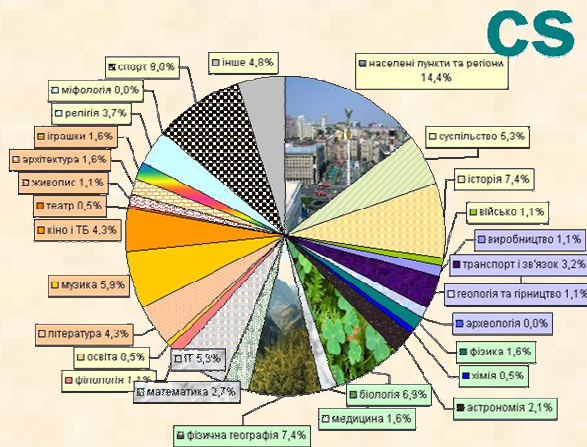
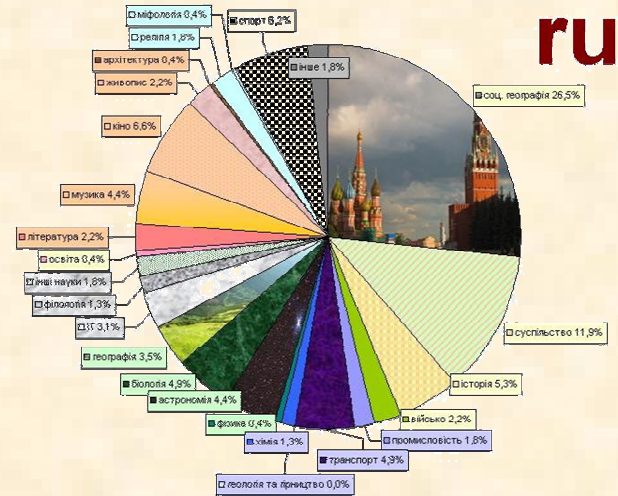
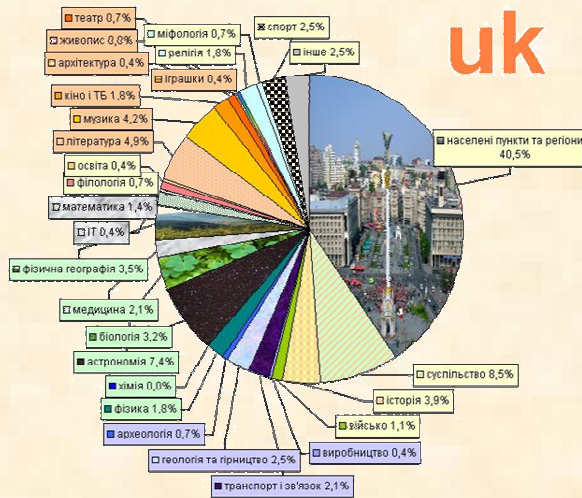


4.2 Analiza tematyczna polskiej Wikipedii

- OKOŁO:
- 24% - geografia społeczna
- 21% - społeczeństwo
- 21% - nauki przyrodnicze i IT
- 20% - sztuka
- 4% - religia i mitologia
- 7% - sport
- 3% - inne
- (oprócz “disambig”)

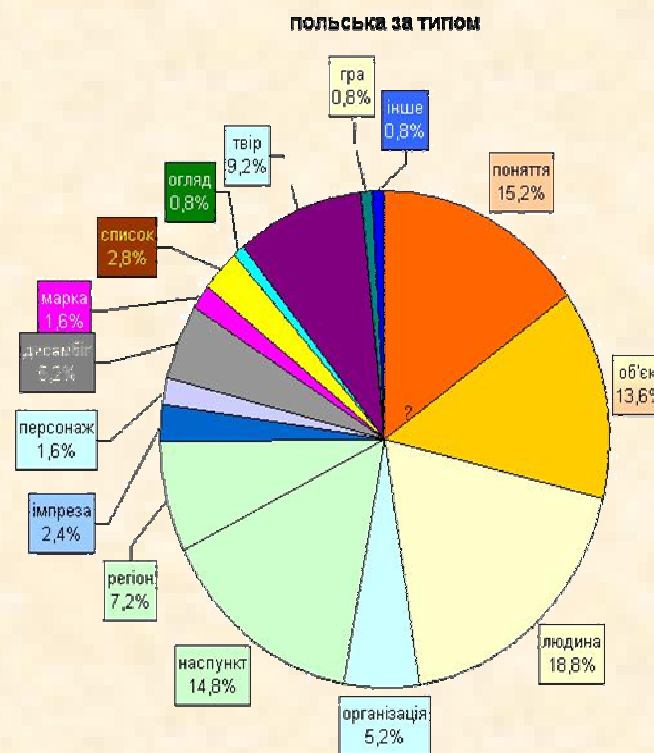


4.3 Analiza tematyczna innych Wikipedii

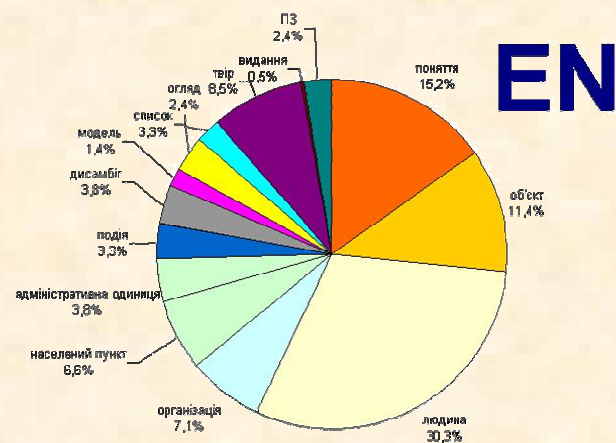
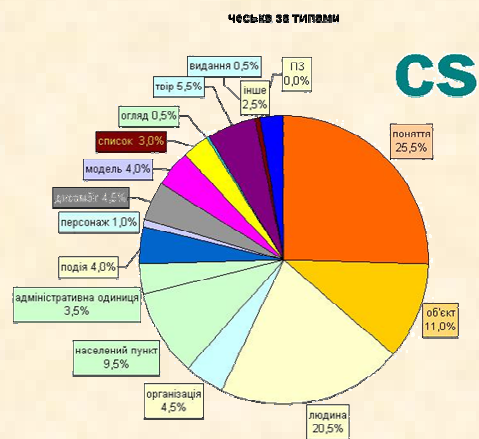
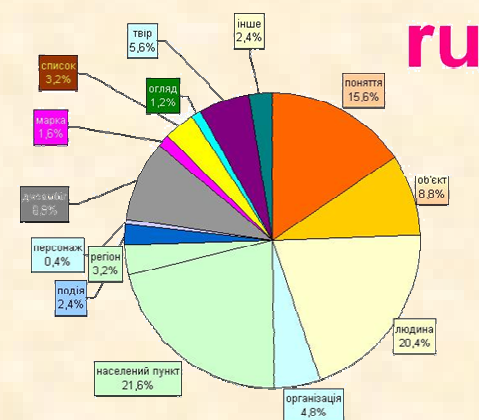
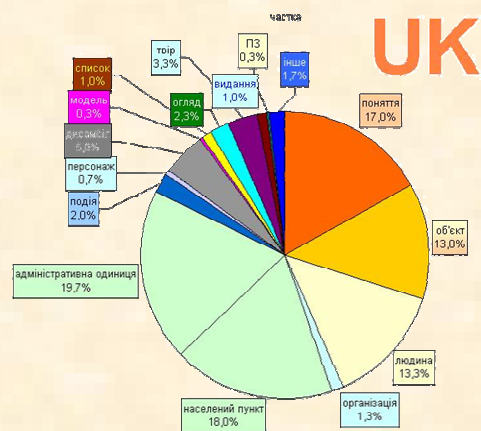


4.4 Analiza typologiczna polskiej Wikipedii

- Мѣдzy innymi:
 - 15% pojęcia
 - 14% obiekty
 - 19% biografii
 - 5% organizacji i kolektywy
 - 22% miejsca zamieszkane, gminy etc.
 - 2% wydarzenia
 - 5% disambig
 - 9% utwory sztuki

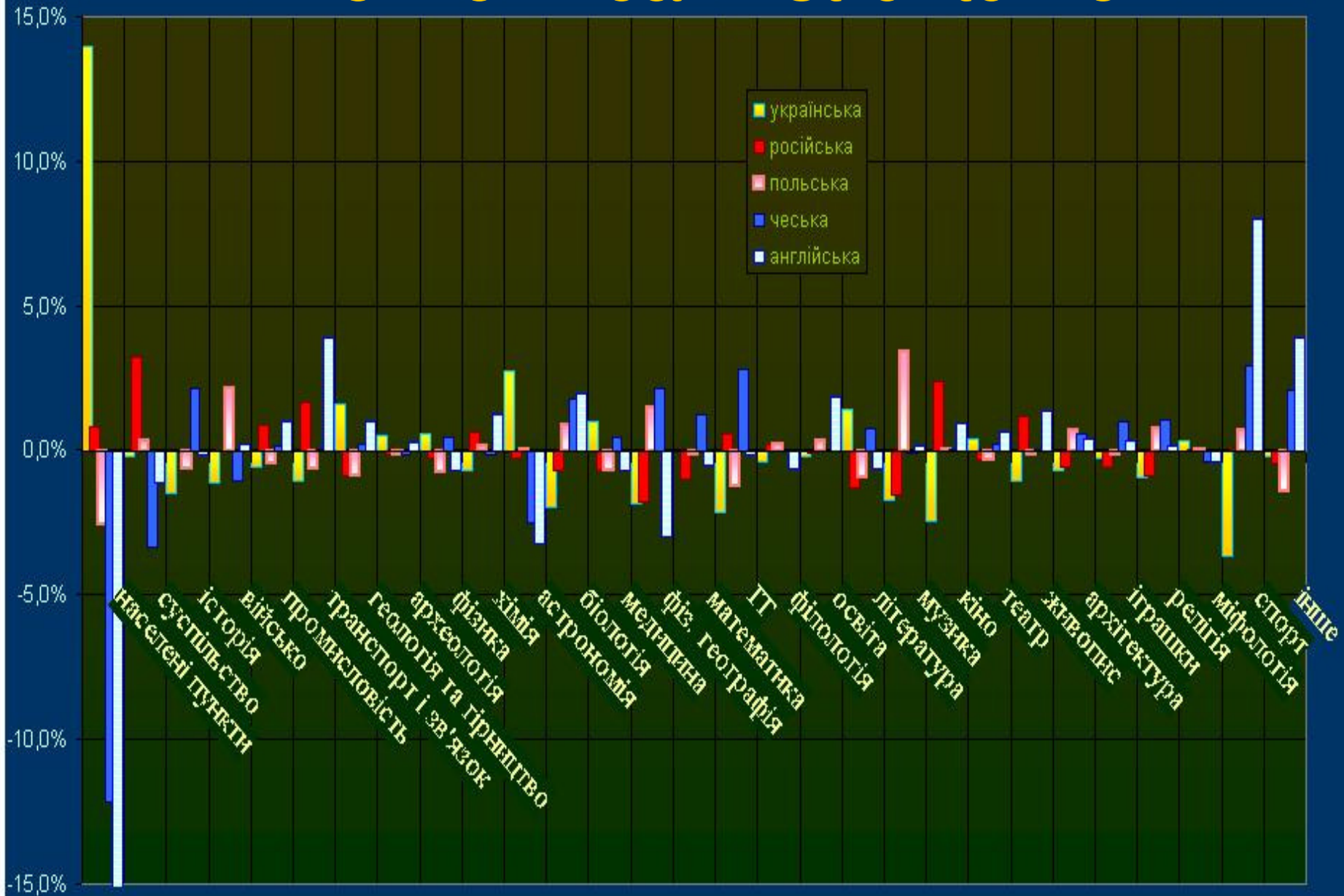


4.4 Analiza tematyczna innych Wikipedii



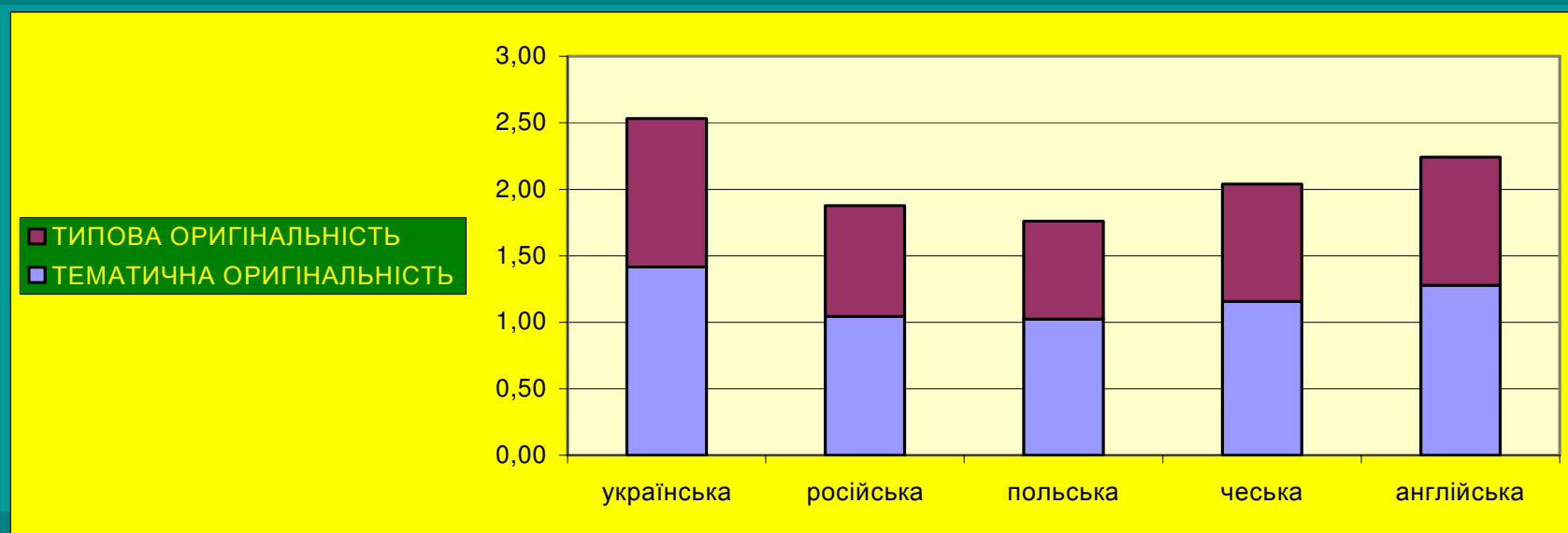
оригінальність вікіпедій

4.5 Różnica w strukturze



4.6 Oryginalność czyli universalność

	українська	російська	польська	чеська	англійська	Разом
ТЕМАТИЧНА ОРИГІНАЛЬНІСТЬ	1,42	1,05	1,02	1,16	1,28	5,92
ТИПОВА ОРИГІНАЛЬНІСТЬ	1,12	0,83	0,74	0,88	0,96	4,53
ПІДСУМКОВА ОРИГІНАЛЬНІСТЬ	2,53	1,88	1,76	2,04	2,24	10,46



	українська	російська	польська	чеська	англійська	разом
українська	0,00	0,53	0,51	0,69	0,80	2,53
російська	0,53	0,00	0,33	0,50	0,51	1,88
польська	0,51	0,33	0,00	0,42	0,50	1,76
чеська	0,69	0,50	0,42	0,00	0,42	2,04
англійська	0,80	0,51	0,50	0,42	0,00	2,24
разом	2,53	1,88	1,76	2,04	2,24	10,46

Dziękuję za uwagę

- Przygotował Andrij Bondarenko
- Konferencja Wikimedia Polska 2011