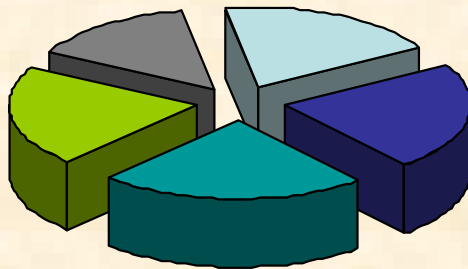


# Зміст Вікіпедії

Досвід порівнянь

# 1. Предмет досліджень

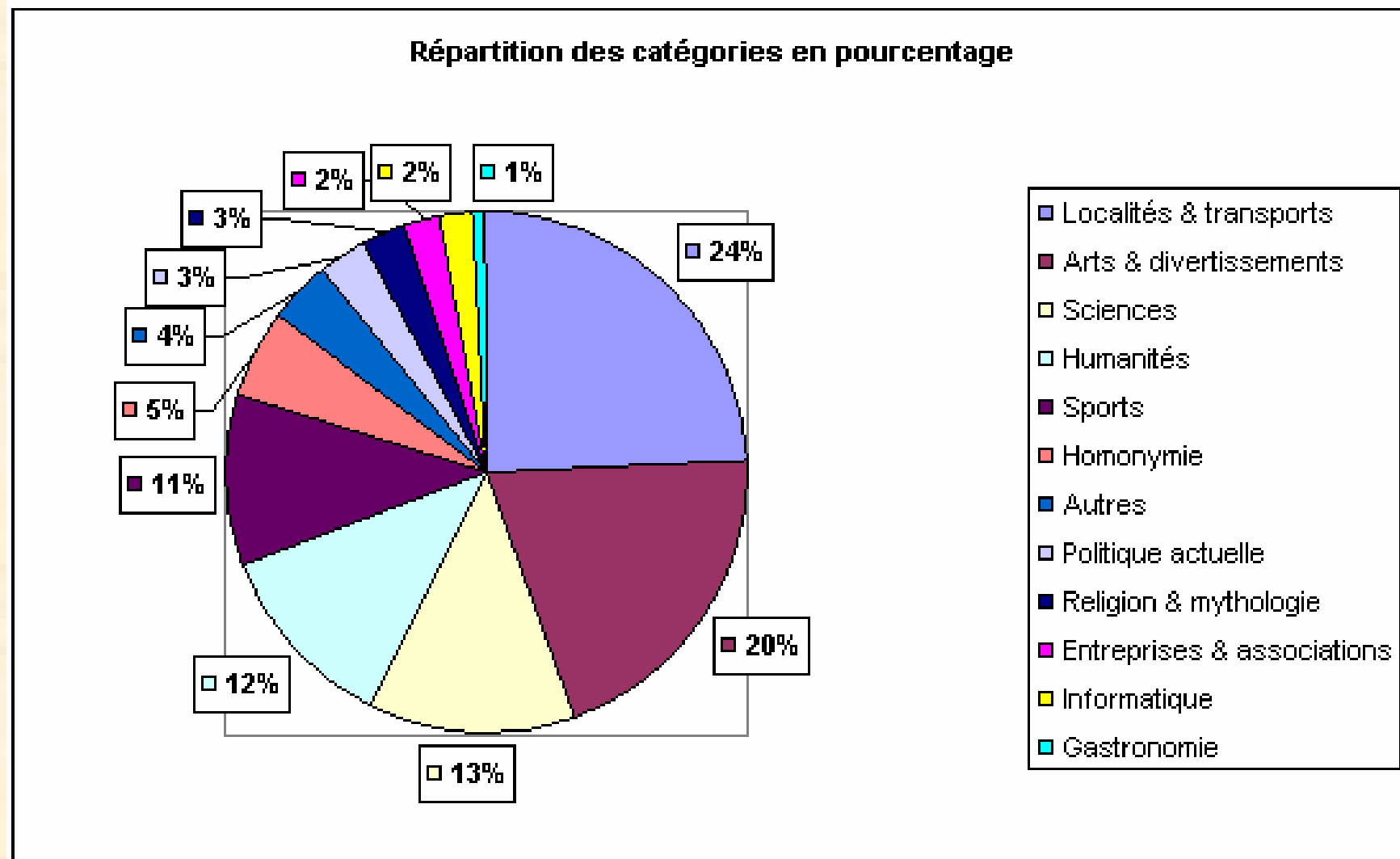
- Джимбо Вейлз:
- “Уявіть, Уявіть планету, на якій кожній окремій особі надано доступ до суми всіх знань людства..”
- **Питання:** Якою є структура знань людства?
- Спробуємо це проаналізувати



dziedziny wiedzy

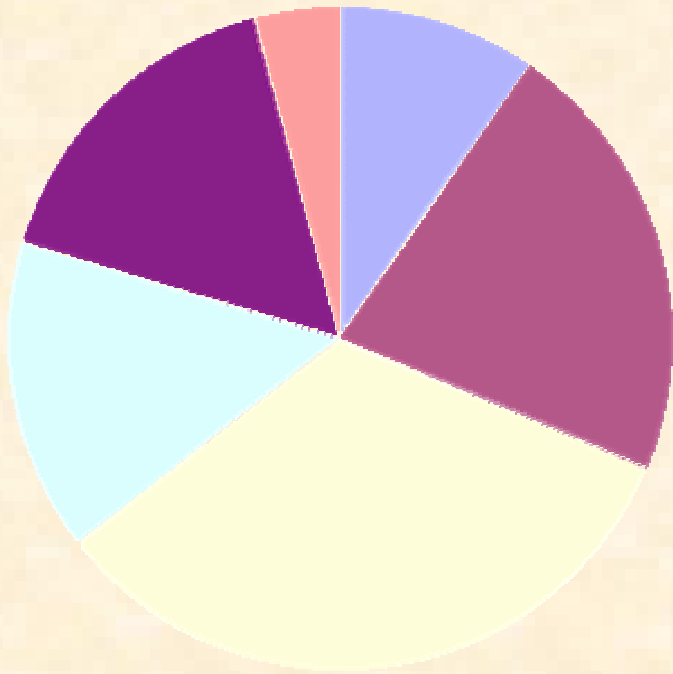
# Історія питання

- Французька вікіпедія 2007



# Історія питання

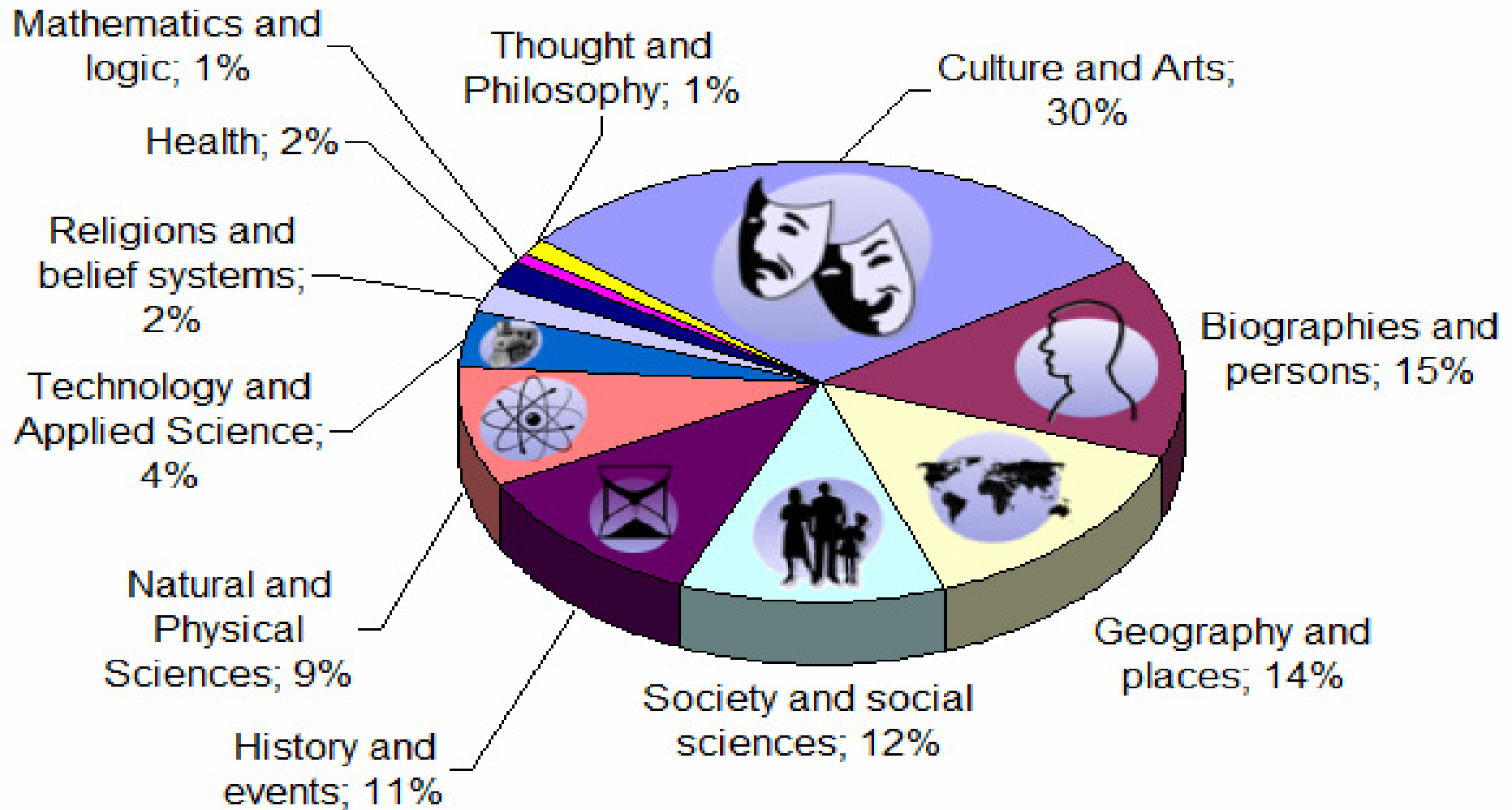
- Шведська вікіпедія 2007



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

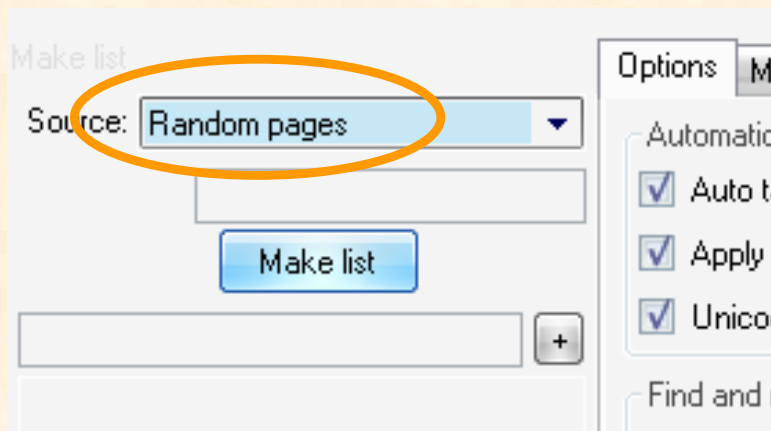
# Історія питання

Англійська Вікіпедія, 2008

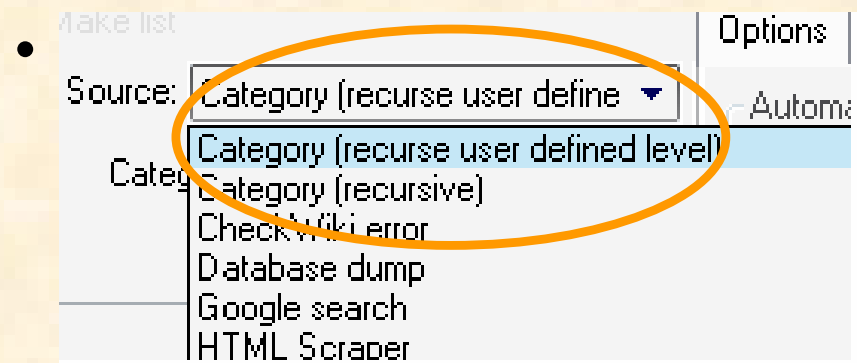


## 2. Методологія

- Статистична вибірка
- За допомогою AWB



- Рекурсивний аналіз категорії



- Або ScanCat

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

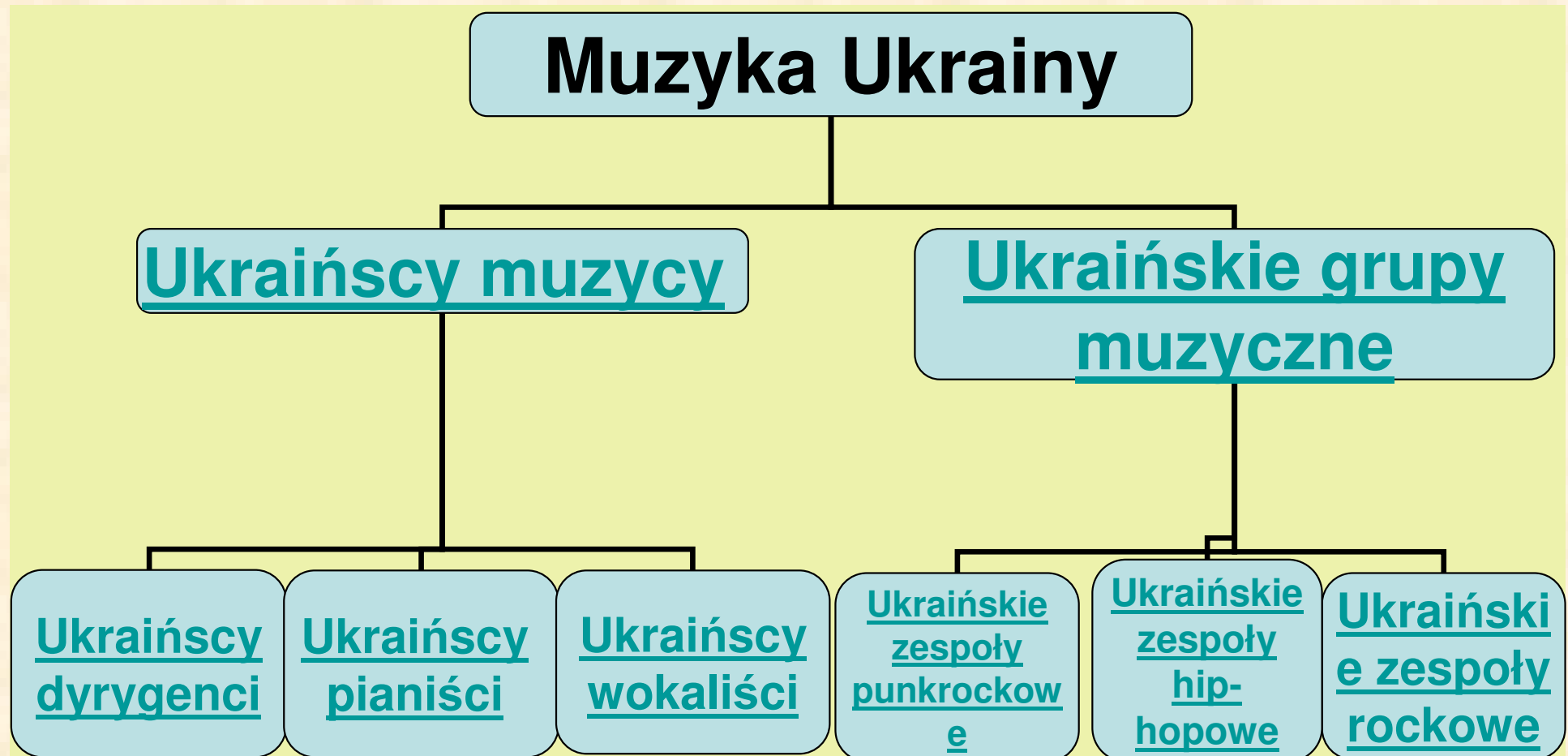
Мова	uk	
Проект	wikipedia	
Глибина	0	(глибина 0 = не шукати в підкатегоріях)
Категорій	Польща	
...		

## 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

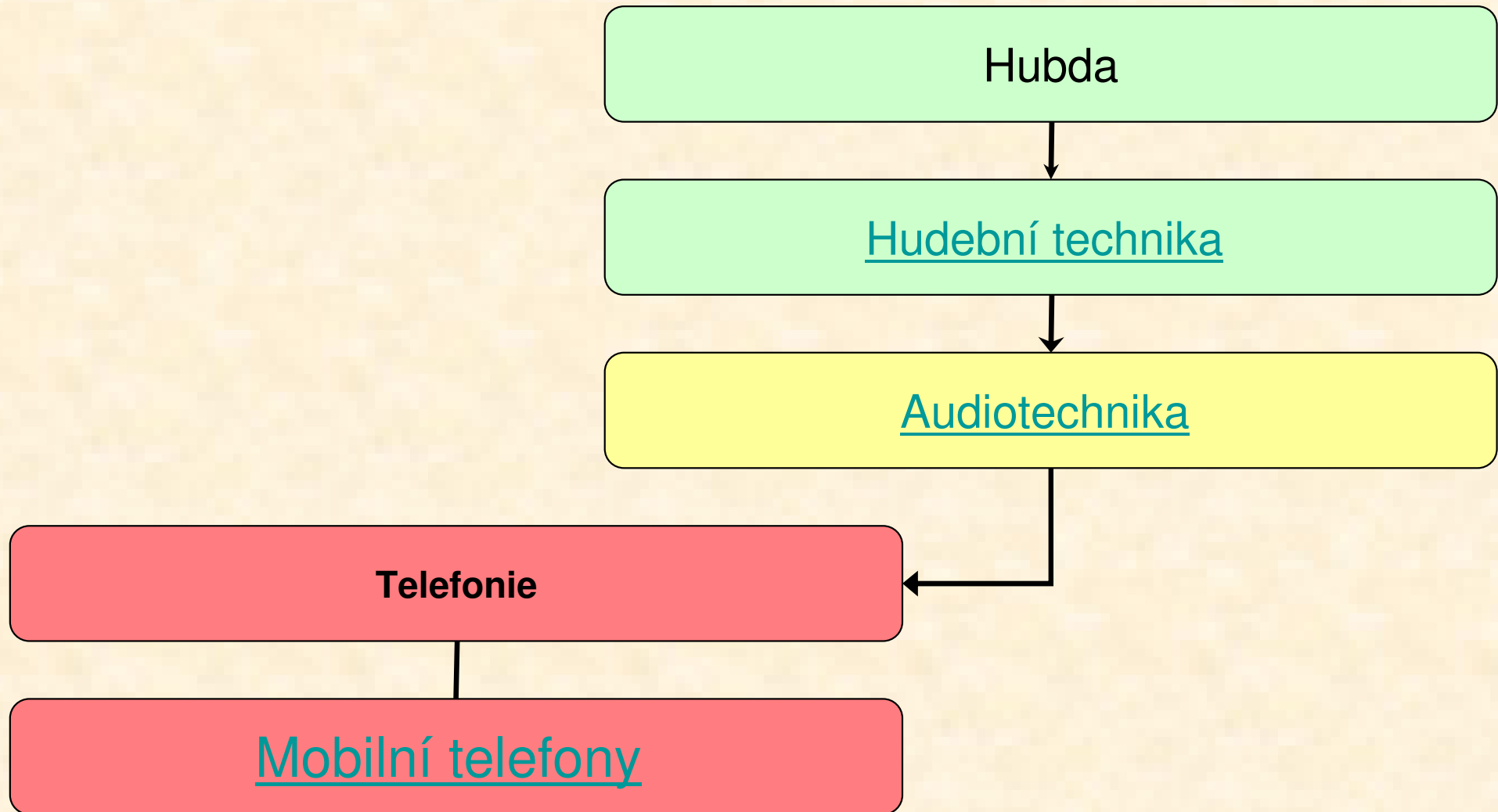
Wikipedia polska, Kategoria:Muzyka Ukrainy





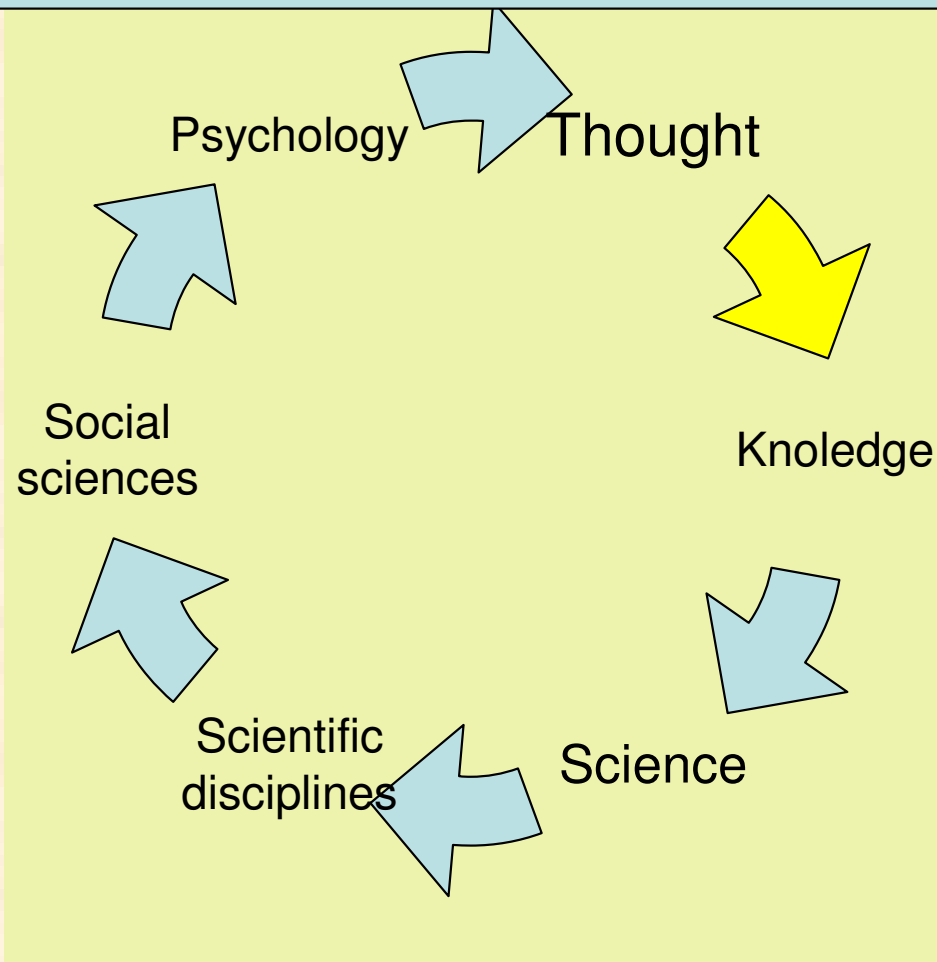
## 2.3 Рекурсивний аналіз – незмістовна транзитивна категоризація

Wikipedia czeska, Kategorie:Hudba

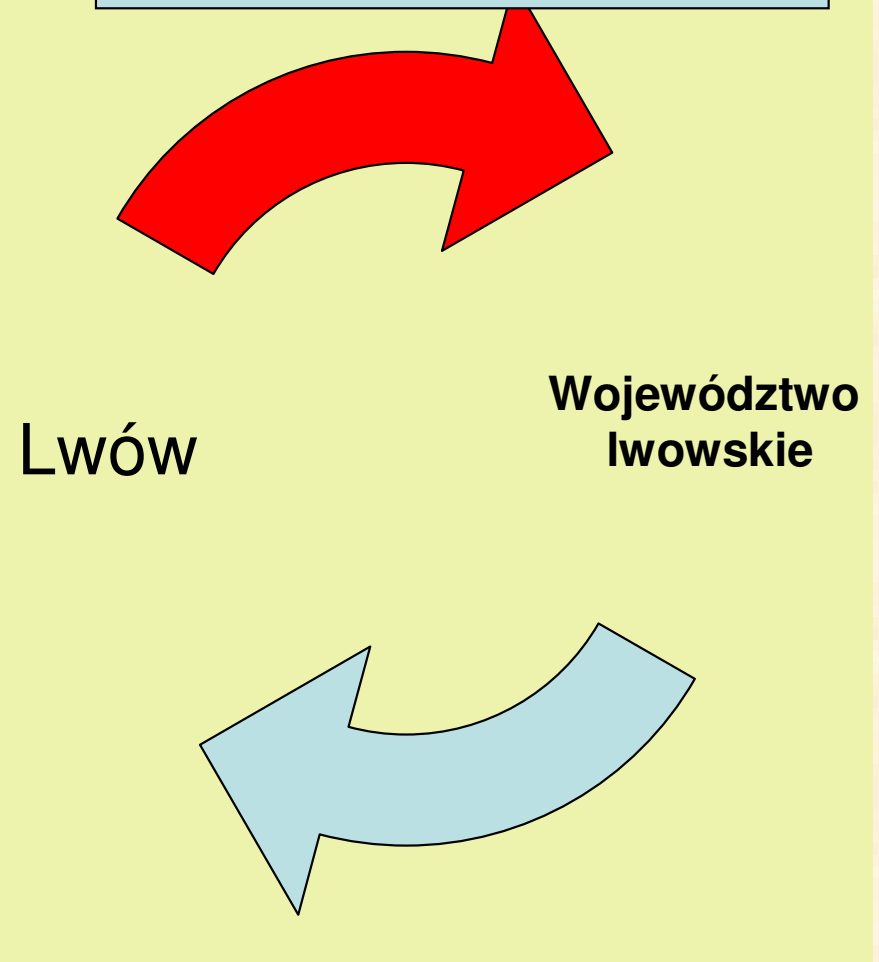


## 2.4 Рекурсивний аналіз – нетранзитивна категоризація

Класична нетранзитивність в англ. вікі



Колись було у польській...

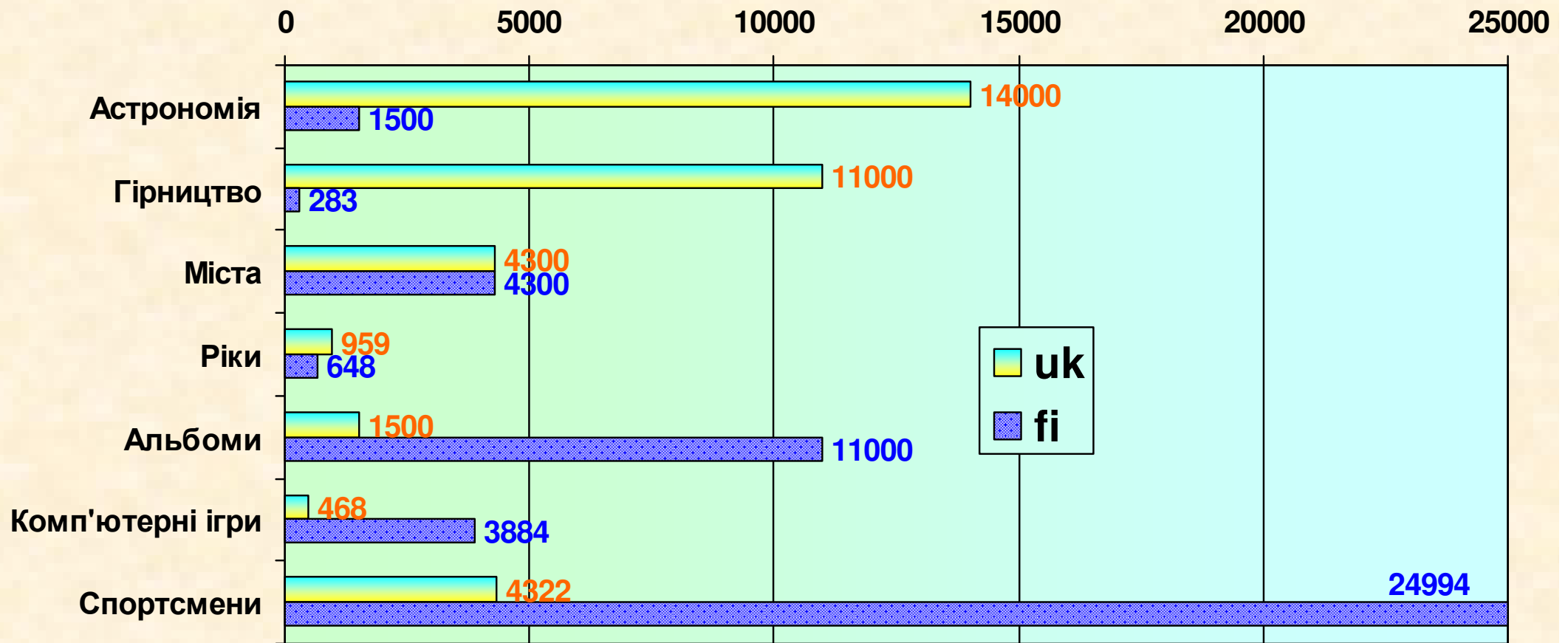


## 2.5 підсумок

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

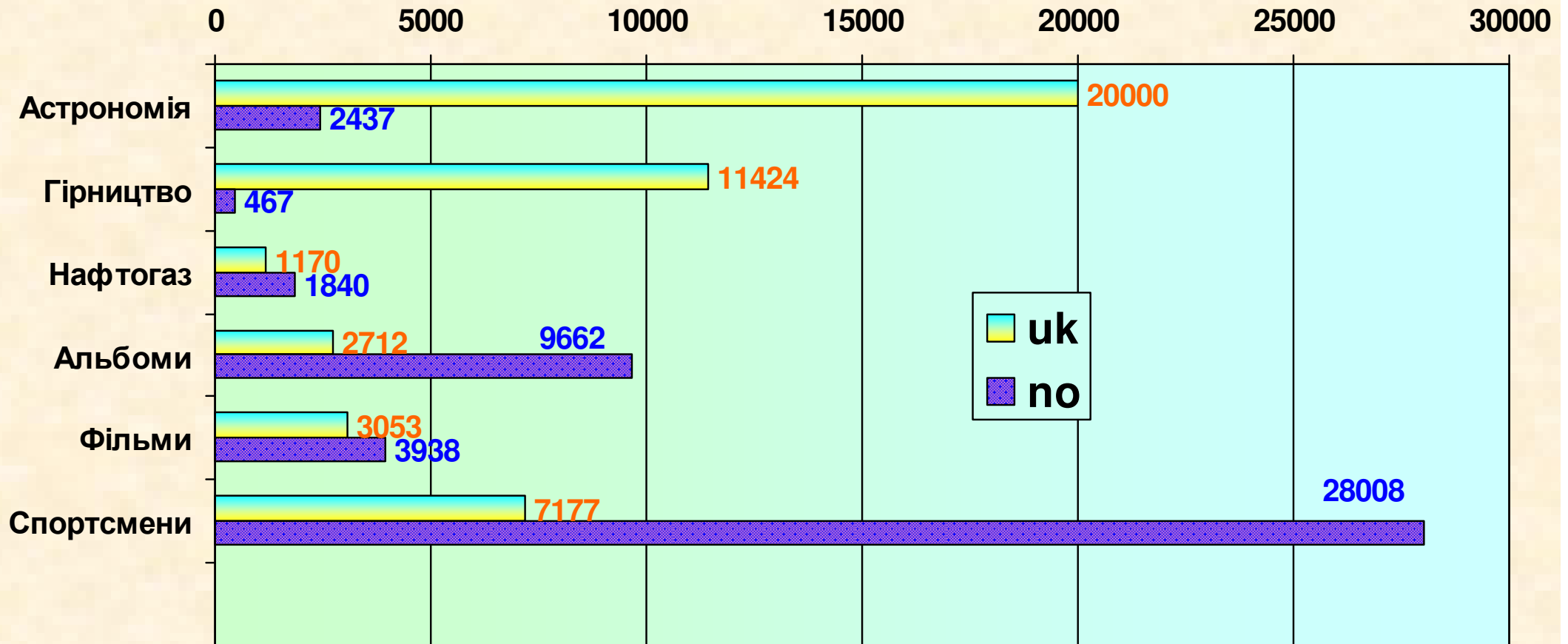
# 3.1. Спеціальні дослідження – Ukr vs Fin

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)

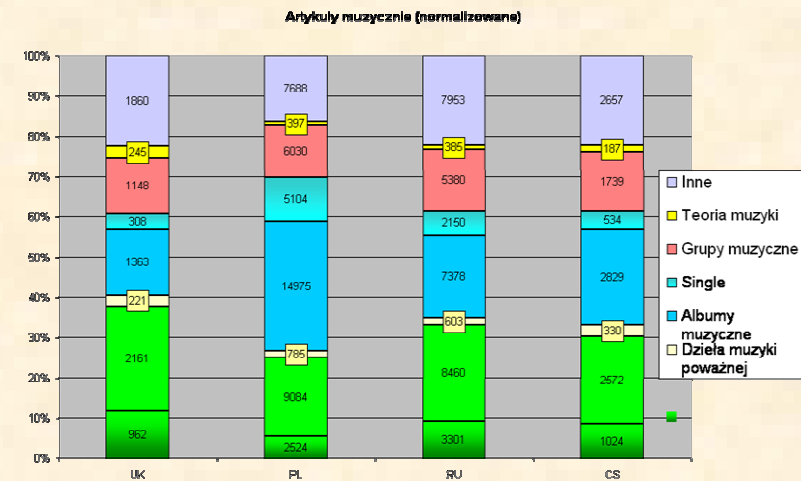
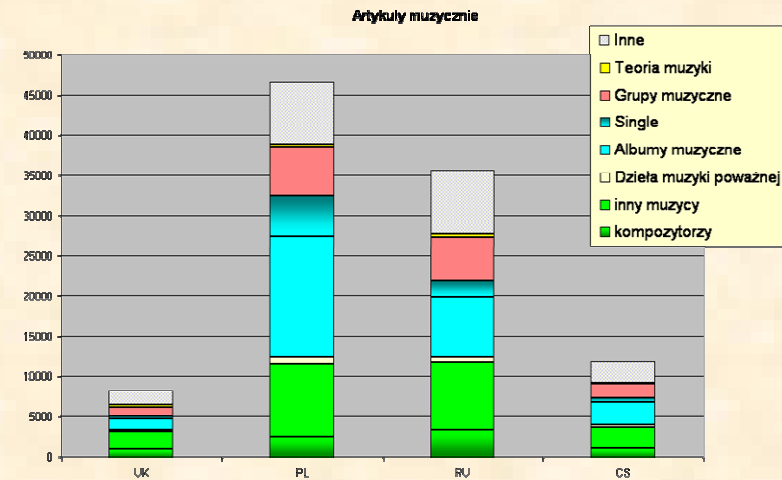


# 3.2. Спеціальні дослідження – Музика

- Кількість статей по музичних тематиках:

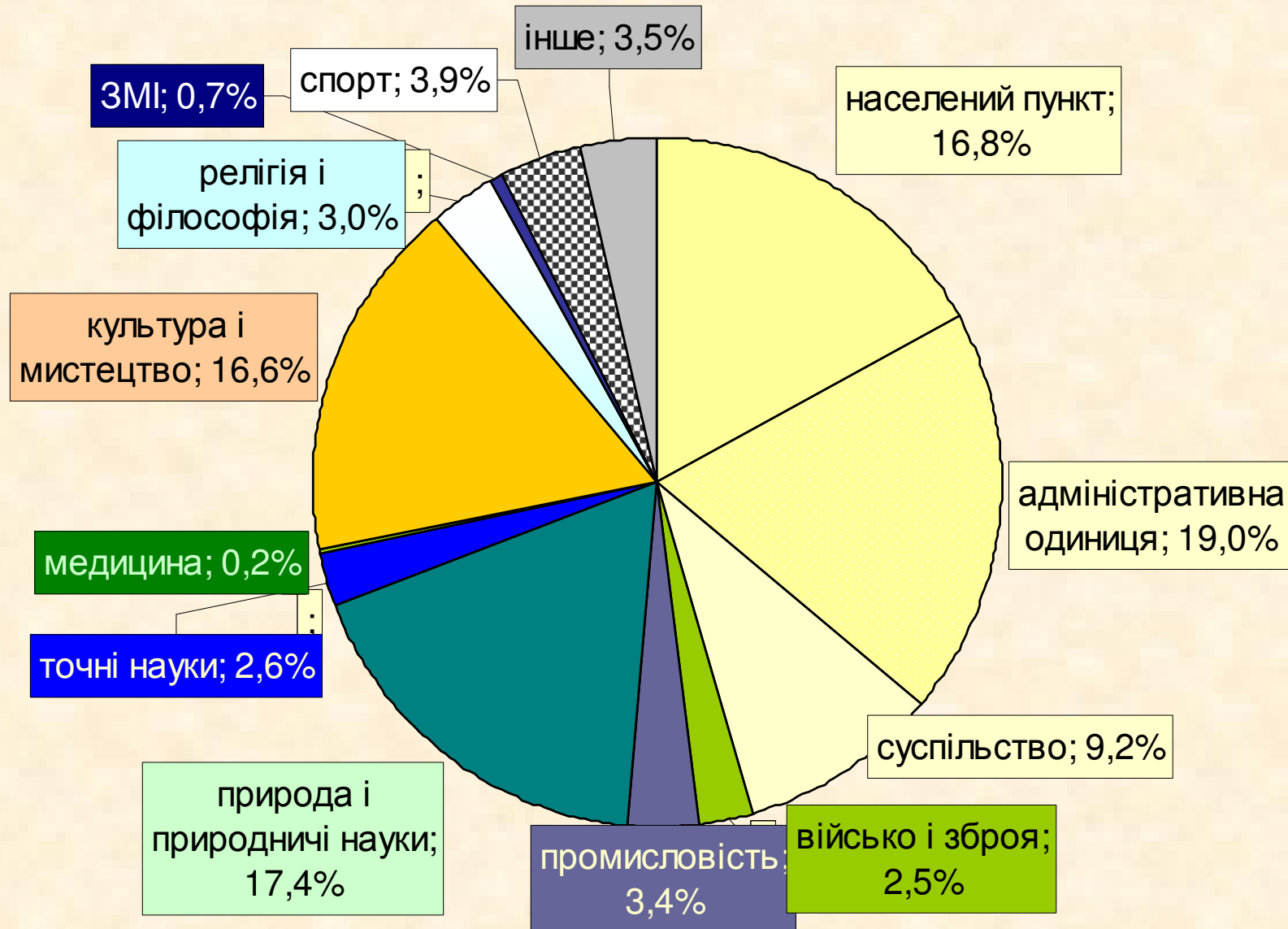
- Абсолютна діаграма
- Нормалізована діаграма
  - Зелений - музиканти
  - Блакитний – альбоми і сингли
  - Червоний – гурти
  - Жовтий – теорія музики

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



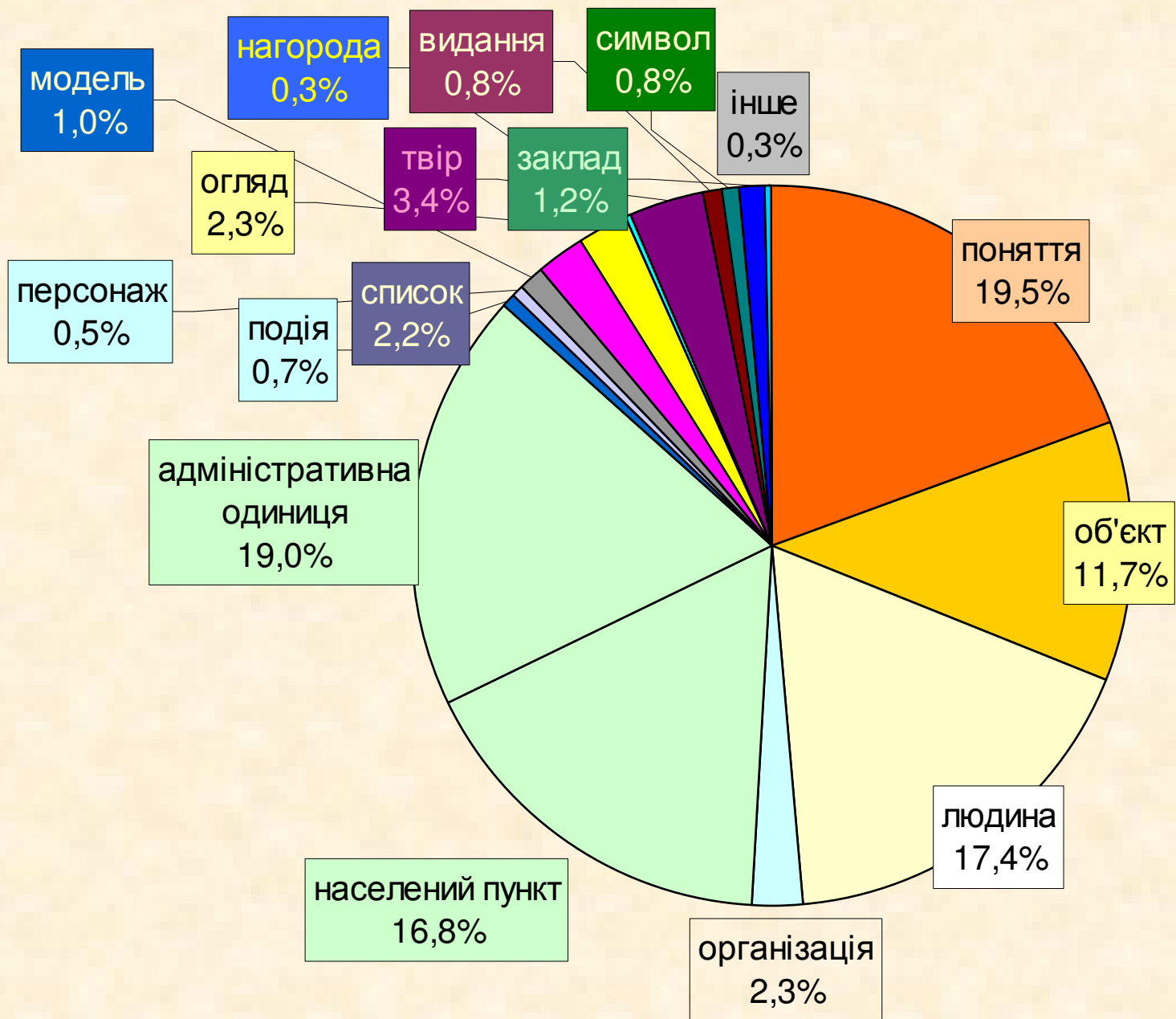


# Вибірка-1000 української Вікіпедії

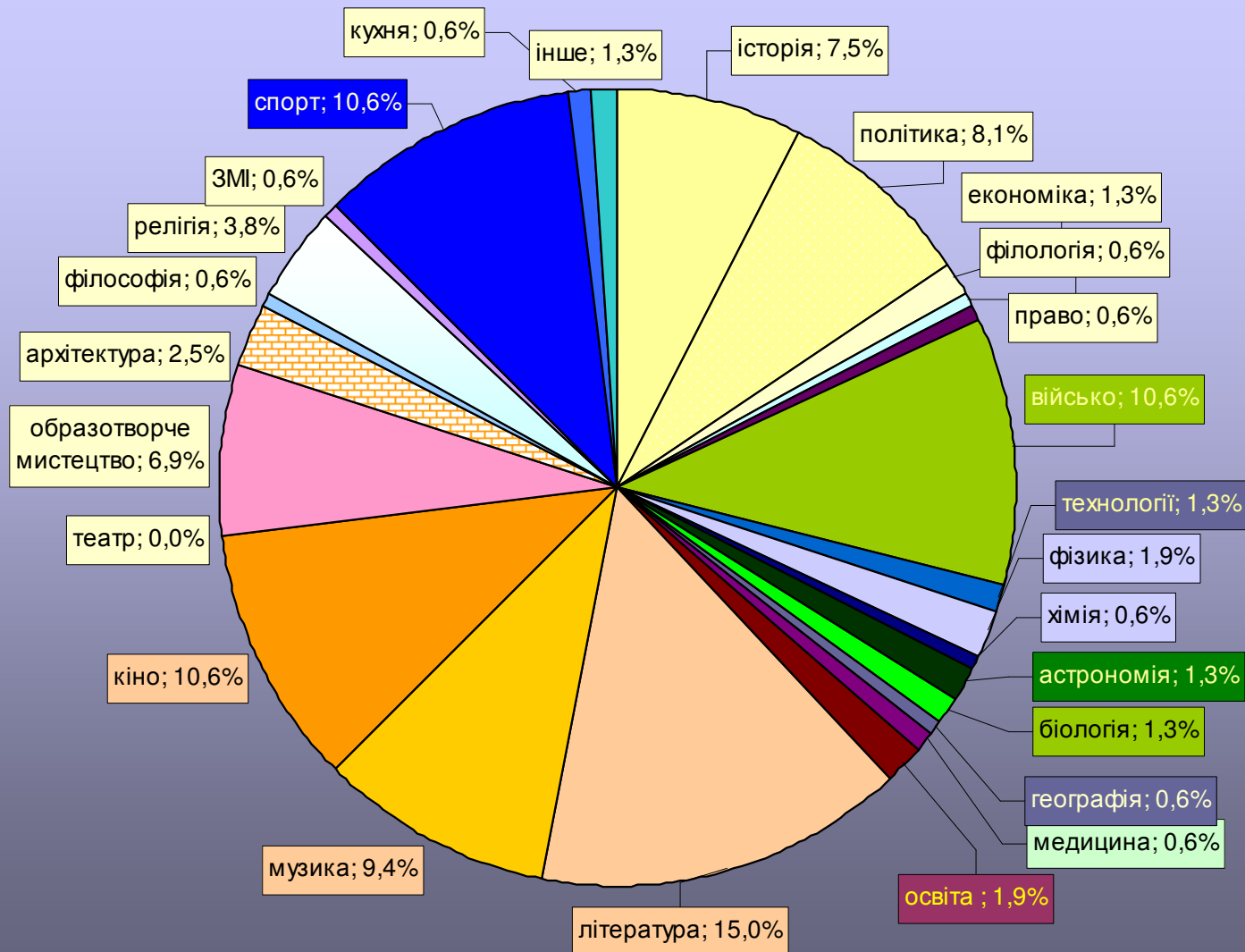








# Персоналії за професією

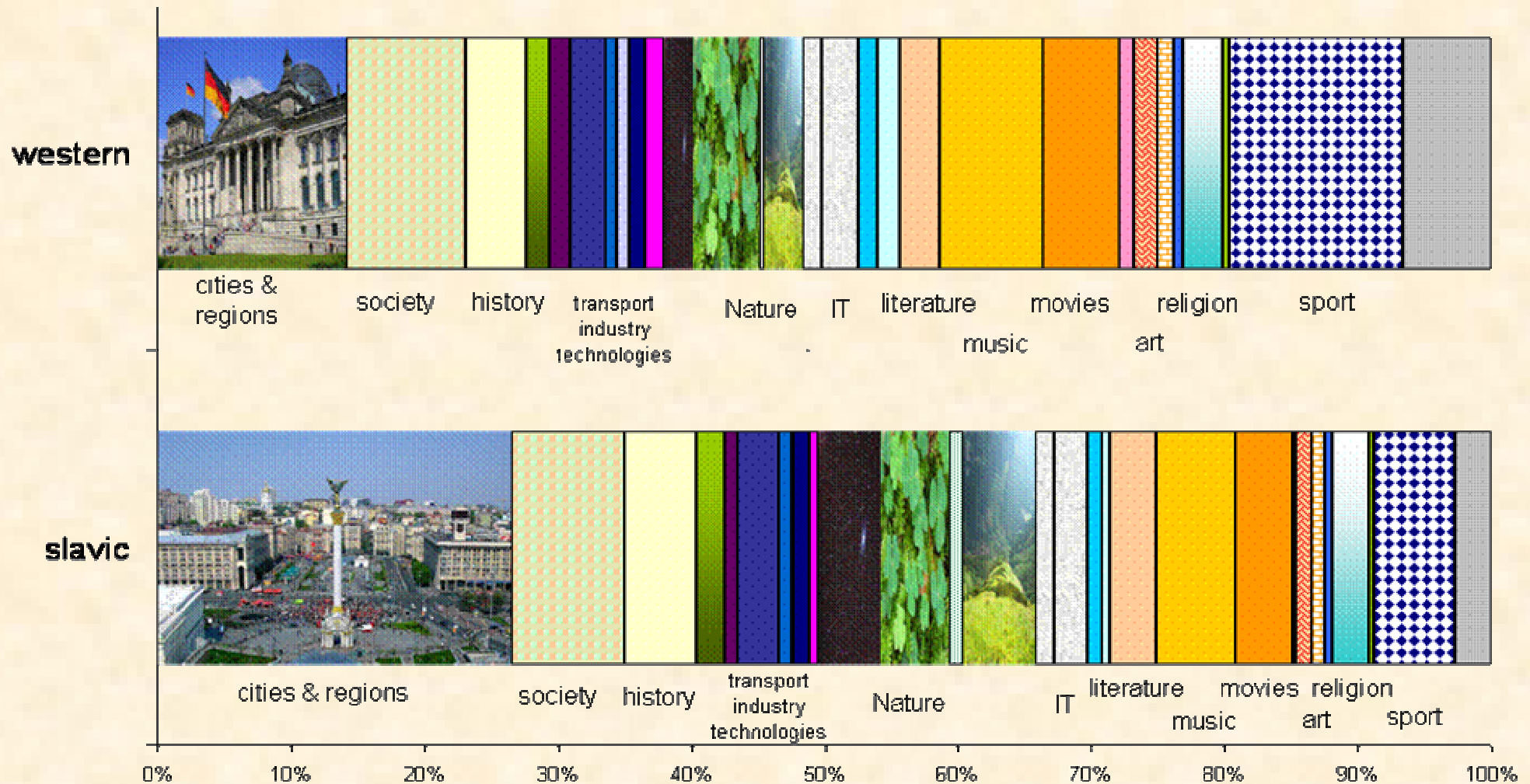




# 4.2 Західні і слов'янські ВП за темою

Західні – більше спорту

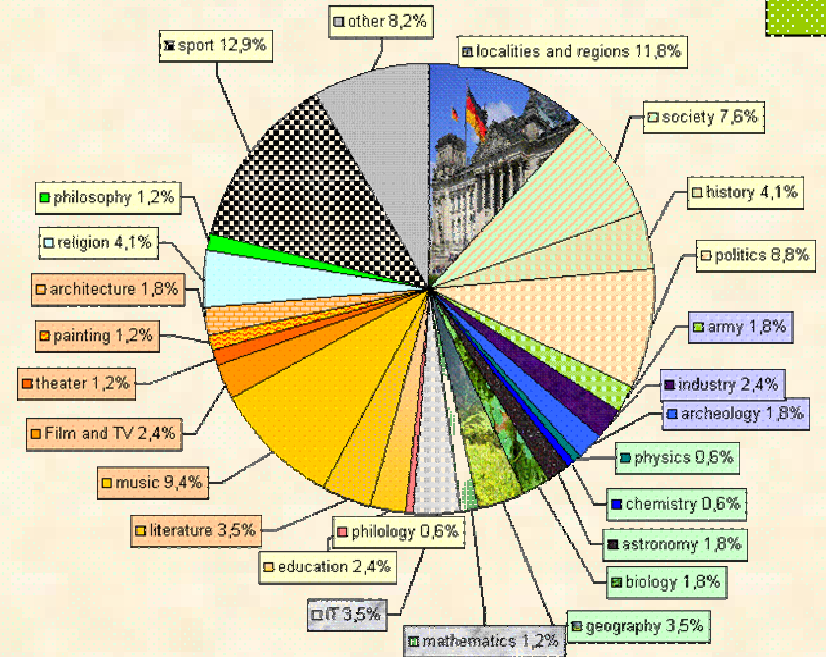
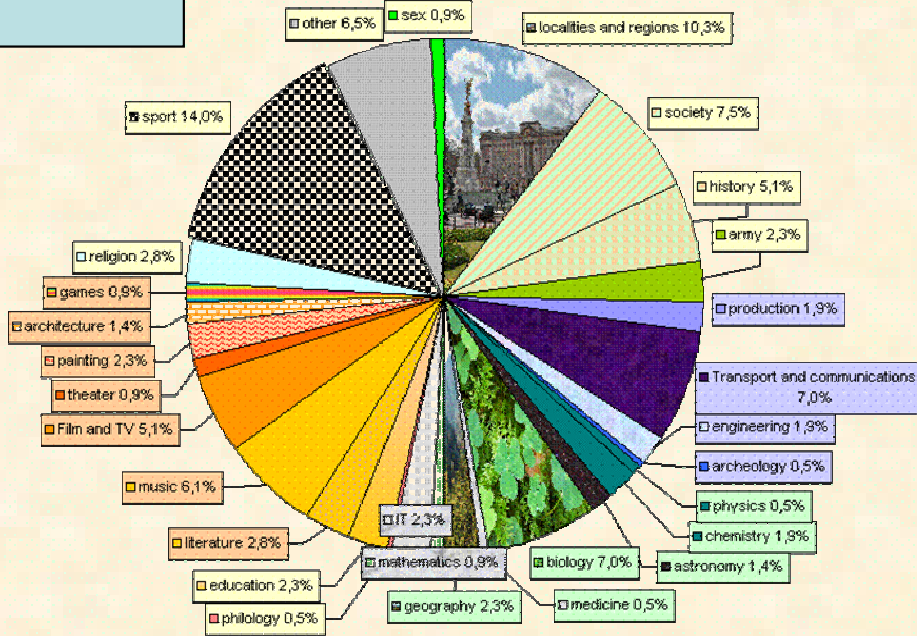
Слов'янські – більше географії



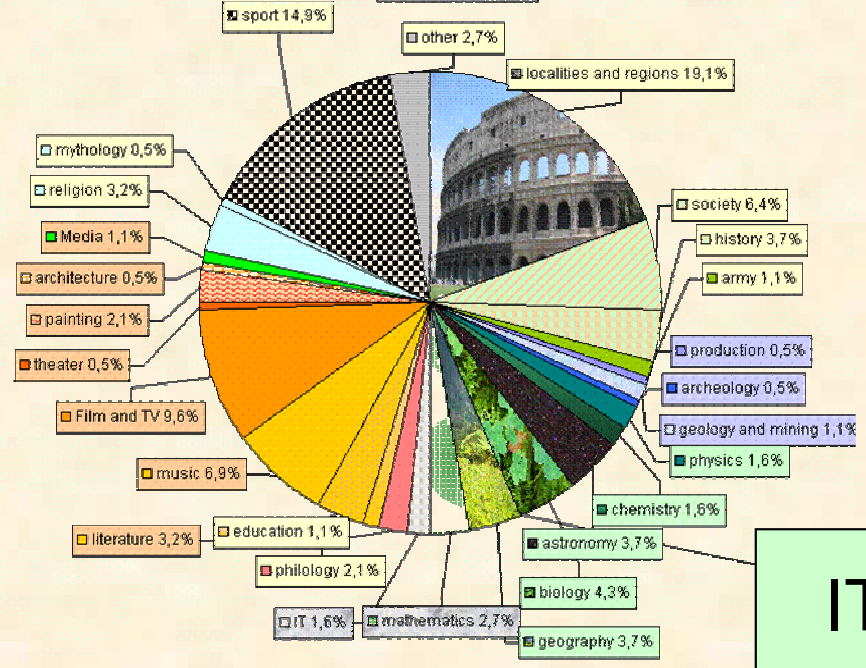
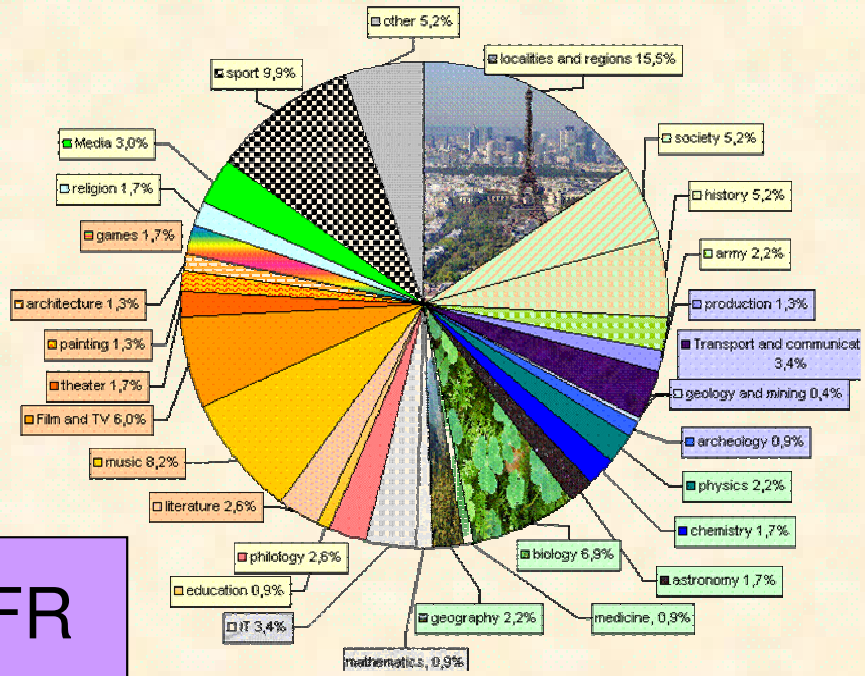
# 4.3 Західні Вікіпедії за тематикою

EN

DE



FR

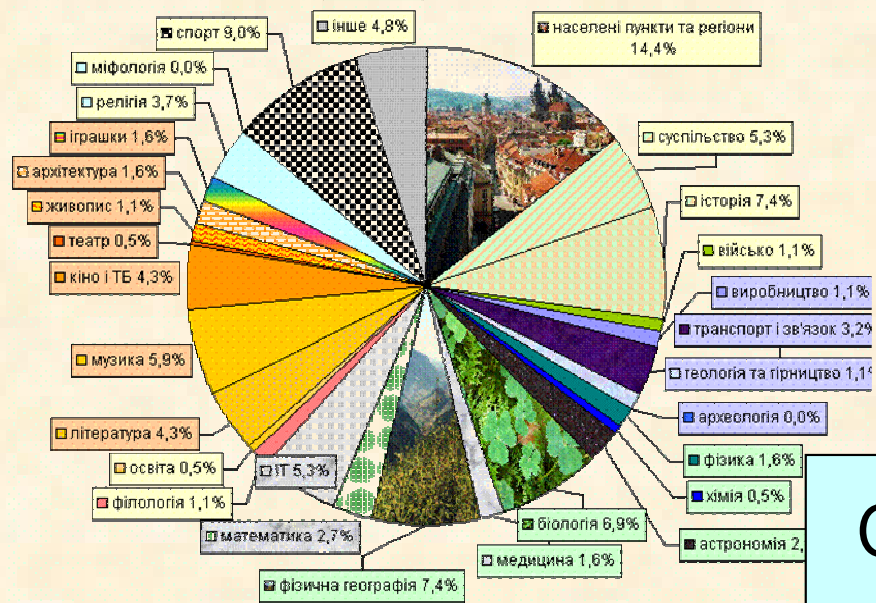
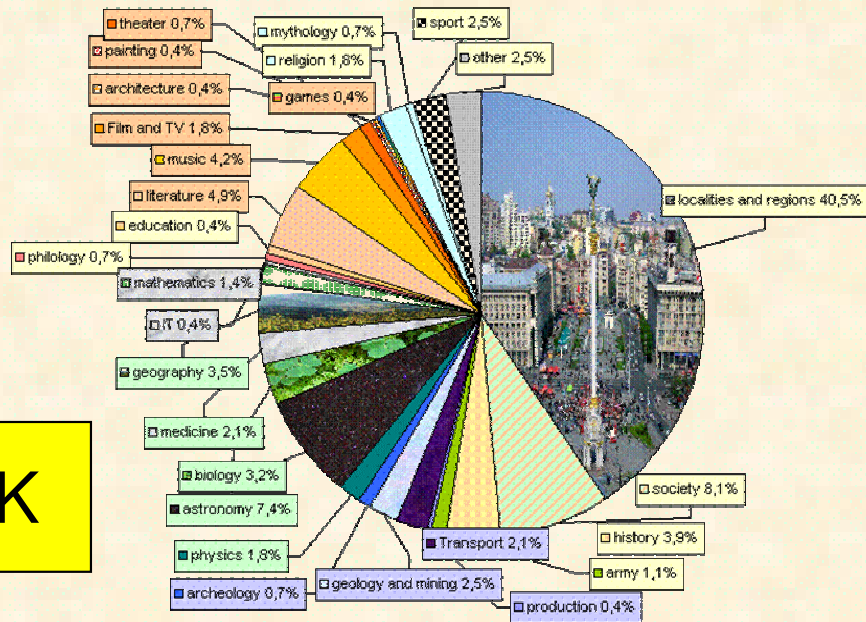
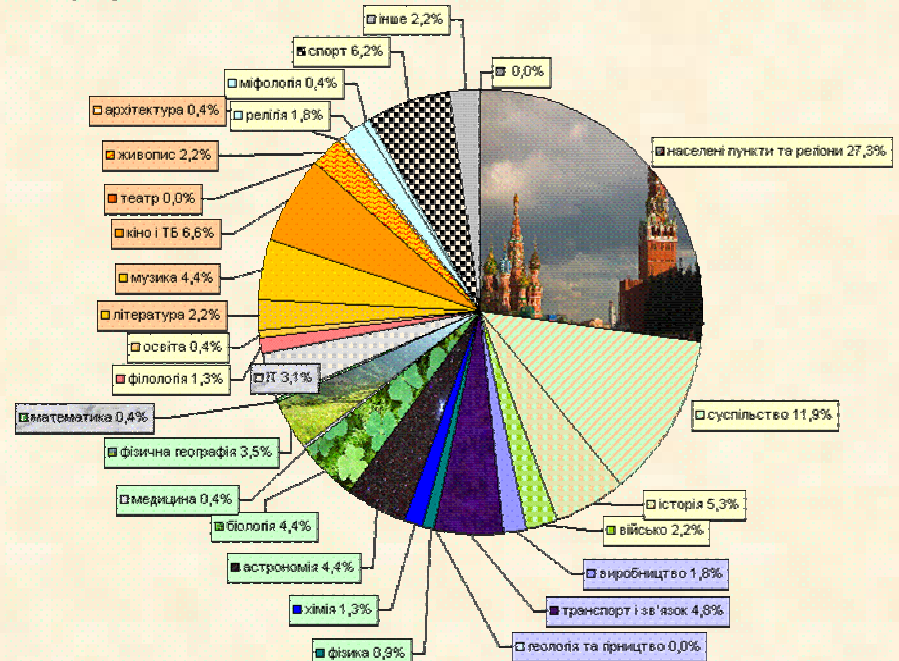
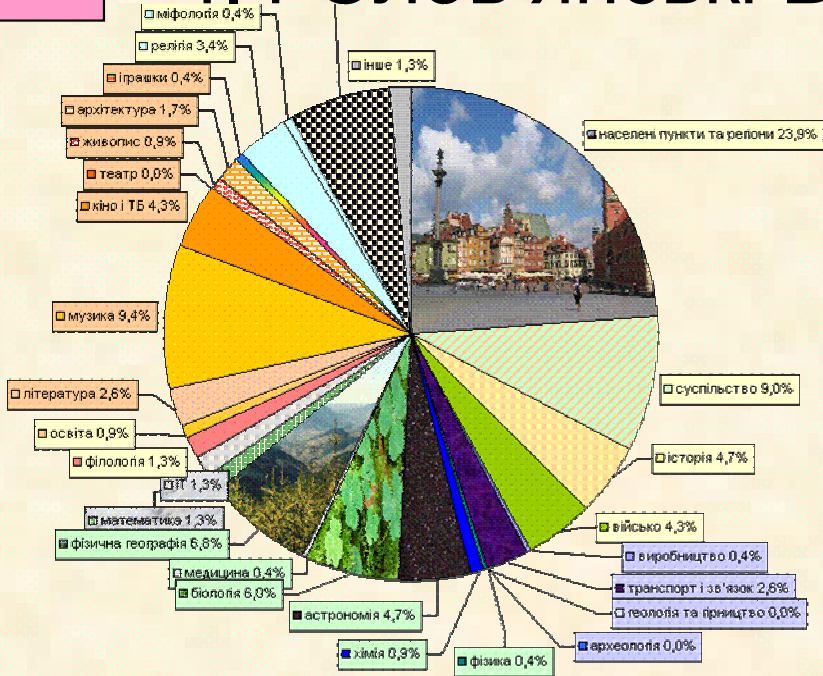


IT

PL

RU

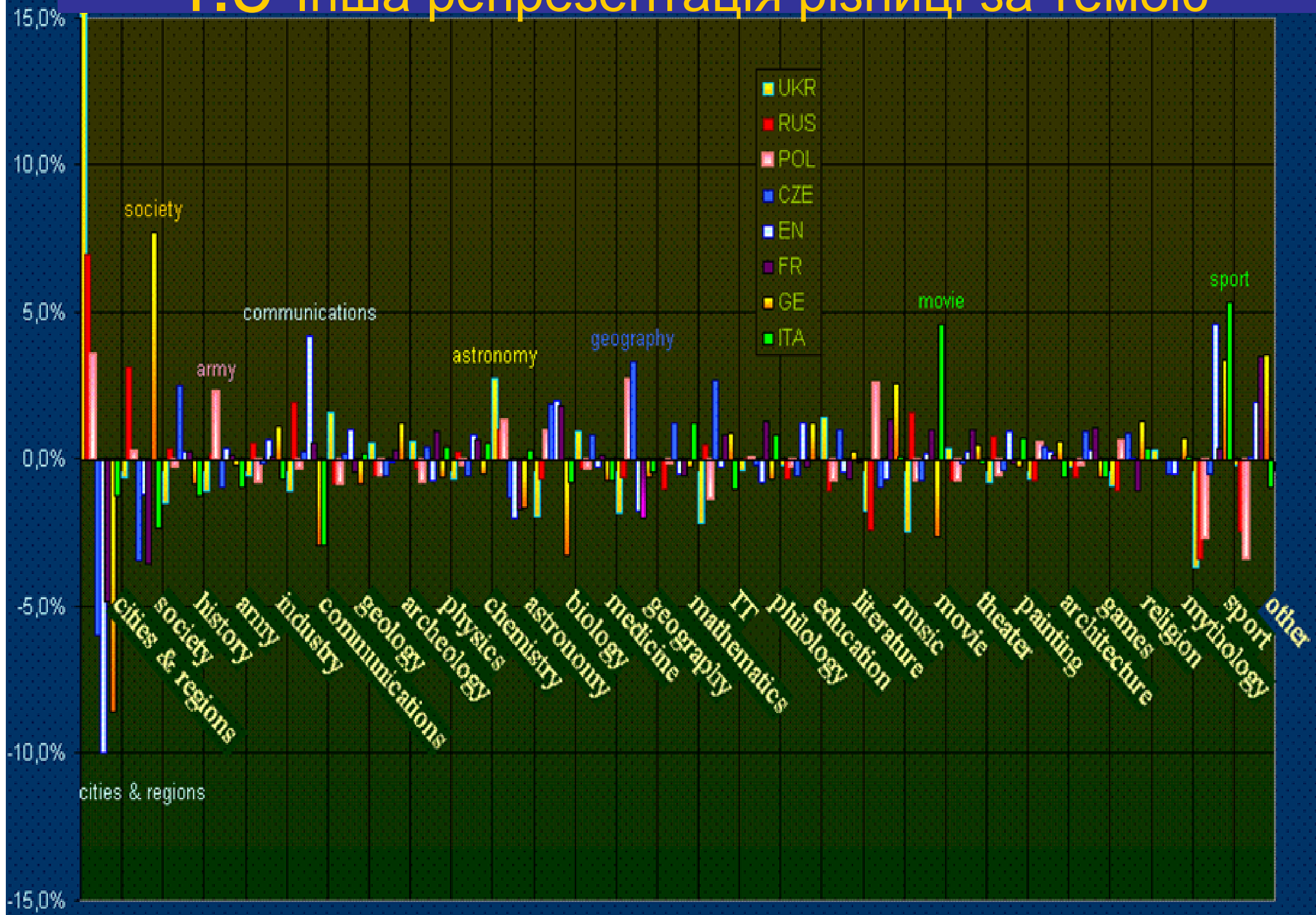
# 4.4 Слов'янські Вікіпедії за тематикою



UK

CS

# 4.5 Інша репрезентація різниці за темою

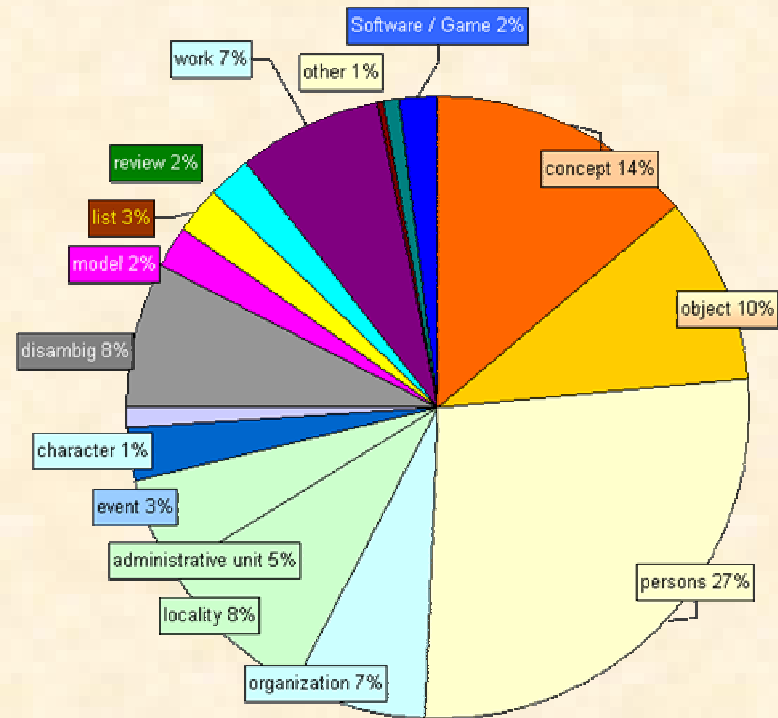
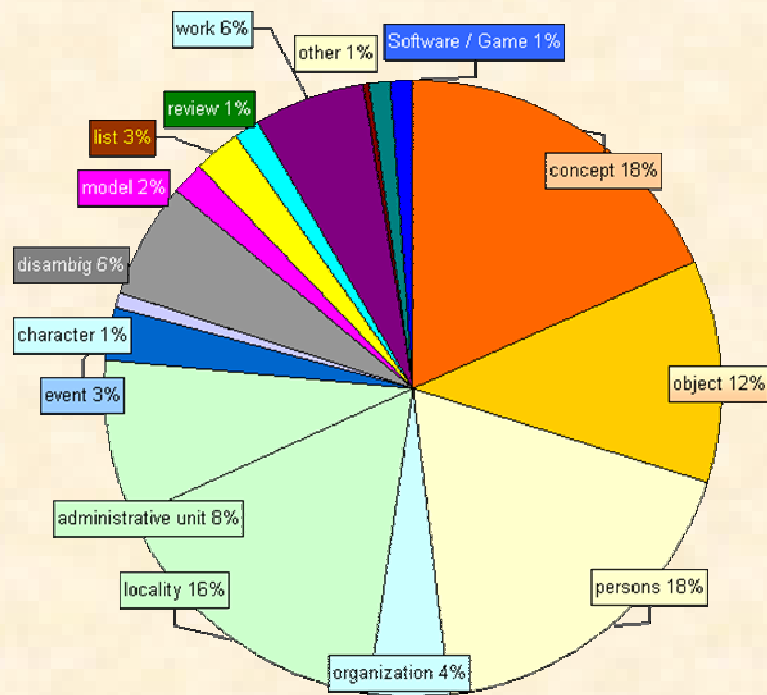




# 4.6

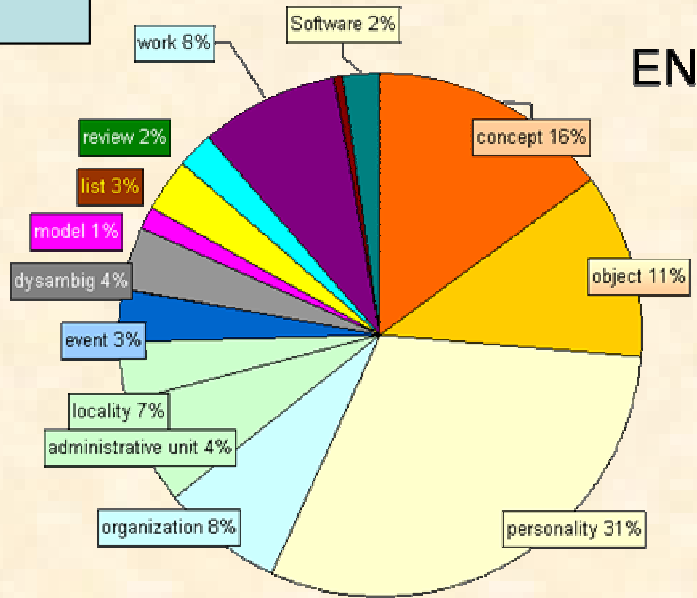
Слов'янські – більше об'єктів і понять

Західні – більше персоналій і компаній

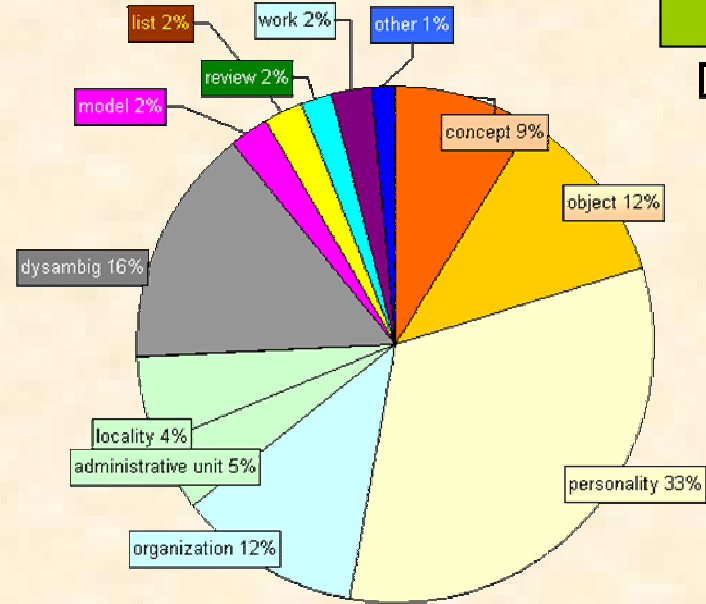


# 4.7 західні Вікіпедії за типом

EN

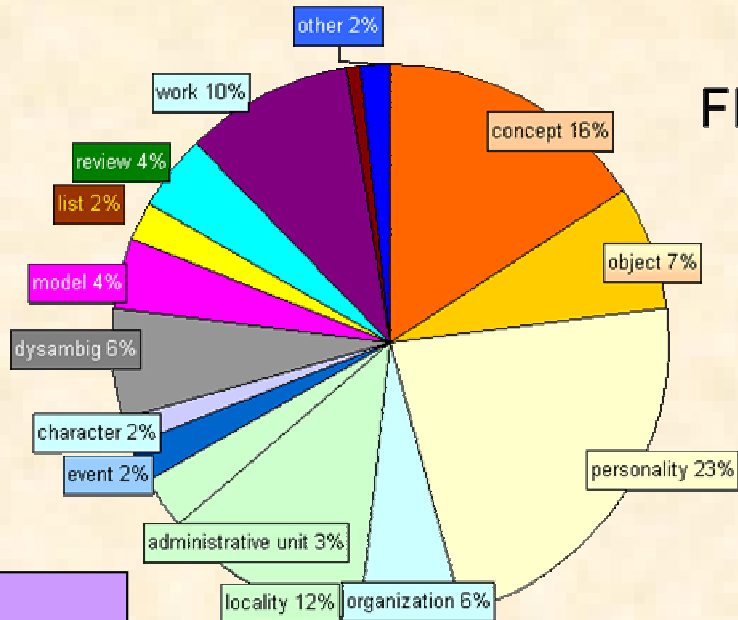


DE



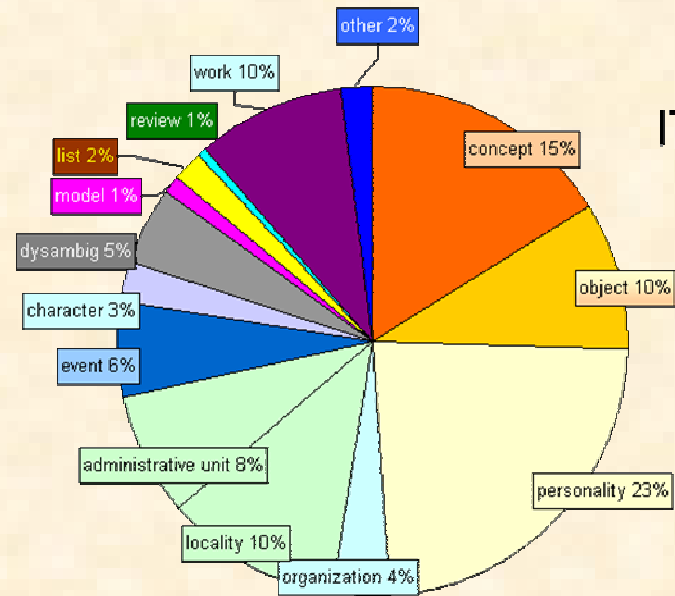
DE

FR



FR

IT

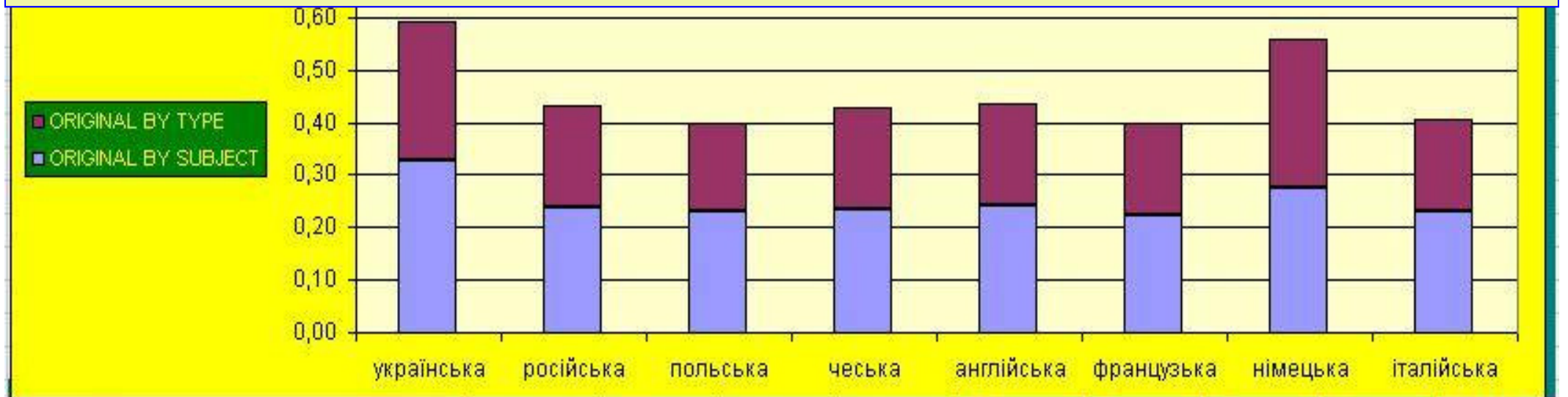


IT



	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!