



Praktične vaje pri predmetu Biokemijska informatika

Študijsko leto 2013/2014

Miha Pavšič
marec 2014

Skupine in termini

skupina	termin	asistent
1.	torek, 13:00-16:00	Miha Pavšič
2.	torek, 16:00-19:00	Miha Pavšič
3.	petek, 13:00-19:00	Aljaž Gaber

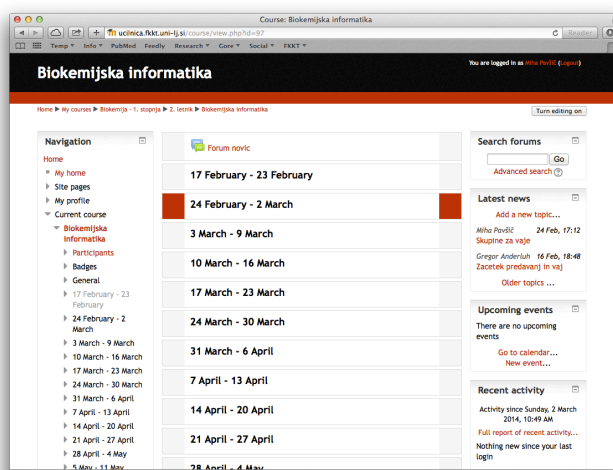
Miha: miha.pavsic@fkkt.uni-lj.si

Aljaž: aljaz.gaber@fkkt.uni-lj.si

- vaje potekajo v **računalniški učilnici**
- vaje so **obvezne**
- opravljene vaje so **pogoj za pristop h kolokviju**
- kolokvij bo **praktične narave** (računalniška učilnica)
- **3 roki** za kolokvij (termini po dogovoru s študenti)

Spletna učilnica UL FKKT (<http://ucilnica.fkkt.uni-lj.si>)

- Biokemija – 1. stopnja > 2. letnik > Biokemijska informatika
- opravljanje vaj je možno zgolj preko spletne učilnice
- vaje in z njimi povezan material bo sproti objavljen
- kategorija “Splošno” (General) – splošno uporabne povezave, ...
- neobvezne domače naloge (+ diskusija na vajah)
- material dostopen tudi od doma (ponovitev vaje, učenje, ...)



Obravnavane teme

vaja	datum*	teme
1.	4. 3.	pisanje besedil, bibliografske baze, citiranje
2.	11. 3.	podatkovne tabele, urejanje in obdelava podatkov, diagrami
3.	18. 3.	prileganje krivulj na podatke, statistična obravnava podatkov
4.	25. 3.	baze nukleotidnih zaporedij, analiza nukleotidnih zaporedij
5.	1. 4.	analiza nukleotidnih zaporedij in molekulska kloniranje
6.	8. 4.	baze aminokislinskih zaporedij, analiza aminokislinskih zaporedij
7.	15. 4.	prileganje zaporedij
8.	22. 4.	iskanje podobnih zaporedij
prvomajski prazniki		
9.	6. 5.	prileganje zaporedij, filogenija
10.	13. 5.	3D struktura: podatkovne baze, vizualizacija in priprava slik
11.	20. 5.	3D struktura: analiza struktur, modeliranje
12.	27. 5.	projektno delo (sinteza znanja na praktičnem primeru)

(*) Datum vaje je zapisan kot datum torika v posameznem tednu.

Vaje v letošnjem študijskem letu oblikujemo na novo; tekom semestra se lahko seznam tem nekoliko spremeni.

Kaj naj bi po opravljenih vajah znali...

Splošna znanja

- iskanje literature po bibliografskih bazah
- oblikovanje seminarske/diplomske/... naloge in citiranje
- urejanje in analiza podatkov v tabelah

Specifična znanja

- iskanje podatkov o makromolekulah in bioloških procesih po različnih bazah
- analiza najdenih/literaturnih in eksperimentalnih podatkov (lokalno, spletna orodja)
 - nukleotidna zaporedja
genom, geni in genske regije, bralni okvir, genski kod in prevajanje zaporedij, eksoni in introni, cDNA, načrtovanje začetnih oligonukleotidov za PCR, restrikcijski encimi, osnove *in silico* molekulskega kloniranja, podobna zaporedja, ...
 - aminokislinska zaporedja
iskanje motivov (glikozilacija, ...), usmerjevalnih zaporedij, transmembranske regije, podobna zaporedja, podobna zaporedja in filogenija, ...
 - strukturni podatki
metode, baze in formati podatkov, programi za vizualizacijo in analizo struktur, modeliranje struktur, umeščanje, ...
- sinteza rezultatov analiz

Praktične vaje pri predmetu
Biokemijska informatika

1. vaja

Pisanje besedil, bibliografske baze in citiranje



ISI Web of
KNOWLEDGE
Transforming Research



 **LibreOffice**
The Document Foundation



 **OpenOffice.org**

 **PubMed**

Pregled vaje

Pisanje / oblikovanje besedil

- oblikovanje naslovov/podnaslovov, ...
- razčlemba dokumenta na dele
- številčenje strani in kazalo
- glava in noga strani
- slike

Bibliografske baze

- primarne in sekundarne bibliografske baza
- iskanje po bibliografskih bazah
- SCI (Science Citation Index)

Citiranje

- programi za delo z referencami oz. "osebnimi" bibliografskimi bazami
- uporaba programa Mendeley za vstavljanje referenc in oblikovanje bibliografije

Urejevalniki besedil

- **lokalno nainstalirani:**
 - Microsoft Word, Write (LibreOffice, OpenOffice), Apple Pages, ...
 - nudijo več možnosti za oblikovanje kot tisti v oblaku
 - številni vtičniki
 - lažja priprava večjih dokumentov
 - deljenje možno preko oblaka ali drugih rešitev (Dropbox)
 - včasih lahko težave pri prenosu med programi pod različnimi OS
- **v oblaku:**
 - Microsoft Office 365, Google Docs, Apple iWork, ...
 - enostavnost uporabe
 - ni potrebna instalacija
 - ponavadi neodvisni od operacijskega sistema; program teče na strežniku, z njim interagiramo preko brskalnika
 - dostopnost do dokumentov z različnih lokacij brez prenašanja pomnilniških medijev
 - enostavna dostopnost tudi z drugih naprav (tablice, telefoni)
 - enostavno deljenje / skupna raba in soustvarjanje dokumentov neposredno iz programa

Praktični del: OBLIKOVANJE BESEDILA

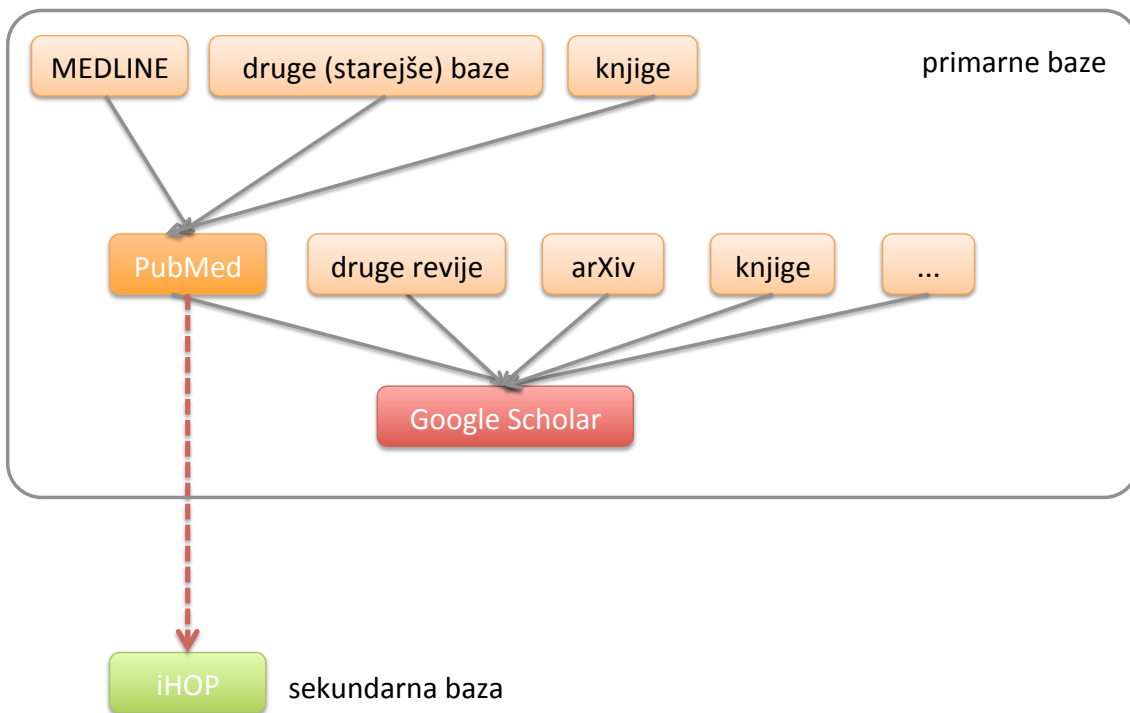
Potek dela

1. S spletne učilnice prenesite datoteko “01 – Oblikovanje besedila in reference.docx”.
2. Datoteko odprite s programom Microsoft Word in si oglejte besedilo. Besedilo še ni oblikovano!
3. Navodila za delo vsebuje datoteka sama.
4. Po končanem oblikovanju shranite spremenjeno datoteko pod istim imenom, saj jo boste potrebovali pri nadaljevanju vaje.
5. Primer oblikovane datoteke je v “01 – Oblikovanje besedila in reference - PRIMER.docx”.

Bibliografske baze podatkov

- so digitalne zbirke referenc iz **objavljene literature** (članki, knjige, prispevki na konferencah, poročila, patenti, ...)
- lahko so **splošne** ali pa pokrivajo **določeno znanstveno disciplino**
 - Google Scholar – splošen iskalnik
 - PubMed – biomedicinska literatura
 - arXiv – matematika, fizika, astronomija, računalništvo, ...
 - IEEE in ACM – tehnika, elektronika, računalništvo, ...
 - ...
- lahko so **prosto dostopne** (npr. PubMed) ali **plačljive** (npr. Web of Science, Scopus)
- vnosi so lahko razdeljeni po **kategorijah**, kar pa je zelo kompleksno – namesto tega se uporabljajo **ključne besede** (“keywords”)
- nekatere podatkovne baze nudijo **orodja za analizo vnosov** (geografski podatki, citiranost, frekvenca objav, ...)
- baze so lahko:
 - **primarne** – vsebujejo posamezne bibliografske enote”, ki so vnešene direktno v posamezno bazo ali v povezane baze; npr. PubMed/MEDLINE, arXiv, ...
 - **sekundarne** – (avtomatsko so) pripravljene so na osnovi primarnih baz; primer je iHOP

Primarne in sekundarne bibliografske baze - primer

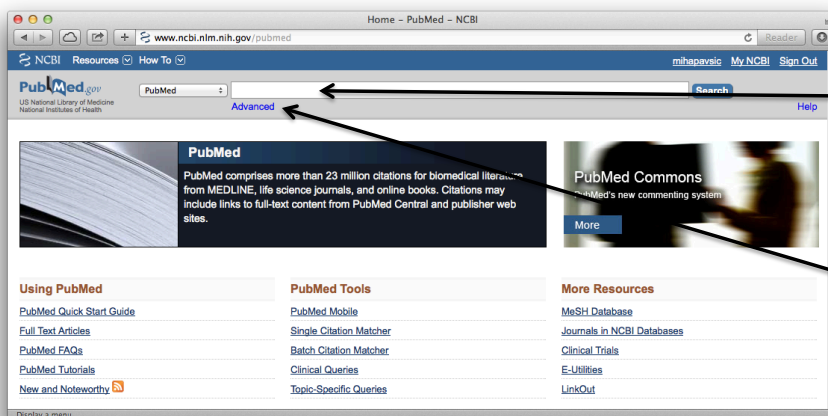


PubMed

<http://www.ncbi.nlm.nih.gov/pubmed>



- **prosto dostopna** bibliografska baza literature iz **biomedicine** in drugih naravoslovnih znanosti
- vzdržuje jo National Library of Medicine v ZDA (US NLM) v okviru NIH (National Institutes of Health)
- omogoča iskanje po bazi MEDLINE in nekaterih drugih bazah
- vsebuje članke iz pribl. 5 600 revij v 39 jezikih (91% v angleščini)
- vsak dan novih 2 000 – 4 000 vnosov



osnovno iskanje
(vnešena gesla niso kategorizirana)

napredno iskanje
(vnešena gesla (avtor, leto, ...) kategoriziramo s pomočjo **Builder**-ja)

PubMed

- primer vnosa v bazo

revija, leto
objave, letnik
(in številka),
strani

naslov

avtor(ji)

povzetek

PMID
(unikatna
številka,
PubMed ID)

(slike)

The screenshot shows a PubMed article page. The title is "Macromolecular complexes in crystals and solutions" by Krissinel E. The abstract discusses methods for analyzing macromolecular interactions. The PMID is 21460456. There are several callout boxes: one for the journal information (Acta Crystallogr D Biol Crystallogr, 2011), one for the author information, one for the abstract, one for the images from the publication, one for the related citations in PubMed, and one for articles cited by PubMed Central.

povezave do članka v celoti

- PubMed Central (prost dostop)
- zunanje povezave

sorodni članki

članki v
PMC, ki ta
članek
citirajo

PubMed – My NCBI

- lahko si ustvarimo **brezplačen uporabniški račun**
- omogoča shranjevanje zadetkov, shranjevanje iskanj, **avtomatsko periodično iskanje po bazi** (novi vnosi v bazo po e-pošti), prilagoditve filtrov, ...

osebne
bibliografije

pretekla
iskanja
(zgodovina)

The screenshot shows the My NCBI user interface. It includes a search bar, a "My Bibliography" section (currently empty), a "Recent Activity" table, a "Saved Searches" table, a "Collections" table, and a "Filters" section. The "Saved Searches" table has the following data:

Search Name	What's New	Last Searched
Testician	0	today
GA733	0	today
EpCAM	0	today
Trop2	0	today

The "Collections" table has the following data:

Collection Name	Items	Settings/Sharing	Type
Favorites	edit 0	Private	Standard
My Bibliography	edit 0	Private	Standard
Other Citations	edit 0	Private	Standard

The "Recent Activity" table has the following data:

Time	Database	Type	Term
3:05 PM	PubMed	record	Macromolecular complexes in crystals and solutions,
3:05 PM	PubMed	search	krissinel
2:55 PM	Books	record	PubMed Help - PubMed Help

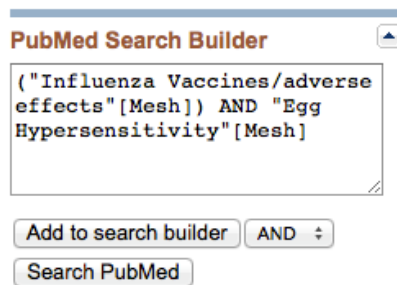
shranjena iskanja
(avtomatsko
periodično
ponavljanje)

zbirke referenc

filtri

PubMed – MeSH (<http://www.ncbi.nlm.nih.gov/mesh>)

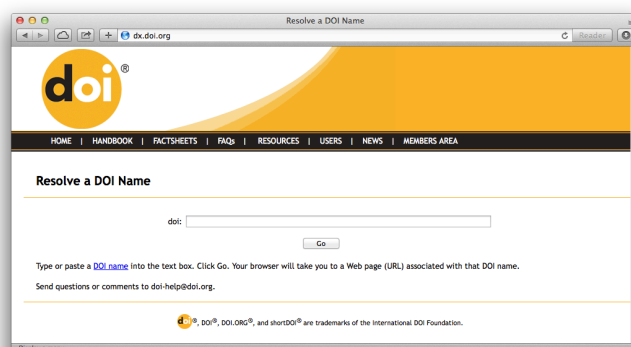
- Medical Subject Headings
- urejen besednjak, ki se uporablja za indeksiranje člankov v PubMed (ni nujno, da so vsi novejši članki že indeksirani) – previdno pri uporabi!
- omogoča gradnjo **popolnejših** in **učinkovitejših** iskanj
- primer: želimo poiskati članke, ki opisujejo povezavo med preobčutljivostjo na jajca (*egg allergy*) in negativnimi stranskimi učinki cepiv proti gripi (*flu vaccine*) → MeSH pojma združimo z logičnim operatorjem AND in izvedemo iskanje



Influenza Vaccines in Egg Hypersensitivity sta osnovna pojma (naslova), adverse effects pa je podpojem (podnaslov), s katerim natančneje omejimo rezultate

DOI (Digital Object Identifier)

- koda, ki **unikatno** opiše nek digitalen objekt, npr. znanstveni članek
- primer: 10.1038/nano.2014.13
- v povezavi s kodo DOI so shranjeni različni **metapodatki** (npr. lokacija – URL)
- **koda DOI je trajna**, metapodatki pa se lahko spremenijo
- praktičnost: koda DOI nekega članka ostane enaka, založniška hiša pa lahko naslov dokumenta PDF spremeni → članek še zmeraj najdemo
- sistem koordinira International DOI Foundation
- kodo DOI določenega članka lahko najdemo na PubMed (ne pri vseh člankih), na strani izdajatelja revije, ...
- več: http://en.wikipedia.org/wiki/Digital_object_identifier



Naslov, povezav s kodo DOI, lahko najdemo preko <http://dx.doi.org> (ali pa v iskalnik vpišemo doi koda_doi).

Web of Science (<http://webofknowledge.com/WOS>) WEB OF SCIENCE™

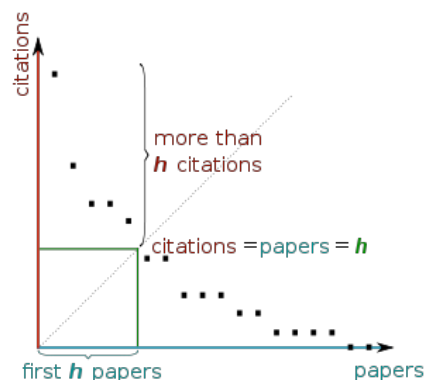
- plačljiv servis (dostop mogoč npr. iz omrežja Univerze v Ljubljani)
- omogoča iskanje referenc/člankov ter člankov, kjer so le-ti citirani
- v okviru Web of Science je baza **Science Citation Index (SCI)**
- **ISI Web of Knowledge** indeksira številne znanstvene revije → indeks vpliva (*impact factor, IF*); to področje imenujemo bibliometrika; dostopen tudi preko COBISS

$$\text{faktor vpliva za neko revijo za leto 2012} = \frac{\text{kolikokrat so bili članki, objavljeni v tej reviji v letih 2010 in 2011, citirani v drugih indeksiranih revijah v člankih v letu 2012}}{\text{število vseh člankov in drugih objav, ki se jih lahko citira, objavljenih v tej reviji v letih 2010 in 2011}}$$

- primerjava revij znotraj enega področja (v nasprotnem primeru izkrivljeno!)
- uporaba faktorja vpliva je nekoliko kontroverzna; več lahko preberete tukaj: http://en.wikipedia.org/wiki/Impact_factor oz. http://sl.wikipedia.org/wiki/Faktor_vpliva

h-indeks (Hirschev indeks)

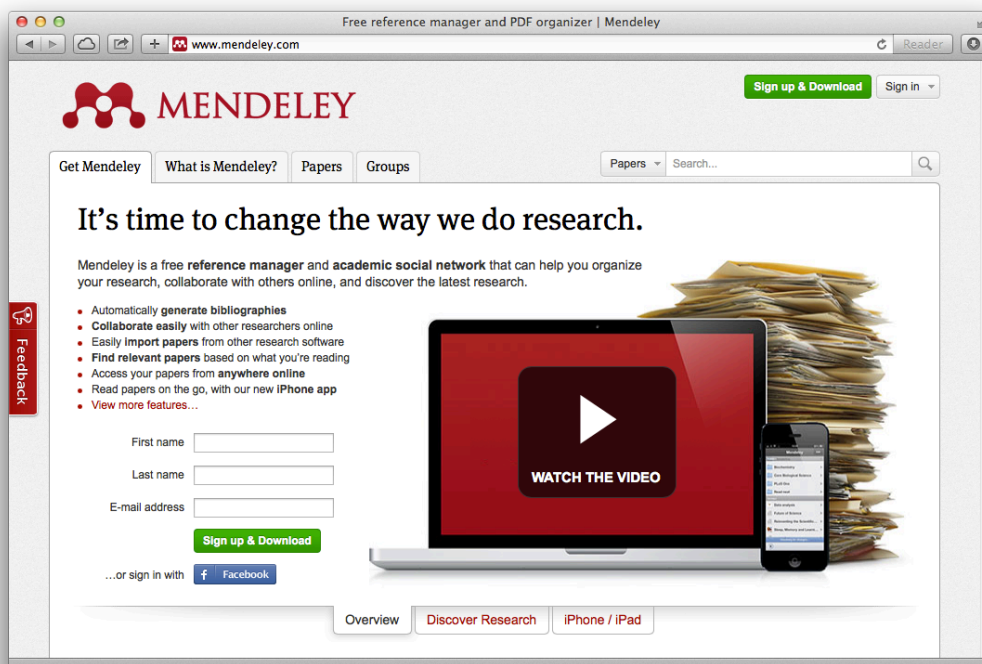
- meri produktivnost in vpliv posameznega znanstvenika, skupine, revije, ...
- temelji na najbolj citiranih člankih in številu citatov, ki so jih le-ti prejeli v drugih publikacijah
- več: <http://en.wikipedia.org/wiki/H-index>
- pomen vrednosti: znanstvenik z osebnim h-indeksom X je objavil X člankov, od katerih je vsak bil najmanj X-krat citiran
- osebni h-indeks je povezan s starostjo (narašča ali ostane enak)



Programi za delo z referencami (*Reference managers*)

- ko najdemo zanimive članke, ki jih bomo morda celo citirali v naših delih (diplomska dela, seminarske naloge, članki, ...), jih shranimo v **osebno bibliografsko bazo**
- osebno bibliografsko bazo najenostavneje gradimo s pomočjo **posebnih programov**
- razvrščanje referenc po kategorijah, zabeležke, shranjevanje člankov v celoti (PDF), ločena polja (avtorji, revija, leto, letnik, številka, strani, ...)
- enostavno **vstavljanje** referenc v nastajajoče besedilo (poljuben vrstni red vstavljanja) in oblikovanje končne **bibliografije** (seznama citiranih del)
- enostavno **preoblikovanje** vstavljenih referenc in bibliografije (različne revije imajo različne zahteve)
- http://en.wikipedia.org/wiki/Comparison_of_reference_management_software
- **programi** so nainstalirani lokalno, **baza podatkov** pa je bodisi lokalna ali v oblaku

Mendeley (<http://www.mendeley.com>)

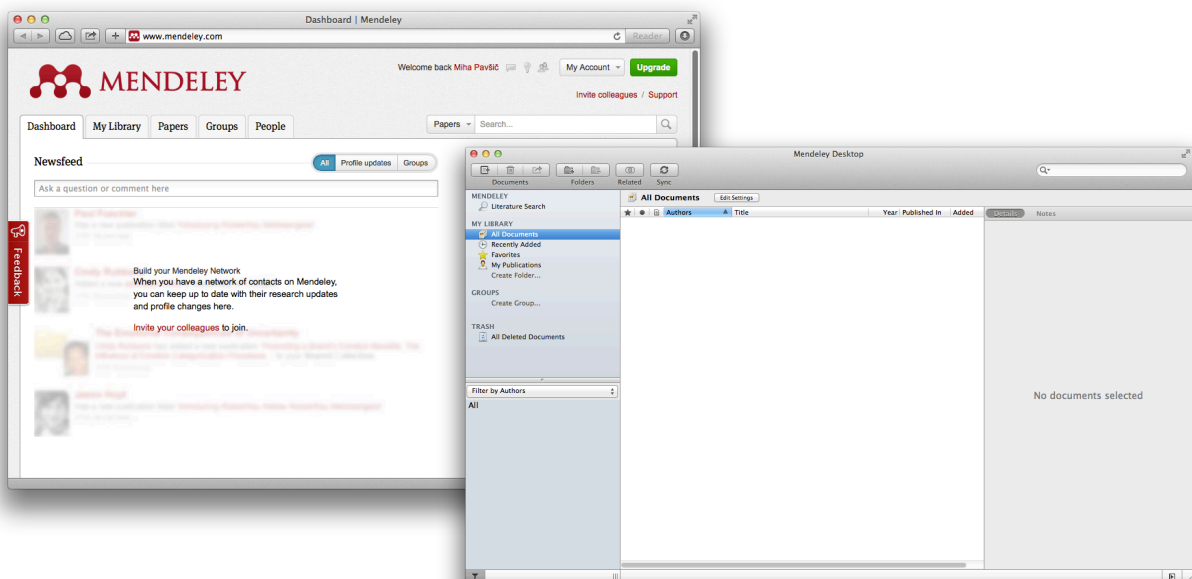


Mendeley (<http://www.mendeley.com>)

- **lokalno** nainstaliran (Windows, OS X, Linux, iOS)
- naša osebna bibliografska baza je **v oblaku** (2 GB brezplačno)
- do bibliografske baze lahko dostopamo tudi preko brskalnika (**Dashboard**)
- vnos v bazo preko lokalnega programa: Mendeley baza ali *import* datoteke (npr. **bib**) ali preko brskalnika (*Mendeley Dashboard* ali preko vtičnika za brskalnik)
- **vstavljanje referenc** v Word, Writer (OpenOffice/LibreOffice) in oblikovanje končne **bibliografije**
- **shranjevanje datotek** PDF, zabeležke & “markiranje” besedila
- **deljenje referenc in beležk** (omejitve glede št. uporabnikov pri brezplačnem računu)
- iskanje po bazi
- vse značilnosti: <http://www.mendeley.com/features/>
- navodila in (video)-tečajji: <http://resources.mendeley.com>

Praktični del: MENDELEY

1. Ustvarite brezplačni uporabniški račun (potrebujete e-poštni naslov).
2. “Instalirajte vtičnik” za brskalnik.
3. Aplikacija je že nainstalirana lokalno – zaženite jo in se vpišite z vašim računom.



Praktični del: MENDELEY

1. V Word-u ponovno odprite datoteko (s prej oblikovanim besedilom) “01 – Oblikovanje besedila in reference.docx”.
2. Vstavite primerne reference v zadnji odstavek, ki vsebuje strokovno besedilo.
 - V PubMed poiščite za vsako poved eno (ali več, kakor je pač smiselno) referenc v PubMed.
 - Uporabite napredni način iskanja, ...
 - Uporabljene reference združite v enotno skupino v Mendeley (z imenom npr. “Biokeminfo”).
 - Na konec besedila dodajte bibliografijo oz. jo oblikujte tako, da bo v skladu s standardi revije:
 - Nature
 - Science
 - PNAS (Proceedings of the National Academy of Sciences)
 - ...
3. Oglejte si, kaj se zgodi, če novo referenco vstavimo med/pred že vstavljene reference. Na koncu pripravite datoteko brez “Mendeley Fields”.
4. Primer oblikovane datoteke je v “01 – Oblikovanje besedila in reference - PRIMER.docx”.

iHOP (<http://www.ihop-net.org>)

- primer **sekundarne** baze oz. iskalnika (postjo dostopen)
- **uporaba imen genov in proteinov kot povezave** med povedmi in povzetki člankov → **navigacija** s povezavami na članke (povzetke) na PubMed
- dobro: enostavnost, hitrost, preglednost
- slabo: nereprezentativnost, nepopolnost
- pomoč/navodila: <http://www.ihop-net.org/UniPub/iHOP/help.html>

The image displays two screenshots of the iHOP (Information Hyperlinked over Proteins) website. The left screenshot shows the main interface with a search bar, navigation menu, and a network diagram of gene interactions. The right screenshot shows a search result page for 'MYO5B' with a list of abstracts and a 'Find in this Page' feature.