

# **Vyhledávání informací**

Operátory, nástroje a strategie

14. 10. 2021

domácí úkol

# Rozehrivací vyhledávačka



V knize Josefa Lady *Kronika mého života* se prý mluví i o Mahenovi. Rád bych věděl, kde to tam najdu. Můžete mi to najít hned tady od stolu?

# Co máme za sebou?

- referenční rozhovor, formulace požadavku
- analýza řešeršního požadavku
- pojmová analýza, tj. hledání klíčových slov, synonym, cizojazyčných ekvivalentů s využitím ŘS atp.
- volba zdroje či zdrojů informací
- volba řešeršní strategie
- formulace řešeršního dotazu

# Proces vyhledávání

- Volba strategie a volba zdrojů

1. pojmová analýza
2. hledání synonym a souvisejících pojmů
3. převedení na výrazy řízeného slovníku
- 4. aplikace operátorů**
5. aplikace dalších vyhledávacích technik

} Formulace  
rešeršního  
dotazu

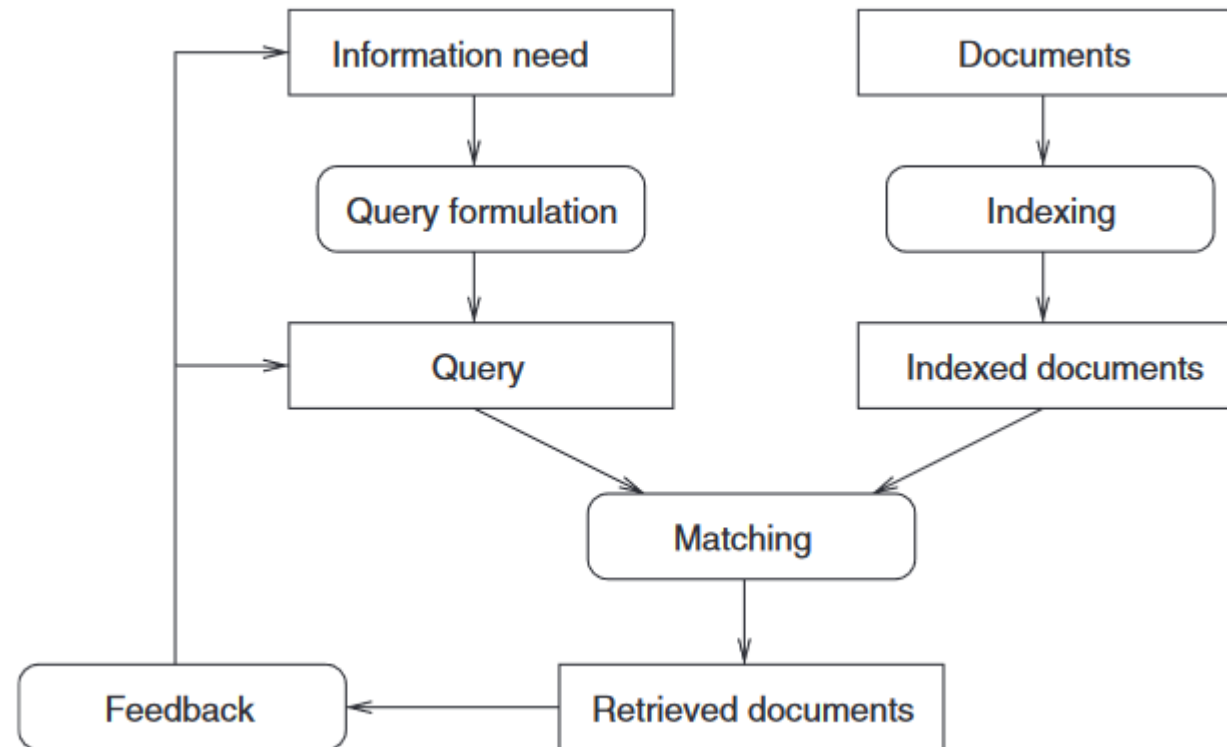
- Vyhledávání, hodnocení, sestavení rešerše



# Aplikace operátorů

- Boolovské operátory

# Indexing-based IR



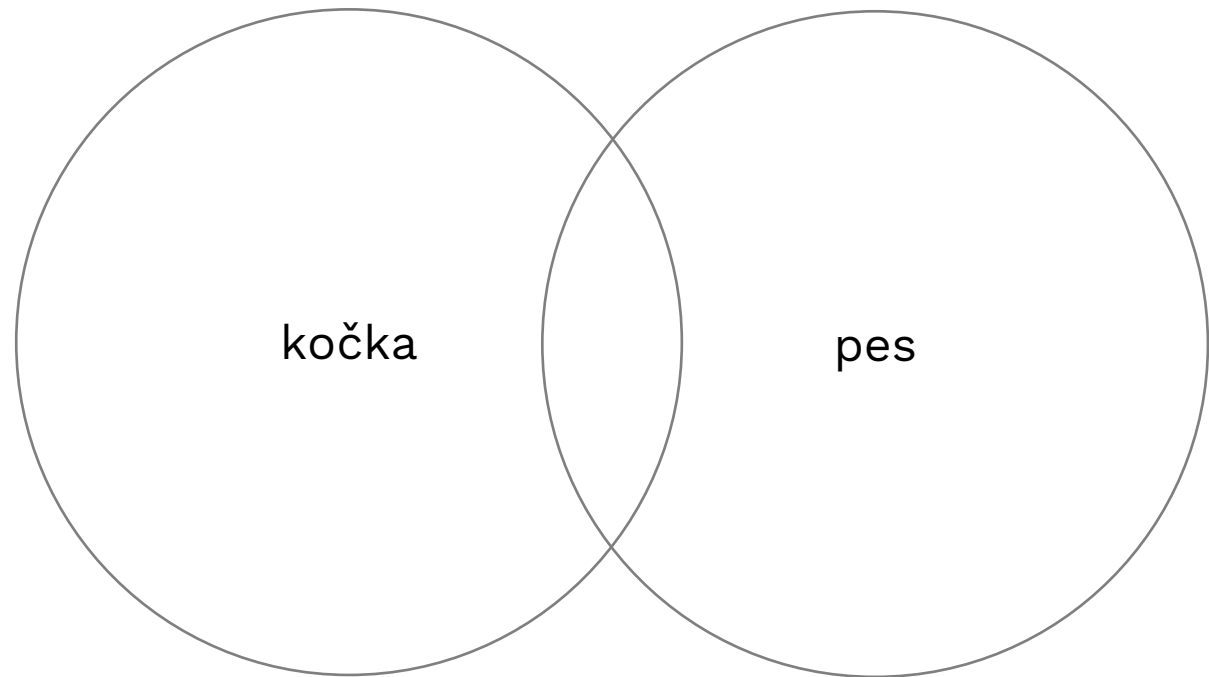
# Jednoduchý boolovský model

- booleovská logika
- první, nejrozšířenější a široce aplikovaný
- dokument i dotaz jsou pojímány jako soubor výrazů
- vyhledání je založeno na výskytu výrazů z dotazu
- vyhledávání „jablko“ jednoduše vrátí množinu dokumentů, kde je „jablko“
- pomocí logických operátorů boolovské logiky lze vytvářet nové množiny dokumentů odpovídající vyhledávacímu dotazu



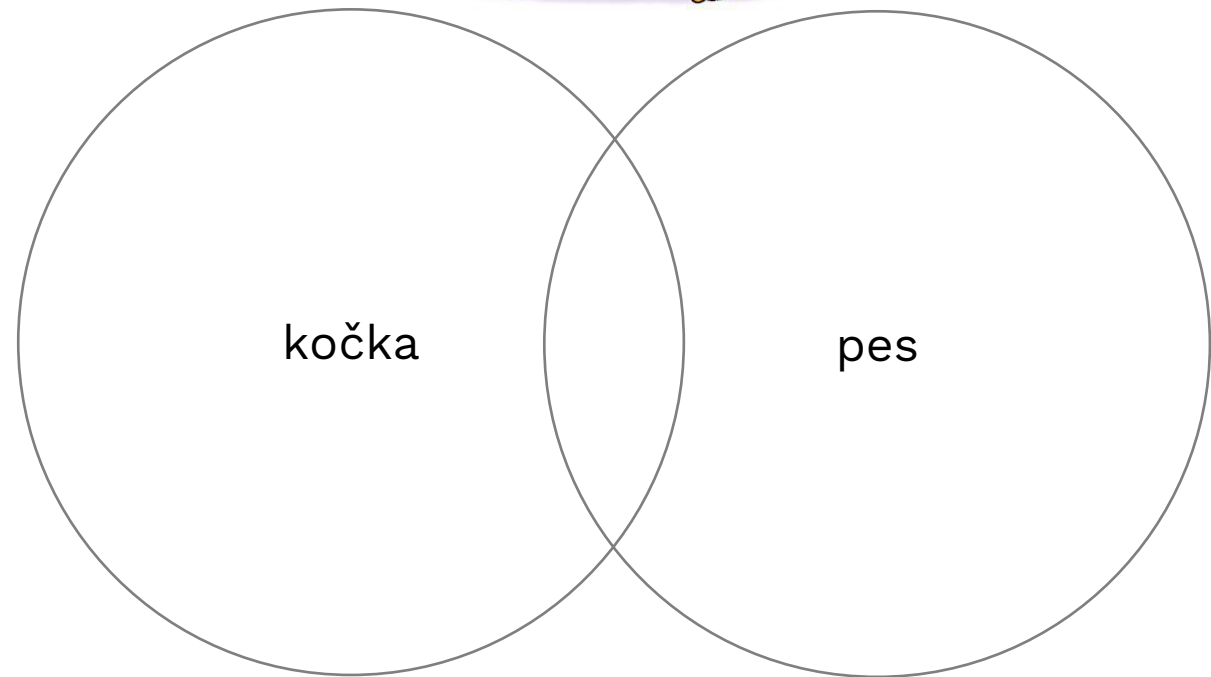
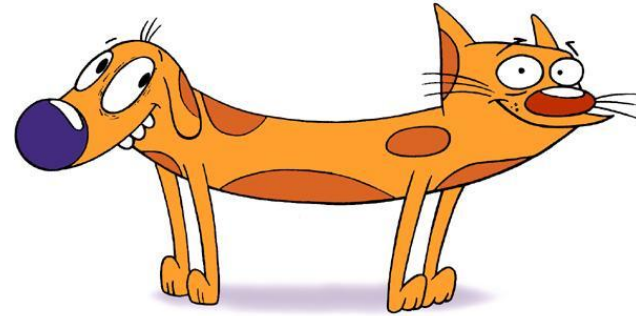
# Boolovské operátory

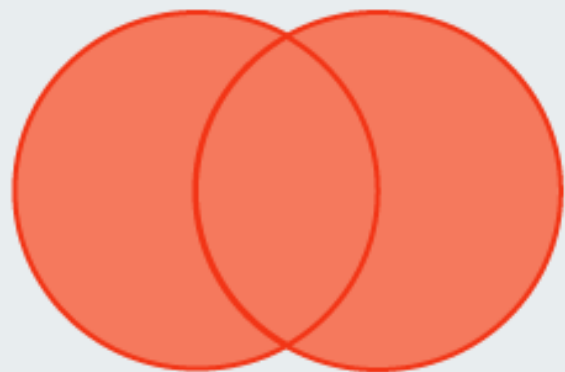
- AND
- OR
- NOT
- XOR



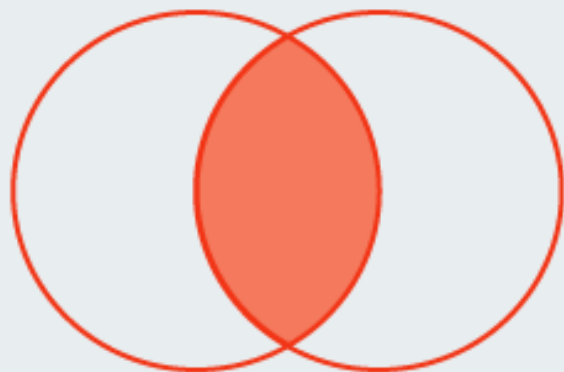
# Boolovské operátory

- AND
- OR
- NOT
- XOR

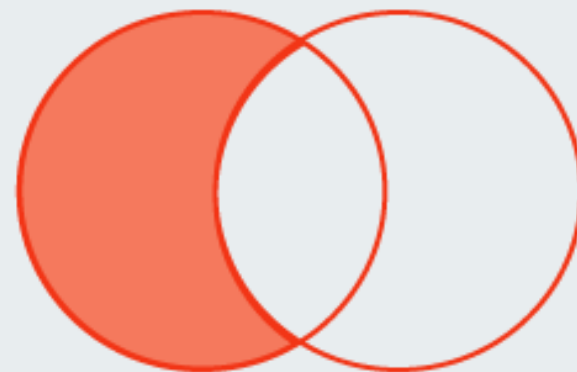




“Notre Dame” OR fire



“Notre Dame” AND fire

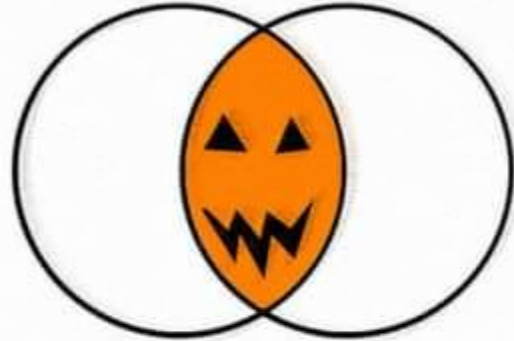


“Notre Dame” NOT fire

Trick OR Treat



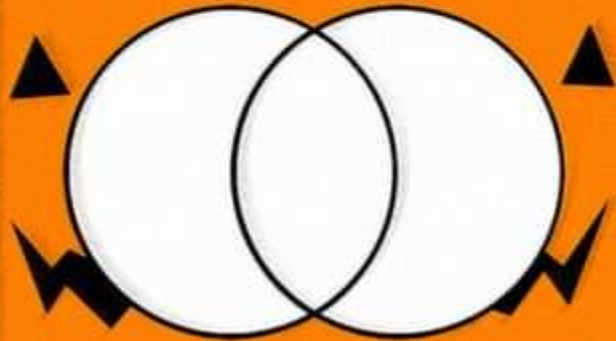
Trick AND Treat



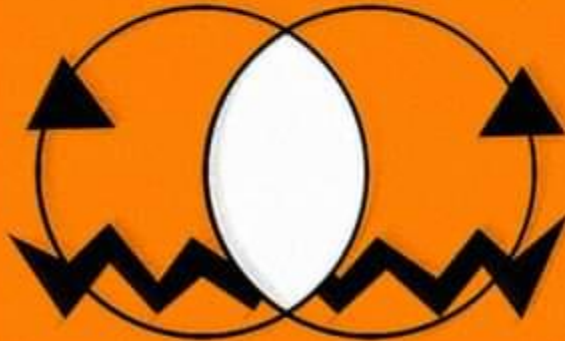
Trick XOR Treat



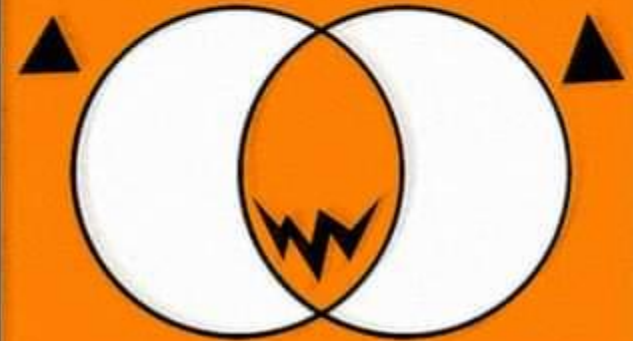
Trick NOR Treat



Trick NAND Treat



Trick XNOR Treat



((diastolický OR systolický OR krevní) AND tlak) OR ...  
malý OR miniaturní OR malinkatý OR drobný OR titěrný  
(ipad OR iphone) AND apple

# Jednoduchý boolovský model

- pocit kontroly - zároveň nejkritizovanější
- výsledky nejsou hodnoceny a řazeny
- model dokument prostě najde nebo ne
- „jablko AND sad AND štrůdl“
- nenajde „letadlo AND pilot AND létání“,
- nenajde „jablko AND štrůdl“ - očividně užitečnější, ale model nemá jak určit, že tomu tak je...

# Modely vyhledávání

- boolovský model
- reprezentace dokumentů skrze vektory
- vektorové modely
- TF-IDF a vážení výrazů
- lematizace, stematizace, stop slova
- indexování latentní sémantiky
- strojové učení, topic modely atp.

# Modely vyhledávání

- boolovský model
- reprezentace dokumentů skrze vektory
- vektorové modely
- TF-IDF a vážení výrazů
- lematizace, stematizace, stop slova
- indexování latentní sémantiky
- strojové učení, topic modely atp.

**SOON**





# Aplikace operátorů

- Boolovské operátory
- Proximitní operátory
- *různé implementace*

# Proximitní operátory

- NEAR (N3 = kolo favorit, kolo značky favorit, favorit je nejlepší kolo)
- WITHIN (W3 = daňová W3 reforma = daňová a sociální reforma, ~~reforma daňová a sociální~~)
- ADJACENT (irská ADJ vlajka = irská vlajka, vlajka irská)
- FOLLOW BY
- PRECEDE BY (PRE3)

# Vyhledávačka



Používá aleph.muni.cz proximitní operátory?  
Pokud ano, jak a jaké?

# Vyhledávačka



Používá [katalog.muni.cz](http://katalog.muni.cz) proximitní operátory? Pokud ano, jak a jaké?

# Vyhledávačka



Používá Google proximitní operátory?  
Pokud ano, pokuste se ho použít  
a vyhodnoťte, jak funguje.

# Proces vyhledávání

- Volba strategie a volba zdrojů

1. pojmová analýza
2. hledání synonym a souvisejících pojmů
3. převedení na výrazy řízeného slovníku
4. aplikace operátorů
- 5. aplikace dalších vyhledávacích technik**

} Formulace  
rešeršního  
dotazu

- **Vyhledávání**, hodnocení, sestavení rešerše

# Další vyhledávací techniky

- zástupné znaky *filozofie OR filosofie -> filo?ofie*
- rozšíření pravostranné, levostranné *filo\* filo?ofi\**
- divoké karty
- vyhledávání celé fráze *(jarní vítr -> katalog.muni.cz)*
- krácení
- stematizace – nalezení kmene slova *(ProQuest - color)*
- lematizace – vyhledání základního tvaru
- vyhledávání podle polí *(EBSCO -> Field Codes)*
- vážení výrazů

# Další vyhledávací techniky

- filtrování
- formální vymezení
- fasetová navigace
- dotaz příkladem
- moderní vyhledávací techniky založené na automatizaci (*JSTOR Text Analyzer*)
- *mnoho dalších...* (*PubChem*)



# Každý IZ má svá specifika

## Google

- ~~AND, OR, XOR, -, AROUND(X)~~
- ()
- “”
- \*
- 0...1
- filtrování
- rozšířené vyhledávání

# **Každý IZ má svá specifika**

- filetype:
- inurl:
- intitle:
- site:
- define:dysfunction
- related:muni.cz
- weather:luhačovice

# **Každý IZ má svá specifika**

Pochopit jak funguje vyhledávání v něm,  
seznámit se s možnostmi,  
číst nápovědy!

# Rešeršní strategie

- specifika IZ se odrážejí mimo jiné i ve volbě RS
- rešeršní strategií se rozumí přístup k vyhledávání položek v informačních zdrojích k jednomu požadavku
  
- strategie stavebních kamenů
- rostoucí perly
- osekávání

# **Stavební kameny**

building blocks

informační chování seniorů

# Stavební kameny

informační chování seniorů



dílčí dotaz 1: informační chování OR informační potřeba

dílčí dotaz 2: senioři OR důchodci OR staří

# Stavební kameny

informační chování seniorů



dílčí dotaz 1: informační chování OR informační potřeba

dílčí dotaz 2: senioři OR důchodci OR staří



(informační chování OR informační potřeba)  
AND (senioři OR důchodci OR staří)

# Stavební kameny



(informační chování OR informační potřeba)  
AND (senioři OR důchodci OR staří)

- > příliš dokumentů?
- > málo dokumentů?



# Stavební kameny



(informační chování OR informační potřeba)  
AND (senioři OR důchodci OR staří)

příliš dokumentů? – mohu odebrat pojmy  
málo dokumentů? – mohu přidat pojmy

krácení (senio?, potřeb?)  
zástupné znaky

# Rostoucí perla

snowballing, citation pearl growing

- vyhledávání podle nejužšího pojmu z požadavku
- podle nejspecifičtějších termínů / jednoho dokumentu
- cílem je alespoň jeden záznam
- vyhledávání se dále upravuje podle něj (ŘS, terminologie)
- intuitivní postup, který většina z nás dělá

autorský zákon AND pouštění hudby AND restaurace

- autorské právo – produkce hudby



# Osekávání

- opak rostoucí perly
- dotaz záměrně široký a postupně osekáván
- osekáván pomocí různých taktik
- NOT, filtrace, druh, jazyk, čas

323.1 – etnické menšiny

# Příliš široká formulace

= příliš velké množství výsledků  
*nové vymezení tématu?*

- řízené termíny v kombinaci s klíčovými slovy
- podřazené termíny
- vyhledávání podle polí
- využití proximitních operátorů
- omezení na určitý typ dokumentu
- NOT pro vyloučení některých záznamů
- jazykové vymezení
- časové rozmezí

# Příliš úzká formulace

*věcné rozšíření*

*přehodnotit hlediska zúžení*

- uvedení synonym
- slovních tvarů
- operátor OR, zástupné znaky, stemming
- hledání řízených termínů jako klíčových slov
- nadřazené řízené termíny
- obecné termíny s vysokým výskytem