

# **Vyhledávání informací**

Hodnocení výsledků vyhledávání

5. 11. 2021

# Rozehrivací vyhledávačka



Jaká zvířata jsou vyobrazena na straně 67 latinského rukopisu od numismatika a historika Chrysostoma Hanthalerera vydaného v roce 1741?

# Rozehrivací vyhledávačka

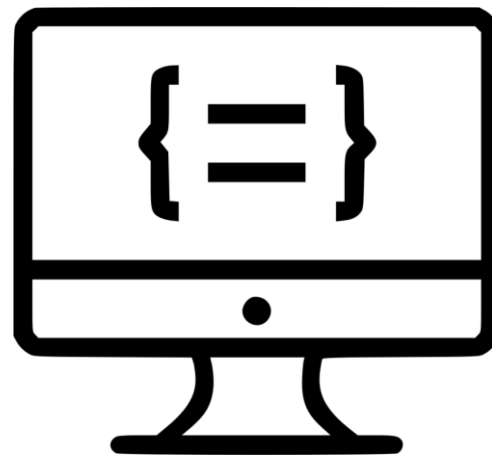


Jaká zvířata jsou vyobrazena na straně 67 latinského rukopisu od numismatika a historika Chrysostoma Hanthamera vydaného v roce 1741?

**Manuscriptorium**

# Jak hodnotit vyhledávání

- systémové – testovací kolekce, R/P
- uživatelské – zadavatel s potřebou a řešeršér
- základní rozměr: relevance
- dílčí hlediska: aktuálnost, včasnost, počet záznamů...



# Relevance

- základní přístup k hodnocení: relevance
- Vůči čemu? -> modelům IR, potřebě atp.
- dokument v kolekci posuzujeme vůči informační potřebě
- relevantní / nerelevantní  $d$
- můžeme hodnotit binárně
- může to být škála...

# Relevance

- relevance se posuzuje vůči potřebě, nikoliv vůči q

Je tmavý cukr zdravější než bílý cukr?

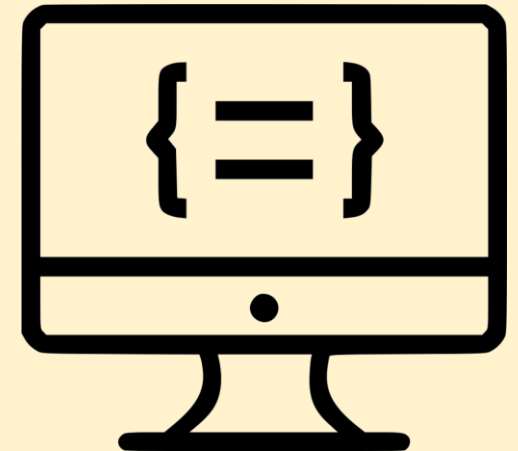
cukr AND tmavý AND bílý AND zdraví AND obezita

relevantní dokument = řeší potřebu

relevantní dokument  $\neq$  obsahuje všechna klíčová slova

# Systemové hodnocení

- viděli jsme, že metod jak vystavět IR systém je kupa
- jak zjistíme, že dávají smysl?
- máme testovací kolekce, *viz minule*
- OK, ale co s nimi?
- jak skutečně říci, že jedna metoda je lepší než jiná?



# Přesnost a úplnost

- přesnost (precision) - P
- úplnost (recall) - R



# Přesnost (precision)

- přesnost je množina vyhledaných relevantních d, vůči všem vyhledaným dokumentům
- $P = (\# \text{ relevantní vyhledané } d / \# \text{ vyhledané } d)$
- 0 až 1 (%)

# Úplnost (recall)

- úplnost je množina relevantních vyhledaných d, vůči celkovému počtu relevantních d v kolekci
- $R = (\# \text{ relevantní vyhledané } d / \# \text{ relevantní } d \text{ v kolekci})$
- jak se to počítá v reálném světě?

# Úplnost + přesnost

	<i>relevantní</i>	<i>nerelevantní</i>
vyhledané	$v_r$	$v_n$
nevyhledané	$n_r$	$n_n$

$$P = v_r / (v_r + v_n)$$

$$R = v_r / (v_r + n_r)$$

System vrátí 8 relevantních dokumentů a 10 nerelevantních dokumentů. V kolekci je celkem 20 relevantních dokumentů. Jaká je přesnost tohoto vyhledávacího systému a jaká je jeho úplnost?

# Úplnost + přesnost

- P a R jsou propojené
- Recall = 1 je jednoduché
- jak toho dosáhneme?



# Úplnost + přesnost

- P a R jsou propojené
- $R = 1$  je jednoduché = vrátíme celou kolekci...
- co to bude znamenat pro P?
  
- *P klesá jak roste R*
- cílem není mít nejvyšší P nebo R
- na webu to bude jinak než v DB právních textů
- jejich vyladěnost se bude lišit i podle odvětví

# Proč máme dvě míry?

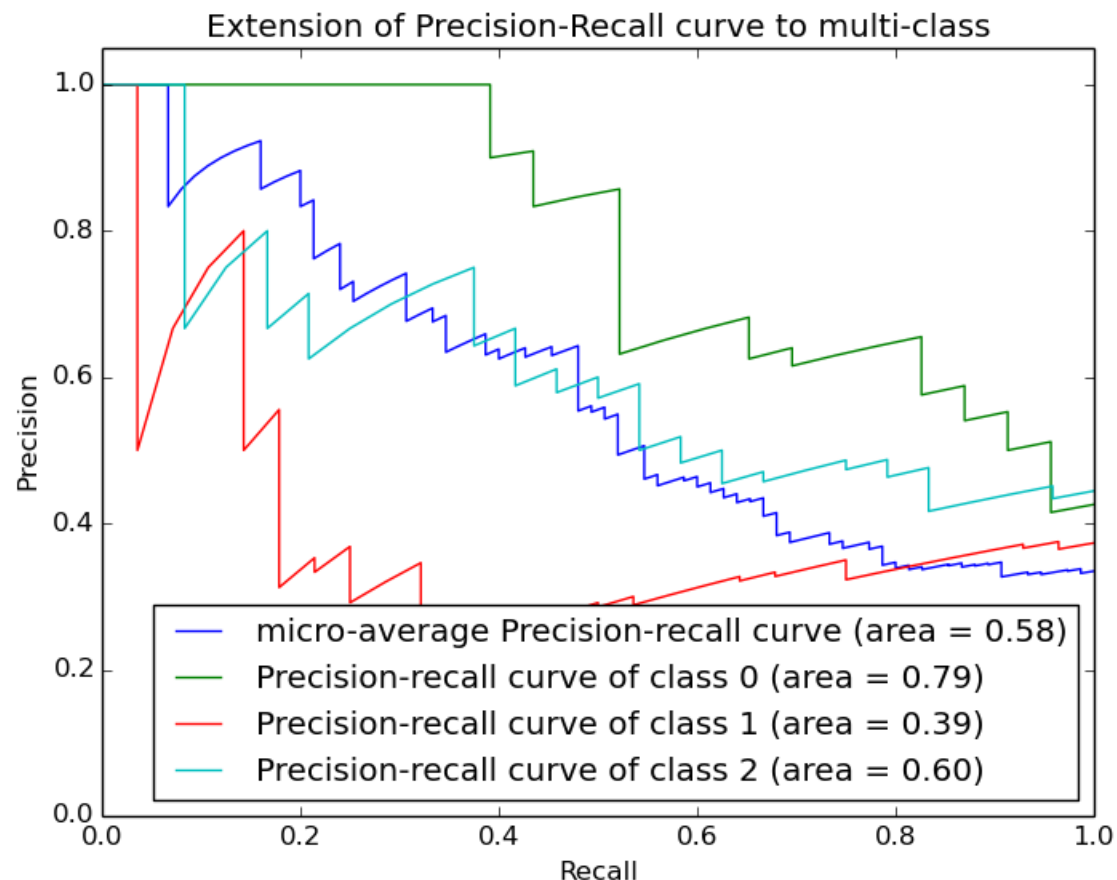
- $accuracy = (v_r + n_n) / (v_r + v_n + n_r + n_n) = \text{„správnost“}$
- většinou je 99 % nerelevantních
- *IR systém vyladěný na maximalizaci accuracy...*
- jak jednoduše dosáhneme **accuracy = 1?** *-> raději nic nemáme*
- vyvažujeme tedy mezi sebou P a R
- *F-measure* (*F1*) – harmonický průměr P a R = jedno číslo
- existují komerční subjekty, které ladí P a R

# Evaluace vážených výsledků

- P a R lze aplikovat na nevážené výsledky
- u řazených výstupů: *precision-recall curve*
- mnohé další metody
- *precision at k*

Více k evaluaci vážených výsledků:

MANNING, Christopher D., Prabhakar RAGHAVAN a Hinrich SCHUTZE. *Introduction to information retrieval*. New York: Cambridge University Press, 2008. ISBN 0521865719.



# Kritika P+R

- hodnocení je binární,
  - nesubjektivní – změna vstupu, změna výstupu
  - informační potřeba se bere jako stabilní, neměnná
  - hodnocení je vázáno na kolekci a doménu
- 
- $d$  se hodnotí vůči  $q$ , nikoliv vůči sobě navzájem
  - *marginal relevance* ([Carbonell and Goldstein, 1998](#))
    - relevantní dokument může být redundantní
    - je třeba hodnotit i novost
    - těžko měřitelné



# Další hodnocení systému

Napadají vás další aspekty kromě  $P$  a  $R$ , které by bylo možné hodnotit z hlediska systému?

# Další hodnocení systému

Napadají vás další aspekty kromě  $P$  a  $R$ , které by bylo možné hodnotit z hlediska systému?

- rychlost indexace
- rychlost porovnávání  $q$  a  $d$
- srozumitelnost vyhledávacího jazyka
- ...

# Udržovací vyhledávačka



S kým si chtěli dopisovat dva inteligentní rotmistři podle inzerátu vyšlého v ranním vydání Lidových novin dne 22. února 1924?

# Uživatelské hodnocení

- hodnotit výstupy vůči q je celkem jednoduché
- co když chceme hodnotit vůči info. potřebě?
- jak měřit naplnění informační potřeby?



# **Uživatelské hodnocení**

- **automatizovaně, statisticky**
- návrat na vyhledávač, čas v systému, A/B testování, proklikovost a analýza logů, analýza clickstreamu,
- **uživatelský výzkum**
- etnografické zkoumání, pozorování, zadávání úkolů, kvalitativní metody...

# Uživatelské hodnocení

- různé metody vylepšování UX (LOL, [Seznam!](#))
- náhledy (snippets)
- statické a dynamické náhledy
- sumarizace textu (NLP)
- KWIC

mortgages



All News Images Maps Shopping More Settings Tools

About 652,000,000 results (0.48 seconds)

MoneySuperMarket › mortgages

## Compare The Best Mortgage Rates | MoneySuperMarket

Compare **mortgages** to find out how much you can borrow and what the repayments will actually cost you. Search for remortgages, buying to let and first time.

How do mortgages work? ▾

How do you get a mortgage? ▾

How much mortgage can you afford? ▾

▾ Show more

Halifax › uk ▾

## Buy a Home With Halifax | Mortgages | Halifax UK

We've helped people buy their own home for over 160 years. With everything under one roof, getting expert advice to find the best **mortgage** for you is easy.

Barclays › uk › mortgages › mortgage-calculator ▾

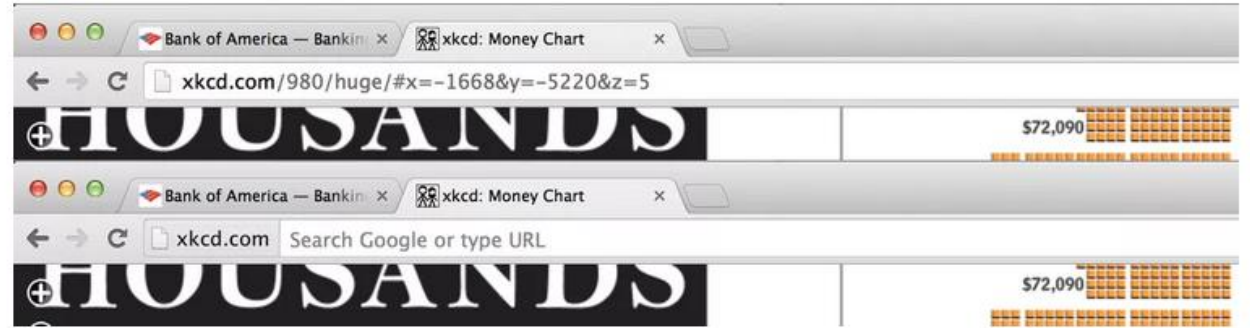
## Mortgage calculator | How much mortgage can I afford ...

**Mortgage** calculators. ... Use our **mortgage** affordability calculators to work out how much you could borrow and what kind of deposit you need for a **mortgage**. ... Use our offset **mortgage** calculator to see how your savings could reduce your **mortgage** term or monthly payments.

### People also ask

What is the meaning of mortgage loan? ▾

Can I borrow 5 times my salary for a mortgage? ▾



# Hodnocení dokumentu/zdroje

- Jak hodnotit kvalitu zdroje?



# **Přístupy k hodnocení *d***

- seznamy tipů, mnohdy nesouvislé
- <https://guides.library.jhu.edu/evaluate/internet-resources>

# Přístupy k hodnocení *d*

- přístupy založeny na komplexnějších checklistech
- vznikly jako vodítka pro výběr knihovnických zdrojů (70. léta)
- *odpovídají prostředí webu?*

- CRAAP, RADCAB,...
- AAOCC (Authority, Accuracy, Objectivity, Currency, Coverage)

CRAAP (2004)	Tate & Alexander (1996)	CHIN (1978)
Currency	Currency	Currency
Relevance	Coverage	Scope of coverage Audience level
Authority	Authority	Accuracy
Accuracy	Accuracy	Accuracy
Purpose	Objectivity	Point of view

# Přístupy k hodnocení *d*

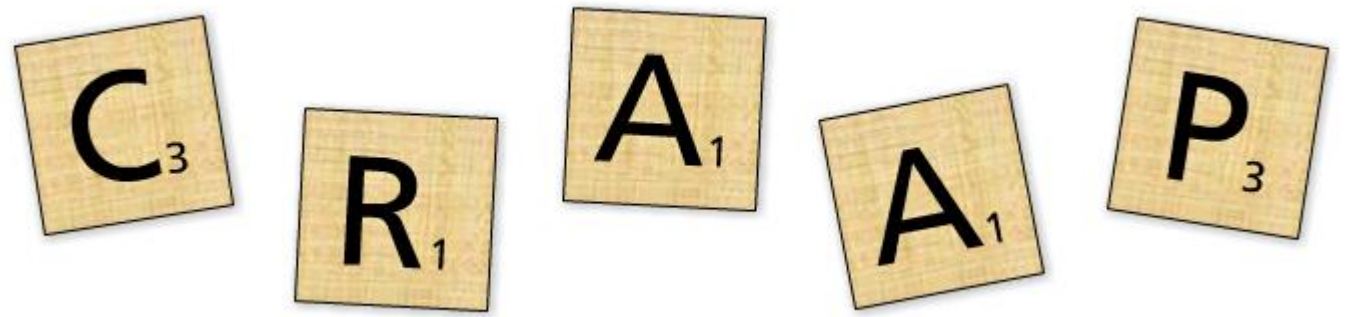
- checklisty pro specifické oblasti
- např. zdravotnictví, žurnalistika...
- <https://nlk.cz/zdroje/medlike/metodika-hodnoceni-kvality-zdroju/>

**Table 2: The Common Dimensions of IQ/DQ**

<b>Dimension</b>	<b># of times</b>	<b>Definitions *1[Wang &amp; Strong; 1996]</b>
1 Accuracy	8	extent to which data are correct, reliable and certified free of error *1
2 Consistency	7	extent to which information is presented in the same format and compatible with previous data *1
3 Security	7	extent to which access to information is restricted appropriately to maintain its security *1
4 Timeliness	7	extent to which the information is sufficiently up-to-date for the task at hand *1
5 Completeness	5	extent to which information is not missing and is of sufficient breadth and depth for the task at hand *1
6 Concise	5	extent to which information is compactly represented without being overwhelming (i.e. brief in presentation, yet complete and to the point) *1
7 Reliability	5	extent to which information is correct and reliable *1
8 Accessibility	4	extent to which information is available, or easily and quickly retrievable *1
9 Availability	4	extent to which information is physically accessible
10 Objectivity	4	extent to which information is unbiased, unprejudiced and impartial *1
11 Relevancy	4	extent to which information is applicable and helpful for the task at hand *1
12 Useability	4	extent to which information is clear and easily used
13 Understandability	5	extent to which data are clear without ambiguity and easily comprehended *1
14 Amount of data	3	extent to which the quantity or volume of available data is appropriate *1
15 Believability	3	extent to which information is regarded as true and credible *1
16 Navigation	3	extent to which data are easily found and linked to
17 Reputation	3	extent to which information is highly regarded in terms of source or content *1
18 Useful	3	extent to which information is applicable and helpful for the task at hand *1
19 Efficiency	3	extent to which data are able to quickly meet the information needs for the task at hand *1
20 Value-Added	3	extent to which information is beneficial, provides advantages from its use *1

# CRAAP test

- Currency, Relevance, Authority, Accuracy, Purpose
- *California State University, Chico*
- <https://library.csuchico.edu/help/source-or-information-good>



**C**

## Currency

### The timeliness of the information

When was the information published or posted?  
Has the information been revised or updated?  
Does your topic require current information, or will older sources work as well?  
Are the links functional?

**R**

## Relevance

### The importance of the information for your needs

Does the information relate to your topic or answer your question?  
Who is the intended audience?  
Is the information at an appropriate level (i.e. not too elementary or advanced for your needs)?  
Have you looked at a variety of sources before determining this is one you will use?  
Would you be comfortable citing this source in your research paper?

**A**

## Authority

### The source of the information

Who is the author/publisher/source/sponsor?  
What are the author's credentials or organizational affiliations?  
Is the author qualified to write on the topic?  
Is there contact information, such as a publisher or email address?  
Does the URL reveal anything about the author or source?  
examples: .com .edu .gov .org .net

**A**

## Accuracy

### The reliability, truthfulness and correctness of the content

Where does the information come from?  
Is the information supported by evidence?  
Has the information been reviewed or refereed?  
Can you verify any of the information in another source or from personal knowledge?  
Does the language or tone seem unbiased and free of emotion?  
Are there spelling, grammar or typographical errors?

**P**

## Purpose

### The reason the information exists

What is the purpose of the information? Is it to inform, teach, sell, entertain or persuade?  
Do the authors/sponsors make their intentions or purpose clear?  
Is the information fact, opinion or propaganda?  
Does the point of view appear objective and impartial?  
Are there political, ideological, cultural, religious, institutional or personal biases?

# RADCAB

- CRAAP používají především univerzity
- RADCAB pro nižší stupně vzdělání
- <https://www.radcab.com/>

**R**ELEVANCY  
**A**PPROPRIATENESS  
**D**ETAIL  
**C**URRENCY  
**A**UTHORITY  
**B**IAS

Putin zápasil s medvědem  
holýma rukama a vyhrál.





# Alternativní přístupy

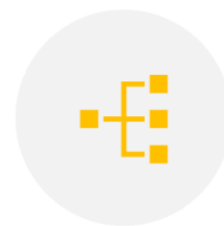
- checklistové vs procesní přístupy
- SIFT ([The Four Moves](#))
- zaměřeno na webové zdroje
- místo checklistu soubor aktivit



STOP



INVESTIGATE THE  
SOURCE



FIND BETTER COVERAGE



TRACE CLAIMS, QUOTES  
AND MEDIA TO THE  
ORIGINAL CONTEXT

# Specifická hlediska

- stav dokumentu: preprinty
- přítomnost dat k ověření
- <https://kosara.net/publications/Kosara-VISShort-2019.html>

# Uklidňovací vyhledávačka



Co působí rozkoš matematikům podle  
*belletristického týdeníku* LUMÍR,  
9. ročník, číslo druhé?