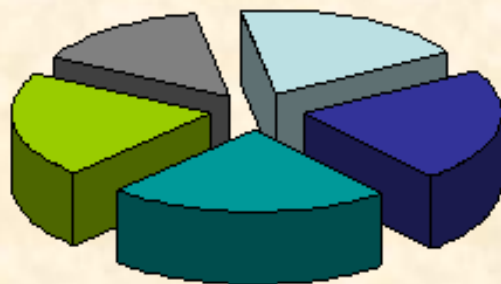


Зміст Вікіпедії

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

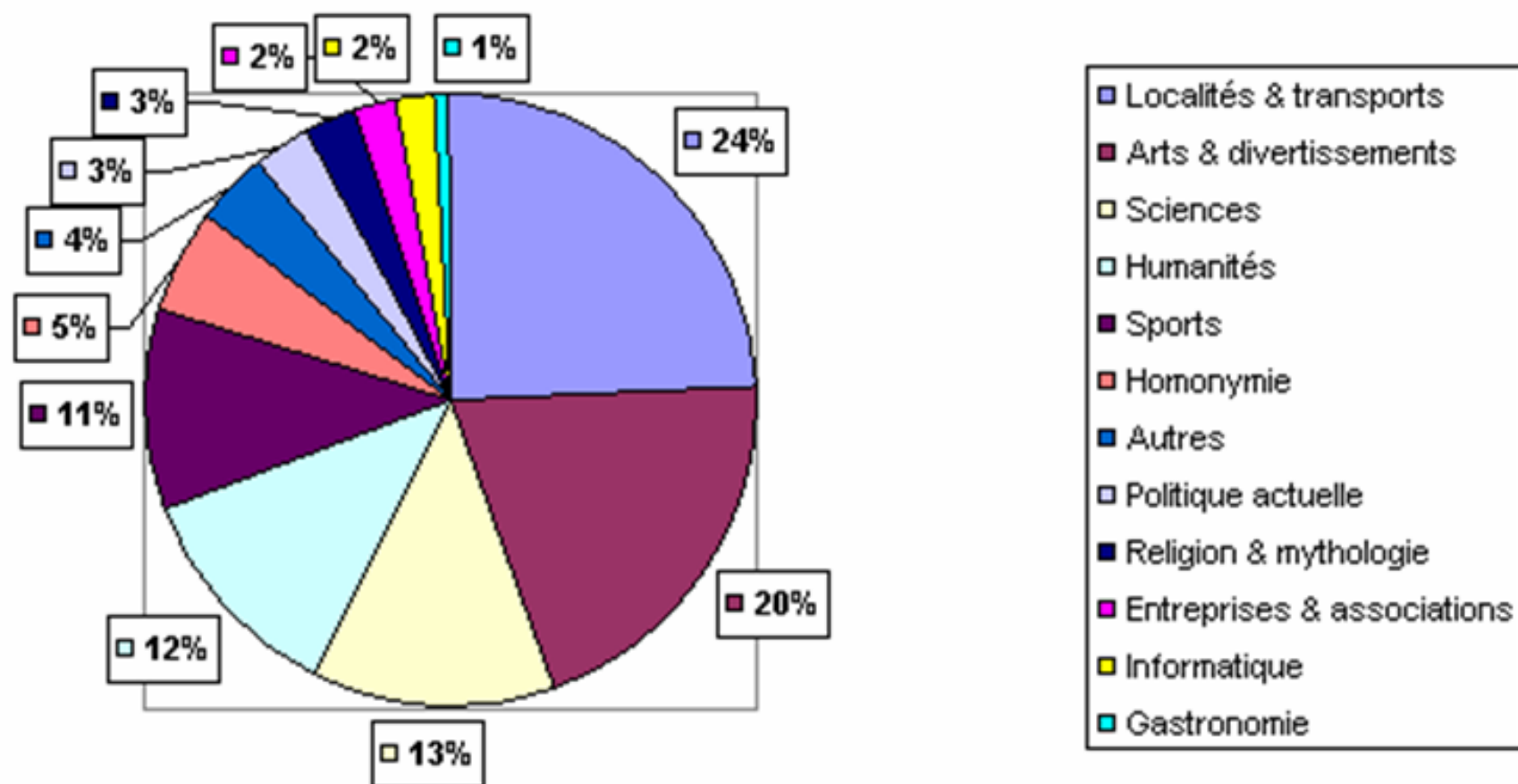


галузі знань

Історія питання

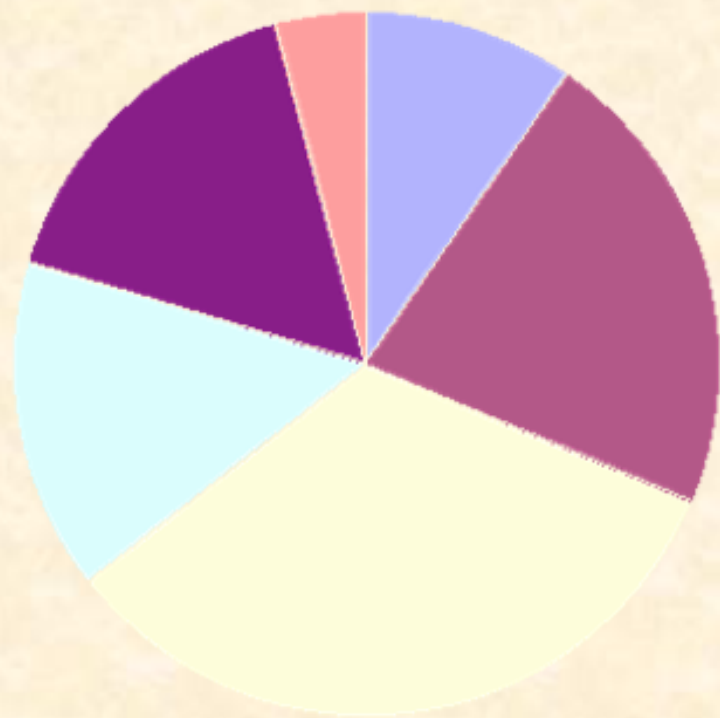
- Французька вікіпедія 2007

Répartition des catégories en pourcentage



Історія питання

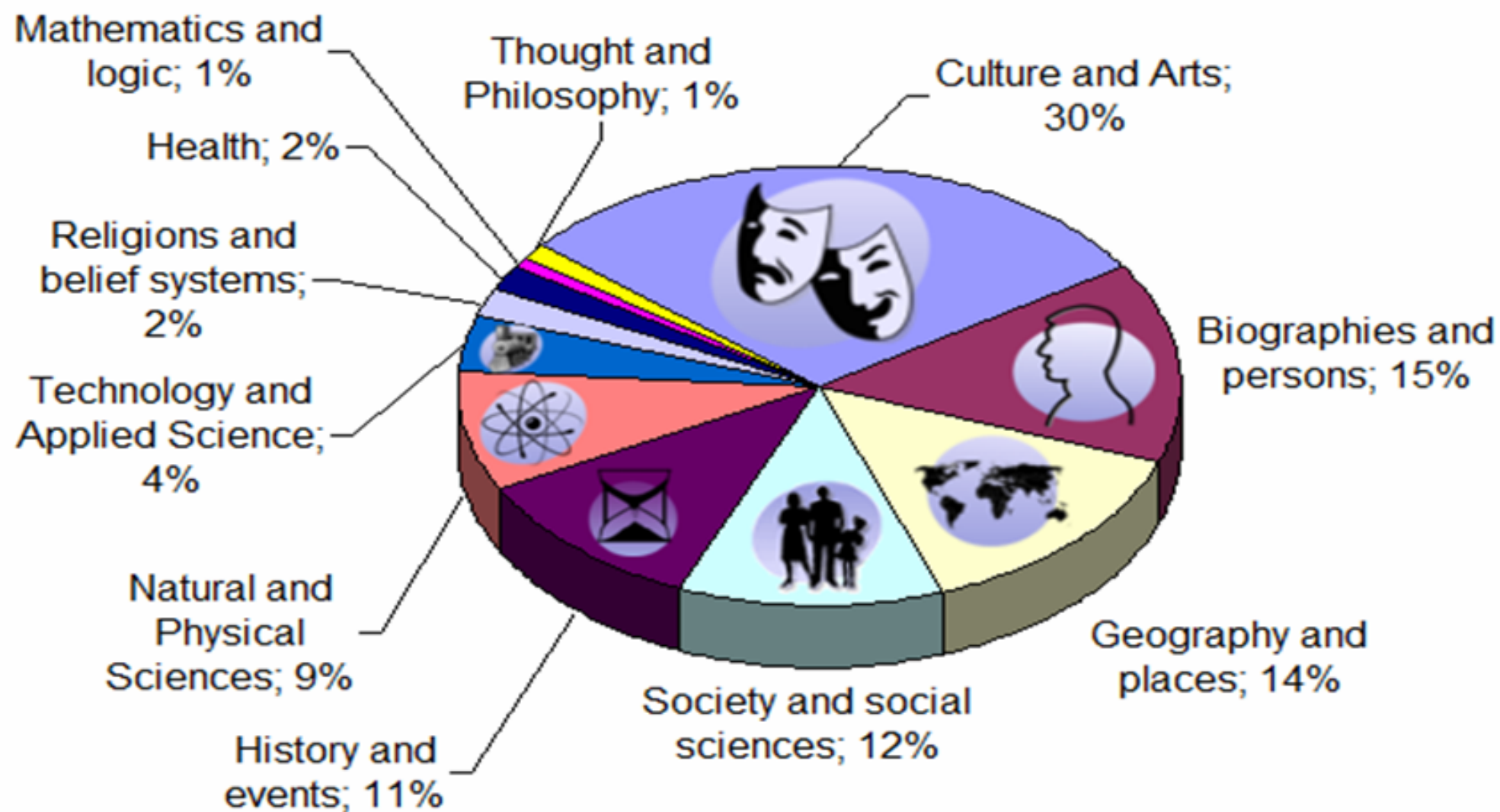
Шведська вікіпедія 2007



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

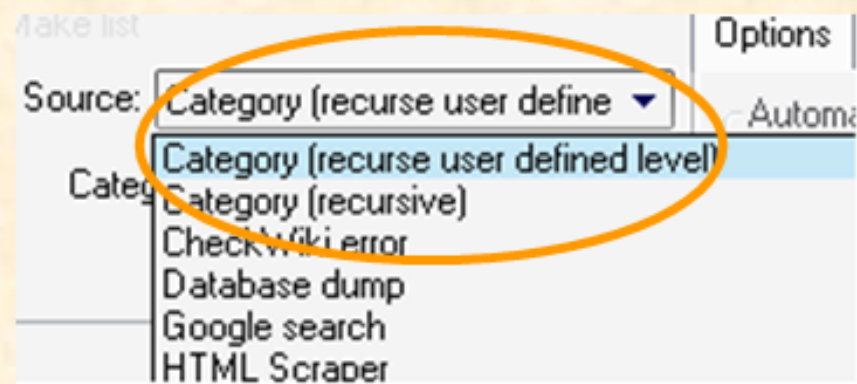
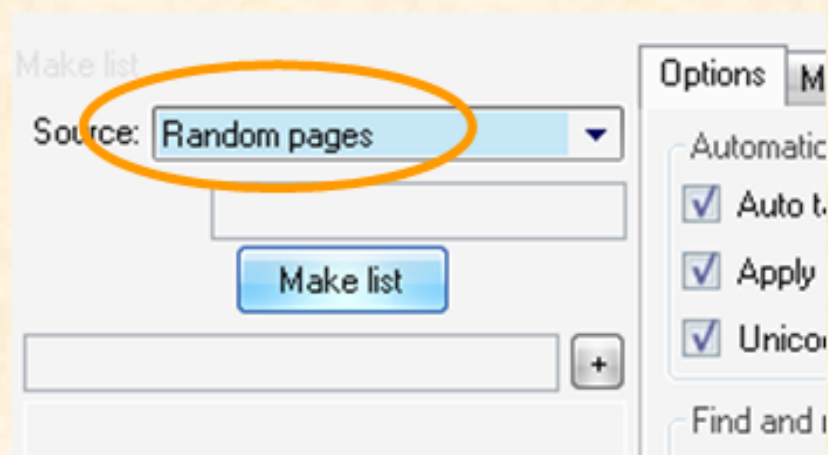
Історія питання

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](#)]

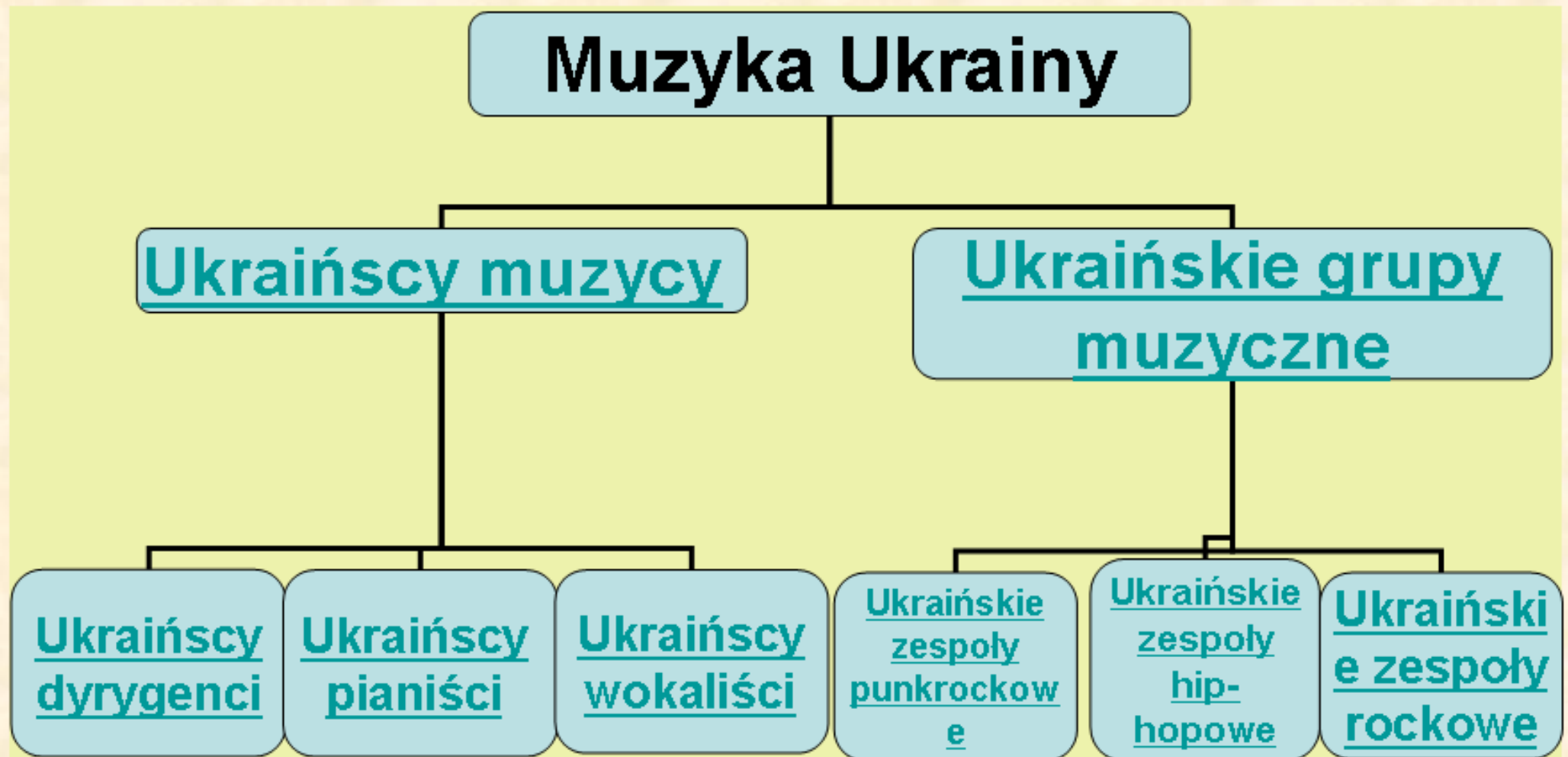
| | |
|-----------|---|
| Мова | uk |
| Проект | wikipedia |
| Глибина | 0 (глибина 0 = не шукати в підкатегоріях) |
| Категорії | Польща |

2.1 Статистична вибірка

| Вибірка і Похибка | En 3,5M+ | PI 800K+ | Ru 750K+ | Uk 300K+ | Cs 200K |
|----------------------|-------------|-------------|-------------|-------------|------------|
| 100 статей 1% | 35K | 8K | 7.5K | 3K | 2K |
| 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, Kategoria:Hudba

Hubda

Hudební technika

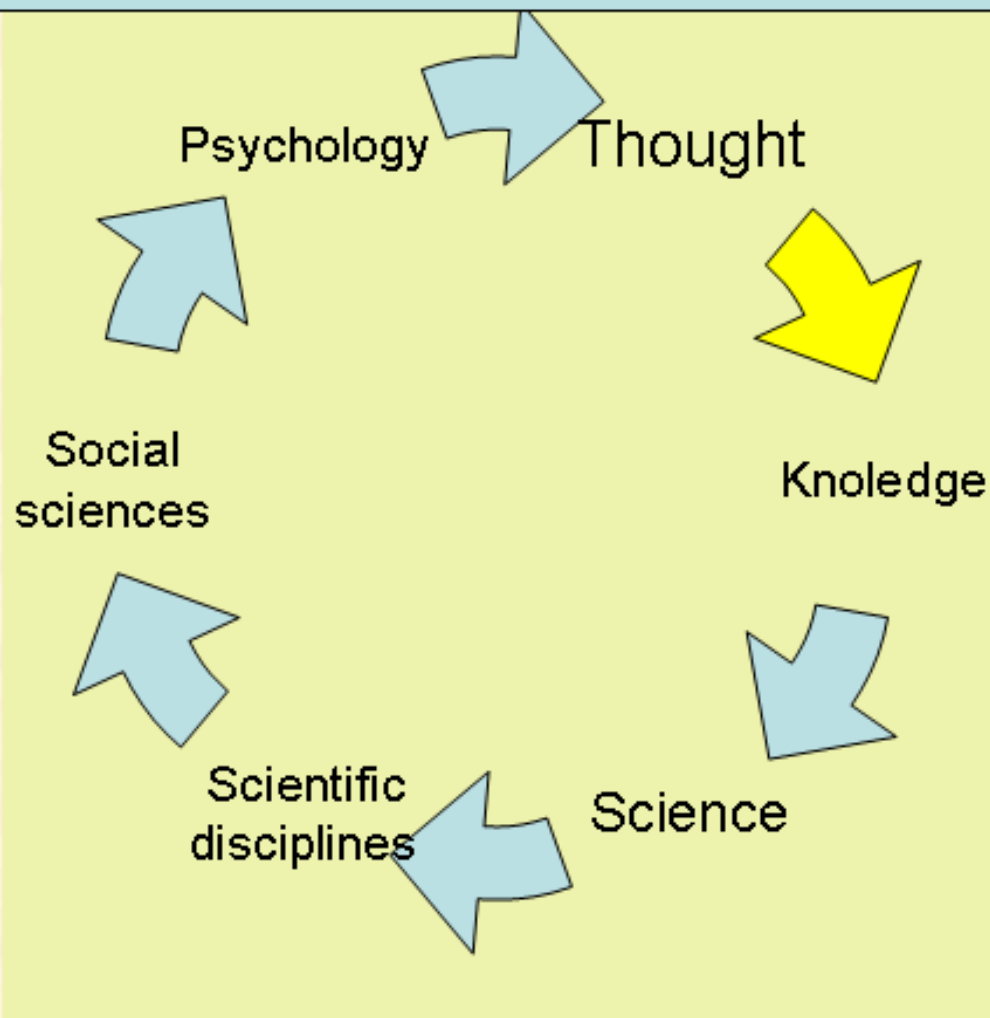
Audiotechnika

Telefonie

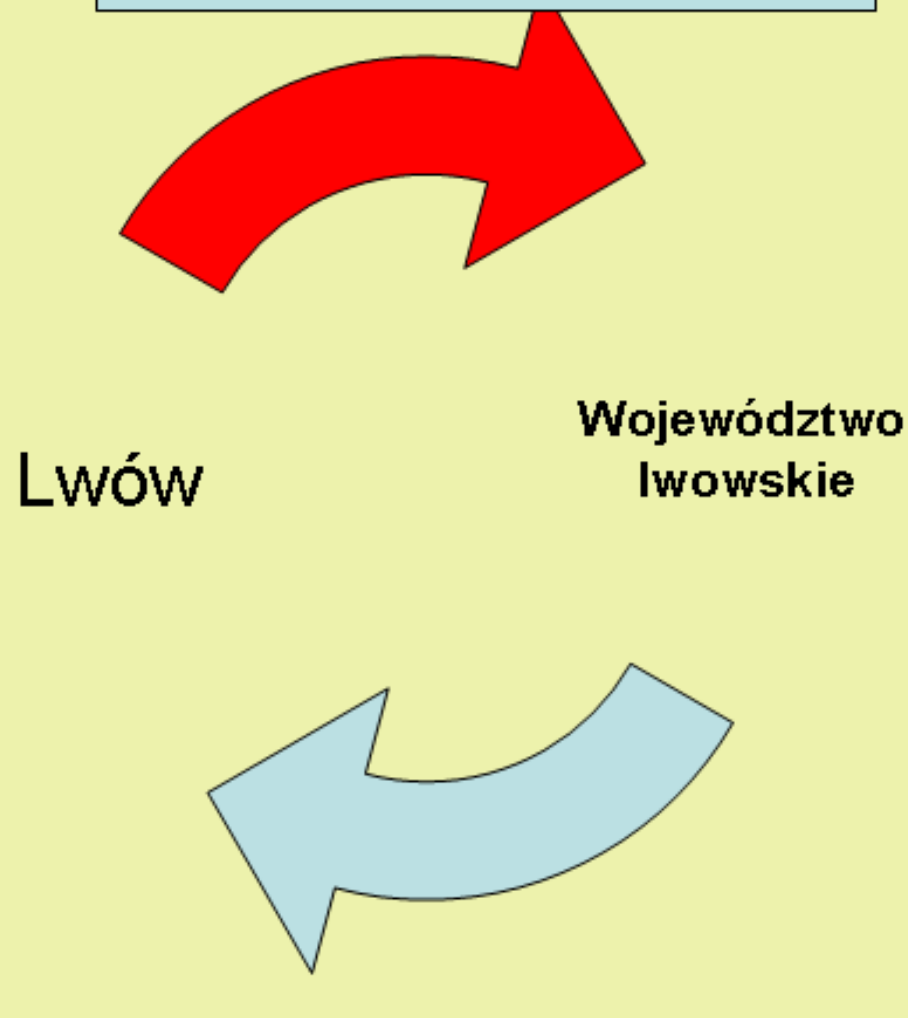
Mobilní telefony

2.4 Рекурсивний аналіз – нетранзитивна категоризація

Класична нетранзитивність в англ. вікі



Колись було у польській...

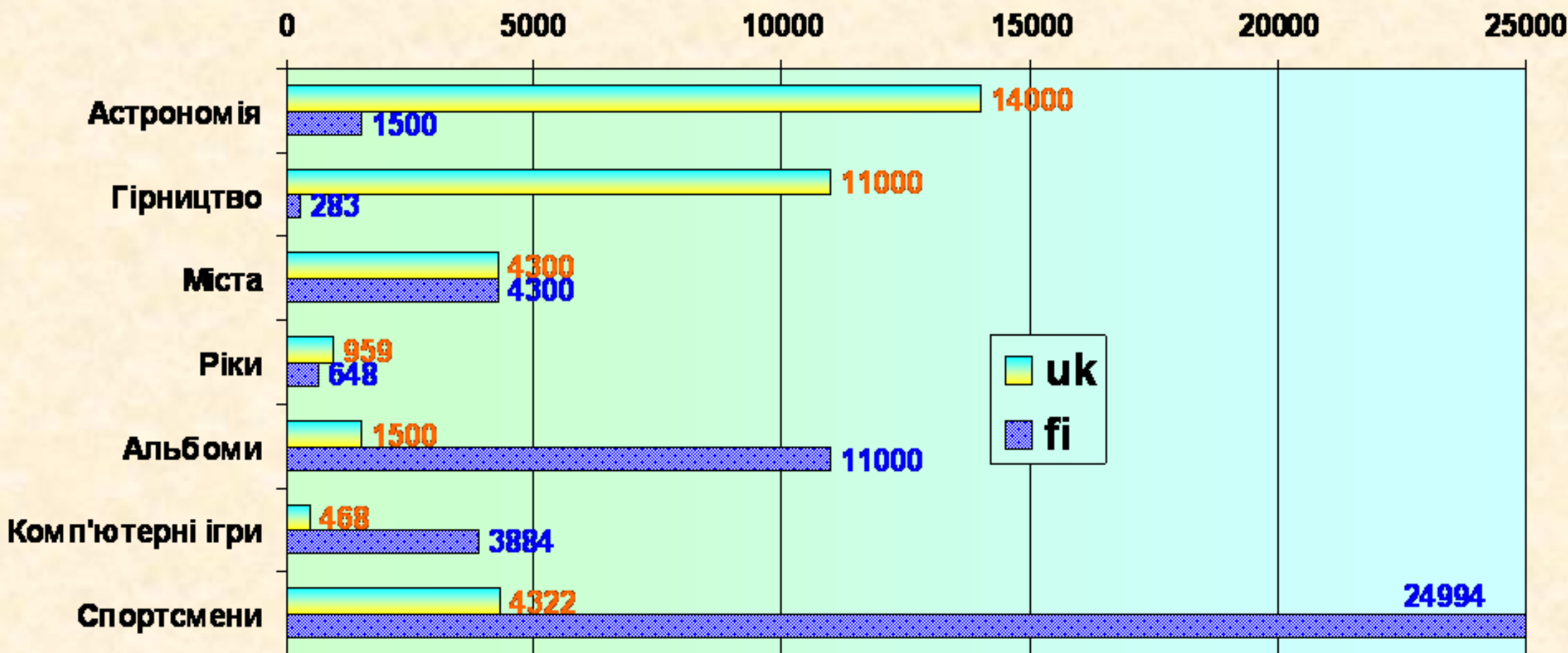


2.5 підсумок

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

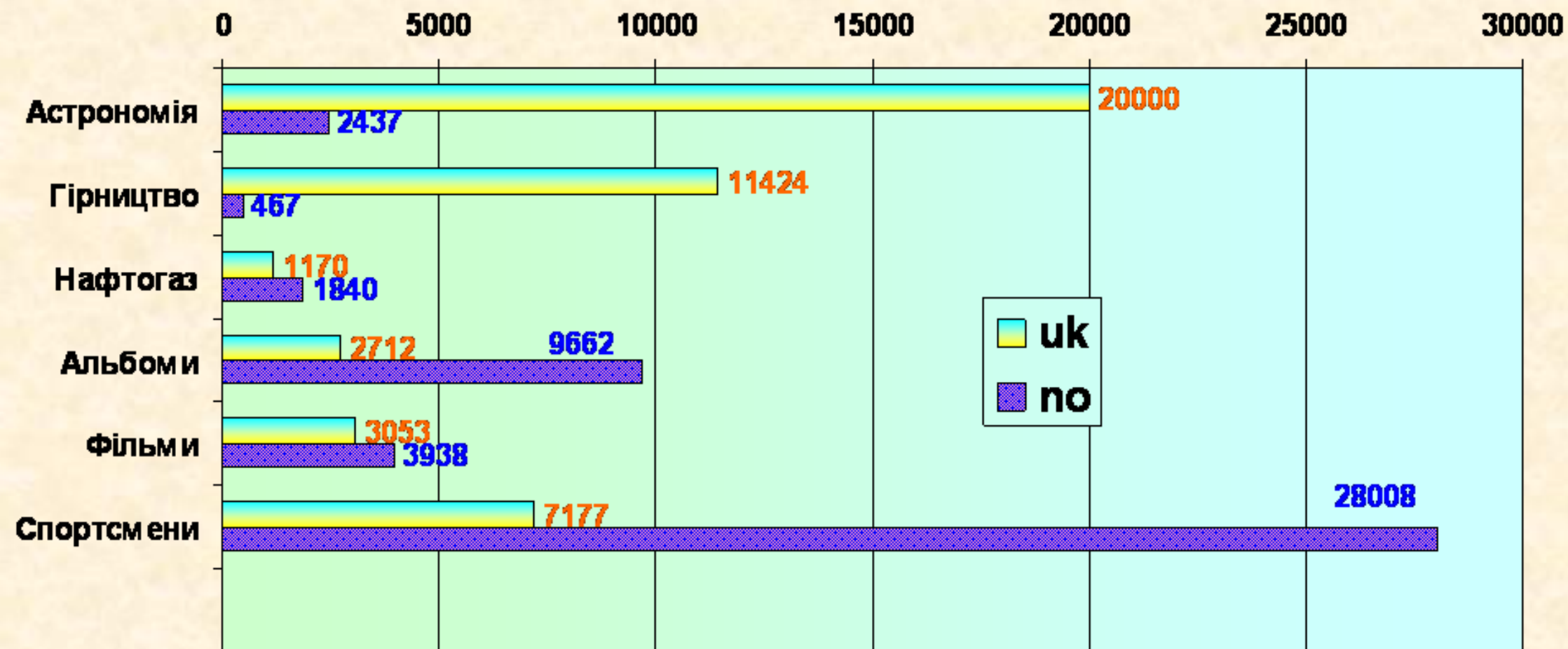
3.1. Спеціальні дослідження – Ukr vs Fin

Порівняння української та фінської вікіпедій на момент обгону
(~ 255K 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:

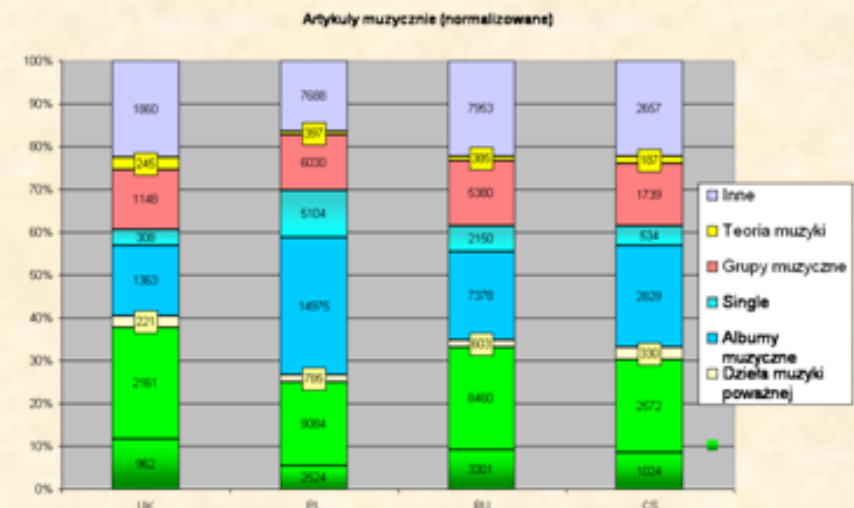
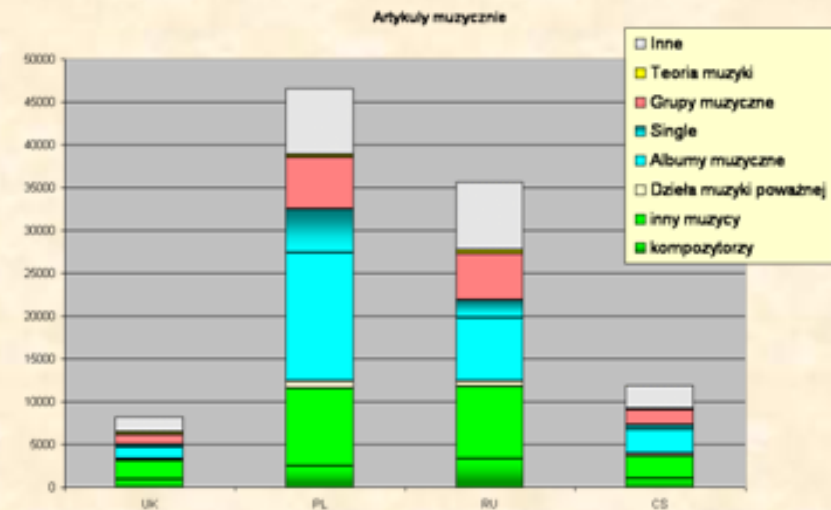
- Абсолютна діаграма
- Нормалізована діаграма
 - Зелений - музиканти
 - Блакитний – альбоми і сингли
 - Червоний – гурти
 - Жовтий – теорія музики

- ❖ UK (Українська)

- ❖ PL (Polska)

- ❖ RU (Русский)

- ❖ CS (Česky)



4.0. Загальні дослідження

1. Створюємо статистичну вибірку

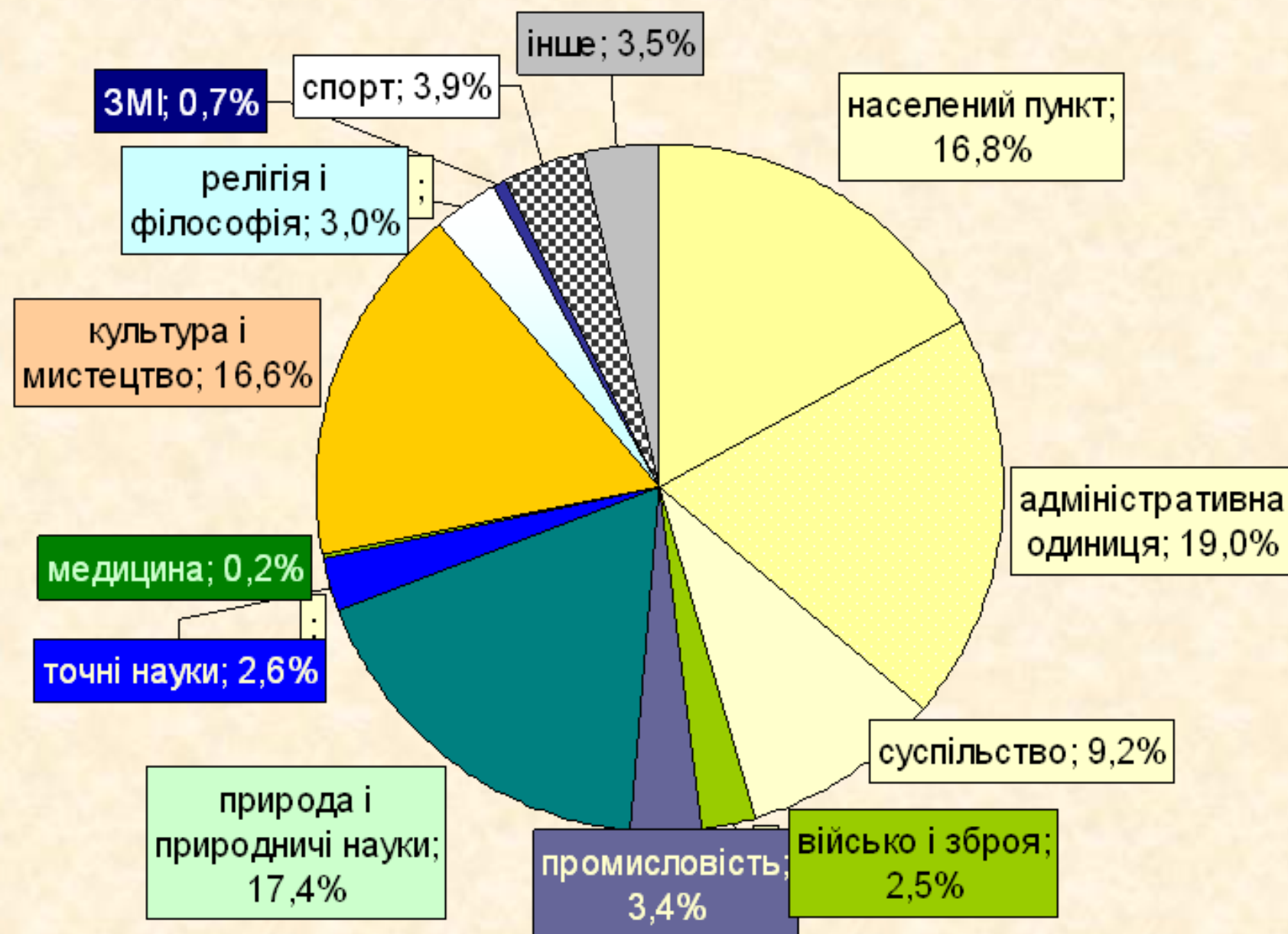
2. Аналізуємо:

- За темою
- За типом

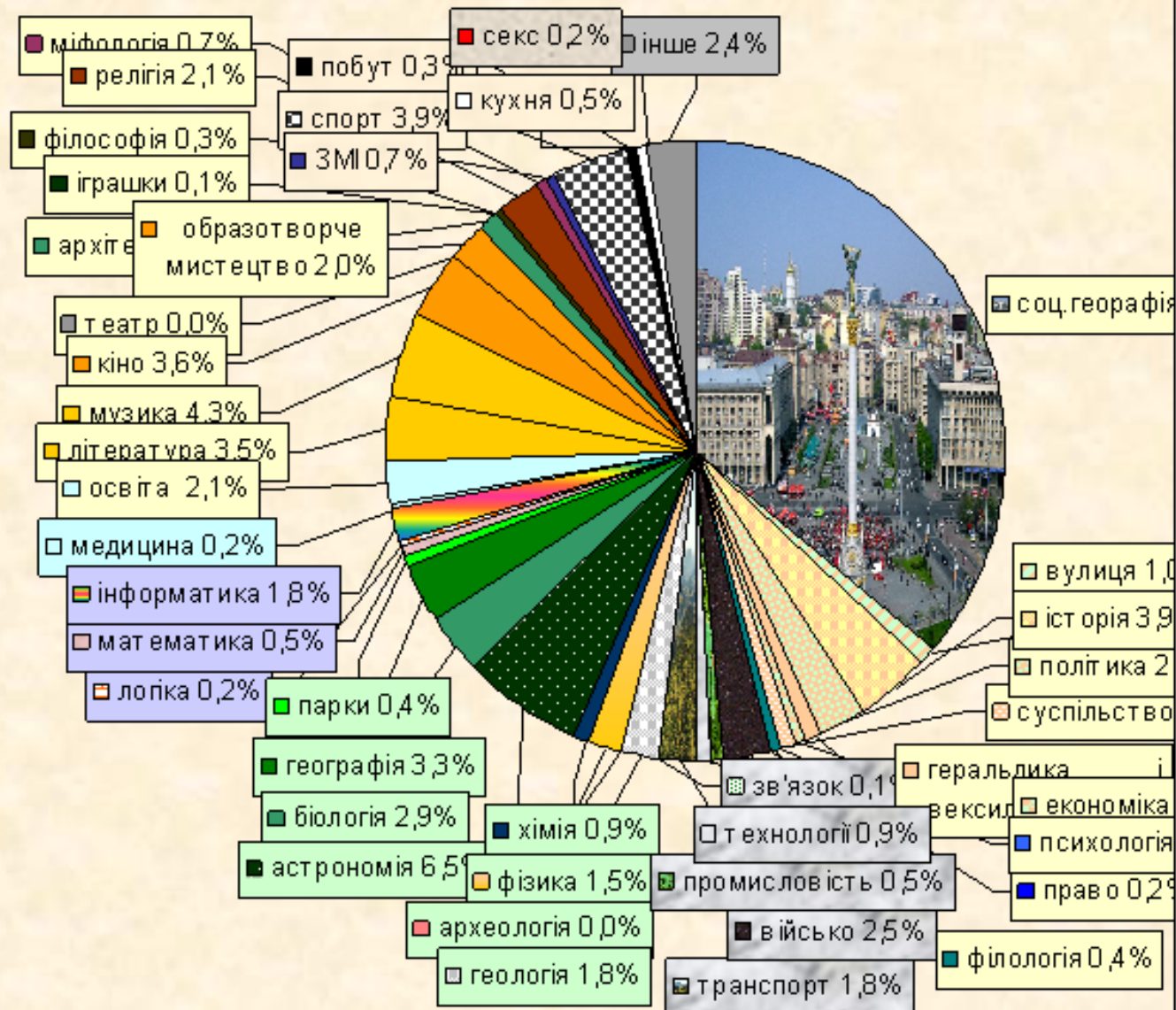
3. Робимо діаграму

| 1 | стаття | тематика | тип предмету статті | повторюваність | повторюваність | зведена таблиця | | | зведена таблиця | | | |
|----|-------------------------------|---------------------|---------------------|----------------|----------------|-------------------|----------|-------------|-----------------|---------------|----------|------|
| 2 | Palác Michny z Václava | архітектура | об'єкт | 22 | 3 | тематика | знайдено | тематика | часта | тематика | знайдено | част |
| 3 | Vinoh | архітектура | початок | 51 | 3 | соц. географія | 27 | населенн | 14.4% | початок | 51 | 2 |
| 4 | Alfred Mosher Butts | архітектура | людина | 41 | 3 | суспільство | 10 | суспільств | 5.3% | об'єкт | 22 | 1 |
| 5 | Sojuz TMA-17 | астрономія | модель | 8 | 4 | історія | 14 | історія | 7.4% | людина | 41 | 2 |
| 6 | Messier 67 | астрономія | об'єкт | 22 | 4 | військо | 2 | військо | 1.1% | організація | 9 | 1 |
| 7 | Národné středisko pro letectv | астрономія | організація | 9 | 4 | промисловість | 2 | вербодит | 1.1% | населений пу | 19 | 1 |
| 8 | Kari (místic) | астрономія | об'єкт | 22 | 4 | транспорт і зв'яз | 6 | транспорт | 3.2% | адміністрація | 7 | 1 |
| 9 | Cardale Babington | біологія | людина | 41 | 13 | геологія та прир | 2 | геологія п | 1.1% | людя | 8 | 1 |
| 10 | Centrální dogma molekulární b | біологія | початок | 51 | 13 | археологія | 0 | археологія | 0.0% | персонаж | 2 | 1 |
| 11 | Kosatec bahenni | біологія | початок | 51 | 13 | фінанса | 3 | фінанса | 1.6% | дисамбі | 9 | 1 |
| 12 | Dynaktin | біологія | початок | 51 | 13 | мова | 1 | мова | 0.5% | модель | 8 | 1 |
| 13 | Homo antecessor | біологія | початок | 51 | 13 | астрономія | 4 | астроном | 2.1% | список | 6 | 1 |
| 14 | Koryjí | біологія | початок | 51 | 13 | біологія | 13 | біологія | 6.9% | співак | 1 | 1 |
| 15 | Mitozom | біологія | початок | 51 | 13 | медицина | 3 | медицина | 1.6% | споруда | 0 | 1 |
| 16 | Památný strom | біологія | початок | 51 | 13 | до. географія | 14 | фінанса та | 7.4% | твор | 11 | 1 |
| 17 | Tegumentum | біологія | початок | 51 | 13 | математика | 5 | математик | 2.7% | видання | 1 | 1 |
| 18 | Dvoudřložné | біологія | початок | 51 | 13 | IT | 10 | IT | 5.3% | лице | 6 | 1 |
| 19 | Sběrač rosný | біологія | початок | 51 | 13 | філософія | 2 | філософія | 1.1% | | | |
| 20 | Lewisuchus | біологія | початок | 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 | Záhar Prálegn | військо | людина | 41 | 2 | кіно | 8 | кіно і ТВ | 4.3% | | | |
| 24 | Hydrodynamická vertikální zon | геологія та принцип | початок | 51 | 2 | театр | 1 | театр | 0.6% | | | |
| 25 | Richard Lydekker | геологія та принцип | людина | 41 | 2 | животні | 2 | животні | 1.1% | | | |
| 26 | | 1586 дата | список | 6 | 3 | архітектура | 3 | архітектура | 1.6% | | | |
| 27 | | 1965 дата | список | 6 | 3 | графика | 3 | графика | 1.6% | | | |
| 28 | Ochotnýj gad | дисамбі | дисамбі | 9 | 9 | реалія | 7 | реалія | 3.7% | | | |
| 29 | Horčička | дисамбі | дисамбі | 9 | 9 | міфологія | 0 | міфологія | 0.0% | усього (без д | 200 | 18 |
| 30 | Obřáhání Bma | дисамбі | дисамбі | 9 | 9 | спорт | 17 | спорт | 9.0% | | | |
| 31 | DeSoto County | дисамбі | дисамбі | 9 | 9 | мист | 9 | мист | 4.8% | | | |
| 32 | Latka | дисамбі | дисамбі | 9 | 9 | усього (без д | 188 | | 100.0% | | | |
| 33 | П | дисамбі | дисамбі | 9 | 9 | дисамбі | 9 | | | | | |
| 34 | інше | дисамбі | дисамбі | 0 | 9 | інше | 9 | | | | | |

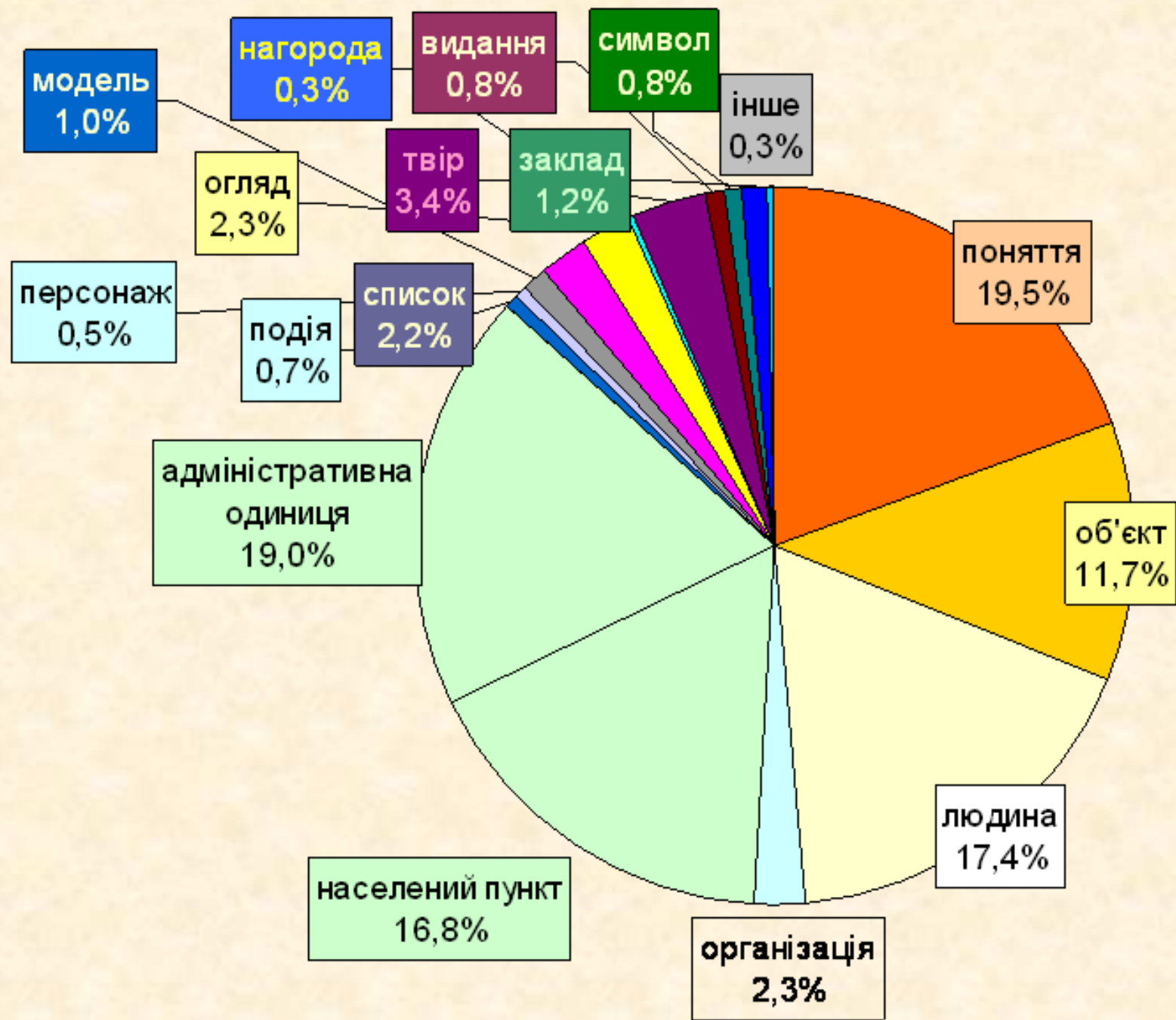
Вибірка-1000 української Вікіпедії



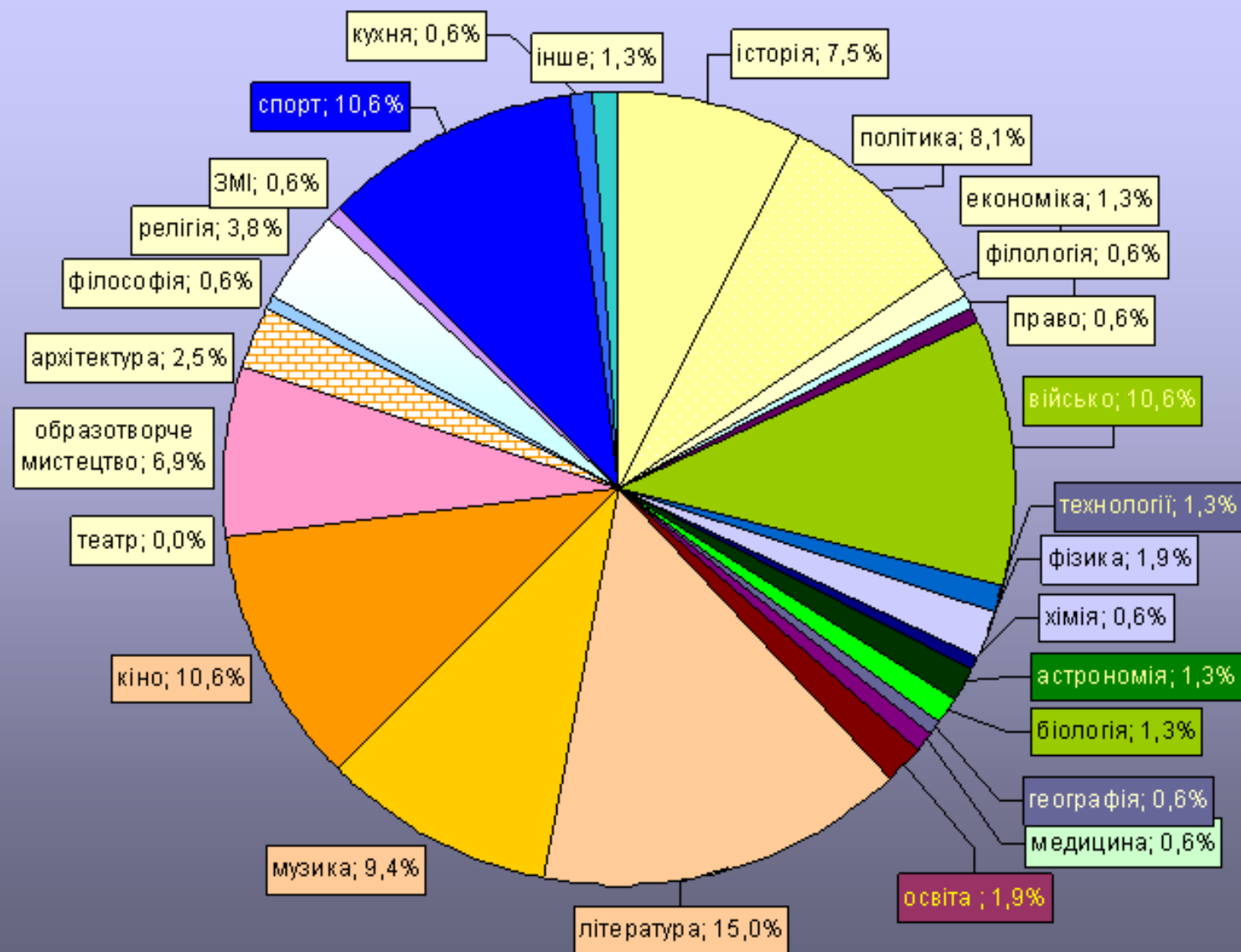
Статистична вибірка української Вікіпедії



- соціогеографія
- вулиця
- історія
- політика
- геральдика і вексиллологія
- 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



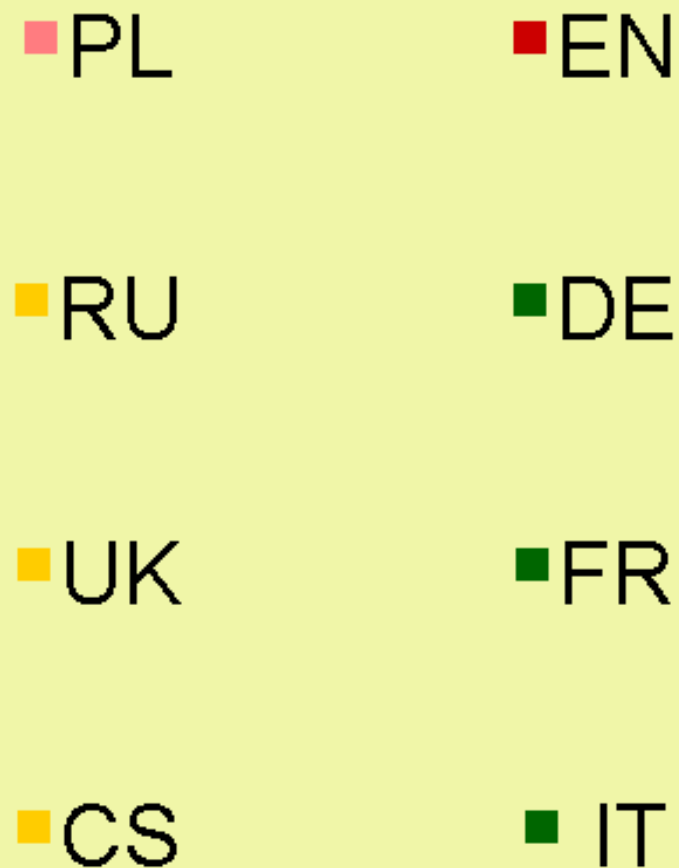
Персоналії за професією



4.1. Міжмовні порівняння

8 Вікіпедій у порівнянні.

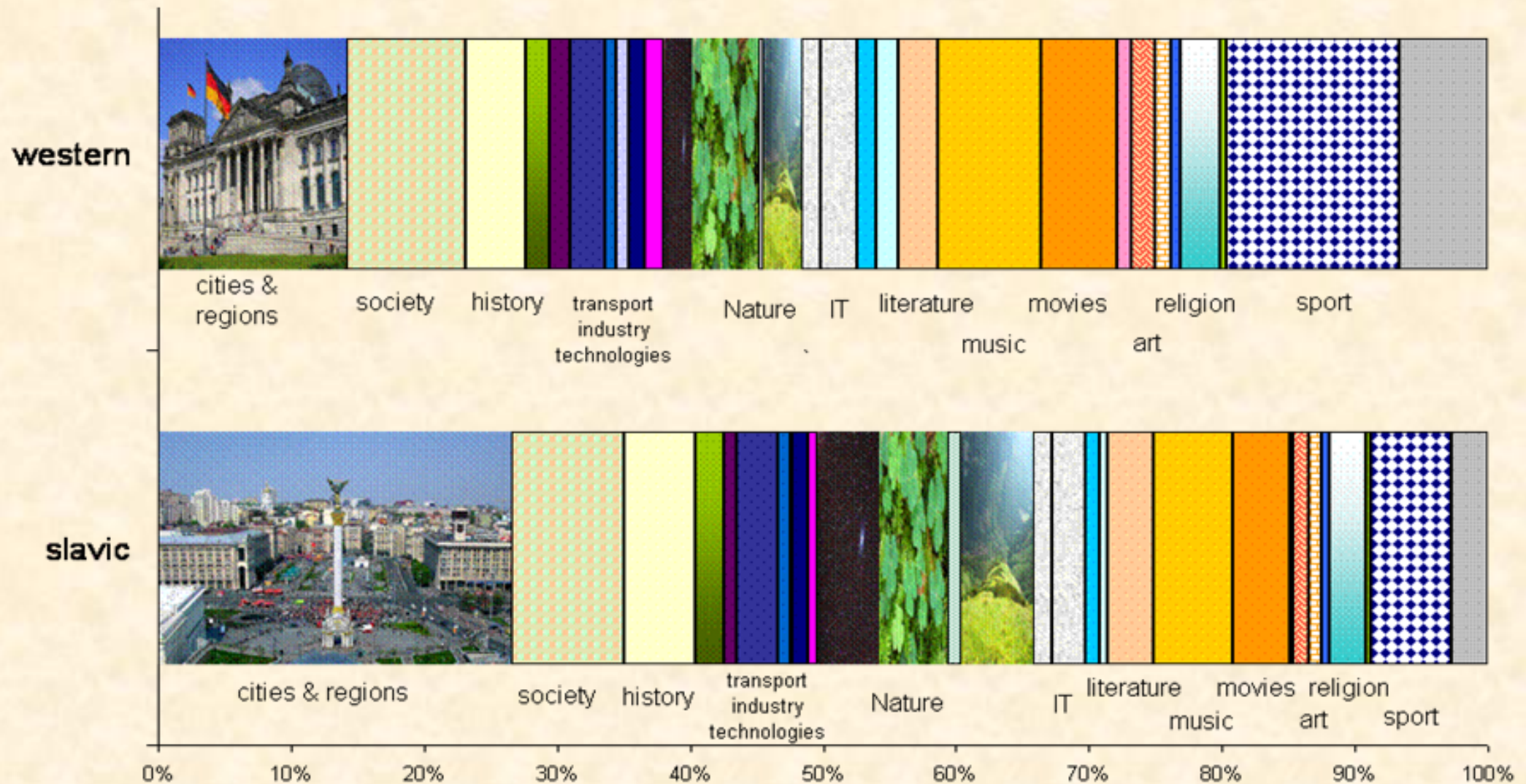
Статистичні вибірки по 200-300 статей, дослідження липня 2011, презентовані на Вікіманії 2011



4.2 Західні і слов'янські ВП за темою

Західні – більше спорту

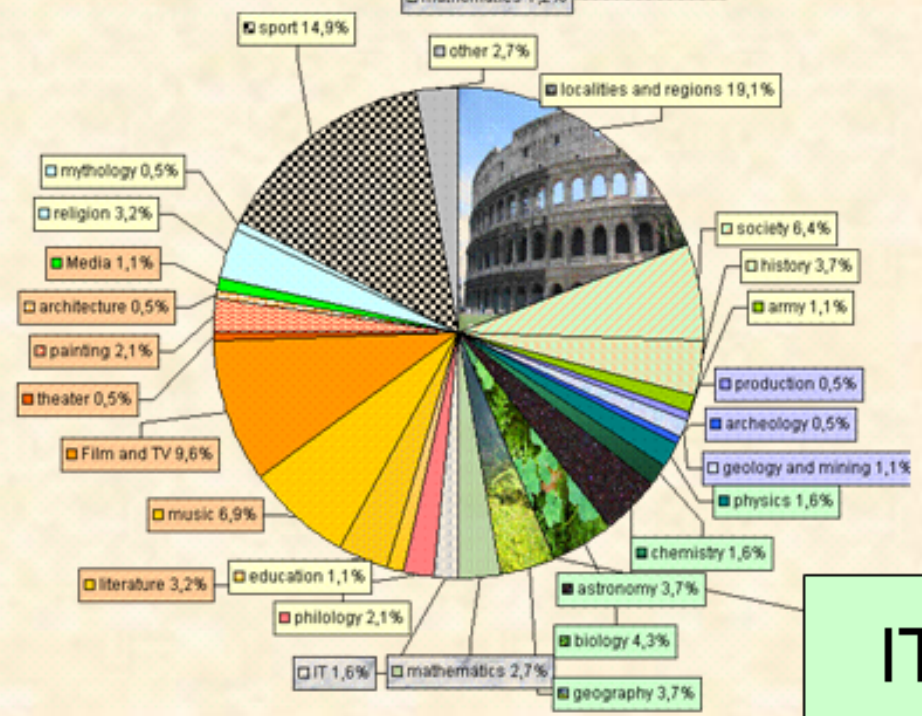
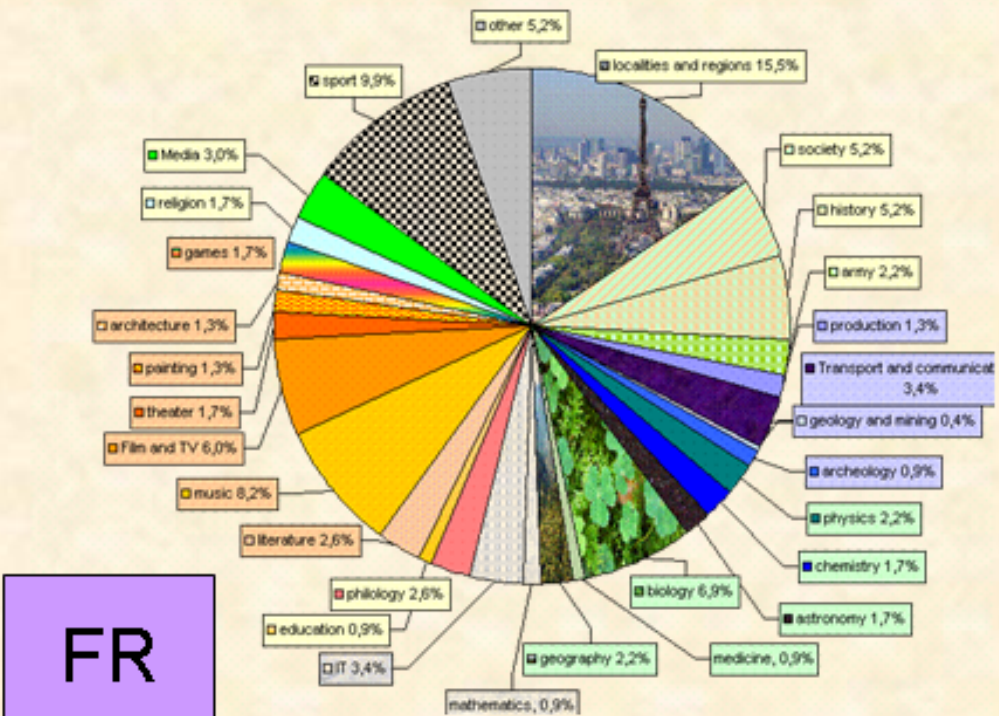
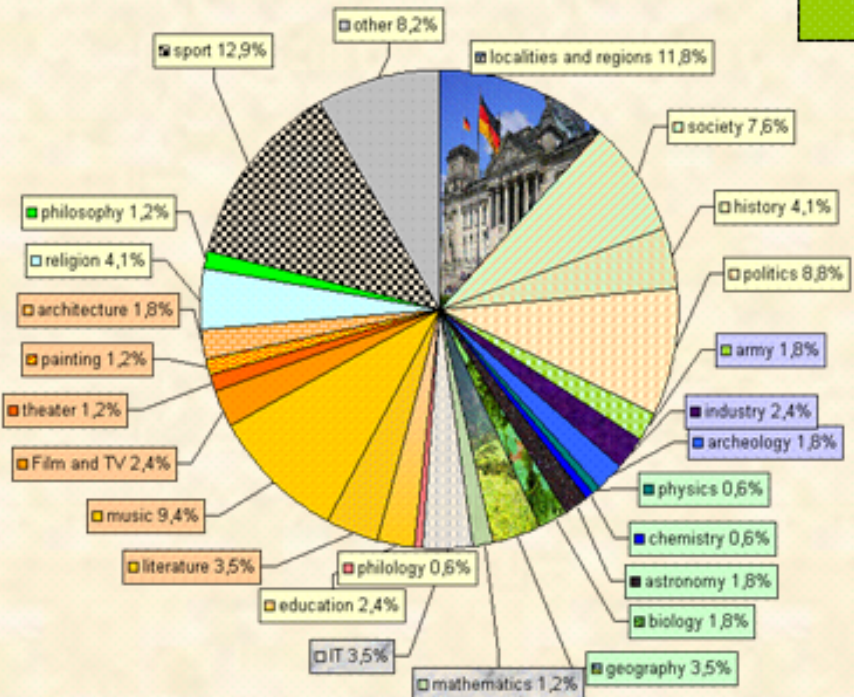
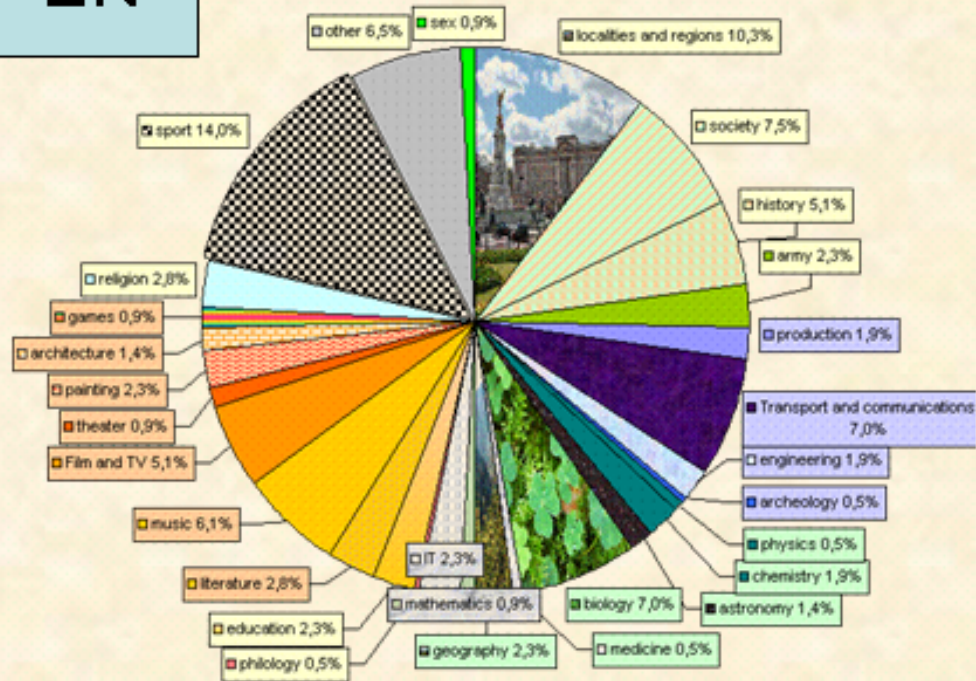
Слов'янські – більше географії



4.3 Західні Вікіпедії за тематикою

EN

DE



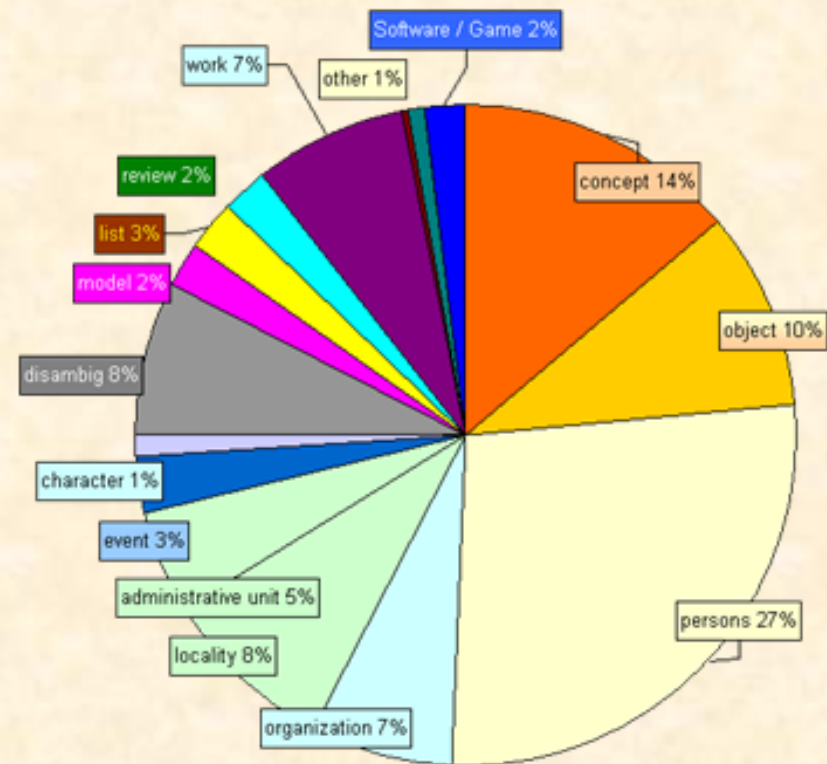
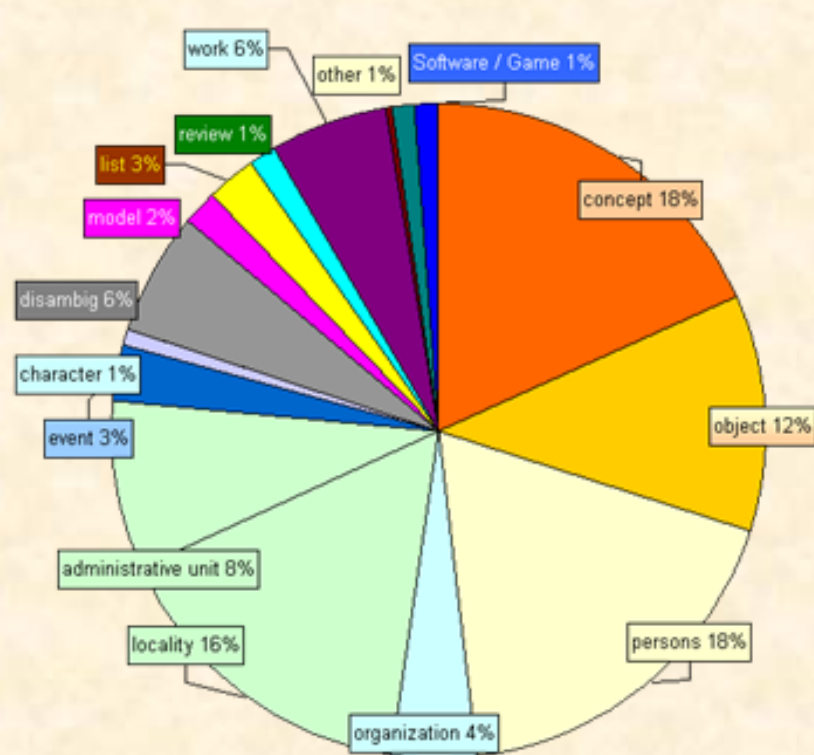
FR

IT

4.6

Слов'янські – більше об'єктів і понять

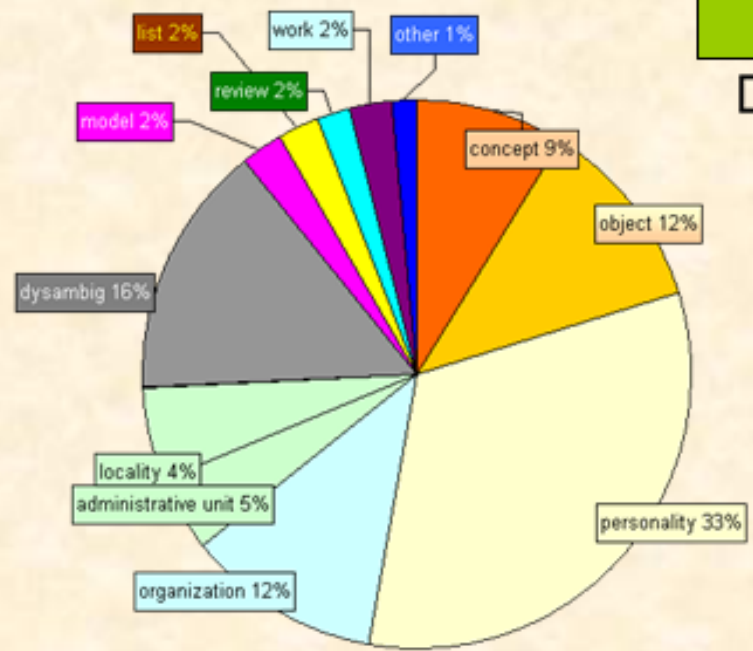
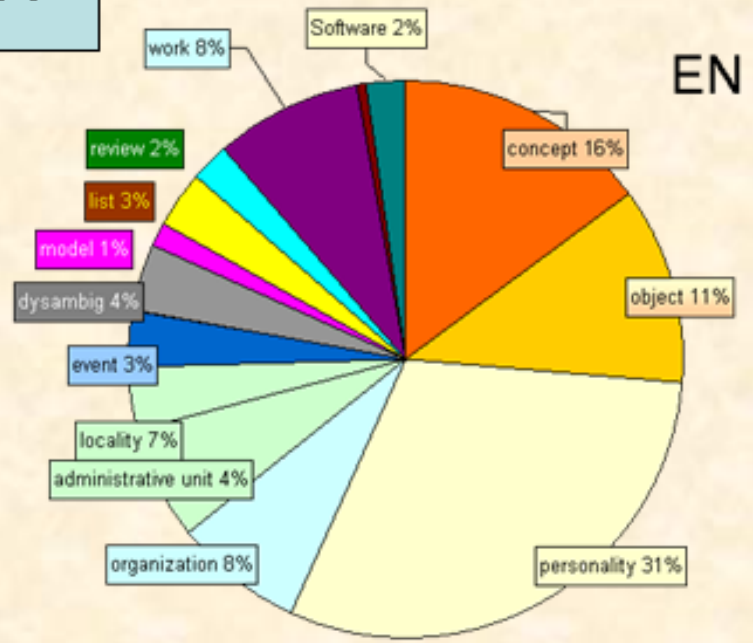
Західні – більше персоналій і компаній



4.7 західні Вікіпедії за типом

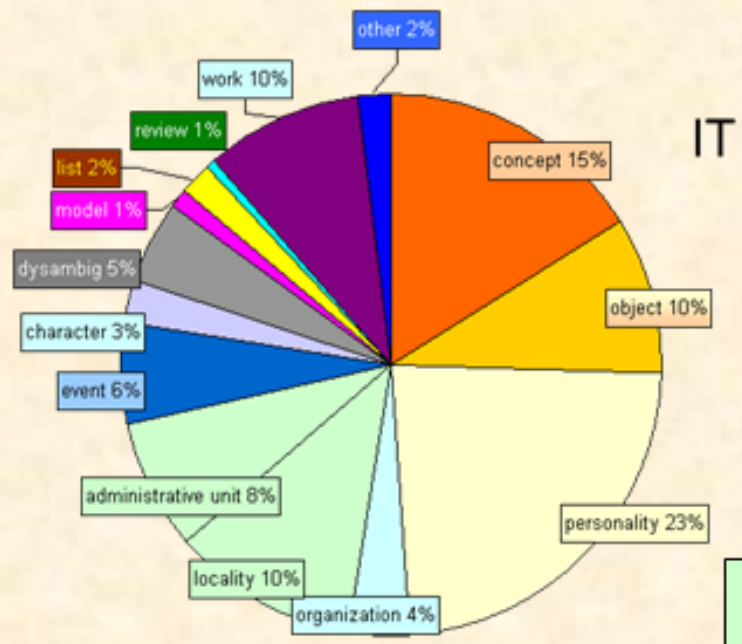
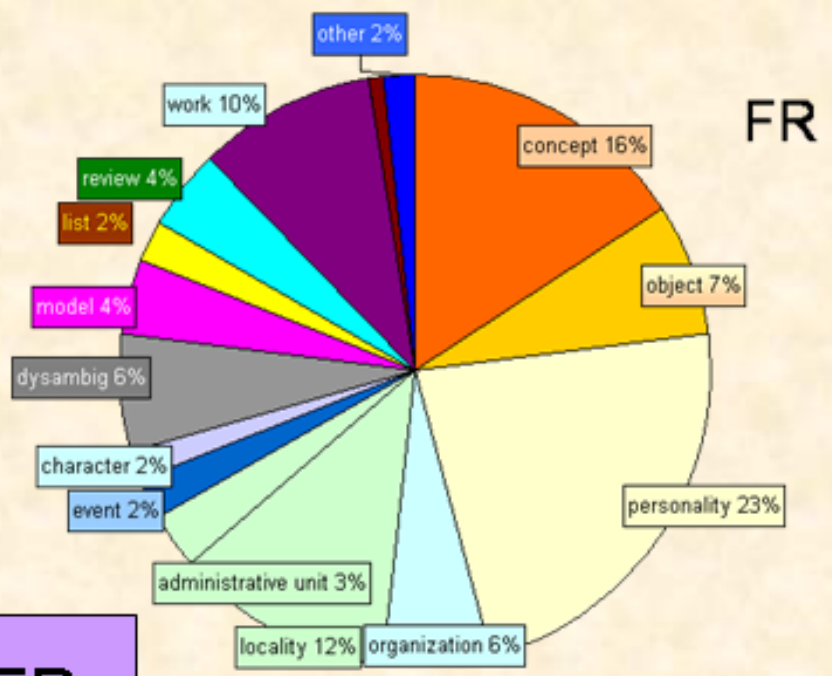
EN

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 |

