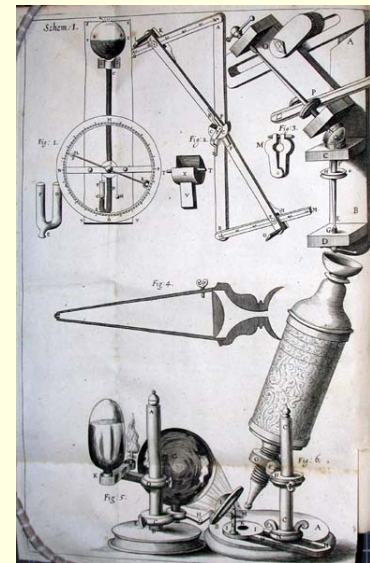


# Kvantitativní metody informační vědy

- každý vědecký obor musí mít vlastní vědecké metody - uplatňuje při zkoumání svého předmětu
- metody – základem exaktnosti. Umožňují opakovat, kontrolovat a srovnávat experimenty
- v informační vědě dvě oblasti – bibliometrie a infometrie, které umožňují kvantifikovat dokumentační a informační fenomény a stávají se tak zdrojem kvantitativních metod a zdrojem určitých pravidelností (s mírnou nadsázkou označovaných jako přírodní zákony)



# Bibliometrie

- **Bibliometrie** – je to aplikace matematických a statistických metod na knihy a další psaná komunikační média
- zabývá se kvantifikovatelnými vlastnostmi znakově zaznamenaného diskurzu a chování
- je vědou studující používání dokumentů a publikačních vzorů pomocí matematických a statistických metod
- vzniká na základě propojení dvou oblastí:

## 1) bibliografie:

- srovnávací a historická studia knih a jejich skladby
- klasifikace knih za účelem šíření znalostí o daném předmětu, autorovi, vědní oblasti apod. formou seznamů těchto knih

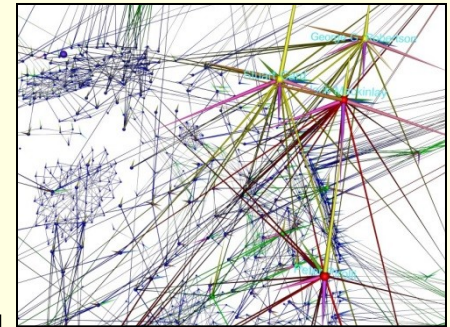
# Bibliometrie

---

- 2) **statistika**: zabývá se množinami faktů, které spolu vzájemně souvisejí. Je to věda, která shromažďuje statistická fakta, třídí je a na jejich základě pak vyvozuje závěry o daném stavu předmětu (stav země, jejího obyvatelstva apod.). Aplikací teorie pravděpodobnosti lze z velkého množství dat činit předpovědi
- **bibliometrie** – název z řeckého biblion (kniha, Byblos – fénické město proslulé exportem papíru) a metricus (měření)
  - průkopníky bibliometrie Cole a Eales – v roce 1917 analyzovali literaturu z anatomie za roky 1543 – 1860. Podobné studie souhrnně označovány jako statistická bibliografie

# Bibliometrie

- roku 1969 Alan Pritchard navrhl alternativní označení bibliometrie, aby se vyhnul záměně původního termínu se statistikou samotnou či s bibliografiemi o statistice
- aplikace bibliometrie:
  - růst a vývoj vědy
  - křížové sítě citací – výzkum vzorů komunikace
  - dle citací určení klíčových časopisů pro knihovny
  - epidemické modely šíření poznatků
  - kocitace – výskyt citace dvou článků → souvisejí spolu → určuje u časopisů teoretické a aplikační jádro oborů



# Infometrie

---

- založil ji v roce 1979 prof. Otto Nacke a definoval ji jako průnik matematiky a informatiky
- lze ji chápat jako využití měřících procedur na informační procesy a funkce
- bibliometrické studie jsou součástí širšího pole studia – infometrie
- základní jednotkou bibliometrie jsou „dokumentační jednotky“ s tematicky vymezenými kategoriemi obsah, autor, jazyk apod., zatímco základní jednotky v infometrii v současnosti přesně stanovit nelze

# Infometrie

---

- infometrie zkoumá množství informací obsažených v dokumentu či jiném sdělení a změny tohoto množství při různých transformacích informací v průběhu společenského informačního procesu
- bibliometrie zkoumá jen jeden druh informační transformace – transformace projevující se v publikační činnosti

# Librametrics

---

- je oblastí širší než infometrie. Česky se překládá jako kvantifikace knihovní činnosti
- zabývá se měřením knihovnických služeb jako celku
- pojem zavedl v roce 1948 Ranganathan, který si uvědomil, že mnohé problémy spojené s knihovní prací souvisí s velkými čísly
- kvantitativní metody jsou aplikovány na informační činnost knihoven a informačních center (chování dokumentů, knihovnického personálu a uživatelů knihoven) při řešení problémů knihoven či jejich jednotlivých oddělení

# Librametrics

---

- aplikace (výběr):
  - » určení optimálního množství zaměstnanců v jednotlivých odděleních
  - » na systematizaci oběhu knihovnických dokumentů
  - » na určení optimální velikosti knihovny
  - » na analýzu potřeb uživatelů
  - » na určení směrování akvizičního systému.  
Měření poměru náklady – efektivnost a náklady – zisky investic v knihovnickém sektoru



# Scientometrie

- dříve nazývána naukometrií je ještě širší oblastí než librametrics. Je součástí scientologie, vědy o vědě
- zkoumá a měří zákonitosti vývoje vědy jako složitého společenského jevu, jde tedy o aplikaci matematických postupů ve scientologii
- je využívána na analýzu a modelování vědecko-sociologických, historických, teoretických a vědecko-ekonomických procesů s cílem řízení vědy
- hodnotí se pokročilost vědy, úroveň jejího rozvoje a vliv a význam pro společnost
- výsledky jsou používány pro vědecký management a plánování vědy

# Bibliometrické zákony

---

- pravidelnosti označované jako zákony nejsou 100% platné, jelikož jde o statistické zákonitosti, tj. pravidelnosti vyskytující se s určitou pravděpodobností
- zákony tvoří teoretickou základnu biblio- ale i infometrie. Jde o tři základní a tři doplňující zákony

# Bibliometrické zákony

---

- **Bradfordův zákon rozptylu** (distribuce publikací) – popisuje rozložení dokumentů (nejčastěji časopisů) ve specifických disciplínách či v určité tematice
- Samuel Clement Bradford zjistil, že pokud bude provádět rešerši článků k určitému tématu, najde několik časopisů které se tématu věnují zcela, několik částečně, několik se mu věnují příležitostně. První časopisy tvoří jádro, další mezikružím kolem jádra
- po sečtení článků v jádře uspořádal časopisy v mezikružím tak, aby obsahovaly stejný počet článků jako časopisy v jádře

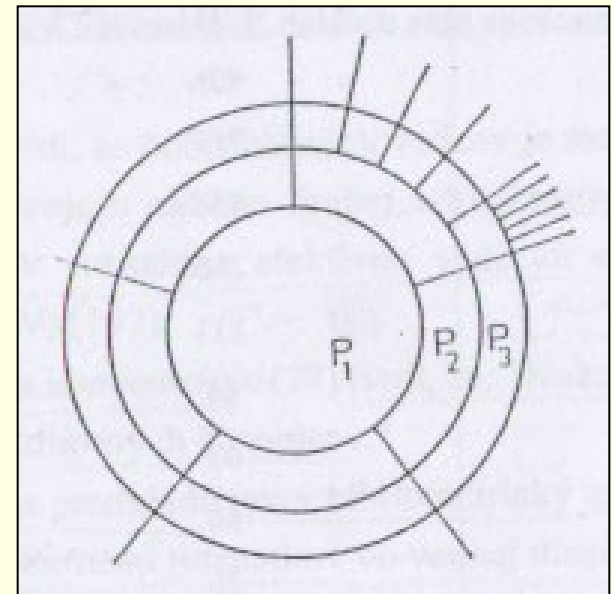


# Bibliometrické zákony

- zákon tvrdí, že v jádře bude  $X$  časopisů, v 1. mezikruží  $n^2X$  časopisů, v 2.  $n^3X$  časopisů atd.

$$\text{Tedy } \mathbf{X : Y : Z = X : n^2X : n^3X = 1 : n : n^2}$$

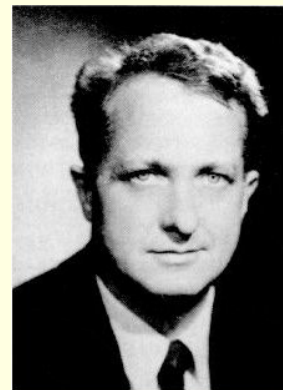
- Bradford dále experimentálně zjistil, že číslo, kterým je třeba násobit počet časopisů je konstanta:  $n = 5$
- zákon je používán např. při hodnocení informačních pramenů v jednotlivých oborech, při určení poločasu stárnutí toku informací, při návrhu designu ekonomičtějších informačních systémů, při predikcích publikačních trendů



# Bibliometrické zákony

---

- **Zipfův zákon výskytu slov** – George Kingsley Zipf, německý lingvista, zjistil, že délka slova je úzce spojena s frekvencí jeho používání
- metoda: uspořádává slova z Joyceova *Odyssea* podle výskytu. Nejčastěji se vyskytující má rang 1, druhé slovo 2 atd. Když vynásobil rang počtem výskytů slova, zjistil, že součiny všech slov mají velmi blízké hodnoty



# Bibliometrické zákony

Rang r	Frekvence f	Součin r.f
10	2653	26530
20	1311	26220
30	926	27780
100	265	26500
500	50	25000
1000	26	26000

- frekvence jakéhokoli slova je inverzně proporční k jeho pořadí ve frekvenční tabulce
- využití: při vytváření slovníků rešeršních nebo nevýznamových slov → nástroje automatického indexování

# Bibliometrické zákony

- **Lotkův zákon vědecké produktivity** – zákon popisuje rozložení počtu autorů podle jejich produkce. Speciální aplikace Zipfova zákona.
- Alfred J. Lotka rozložil soubor odborníků a podle profesí zjišťoval, kolik publikují vědeckých článků za rok a dlouhodobě
- několik autorů publikuje mnoho článků, víc autorů středně a většina málo článků
- Lotka odvodil vztah mezi frekvencí autorů  $y$ , přispívajících  $x$  příspěvků:  **$x^n y = \text{konstanta}$** . Pokud bude 6 autorů publikovat 4 články, 11 autorů 3 články, 25 autorů 2 články a 100 autorů jeden, pak dosazením do vzorce (s Lotkou zvolíme  $n=2$ ) získáme:



# Bibliometrické zákony

---

$$K_1 : 6.4^2 = 99$$

$$K_2 : 11.3^2 = 99$$

$$K_3 : 25.2^2 = 100$$

$$K_4 : 100.1^2 = 100$$

- když známe počet autorů publikujících 1 článek, můžeme předpovídat, kolik autorů publikuje více článků
- **Lotka's Law**
- <http://youtu.be/JxsweVr1dGg>



# Bibliometrické zákony

- **Priceův zákon druhé odmocniny vědecké produktivity** – celá populace vědců s velikostí  $N$  obsahuje efektivní vědeckou elitu rovnou velikosti druhé odmocniny čísla  $N$ . **Elita =  $\sqrt{N}$**
- **Garfieldův zákon koncentrace** – každá tematická oblast má svoji množinu základních odborných časopisů
- **Senguptův zákon bibliometrie** – během fáze rychlého růstu poznatků ve vědní disciplíně se články zabývající se touto vědní disciplínou objevují v rostoucím počtu v časopisech, které jsou daleko od jádra časopisů vědní oblasti

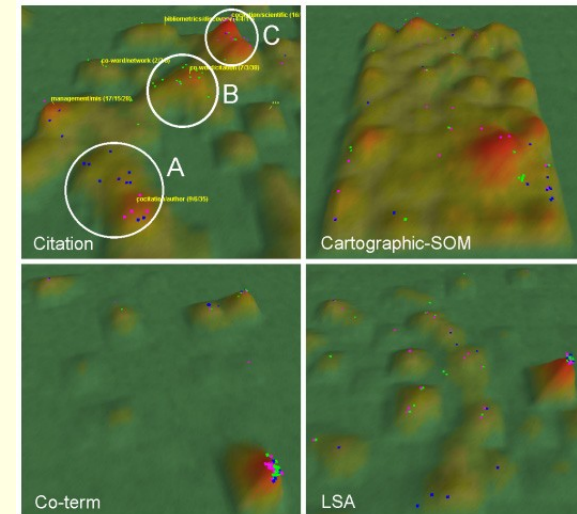
# Citační analýza

---

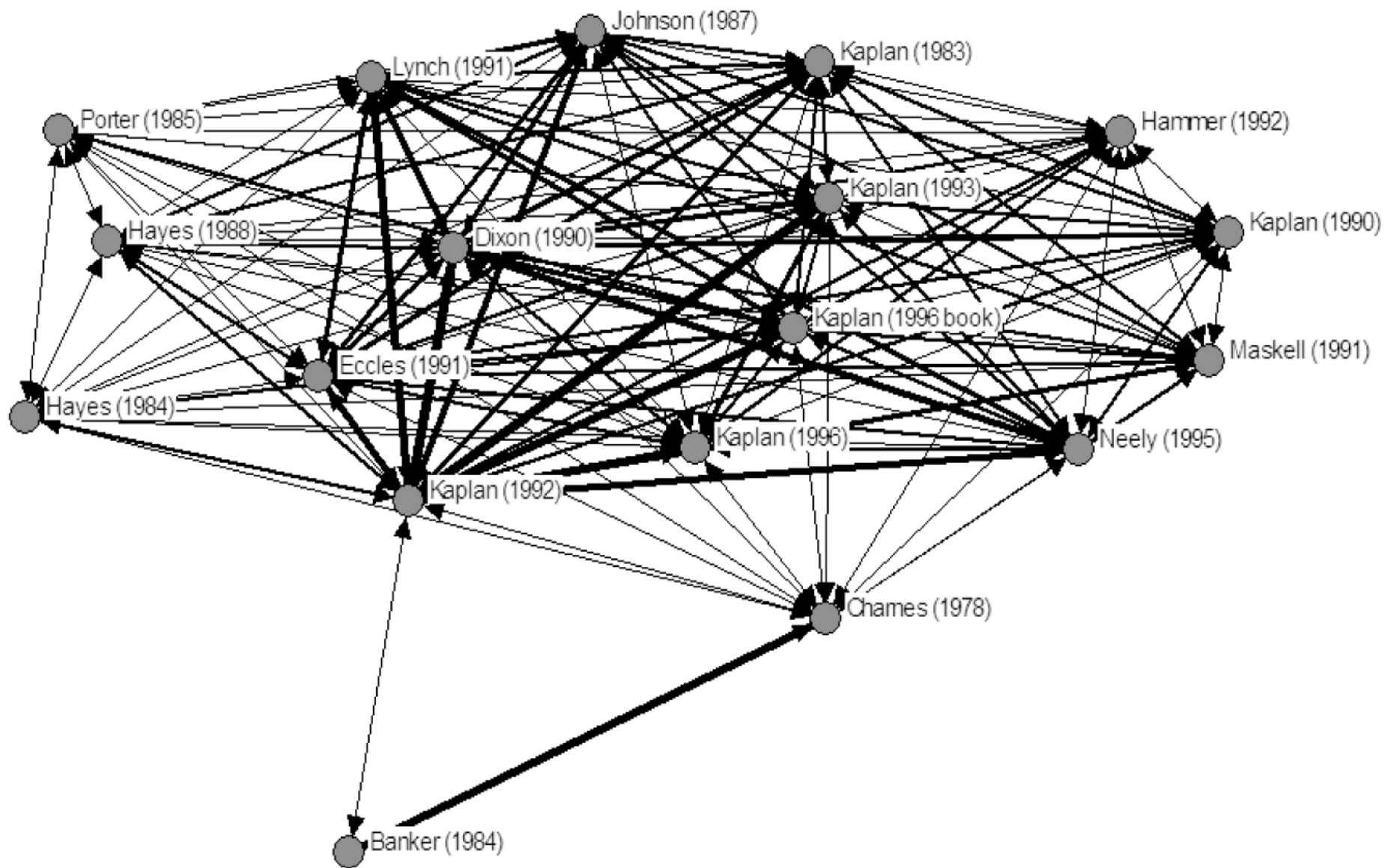
- bibliografická metoda, která vychází z faktu, že libovolná citace práce autora je významný fakt
- kvantifikuje vztahy mezi autory, dokumenty a vědními obory na základě bibliografických citací a bibliografických referencí
- citace – odkaz, který dokument obdrží od jiného dokumentu
- reference – odkaz, který jeden dokument uvádí na jiný dokument
- 2 typy studií- produktivita vědecké literatury (založeno na počítání vstupů)
  - použití literatury autory

# Citační analýza

- umožňuje zjišťovat:
  - » význam literatury z různých oblastí pro zkoumaný předmět
  - » časové rozpětí této literatury – od data publikování po její využití
  - » základní formy použité literatury a jejich relativní význam
  - » národnostní původ literatury používané v určité oblasti
  - » nejdůležitější periodické tituly pro každou oblast
  - » sestavování citačních sítí za účelem citačního mapování vědy (např. kocitace)



● Palmer (1996)



# Citační indexy

---

- citační analýza využívá pro své studie údaje z bibliografií a z citačních registrů
- sestavovat citační indexy navrhl Eugene Garfield v článku „Citation Indexes for Science“. Doufal, že lze eliminovat nekritické citování falešných, nekompletních a zastaralých dat zpřístupněním dřívějších prací a jejich kritik
- citační index je „index asociovaných idejí“, je to soupis všech publikovaných materiálů citovaných v sledovaných pramenech

# Citační indexy

---

- je uspořádán abecedně podle citovaných autorů a pod každým jménem je chronologický seznam citovaných prací
- dozvíme se z něj: název časopisu v němž práce vyšla, ročník a stranu. Dále údaje o článku a citující pramen, tj. kde je práce citována a jakými autory
- má tři části: 1) seznam citované literatury (kde je autor citován)  
2) seznam citující literatury (z čeho vychází)  
3) permutovaný předmětový index

# Citační indexy

- nejznámější citační indexy jsou:
  - » *Science Citation Index* (SCI)
  - » *Social Science Citation Index* (SSCI)
  - » *Arts&Humanities Citation Index* (A&HCI)

- jsou vydávány na komerční bázi Institutem pro vědecké informace (ISI) ve Filadelfii

  - » *Journal Citation Reports* (JCR) – soubor statisticky zpracovaných informací o citovanosti časopisů, počtu publikovaných článků atd., vydávaný ročně
  - » *Web of Science* – databáze založená na SCI
- citační indexy slouží k:
  - zjišťování citovanosti prací
  - vlivu těchto prací pomocí tzv. impakt faktorů

# Impakt faktor (IF)

- udává, kolikrát je průměrný článek publikovaný v určitém časopise citován, během dvou předcházejících let
- podle této hodnoty je porovnávána míra vědecké důležitosti časopisů
- číslo je výsledkem poměru součtu citací za roky X-1 a X-2 k počtu článků publikovaných v letech X-1 a X-2

Matematické vyjádření: 
$$\mathbf{IF (X) = \frac{Cit_{x-1}(X) + Cit_{x-2}(X)}{Pub (X-1) + Pub (X-2)}}$$

- Dnes rozšiřování IF na délku pěti let



# Impakt faktor

- impakt faktor se používá při hodnocení produktivity vědců, výzkumných týmů či výzkumných ústavů – sečtou se IF jejich publikací za určité období a porovnají se. Tyto údaje vyžadovány při udělování titulů, při konkurzech na významná vědecko-administrativní místa, při žádostech o granty
- kritika používání impakt faktoru:
  - nerovnoměrné geografické rozložení titulů – nedá se použít pro některá etnika
  - problém autocitací – autor může uměle zvyšovat vlastní IF – třeba eliminovat
  - obory se liší citačními zvyklostmi – nelze srovnávat mezioborově. Signifikantní až řádové rozdíly v citovanosti, ne drobné

# Impakt faktor

---

- připisování lidí na práce, na nichž se nepodíleli (šéfové apod.)
- práce příliš geniální → předběhla dobu → necitována
- negativní citovanost prací – nelze zjistit
- ve vědě důležití i lidé, kteří nepíší, ale jsou špičkovými pedagogy a manažery vědy. Nehodnotí ani pracovníky v aplikovaném výzkumu
- přednost časopisů v angličtině → rodilí mluvčí ve výhodě
- malé zastoupení monografií a konferenčních materiálů
- módnost ve vědě – určitá témata stranou → skvělým pracím v těchto tématech chybí citace

# Impakt faktor

---

- zkratky časopisů se liší od užívaných zkratk v ostatních referátových časopisech, nejednotná transkripce u jmen → jméno autora na více místech
- nejednotnost užívání názvů institucí → jméno na více místech
- lékařství – nejvyšší citovanost u klinických studií → užívají praktici, menší citovanost vědeckých prací
- zpoždění ohlasů práce
- nové časopisy těžko shánějí autory, kterým se v nich nevyplatí publikovat
- chyby při přepisování údajů – lidský faktor, nezachycení všech článků autora
- problémy mezioborových prací

# H-index

---

- Hirschův index, někdy označován jako highly-cited index
- kolik článků daného autora dosahuje citovanosti vyšší, než je pořadové číslo článku
- měří citační ohlas jednoho autora
- př. h-index autora = 5 – autor publikoval pět článků, které jsou každý pětkrát citovány
- autor napsal pravděpodobně více článků, ovšem právě pět z nich bylo pětkrát citováno

# Databáze

---

- Databáze dostupné přes knihovnu FF (i přes vzdálený přístup z domů pomocí proxy serveru):
  - Web of Science (<http://isiknowledge.com/>)
  - Scopus

Mapování vědy: <http://www.eigenfactor.org/>

Úkol: zjistěte v databázích tři časopisy, které v mají nejvyšší dopad na vývoj našeho oboru (dle impakt faktoru)

