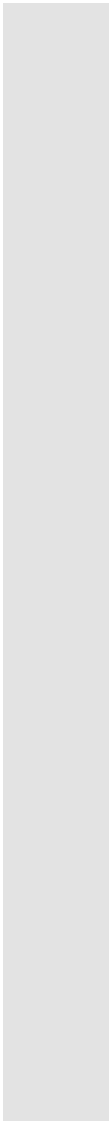


# Biomedical information retrieval (BIR)

---

Dohvaćanje  
informacija iz  
područja Biomedicine  
i zdravstva



## Information retrieval



# VEKTORSKI MODEL

- način predstavljanja dokumenata i upita kao vektora u višedimenzionalnom prostoru
  - Svaka riječ u dokumentu ima brojčanu vrijednost
  - Dokumenti i upiti se predstavljaju kao vektori
  - Sličnost između dokumenta i upita mjeri se kao "blizina" njihovih vektora

Primjer:

3 riječi: "Zagreb", "hotel", "restoran"

Dokument: "Zagreb ima mnogo hotela"Možemo ga predstaviti kao vektor: [1, 1, 0]

Upit: "hoteli u Zagrebu" [1, 1, 0]

# VEKTORSKI MODEL

## Što predstavljaju brojevi u vektoru?

### 1. Binarni pristup [0,1]

Dokument: "Zagreb hotel restoran"

Vektor: [1, 1, 1, 0, 0] znači:

1 = riječ postoji u dokumentu, 0 = riječ ne postoji

### 2. Brojanje pojavljivanja

Dokument: "Zagreb hotel hotel restoran hotel"

Vektor: [1, 3, 1, 0, 0] znači:

Brojevi pokazuju koliko se puta riječ pojavila

### 3. TF-IDF težine

Dokument: "Zagreb hotel hotel restoran hotel"

Vektor: [0.2, 0.5, 0.3, 0, 0] znači:

Brojevi pokazuju važnost riječi u dokumentu

Veća vrijednost = riječ je važnija za dokument

# VEKTORSKI MODEL

## TF (Term Frequency)

Koliko puta se riječ pojavljuje u dokumentu

Primjer: riječ "hotel" se pojavljuje 3 puta u dokumentu →  $TF = 3$

## IDF (Inverse Document Frequency)

Koliko je riječ rijetka u svim dokumentima

Primjer: "hotel" se pojavljuje u malo dokumenata → IDF velik

## BM25

Poboljšani TF-IDF koji:

1. Ograničava utjecaj previše ponavljanja riječi
2. Uzima u obzir duljinu dokumenta

# AI vektorska metoda

## Vektorski modeli u AI

### Klasični vektorski model

- Svaka riječ = jedna vrijednost (0 ili 1)
- Brojevi pokazuju samo pojavljivanje riječi
- Ne razumije značenje riječi
- Vektor: [1,1,0] = ["hotel", "Zagreb", "restoran"]

Problem: "hotel" i "apartman" su potpuno različiti

### AI vektorski model (embeddings)

- Svaka riječ = vektor sa stotinama vrijednosti
- Svaka vrijednost predstavlja neku značajku riječi
- Razumije značenje i kontekst
- Vektor: [0.2, 0.8, 0.1, ...] = semantički opis riječi

Prednost: "hotel" i "apartman" imaju slične vrijednosti

Koriste ga: BERT, Word2Vec, GPT modeli...

# Načini AI procesuiranja

## Hijerarhija AI Tehnologija

### 1. Artificial Intelligence (AI)

- Najširi pojam
- Uključuje sve sustave koji simuliraju ljudsku inteligenciju

### 2. Machine Learning

- Podskup AI-ja
- Sustavi koji uče iz podataka i poboljšavaju se s iskustvom

### 3. Deep Learning

- Podskup Machine Learninga
- Koristi neuronske mreže s više slojeva za obradu

### 4. Natural Language Processing (NLP)

- Specifično područje AI-ja fokusirano na obradi jezika
- Kombinira različite AI tehnike

### 5. LLM (Large Language Models)

- Specifična implementacija (GPT, BERT, Claude)
- Koristi Deep Learning (Transformere)
- Najnapredniji oblik AI procesuiranja jezika

# AI vs. Klasično

## Uloga AI u Pretraživanju

### 1. Razumijevanje Upita

- Prepoznaje sinonime ("auto" = "automobil")
- Razumije kontekst ("jaguar" - životinja ili auto?)

### 2. Pametnije Rangiranje

- Uzima u obzir kvalitetu sadržaja
- Personalizira rezultate prema korisniku

### 3. Semantička Analiza

- Razumije značenje, ne samo riječi
- Povezuje slične koncepte

Klasično: "hotel Zagreb" → traži točne riječi

AI: "smještaj blizu centra Zagreba" → razumije namjeru i kontekst



# Boolean

## Boolean Pretraživanje

### Dokumenti:

D1: "Luksuzni hotel u Zagrebu"  
D2: "Zagreb je glavni grad"  
D3: "Hotel nudi parking"

### Operatori:

AND: obje riječi moraju postojati  
OR: barem jedna riječ mora postojati  
NOT: riječ ne smije postojati

### Upit: "hotel AND Zagreb"

D1: "Luksuzni hotel u Zagrebu"

✓ Sadrži obje riječi → ODGOVARA

D2: "Zagreb je glavni grad"

X Nema riječ "hotel" → NE ODGOVARA

D3: "Hotel nudi parking"

X Nema riječ "Zagreb" → NE ODGOVARA

# Evaluacija pretraživanja

## Preciznost i Odziv

### Rezultati za upit "Hotel u Zagrebu"

Sustav je vratio 5 dokumenata:

1. Hotel Westin Zagreb ✓

2. Hotel Dubrovnik Zagreb ✓

3. Apartman u Zagrebu X

4. Restoran Zagreb X

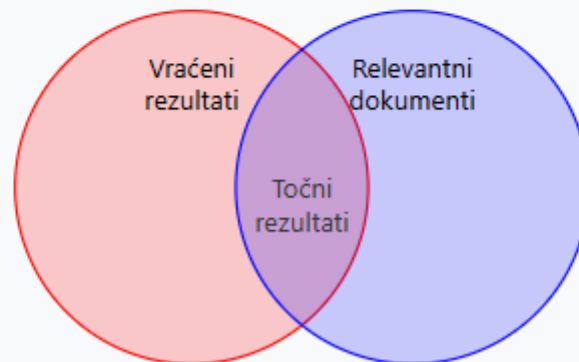
**Preciznost = Točni/Vraćeni = 2/4 = 0.5 (50%)**

→ Koliko vraćenih rezultata je relevantno

**Odziv = Točni/Ukupno\_relevantni = 2/3 = 0.67 (67%)**

→ Koliko relevantnih dokumenata smo pronašli

### Vennov dijagram



Preciznost = Točni/Vraćeni

Odziv = Točni/Relevantni

# Evaluacija pretraživanja

## F-mjera i MAP

### F-mjera (F1 Score)

Formula:

$$F = 2 \times (\text{Preciznost} \times \text{Odziv}) / (\text{Preciznost} + \text{Odziv})$$

Primjer:

- Preciznost = 0.8 (80%)
- Odziv = 0.6 (60%)
- F-mjera =  $2 \times (0.8 \times 0.6) / (0.8 + 0.6) = 0.686$  (69%)

### MAP (Mean Average Precision)

Rangirani rezultati (✓ = relevantno, X = nije relevantno):

1. ✓ (P=1/1=1.0)
2. X
3. ✓ (P=2/3=0.67)
4. ✓ (P=3/4=0.75)

$$AP = (1.0 + 0.67 + 0.75) / 3$$

$$AP = 0.81 \text{ (81\%)}$$

MAP je prosjek AP-a za sve upite

# Baze podataka

Izvori za dohvaćanje

Ključan aspekt  
biomedicinskih  
istraživanja i  
medicine  
utemeljene na  
dokazima

- **Izvori za dohvaćanje / baze podataka**
  - PubMed
  - Google Scholar
  - Web of Science
  - Scopus
- **Alati za dohvaćanje**
  - Klasični upiti (Queries)
  - Sustavi umjetne inteligencije

# Baze podataka vrste

- Osnovne vrste:
  - Bibliografske
    - podaci o primarnim izvorima: autor/i, naslov rada, naslov knjige, naslov časopisa, godina/svezak/stranice, naziv ustanove i sl., ključne riječi, sažetak...
    - Na području Biomedicine: MEDLINE, EMBASE, CURRENT CONTENTS, PsycINFO, CINAHL
  - Citatne
    - podaci o broju citata
    - WOS, SCOPUS, Google Scholar
  - Baze podataka s cjelovitim tekstovima
    - [Nature Journals Collection](#)
    - [Springer Journals](#)
    - [Cambridge Journals](#)
    - [IOPscience](#)
    - [Taylor & Francis Subject Collections](#)
    - [The Cochrane Database of Systematic Reviews](#)
    - ...
  - Kombinirane



# National Library of Medicine

## NLM

najpoznatije  
informacijsko  
čvorište za  
diseminaciju  
biomedicinskih  
informacija

### National Library of Medicine (NLM)

#### Lokacija:

- Bethesda, Maryland, SAD
- Dio National Institutes of Health (NIH) kampusa
- 8600 Rockville Pike, Bethesda, MD 20894

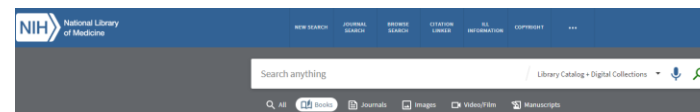
#### Institucija:

- Najveća biomedicinska knjižnica na svijetu
- Dio američkog ministarstva zdravstva (U.S. Department of Health)
- Osnovana 1836. godine

#### Značaj:

- Središte biomedicinskih informacija za cijeli svijet
- Upravlja ključnim bazama podataka (PubMed, MEDLINE)
- Razvija standarde za medicinsku informatiku

Katalog →



## National Library of Medicine (NLM)

### Glavni Resursi:



MEDLINE



PubMed



MeSH



Clinical  
Trials



PMC

### Pretraživanje:

- Boolean operatori (AND, OR, NOT)
- MeSH tezaurus za precizno pretraživanje
- Filteri (tip publikacije, godina, jezik...)

### Značaj:

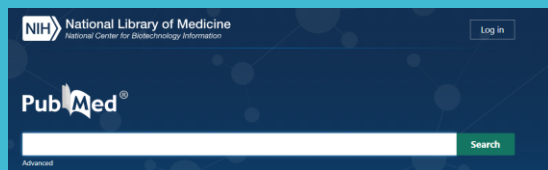
- Najveća biomedicinska knjižnica na svijetu
- Besplatan pristup medicinskoj literaturi



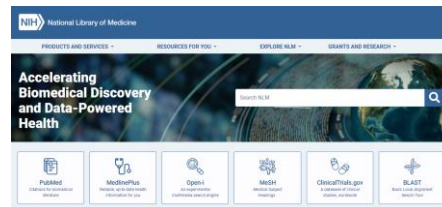


# National Library of Medicine

## Pubmed sučelje

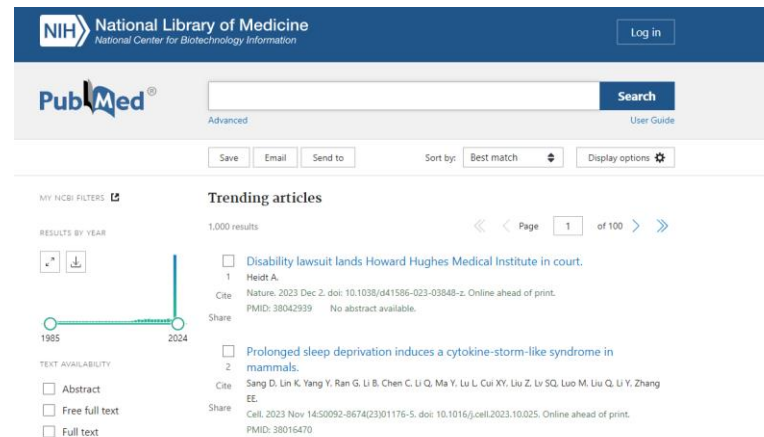


- održava je NLM (National Library of Medicine)



- Sadrži:
  - Članke u časopisima indeksiranim u NLM bazi [MEDLINE](#)
    - uvršteno više od 5000 biomedicinskih časopisa
    - dostupna preko platformi: WOS, OVID, PUBMED, EBSCOhost
    - dnevno ažuriranje
  - online [knjige i dokumente](#) iz područja Biomedicine
  - cjelovite tekstove (PUBMED CENTRAL [PMC](#))
    - članci o istraživanjima financiranim javnim novcem (SAD)

# Pubmed sučelja



Clarivate

Web of Science™

# WEB OF SCIENCE

## sučelje

- zapisi/reference unose se od 1945.
- alati za analizu uspješnosti i kvalitete istraživanja
  - Broj citata
- oko 34,000 časopisa
- jedinstveno sučelje za **više baza**

The screenshot shows a search interface with two main tabs: "DOCUMENTS" and "RESEARCHERS". Below the tabs, there is a search bar with the text "Search in: All Databases ^ Collections: All v". A dropdown menu is open, listing several databases: "Web of Science Core Collection", "BIOSIS Citation Index", "Current Contents Connect", "Data Citation Index", "Derwent Innovations Index", "KCI-Korean Journal Database", "MEDLINE®", "Preprint Citation Index" (with a "New" badge), and "ProQuest™ Dissertations & Theses Citation Index" (with a "New" badge). To the right of the dropdown, there is a section titled "All Databases (1637-present)" with a description: "For the most comprehensive results, search across all subscribed resources using a common set of search fields." Below this description is a search input field and a "Search" button.

# WOS (All databases)

# Current Contents Connect

# CCC

- dio WOS kolekcije
- Vlasnici u prošlosti:
  - Institute for Scientific Information
  - Thomson Reuters
  - **Clarivate Analytics** (danas)
- Indeksira sadržaje preko 10.000 znanstvenih časopisa i drugih publikacija
- Publikacije objavljene od 1989. do danas
- Baza se ažurira jednom tjedno

# WoSCC Core Collection

Science Citation Index Expanded (SCIE)  
Social Sciences Citation Index (SSCI)  
Arts & Humanities Citation Index (AHCI)  
Conference Proceedings Citation Index-  
Science (CPCI-S) – od 1990.  
Conference Proceedings Citation Index-  
Social Science & Humanities (CPCI-  
SSH)  
Book Citation Index – Science  
Book Citation Index – Social Sciences &  
Humanities (BKCI-SSH)  
Emerging Sources Citation Index (ESCI)  
Current Chemical Reactions (CCR-  
EXPANDED)  
Index Chemicus (IC)

- Indeksira sadržaje preko 24 000 znanstvenih časopisa
- Publikacije objavljene od 1955. do danas
- Pretraživanje
  - General Search (Documents)
    - zapisa o indeksiranim publikacijama
  - Cited References
    - popisa korištene literature u indeksiranim publikacijama
  - Structure
    - pretraživanje struktura (npr. kemijskih)
- Časopisima u WoSCC-u na temelju citiranosti članaka izračunava se svake godine čimbenik utjecaja **Impact Factor**

# KLINIČKI POKUSI



The **Epistemonikos Database of Trials** contains over **1,350,000** records of randomized trials sourced from multiple electronic databases, preprint servers, trial registries, and other sources.

A notable source are randomized trials included in systematic reviews from the Epistemonikos Database of Systematic Reviews.

## Sučelje

nizozemskog izdavača Elseviera

Bibliografski zapisi publikacija  
objavljenih od 1790.

### Članci iz časopisa

Knjige

Zbornici

alati za analizu  
uspješnosti i kvalitete istraživanja

### Broj citata

<input type="checkbox"/>	Article	5,532,159
<input type="checkbox"/>	Conference paper	771,632
<input type="checkbox"/>	Review	546,434
<input type="checkbox"/>	Book chapter	362,751
<input type="checkbox"/>	Book	114,145
<input type="checkbox"/>	Editorial	50,260
<input type="checkbox"/>	Letter	48,762
<input type="checkbox"/>	Note	41,791
<input type="checkbox"/>	Short survey	17,008
<input type="checkbox"/>	Erratum	15,487
<input type="checkbox"/>	Conference review	7,425
<input type="checkbox"/>	Retracted	2,065
<input type="checkbox"/>	Data paper	2,046
<input type="checkbox"/>	Report	645
<input type="checkbox"/>	Abstract report	164
<input type="checkbox"/>	Undefined	49
<input type="checkbox"/>	Business article	33



# Baze s podacima o utjecajnosti časopisa

- važne pri odabiru časopisa za objavljivanje
- pokazatelji kvalitete časopisa
- polazište za vrednovanje produktivnosti i utjecajnosti znanstvenika → NAPREDOVANJE

Clarivate

Journal Citation Reports™

Journals

Categories

Publishers

Countries/Regions

SJR

Scimago Journal & Country Rank

# MESH baza

## MEDICINSKI TEZAURUS

- održava je [NCBI](#) (National Center for Biotechnology Information)
- MeSH izraze ručno indeksiraju eksperti koji čitaju znanstvene publikacije i odabiru relevantne MeSH pojmove na temelju sadržaja članaka
- Datoteke se ažuriraju svaki tjedan od ponedjeljka do petka do 20:00 h
- MESH izrazi povećavaju kvalitetu rezultata pretraživanja
- razvijaju se alati koji u tekstu automatski prepoznaju MESH pojmove
  - [MeSH on Demand](#)

# The Cochrane Library

- Cochrane Collaboration
  - Međunarodna neprofitna organizacija koja izrađuje, objavljuje i održava sustavne preglede (*systematic reviews*) o učincima intervencija u zdravstvu
- Cochraneova knjižnica je dostupna hrvatskoj akademskoj zajednici i zdravstvenim ustanovama preko Ovidovog sučelja
  - Ostali je mogu pretraživati do razine sažetaka

# The Cochrane Library

- Cochrane knjižnica sadrži različite baze podataka, najvažnije su:
  - **The Cochrane Database of Systematic Reviews**
    - sustavni pregledni članci i protokoli za članke koji su u postupku izrade
    - najbolji izvor podataka o učinkovitosti terapije
  - **Database of Abstracts of Reviews of Effects (DARE)**
    - strukturirani sažetci kvalitativno procijenjenih sustavnih pregleda koji nisu izrađeni u okviru Cochrane kolaboracije
  - **The Cochrane Central Register of Controlled Trials (CENTRAL)**
    - nezaobilazan izvor randomiziranih kontroliranih pokusa koji se preuzimaju iz različitih baza (MEDLINE, EMBASE...) ili su pronađeni ručnim pretraživanjem literature

## Select Resource(s) to search:

▼  **All Resources**

Books@Ovid January 17, 2017 i

Journals@Ovid Full Text January 19, 2017 i

Your Journals@Ovid i

EBM Reviews - ACP Journal Club 1991 to January 2017 i

EBM Reviews - Cochrane Central Register of Controlled Trials November 2016 i

EBM Reviews - Cochrane Database of Systematic Reviews 2005 to January 18, 2017 i

EBM Reviews - Cochrane Methodology Register 3rd Quarter 2012 i

EBM Reviews - Database of Abstracts of Reviews of Effects 1st Quarter 2015 i

EBM Reviews - Health Technology Assessment 4th Quarter 2016 i

EBM Reviews - NHS Economic Evaluation Database 1st Quarter 2015 i

EBM Reviews Full Text - Cochrane DSR, ACP Journal Club, and DARE i

All EBM Reviews - Cochrane DSR, ACP Journal Club, DARE, CCTR, CMR, HTA, and NHSEED i

AGRICOLA 1970 to December 2016 i

FRIC 1965 to October 2016 i

C  
O  
C  
H  
R  
A  
N  
E  
  
L  
I  
B  
R  
A  
R  
Y

OK

Add Group

Delete Group

# INFO

## sustavi

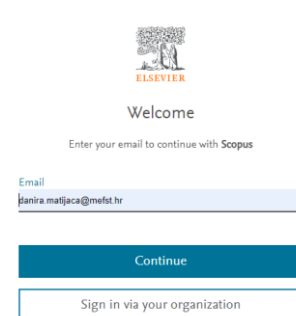
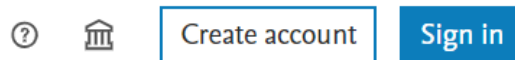
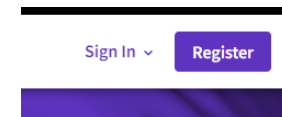
- SUMSearch
  - Istovremeno pretražuje različite medicinske izvore (DARE, Medline, National Guideline Clearinghouse)
  - pristup slobodan
- TRIPDatabase
  - Pretražuje veliki broj EBM-izvora
  - pristup slobodan do određene razine
- UpToDate
  - Komercijalni servis koji daje odgovore na klinička pitanja

# PRISTUPI izvorima

- a) slobodni
- b) uz licencu

NACIONALNE  
LICENCE

- Registrirajte se na fakultetskom računalu
- Možete koristiti bilo koju e-mail adresu (WOS)
- Potvrdite svoju račun putem e-maila (WOS)
- Koristite adresu u sustavu AAI@EduHr (SCOPUS)



- Od kuće slobodno pristupate putem prijave!

# Klasični upiti

## WoS Query

### Karakteristike WoS pretraživanja:

#### Klasične značajke:

- Boolean operatori (AND, OR, NOT)
- Egzaktno podudaranje izraza
- Korištenje zagrada za grupiranje

#### Napredne značajke:

- NEAR/n operator za blizinu riječi
- Filtriranje po metapodacima
- Pretraživanje po citatima

### Primjer WoS upita:

```
TS=(("artificial intelligence" NEAR/3 "information retrieval")  
AND ("search engine*" OR "web search"))
```

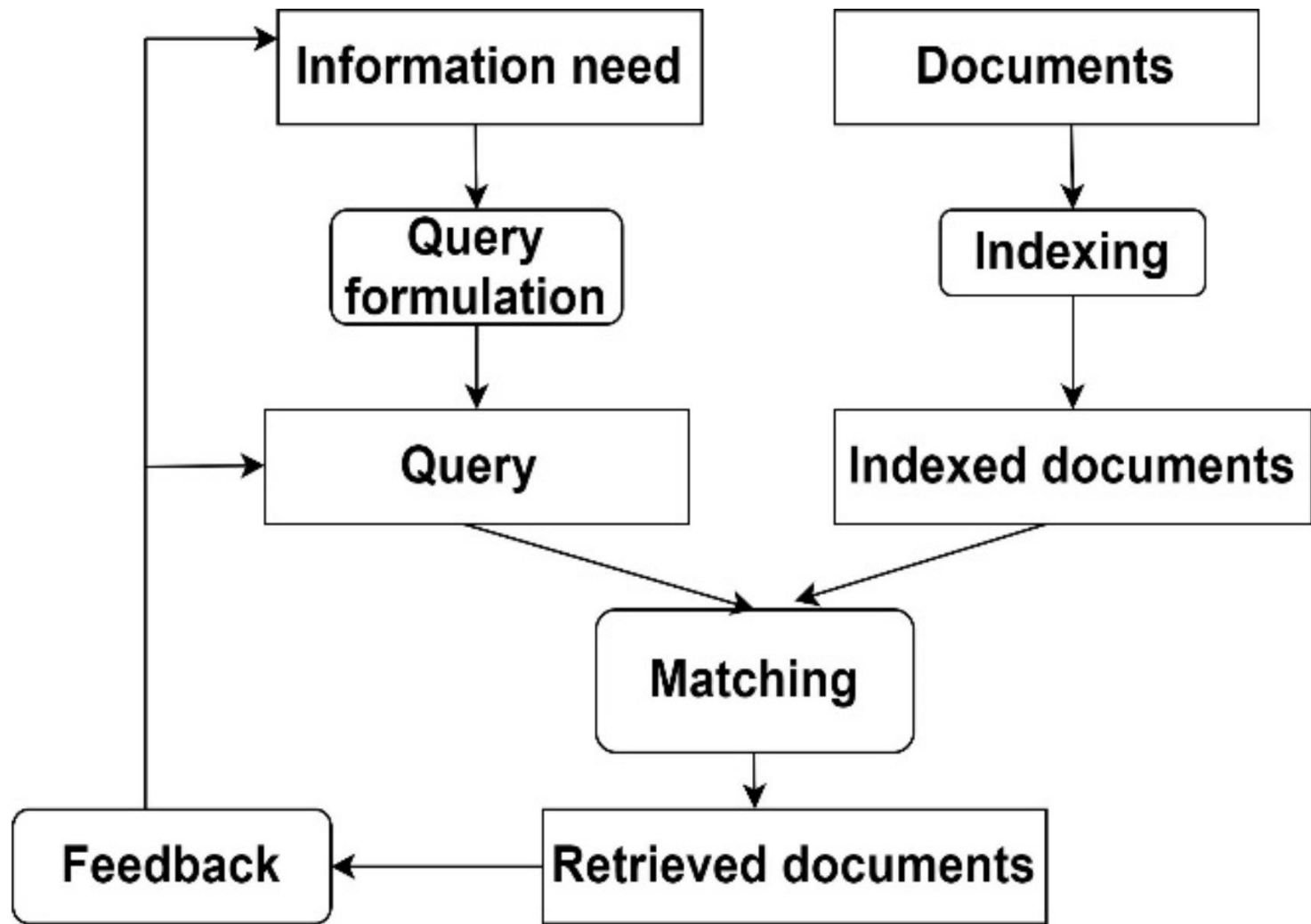
#### Objašnjenje:

- TS = pretraživanje po temi (topic search)
- NEAR/3 = riječi su udaljene najviše 3 riječi



1. Mhawi DN, Olewi HW, Saed NH, Al-Taie HL. An Efficient Information Retrieval System Using Evolutionary Algorithms. Network. 2022 Dec;2(4):583–605.

Ponavljanje  
postupka



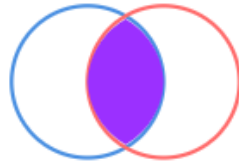
# Oblikovanje upita

- Početna formulaciju upita
  - predviđanje riječi koje će biti uključene u upit
- Preoblikovanje upita
  - izmjena početnog upita radi boljeg odaziva i preciznosti
- Proširenje upita (Query Expansion)
  - preoblikuje početni upit dodavanjem sličnih izraza, izvučenih iz nekoliko izvora
- Optimizacija upita
  - prilagodba parametara pretraživanja

# Bazični Booleanovi operatori

## Booleanovi Operatori

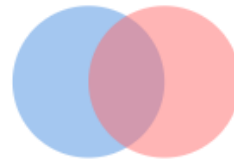
**AND**



Oba pojma moraju  
biti prisutna

A AND B

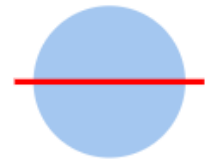
**OR**



Barem jedan pojam  
mora biti prisutan

A OR B

**NOT**



Pojam ne smije  
biti prisutan

NOT A

### Primjer:

"hotel" AND "Zagreb" NOT "hostel"

→ Traži dokumente koji sadrže "hotel" i "Zagreb", ali ne sadrže "hostel"

# Model PICO

## PICO Framework

**P**

Population  
Pacijenti  
Problem

**I**

Intervention  
Intervencija  
Tretman

**C**

Comparison  
Usporedba  
Kontrola

**O**

Outcome  
Ishod  
Rezultat

### Primjer:

P: dijabetes tip 2 | I: metformin | C: placebo | O: razina glukoze u krvi

# AI ALATI

**SR Accelerator**: Alat namijenjen ubrzavanju i optimizaciji procesa revizije i analize istraživanja.

**Notepad LM**: Alat za vođenje bilješki i organizaciju misli, često korišten za brzo bilježenje ideja

• **Chat GPT**: AI asistent za generiranje teksta, koristan za podršku pri učenju, pisanju i istraživanju

**Claude**: Napredni AI model dizajniran za poslovne primjene i analizu podataka

**Scite AI**: Alat za analizu i procjenu znanstvenih radova, koji omogućava korisnicima da vide kako su drugi istraživači citirali radove.