

Report k BP-Wikipeople 31.3.2021

Milan Welser

Procházení stránek

Stránky z kategorií jsem procházel pomocí třídícího klíče

"!ABCDEFGHIJKLMNPQRSTUVWXYZ", kdy jsem bral vždy N stránek, jejichž názvy byly lexikograficky větší než každý znak z tohoto klíče.

Znak "!" symbolizuje stránky začínající speciálními znaky a číslicemi.

Pro N=100 bych tedy prošel 100 stránek začínající speciálními znaky či číslicemi a následně také 100 stránek začínajících každým písmenem abecedy.

Living people (N = 1000)

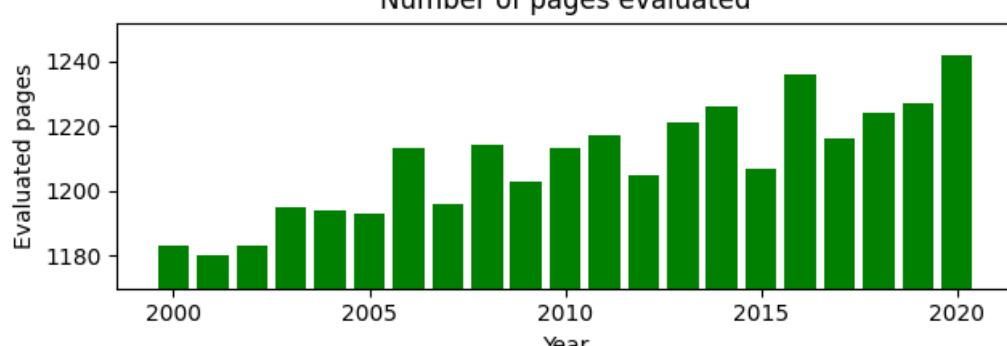
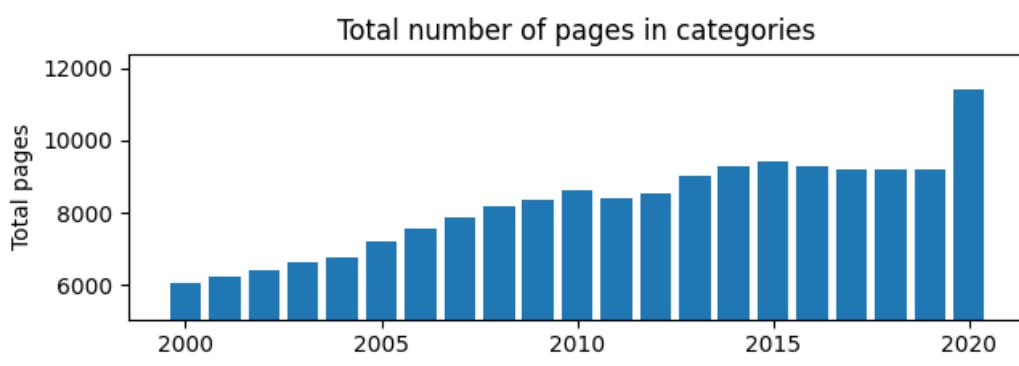
Kategorie Living_people obsahuje zhruba **993,865** stránek, z čehož jsem prohledal **26063** stránek (~2.5%), číslící či speciálním znakem začínalo 63 stránek, od všech písmen abecedy jsem prohledal 1000 stránek.

Z těchto **26063** stránek obsahovalo obrázek **8185 (31.4%)** a prohledání zabralo **5326.22s**.

Průměrná doba zpracování jedné stránky v tomto případě byla **0.204s**.

Dead people (N = 50)

Zesnulí lidé jsou na wikipedii kategorizování podle roku úmrtí do kategorí stylu "2020_deaths". Zpracovával jsem roky **2000-2020**. Nejvíce stránek obsahuje kategorie 2020_deaths (**11414**), nejméně potom 2000_deaths (**6057**). Prohledával jsem nejvýše **N=50** stránek pro každé písmeno (celkem max. 1350), kvůli rozložení se však počet prohledaných stránek pohyboval v rozmezí **1180-1242** stránek.



Z prohledaných stránek z kategorií <Year>_deaths obsahovalo průměrně **24.78%** stránek obrázek, rozsah byl **20.96%** (2005) až **30.15%** (2019).

Průměrný čas zpracování stránky se pohyboval v rozmezí 0.20s až 0.23s na stránku, průměrně to bylo **0.218s**.

