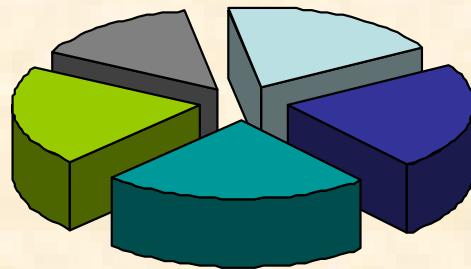


Зміст Вікіпедії

Досвід порівнянь

1. Предмет досліджень

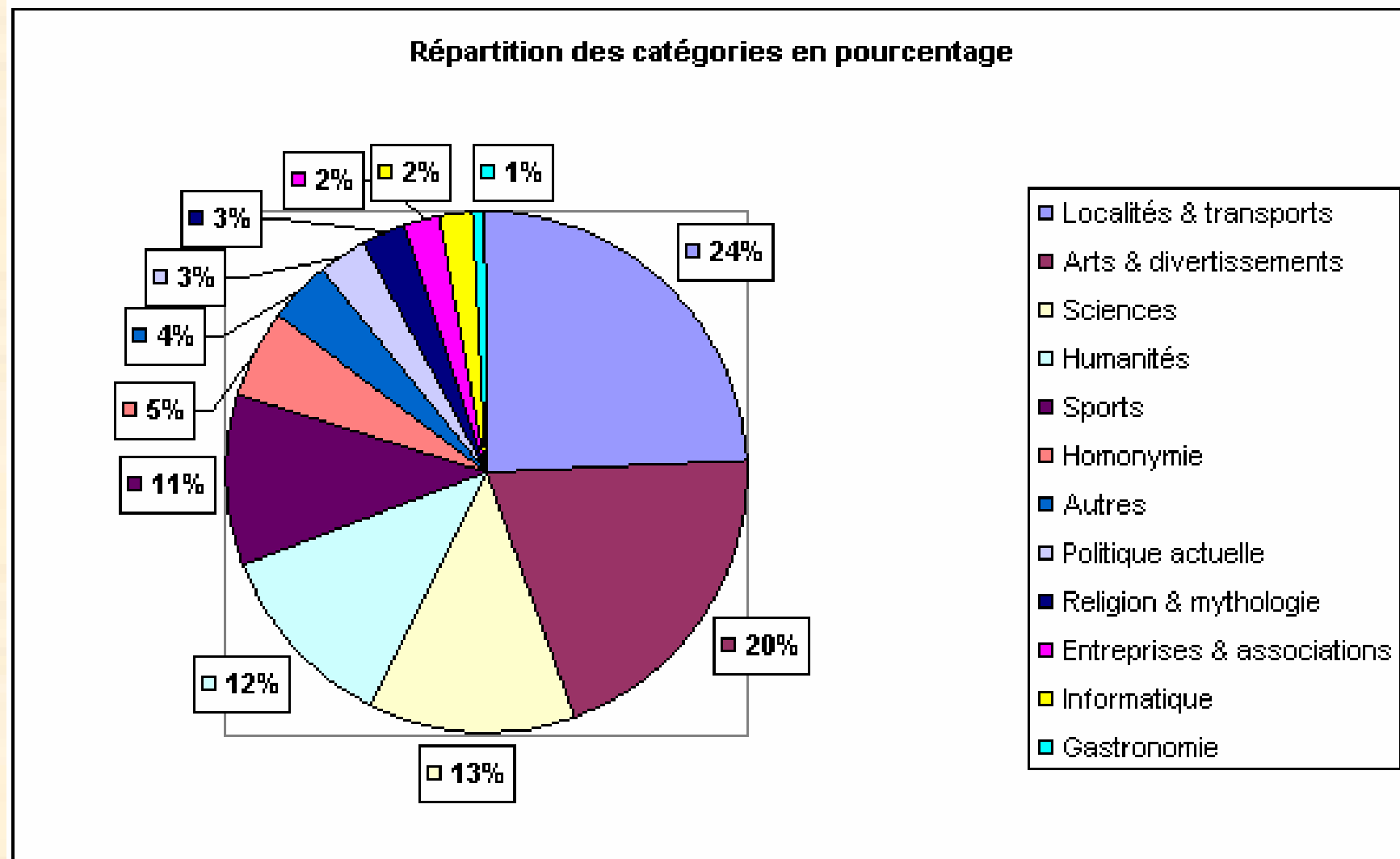
- Джимбо Вейлз:
- “Уявіть, Уявіть планету, на якій кожній окремій особі надано доступ до суми всіх знань людства..”
- **Питання:** Якою є структура знань людства?
- Спробуємо це проаналізувати



dzedziny wiedzy

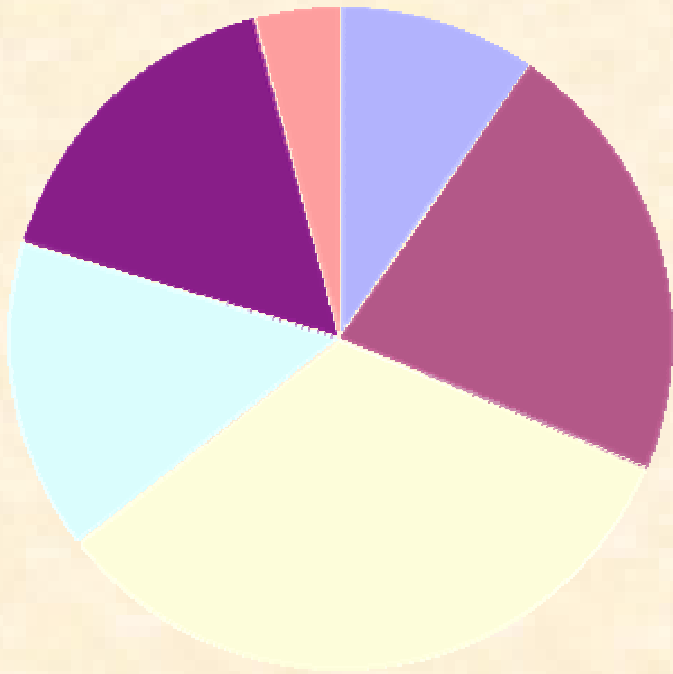
Історія питання

- Французька вікіпедія 2007



Історія питання

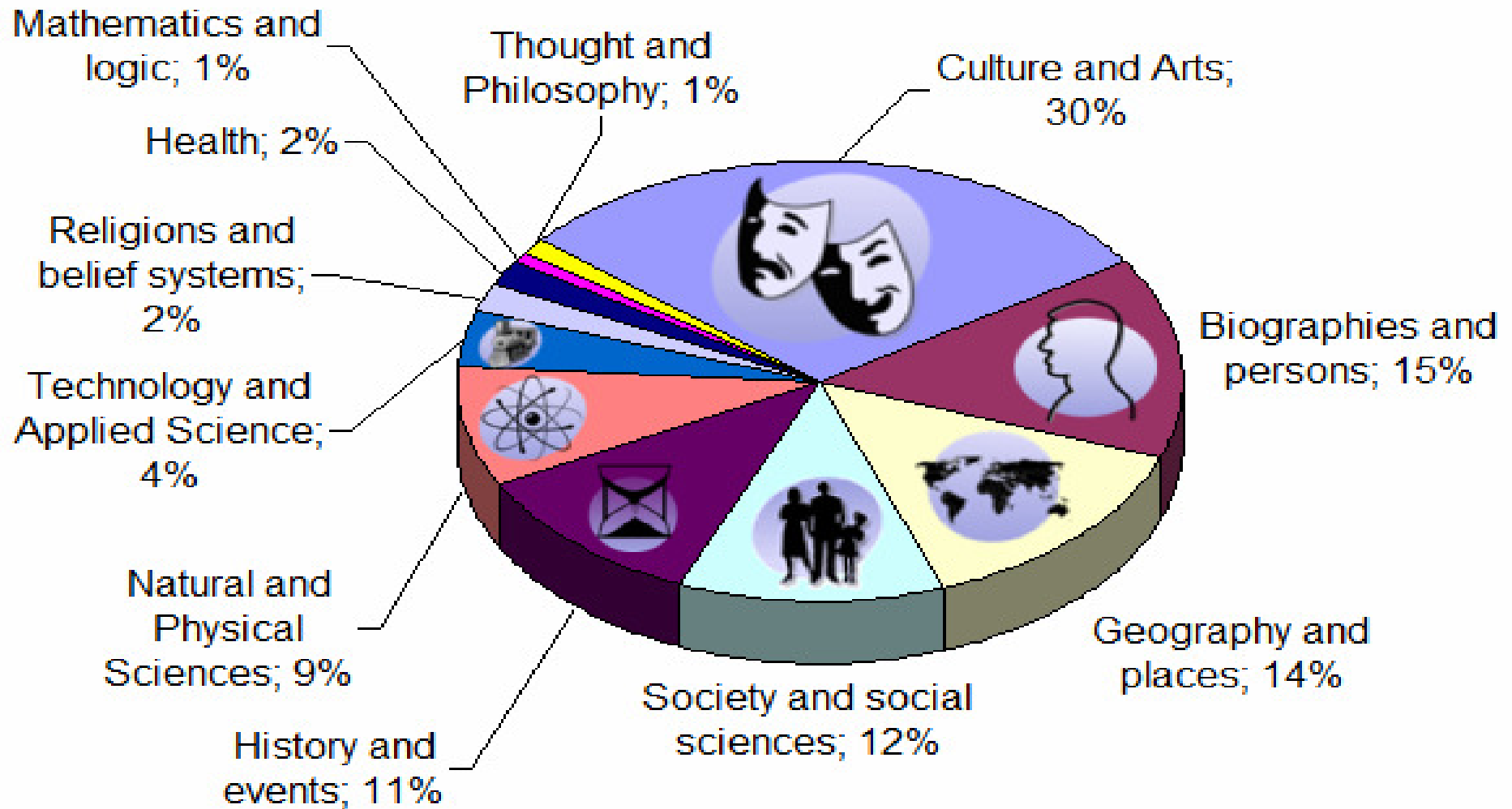
- Шведська вікіпедія 2007



- Географія - 31
- Історія – 68
- Культура – 104
- Природа – 48
- Суспільство – 51
- Техніка – 13

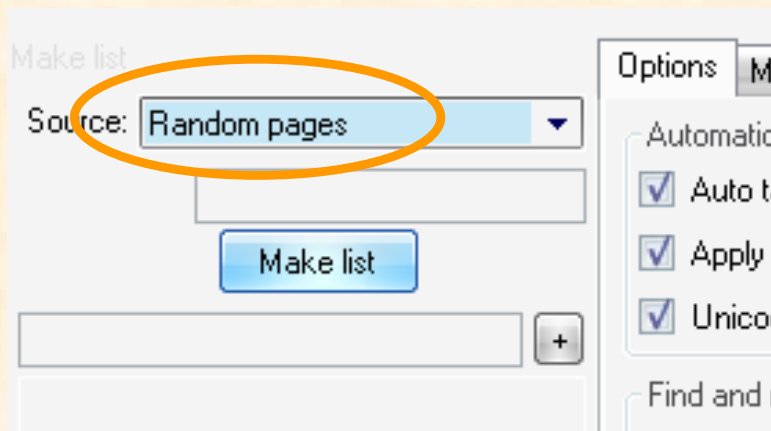
Історія питання

Англійська Вікіпедія, 2008

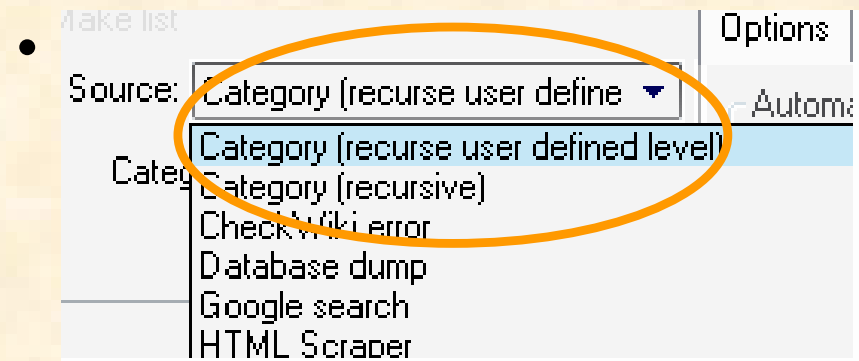


2. Методологія

- Статистична вибірка
- За допомогою AWB



- Рекурсивний аналіз категорії



- Або ScanCat

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

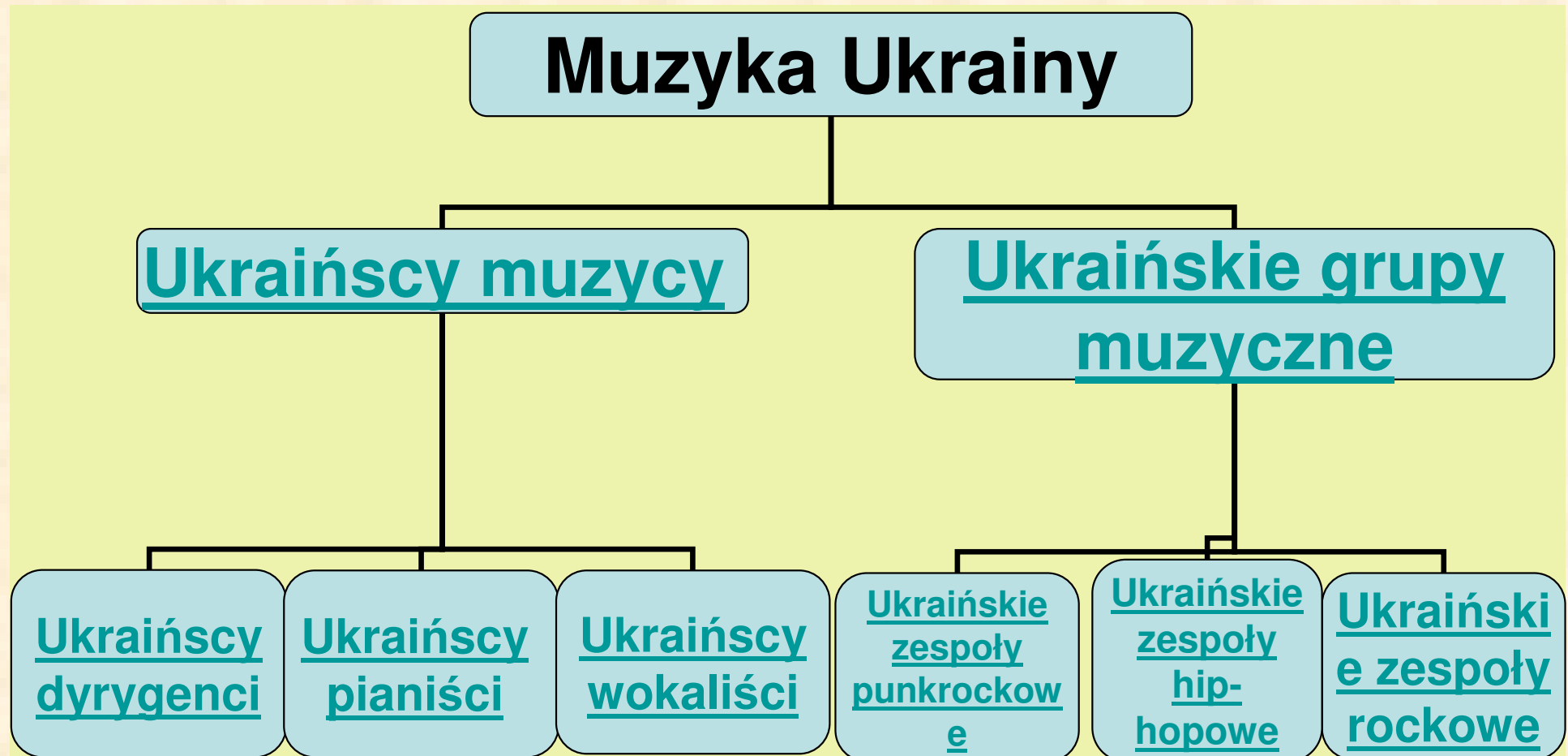
Мова	uk	
Проект	wikipedia	
Глибина	0	(глибина 0 = не шукати в підкатегоріях)
Категорій	Польща	
...		

2.1 Статистична вибірка

Вибірка і Похибка	En	Pl	Ru	Uk	Cs
100 статей 1%	3,5M+	800K+	750K+	300K+	200K
250 статей 0,4%	14K	3,2K	3K	1,2K	800
1000 статей 0,1%	3,5K	800	750	300	200

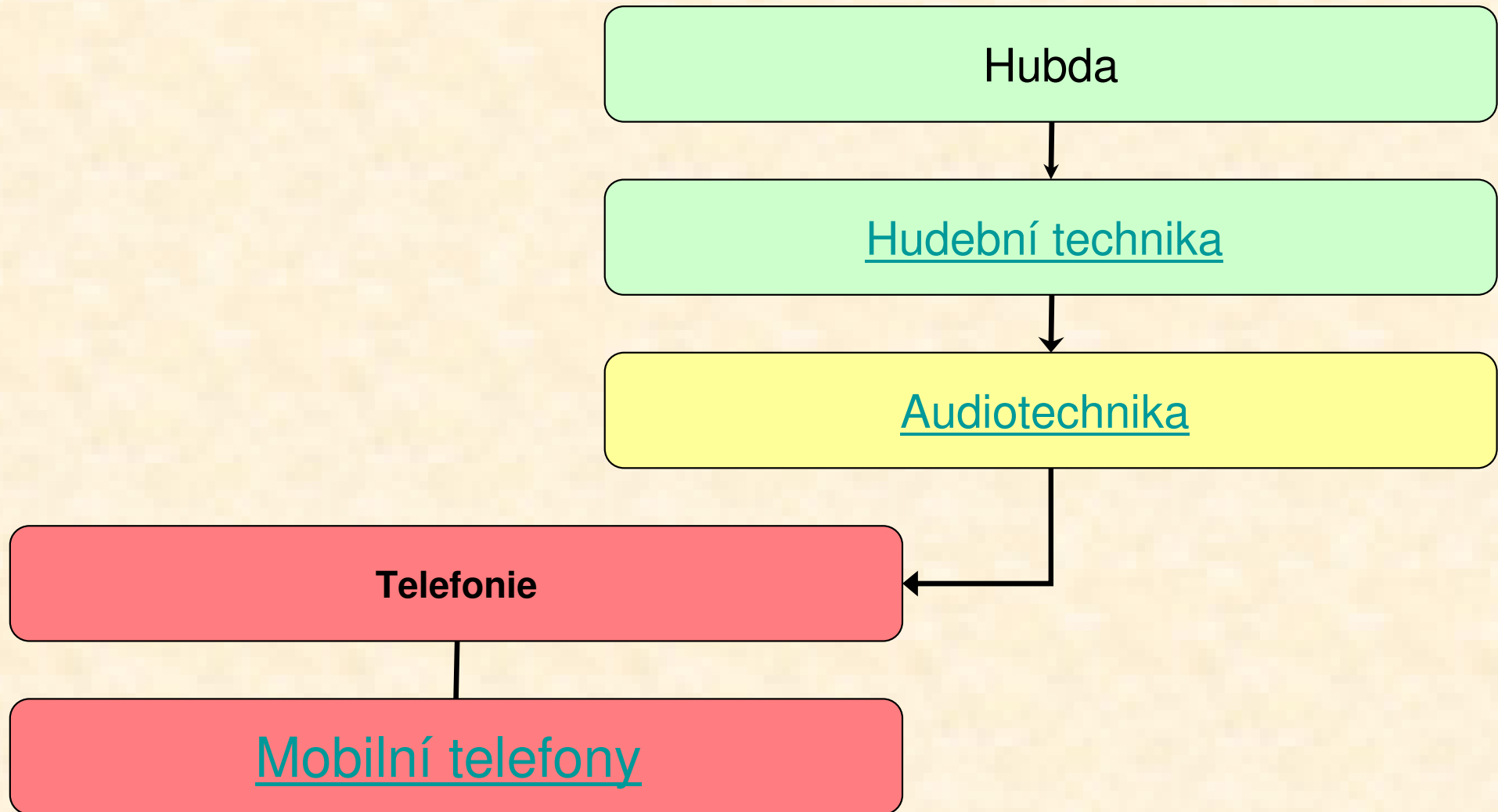
2.2 Рекурсивний аналіз – змістовна транзитивна категоризація

Wikipedia polska, Kategoria:Muzyka Ukrainy



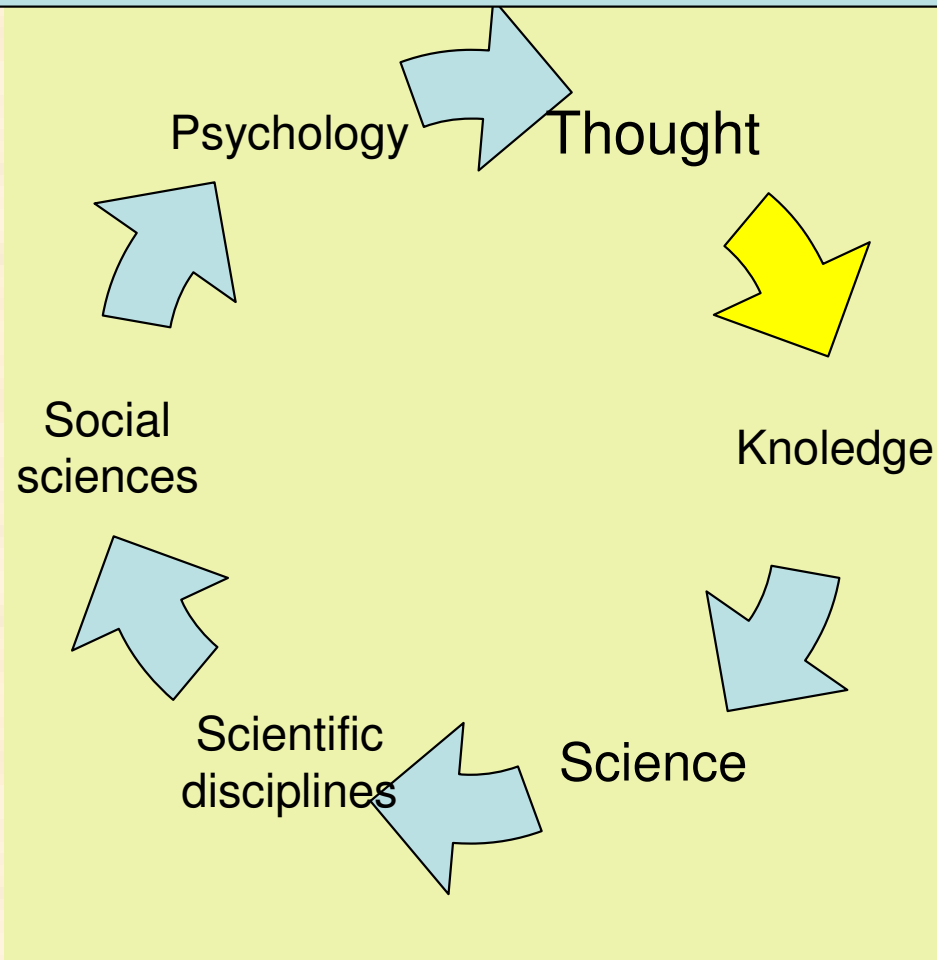
2.3 Рекурсивний аналіз – незмістовна транзитивна категоризація

Wikipedia czeska, Kategorie:Hudba

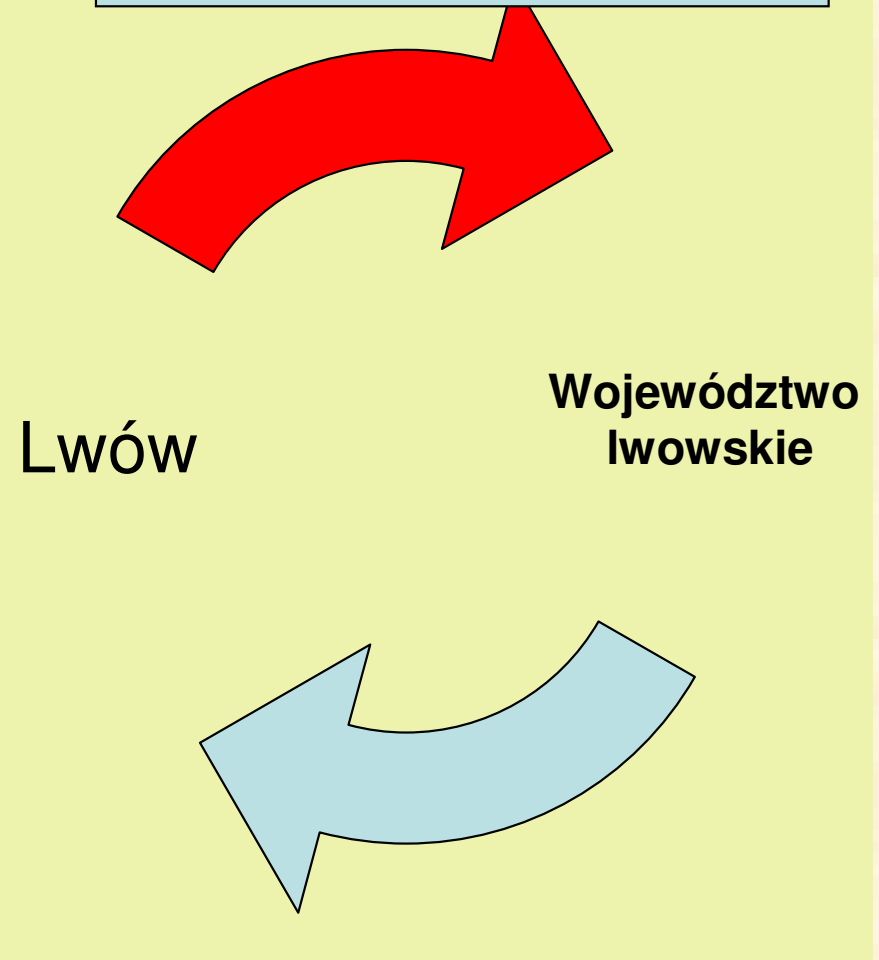


2.4 Рекурсивний аналіз – нетранзитивна категоризація

Класична нетранзитивність в англ. вікі



Колись було у польській...

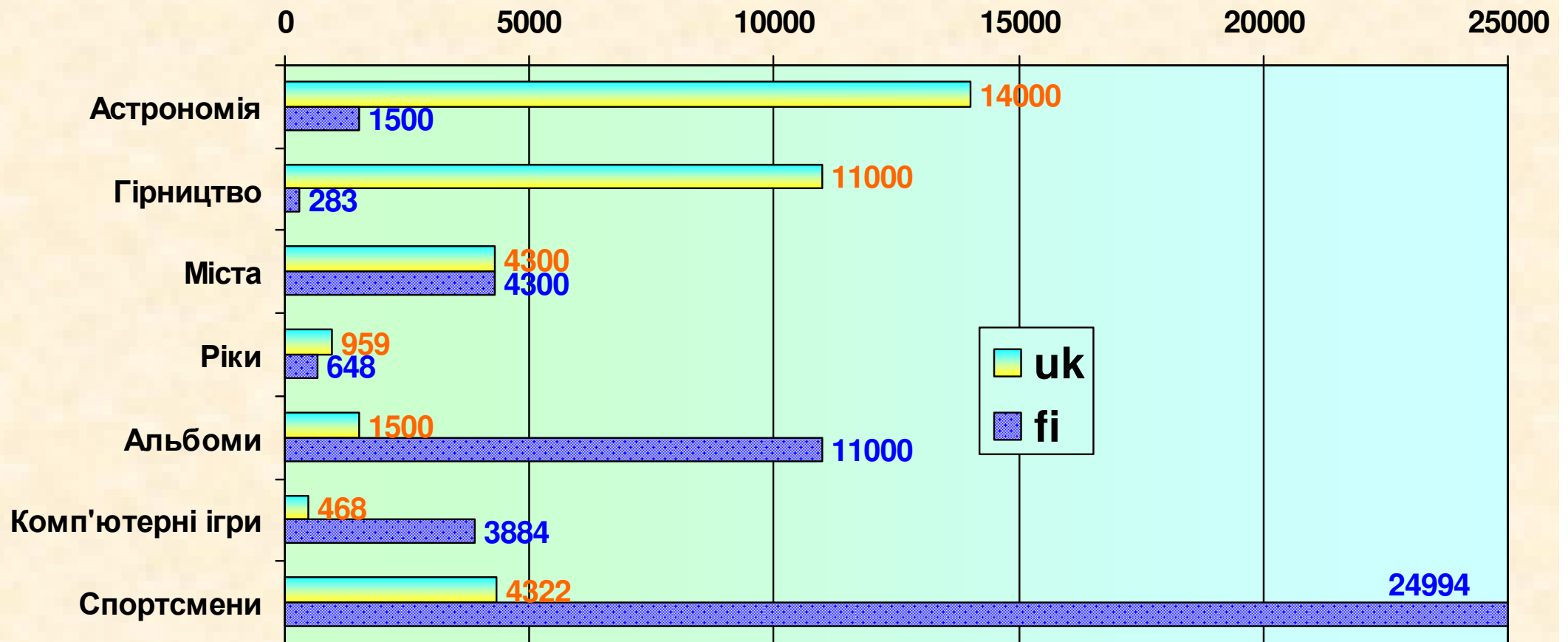


2.5 підсумок

- Статистична вибірка – для більших обсягів
- Рекурсивний аналіз – для вузьких обсягів

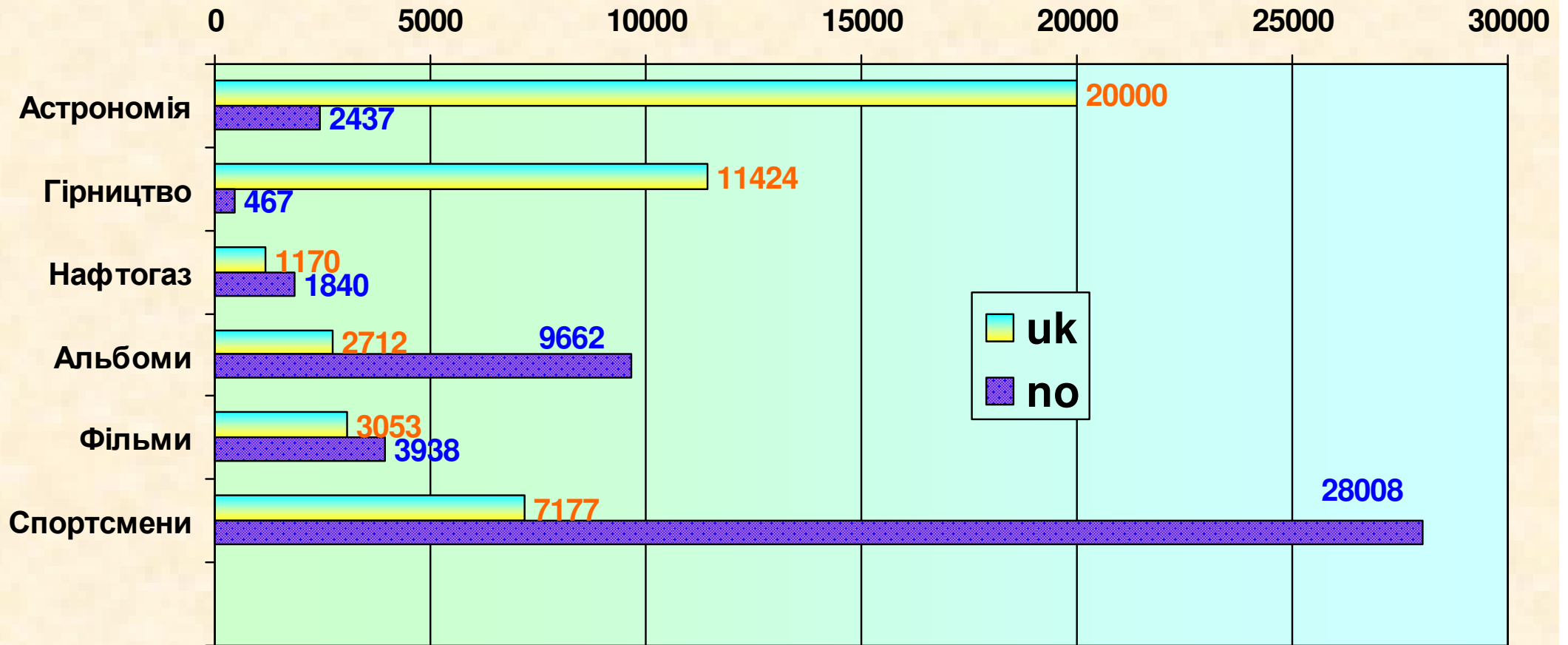
3.1. Спеціальні дослідження – Ukr vs Fin

Порівняння української та фінської вікіпедій на момент обгону
(~ 255К статей)



3.2. Спеціальні дослідження – Ukr vs Nor

Порівняння української та фінської вікіпедій на момент обгону
(~ 311К статей)

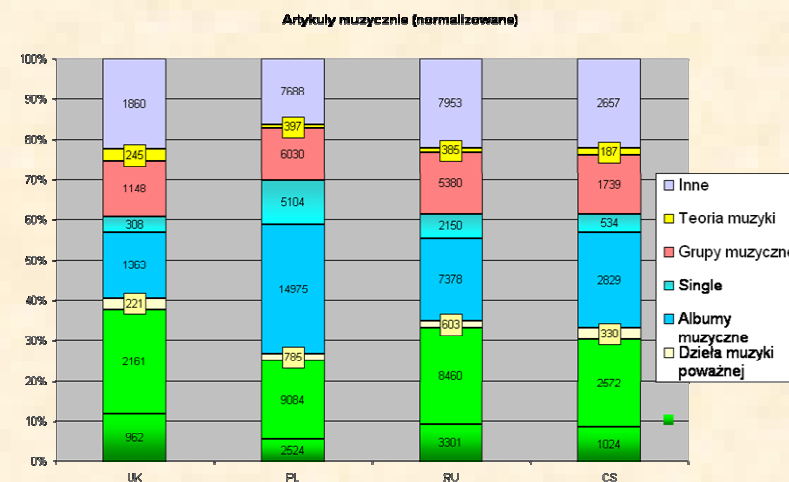
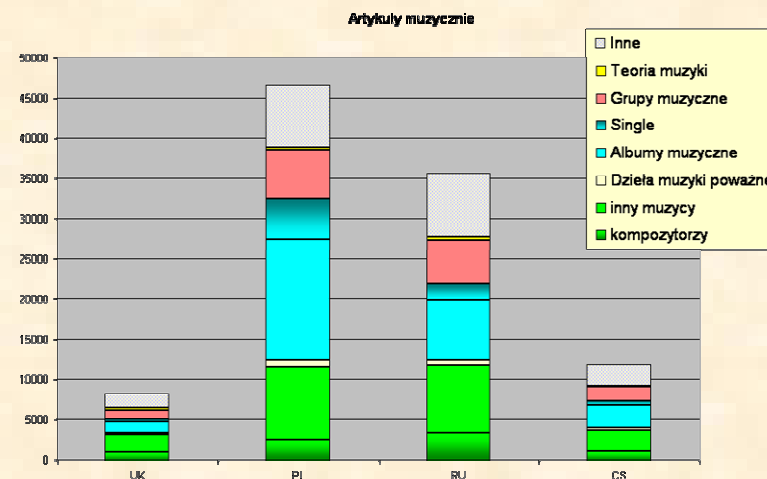


3.2. Спеціальні дослідження – Музика

- Кількість статей по музичних тематиках:

- Абсолютна діаграма
- Нормалізована діаграма
 - Зелений - музиканти
 - Блакитний – альбоми і сингли
 - Червоний – гурти
 - Жовтий – теорія музики

- ❖ UK (Українська)
- ❖ PL (Polska)
- ❖ RU (Русский)
- ❖ CS (Česky)



4.0. Загальні дослідження

1. Створюємо статистичну вибірку

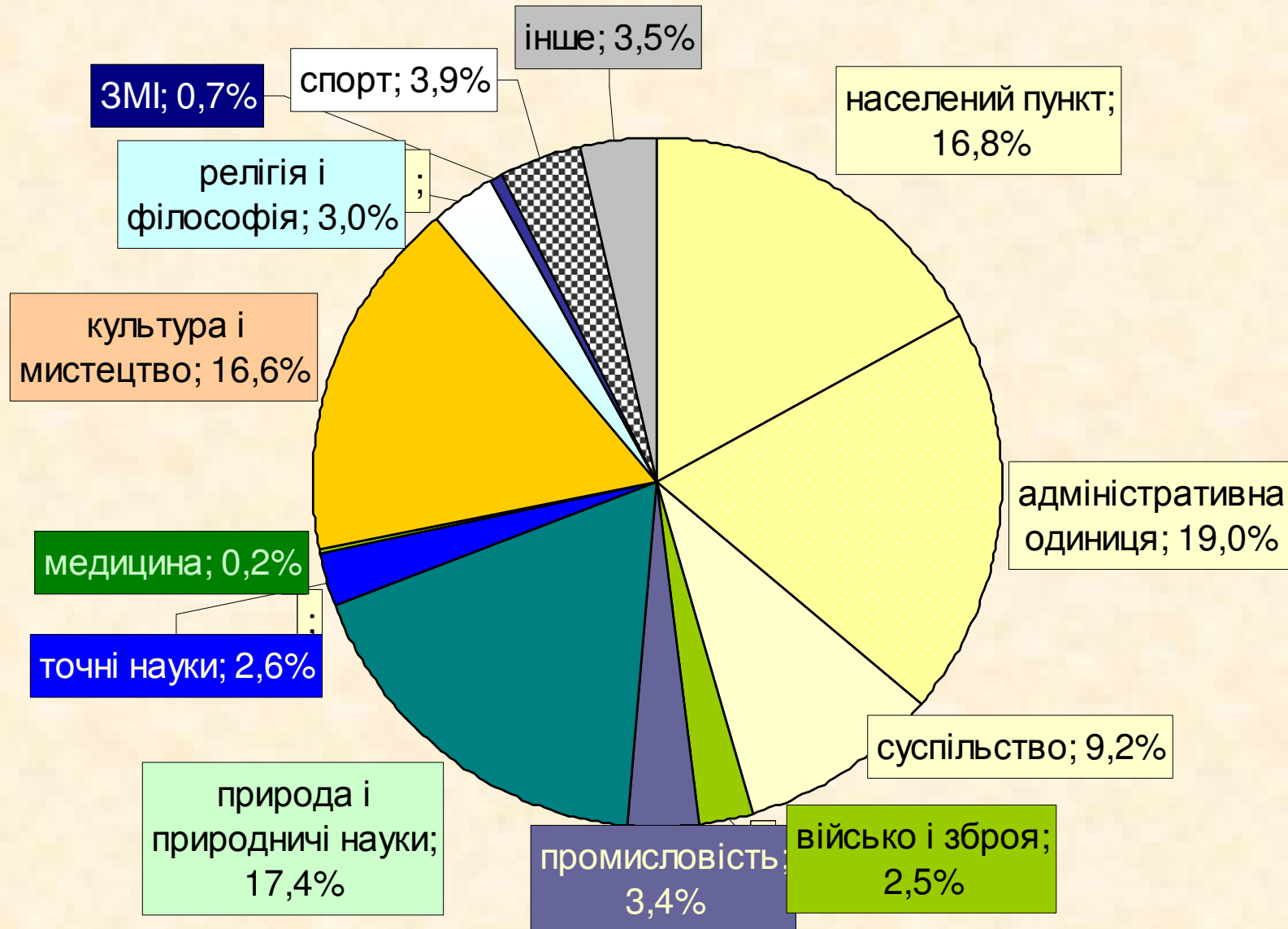
2. Аналізуємо:

- За темою
- За типом

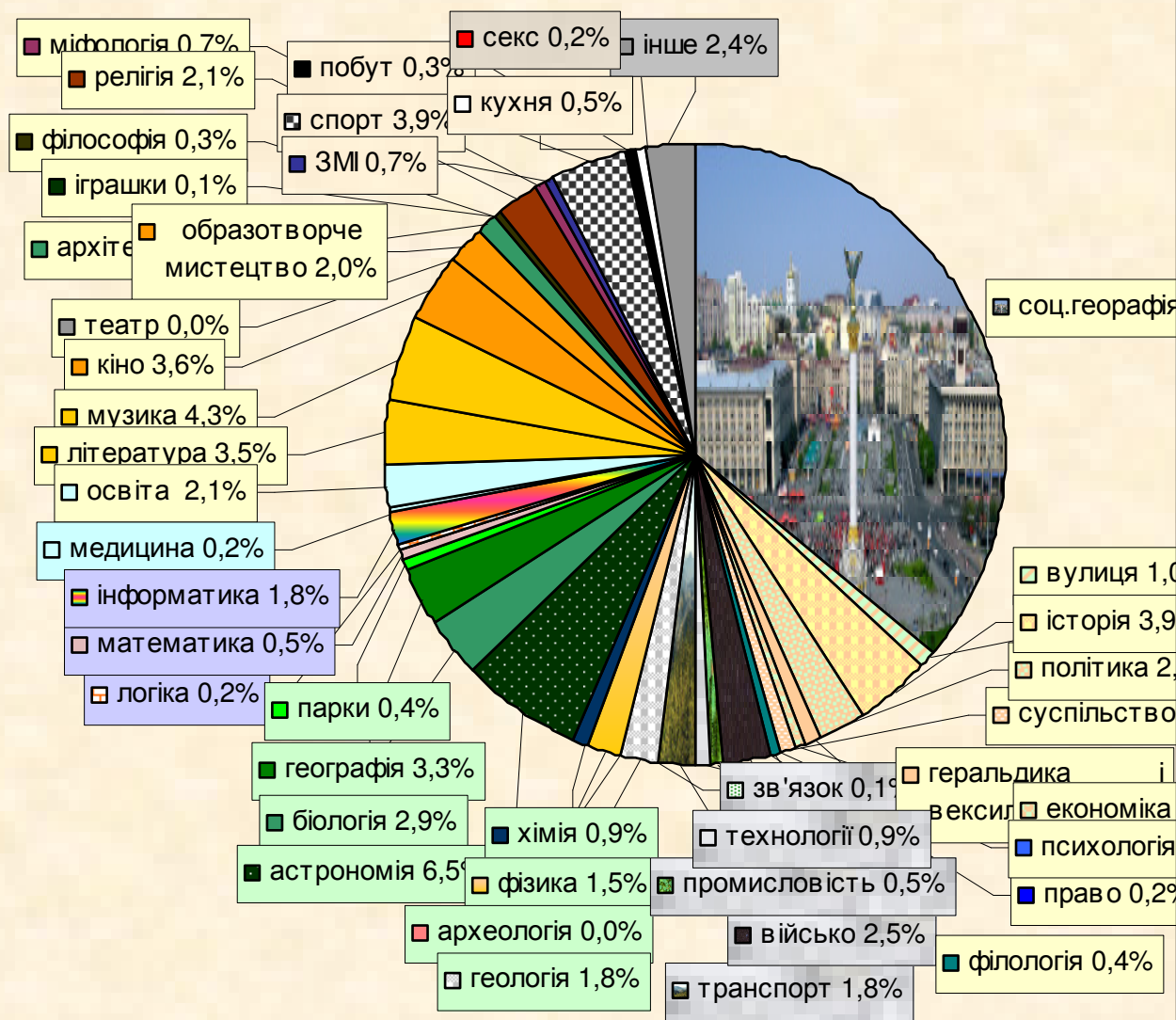
3. Робимо діаграму

стаття	тематика	тип предмета статті	повторюваність	повторюваність	зведена таблиця				зведена таблиця					
1														
2	Palác Michny z Vácínova	архітектура	об'єкт	22	3	тематика	знайдено	тематика	частка	тематика	знайдено	частка	тематика	знайдено
3	Vitřáž	архітектура	поняття	51	3	соц. географія	27	населені п	14,4%	поняття	51	24	соц. географія	27
4	Alfred Moshier Butts	архітектура	людина	41	3	суспільство	10	суспільств	5,3%	об'єкт	22	11	суспільство	10
5	Sojuz TMA-17	астрономія	модель	8	4	історія	14	історія	7,4%	людина	41	24	історія	14
6	Messier 67	астрономія	об'єкт	22	4	військо	2	військо	1,1%	організація	9	4	військо	2
7	Německé středisko pro letectv	астрономія	організація	9	4	промисловість	2	виробницт	1,1%	населений пу	19	9	промисловість	2
8	Kari (měsíc)	астрономія	об'єкт	22	4	транспорт і зв'яз	6	транспорт	3,2%	адміністратив	7	3	транспорт і зв'яз	6
9	Cardale Babington	біологія	людина	41	13	геологія та гірницт	2	геологія та	1,1%	подія	8	4	геологія та гірницт	2
10	Centrální dogma molekulární b	біологія	поняття	51	13	археологія	0	археологія	0,0%	персонаж	2	1	археологія	0
11	Kosatec bahenní	біологія	поняття	51	13	фізика	3	фізика	1,6%	дисамбіг	9	4	фізика	3
12	Dynaktin	біологія	поняття	51	13	хімія	1	хімія	0,5%	модель	8	4	хімія	1
13	Homo antecessor	біологія	поняття	51	13	астрономія	4	астрономі	2,1%	список	6	3	астрономія	4
14	Koryši	біологія	поняття	51	13	біологія	13	біологія	6,9%	огляд	1	0	біологія	13
15	Mitozom	біологія	поняття	51	13	медицина	3	медицина	1,6%	споруда	0	0	медицина	3
16	Památný strom	біологія	поняття	51	13	фіз. географія	14	фізична ге	7,4%	твір	11	5	фіз. географія	14
17	Tegmentum	біологія	поняття	51	13	математика	5	математик	2,7%	видання	1	0	математика	5
18	Dvouděložné	біологія	поняття	51	13	ІТ	10	ІТ	5,3%	інше	5	2	ІТ	10
19	Sběrač rosný	біологія	поняття	51	13	філологія	2	філологія	1,1%				філологія	2
20	Lewisuchus	біологія	поняття	51	13	освіта	1	освіта	0,5%				освіта	1
21	Korandův smrk	біологія	об'єкт	22	13	література	8	література	4,3%				література	8
22	Canadian War Museum	військо	об'єкт	22	2	музика	11	музика	5,9%				музика	11
23	Zahar Prilepin	військо	людина	41	2	кіно	8	кіно і ТВ	4,3%				кіно	8
24	Hydrodynamická vertikální zon	геологія та гірницт	поняття	51	2	театр	1	театр	0,5%				театр	1
25	Richard Lydekker	геологія та гірницт	людина	41	2	живопис	2	живопис	1,1%				живопис	2
26	1586	дата	список	6	3	архітектура	3	архітектур	1,6%				архітектура	3
27	1985	дата	список	6	3	іграшка	3	іграшки	1,6%				іграшка	3
28	Ochotnýj rjad	дисамбіг	дисамбіг	9	9	релігія	7	релігія	3,7%				релігія	7
29	Horčička	дисамбіг	дисамбіг	9	9	міфологія	0	міфологія	0,0%	усього (без д	200	100	міфологія	0
30	Obléhání Bma	дисамбіг	дисамбіг	9	9	спорт	17	спорт	9,0%				спорт	17
31	DeSoto County	дисамбіг	дисамбіг	9	9	інше	9	інше	4,8%				інше	9
32	Letka	дисамбіг	дисамбіг	9	9	усього (без д	188		100,00%				усього (без д	188
33	П	дисамбіг	дисамбіг	9	9	дисамбіг	9						дисамбіг	9
34	Імперія	дисамбіг	дисамбіг	9	9	дата	3						дата	3

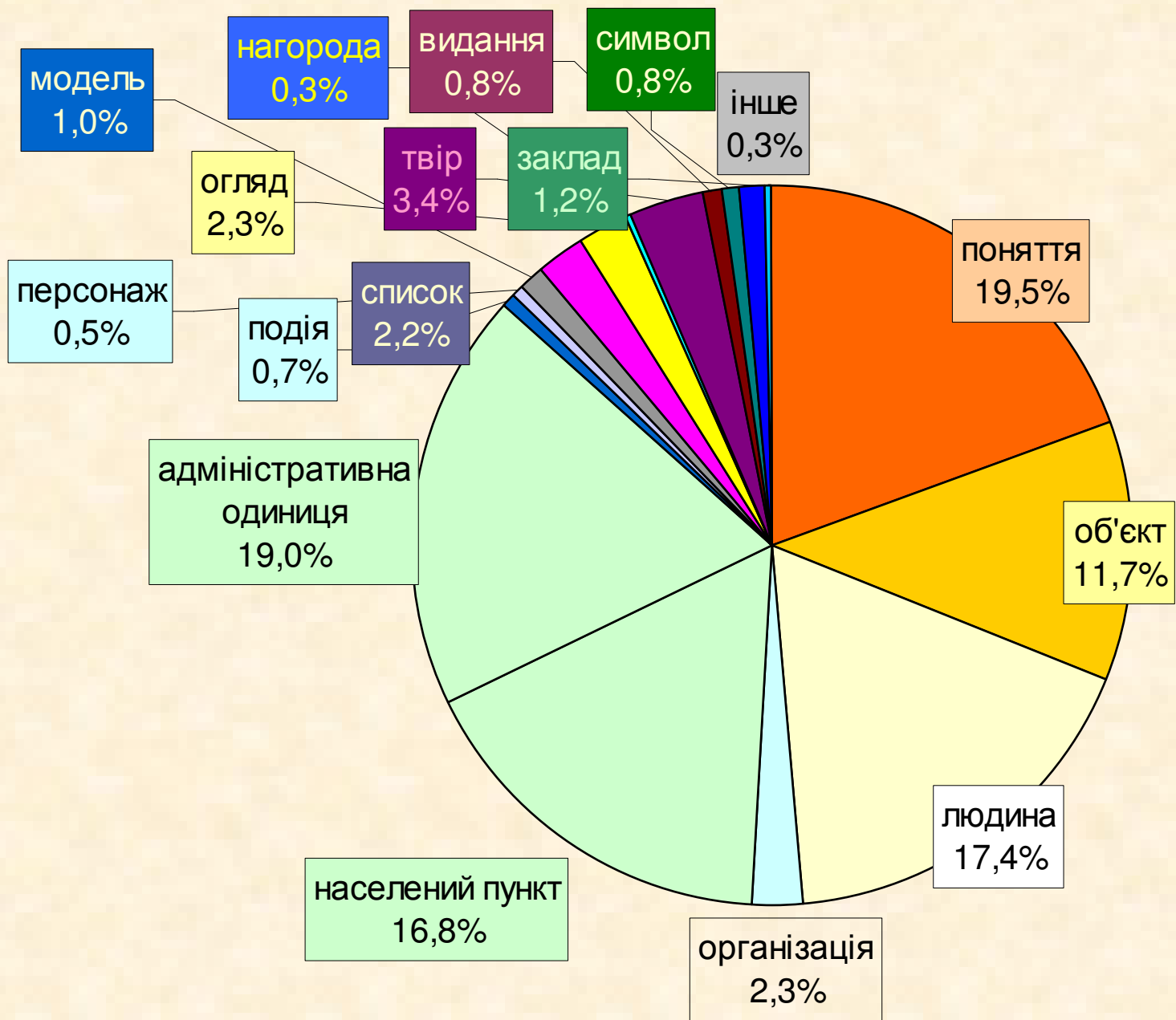
Вибірка-1000 української Вікіпедії



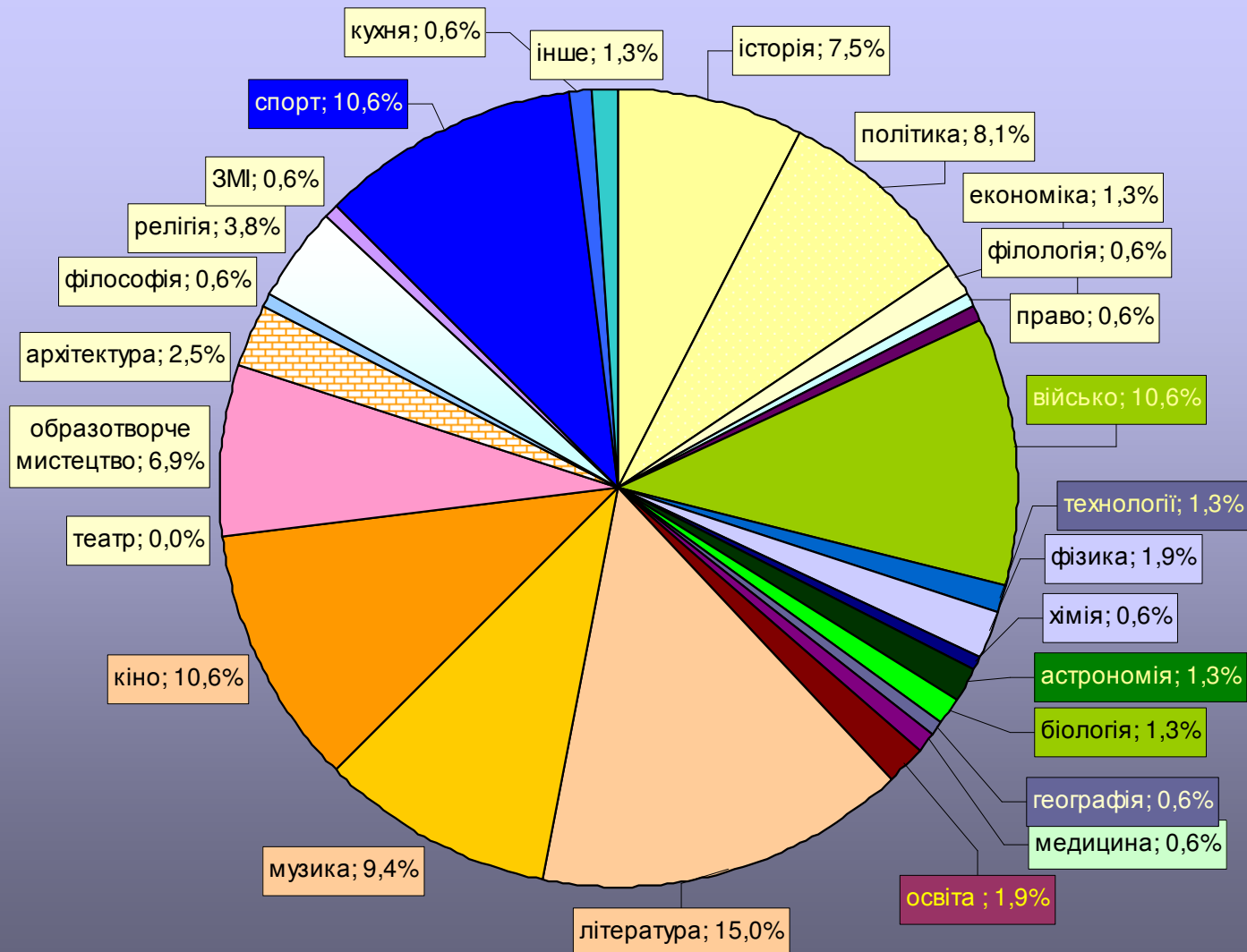
Статистична вибірка української Вікіпедії



- соц.географія
- вулиця
- історія
- політика
- геральдика і вексилологія
- економіка
- суспільство
- психологія
- філологія
- право
- військо
- промисловість
- технології
- транспорт
- зв'язок
- геологія
- археологія
- фізика
- хімія
- астрономія
- біологія
- географія
- парки
- математика
- логіка
- інформатика
- медицина
- освіта
- література
- музика
- кіно
- театр
- образотворче мистецтво
- архітектура
- іграшки
- філософія
- релігія
- міфологія
- ЗМІ
- спорт
- побут



Персоналії за професією



4.1. Міжмовні порівняння

8 Вікіпедій у порівнянні.

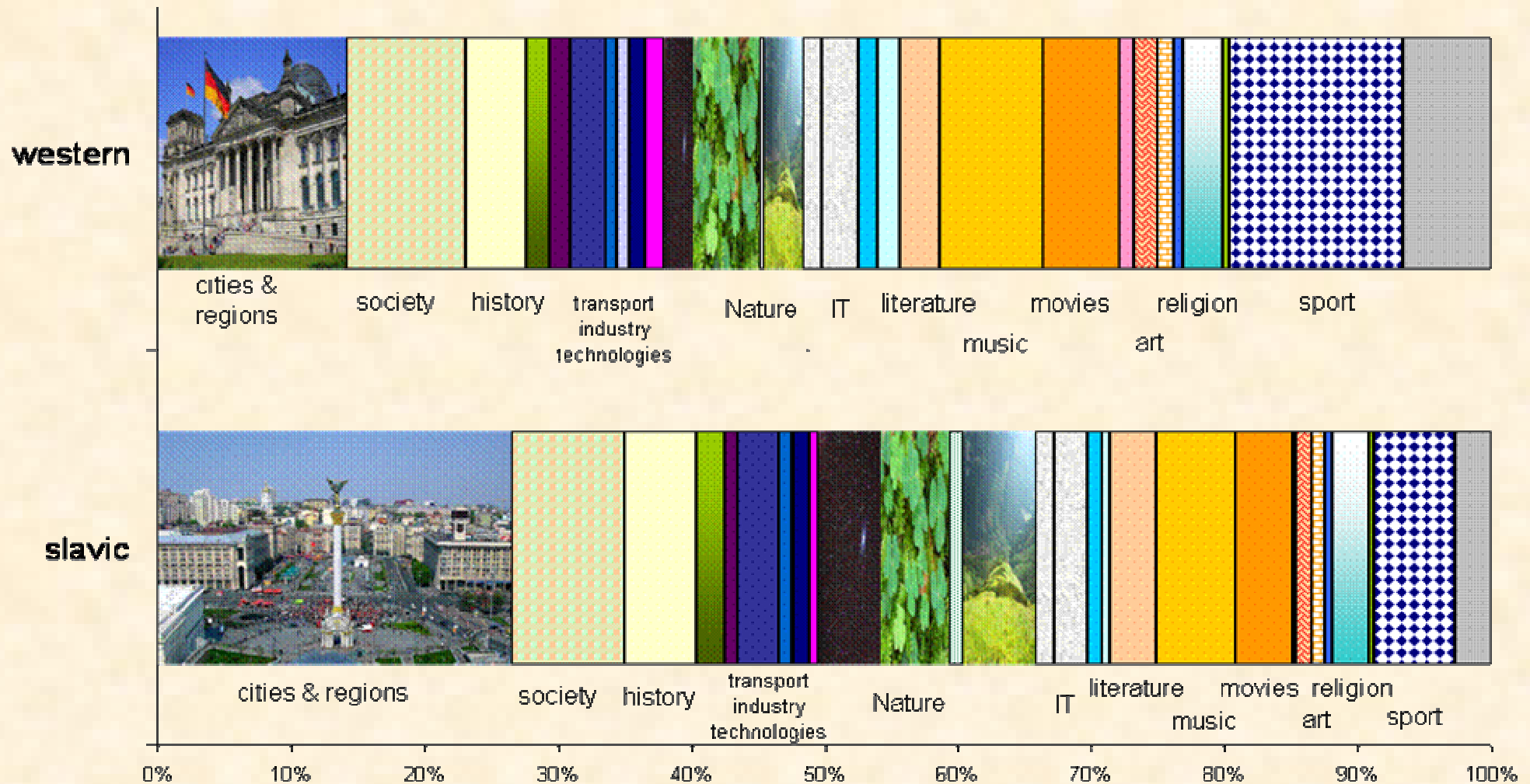
Статистичні вибірки по 200-300 статей, дослідження липня 2011, презентовані на Вікіманії 2011

■ PL	■ EN
■ RU	■ DE
■ UK	■ FR
■ CS	■ IT

4.2 Західні і слов'янські ВП за темою

Західні – більше спорту

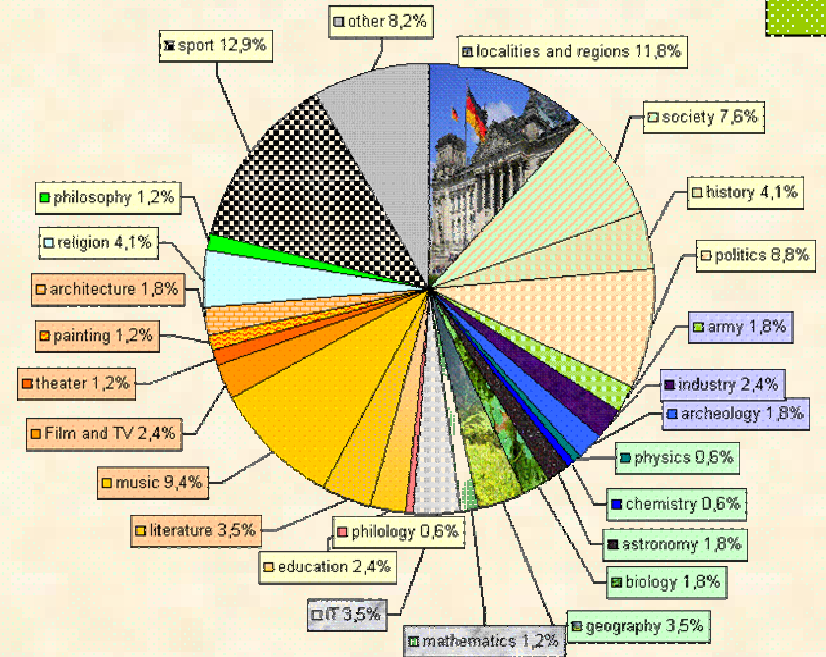
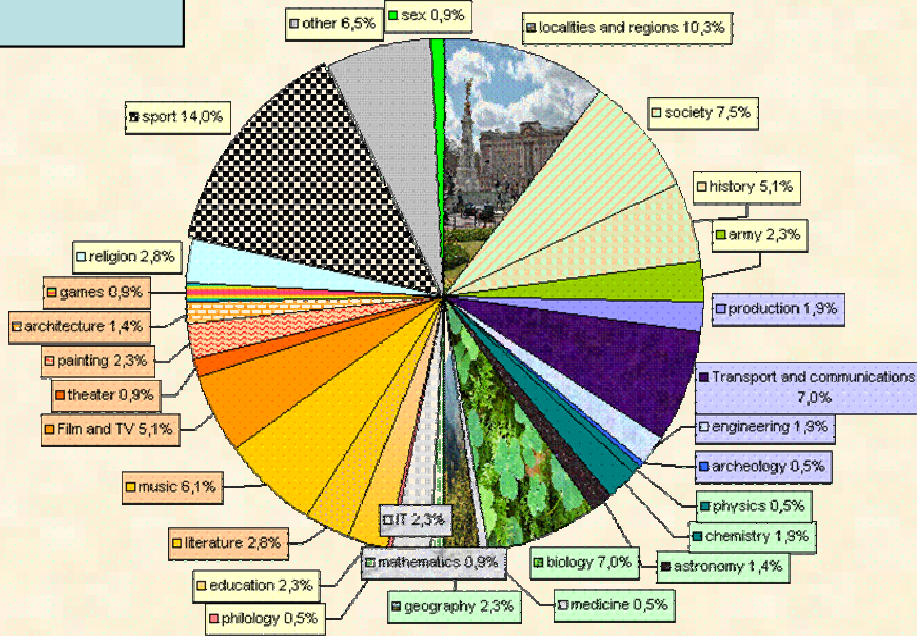
Слов'янські – більше географії



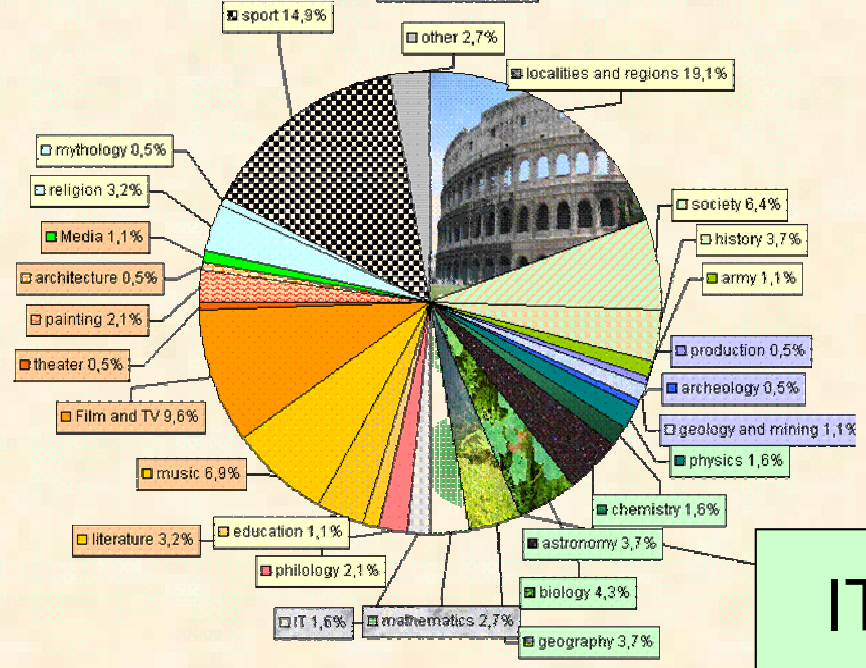
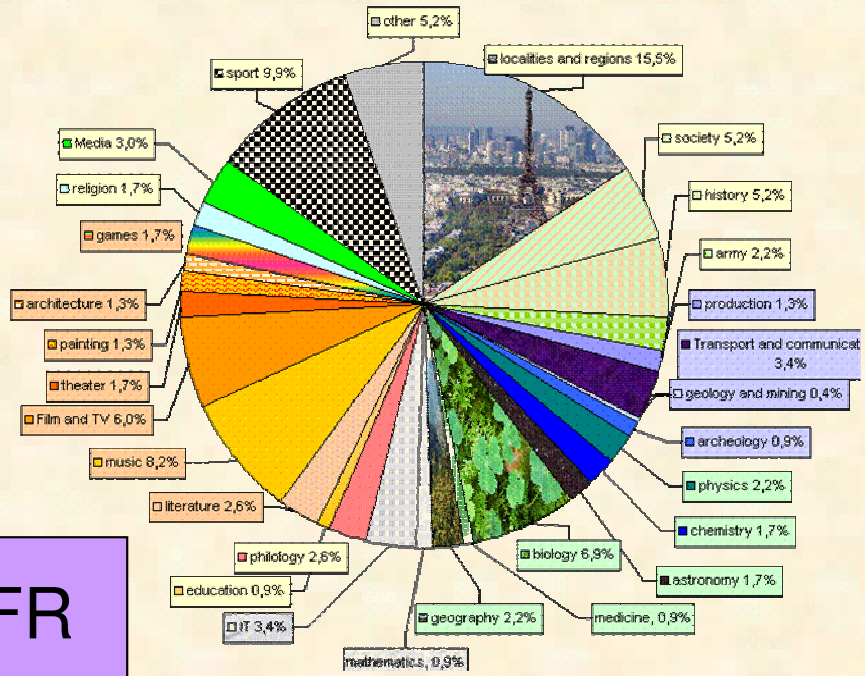
4.3 Західні Вікіпедії за тематикою

EN

DE



FR

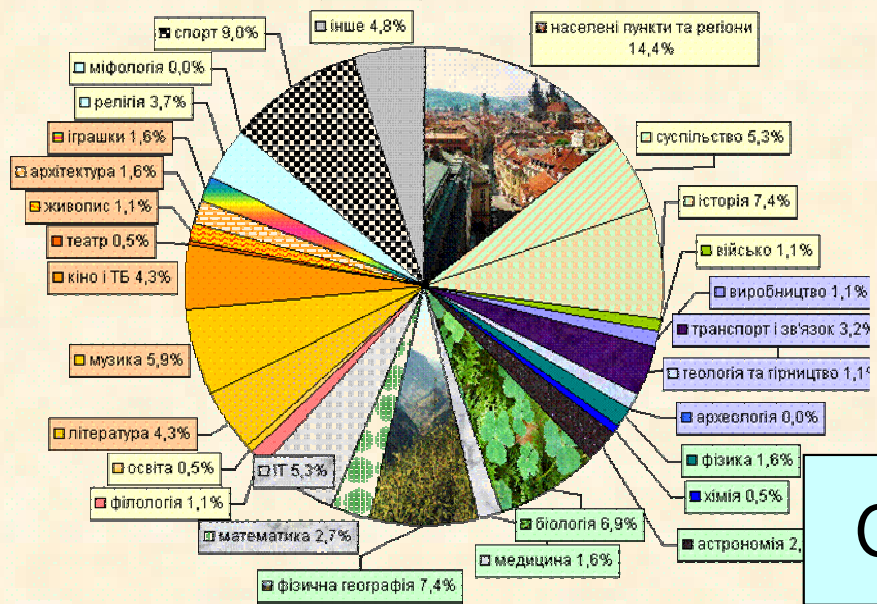
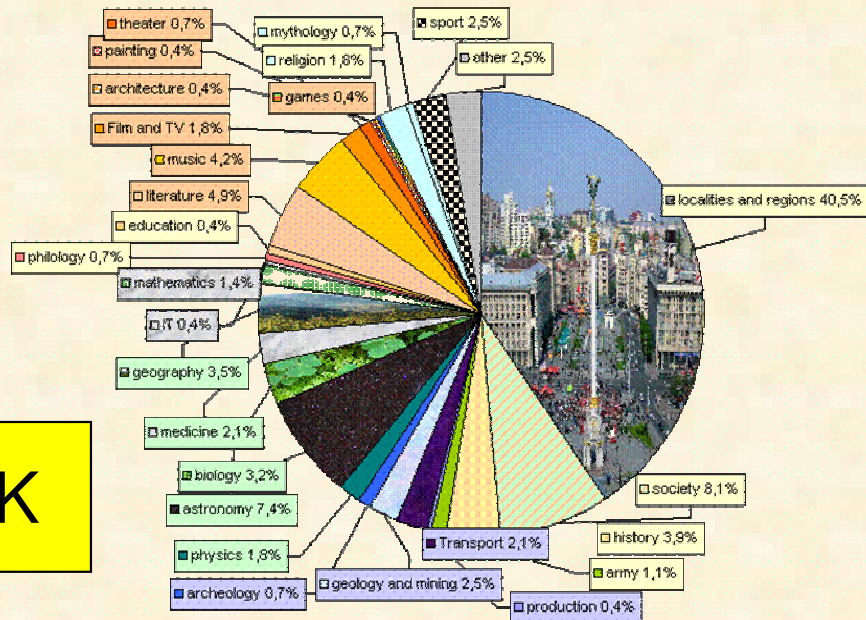
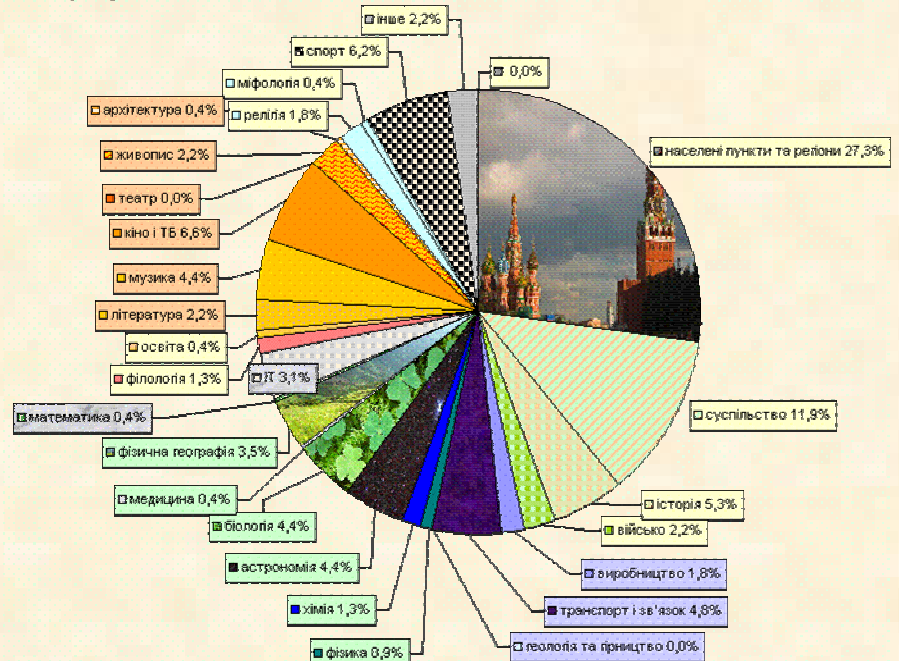
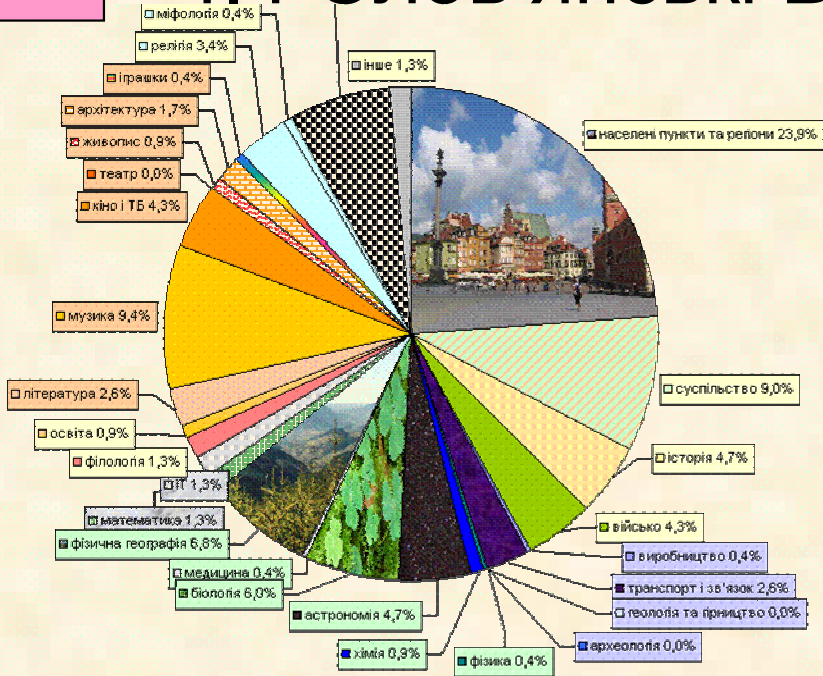


IT

PL

RU

4.4 Слов'янські Вікіпедії за тематикою



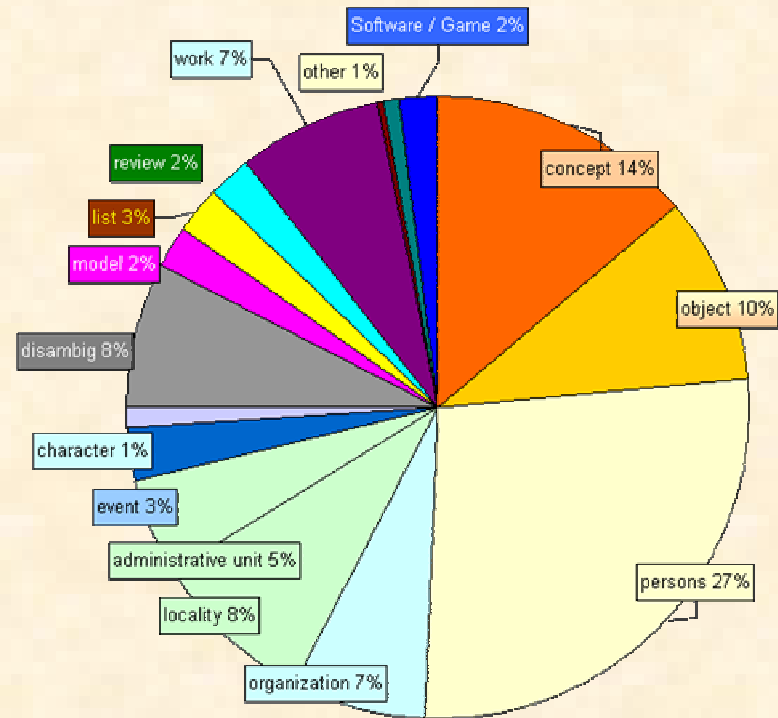
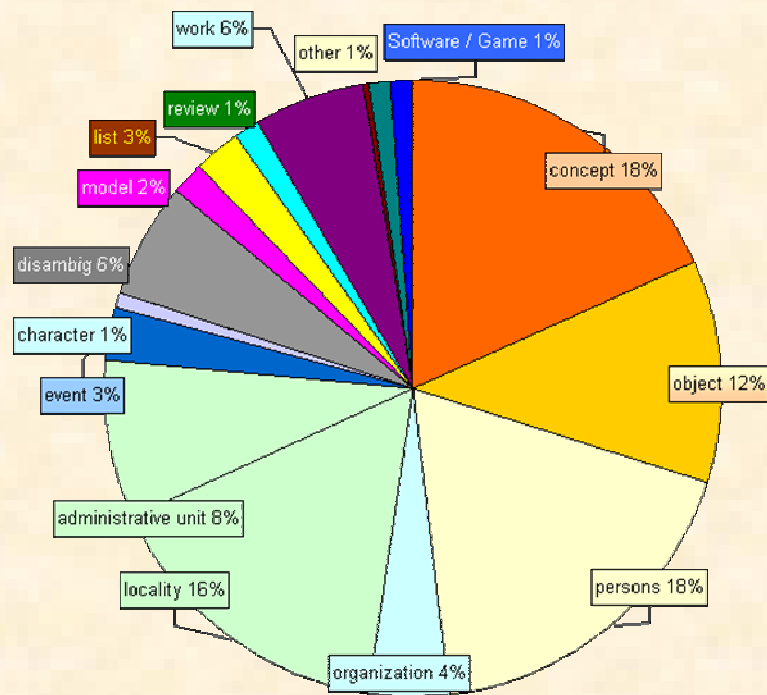
UK

CS

4.6

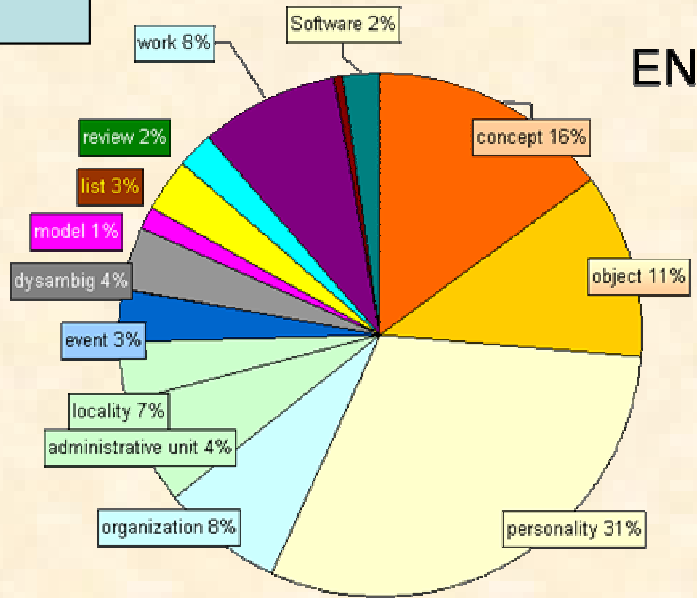
Слов'янські – більше об'єктів і понять

Західні – більше персоналій і компаній

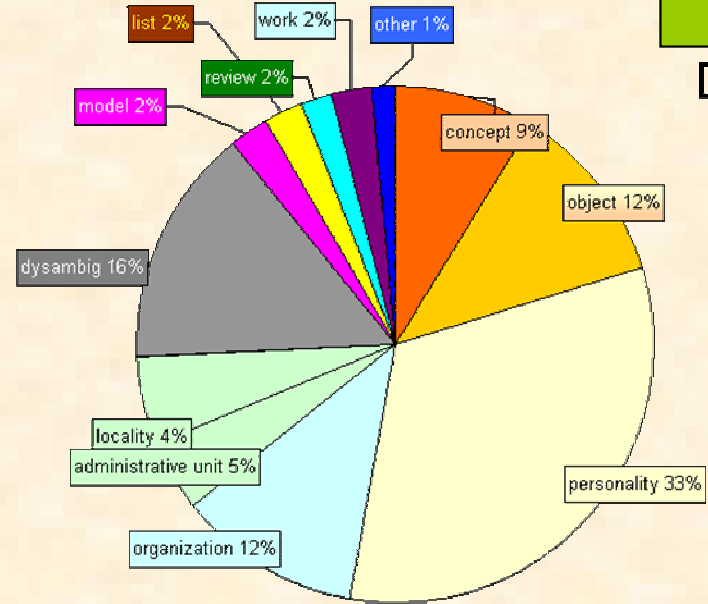


4.7 західні Вікіпедії за типом

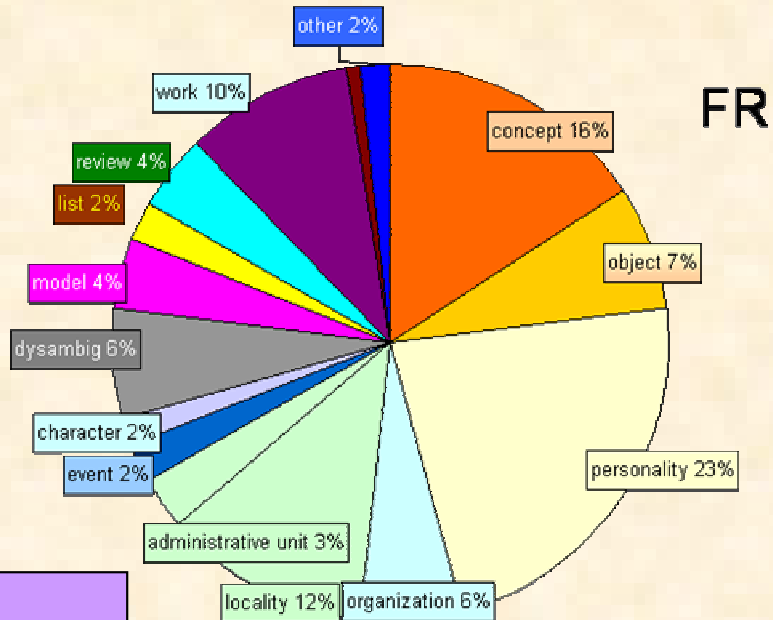
EN



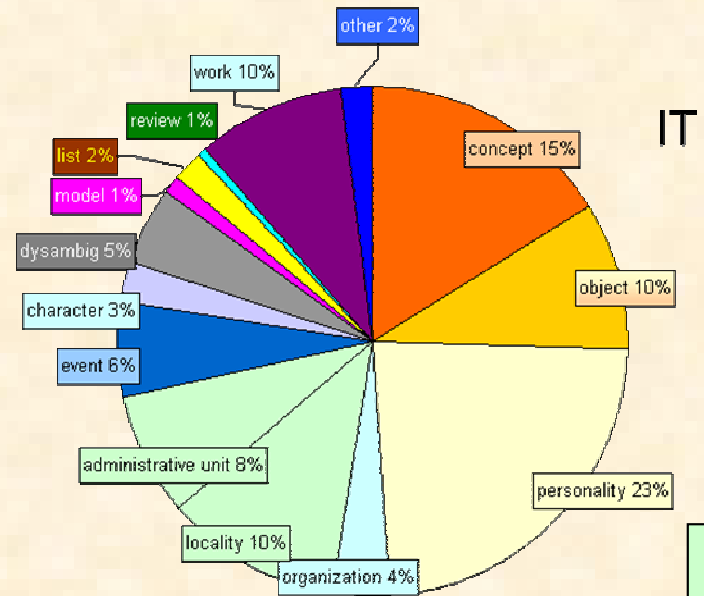
DE



FR

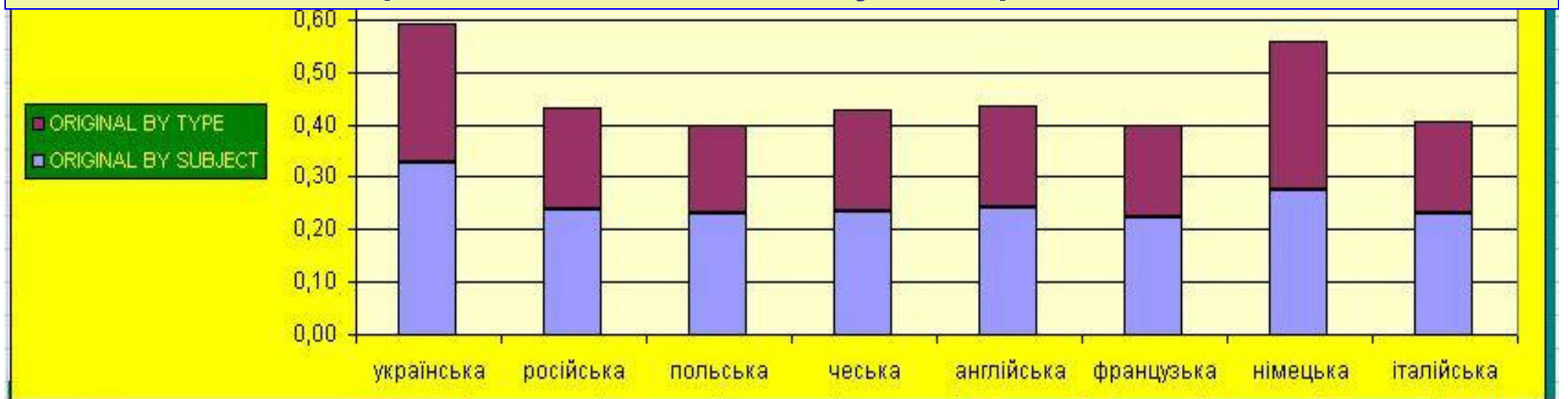


IT



	українська	російська	польська	чеська	англійська	французька	німецька	італійська	середнє
ORIGINAL BY SUBJECT	0,33	0,24	0,23	0,23	0,24	0,22	0,28	0,23	0,24
ORIGINAL BY TYPE	0,26	0,19	0,17	0,19	0,19	0,18	0,28	0,17	0,20
TOGETHER	0,59	0,43	0,40	0,43	0,43	0,40	0,56	0,41	0,44

5.0 Оригінальність чи універсальність?



	UK	RU	PL	CS	EN	FR	DE	IT	разом
Ukrainian	0,00	0,53	0,51	0,68	0,80	0,72	0,88	0,62	0,59
Russian	0,53	0,00	0,33	0,50	0,51	0,44	0,68	0,46	0,43
Polish	0,51	0,33	0,00	0,42	0,50	0,41	0,64	0,37	0,40
Czech	0,68	0,50	0,42	0,00	0,42	0,35	0,63	0,41	0,43
English	0,80	0,51	0,50	0,42	0,00	0,36	0,46	0,43	0,44
French	0,72	0,43	0,40	0,34	0,35	0,00	0,56	0,36	0,40
German	0,88	0,68	0,64	0,63	0,46	0,56	0,00	0,60	0,56
Italian	0,62	0,46	0,37	0,41	0,43	0,36	0,60	0,00	0,41
together	0,59	0,43	0,40	0,43	0,43	0,40	0,56	0,41	0,45
coefficient of Slavic WP compatibility	1,74	2,20	2,38	1,86	1,79	2,09	1,41	2,15	2,16
coefficient Western WP compatibility	1,33	1,92	2,08	2,20	2,42	2,34	1,85	2,16	2,23
direction	1,31	1,14	1,14	0,85	0,74	0,89	0,76	0,99	0,97

ДАЛІ БУДЕ.....

- Дякую за увагу!
- Склав користувач А1
- А1 вдячний Епп7 за допомогу в аналізі Вибірки-1000
- Всі матеріали суворо під копілефтом!