Vyhledávání informací

Hodnocení výsledků vyhledávání

10. 11. 2023

Rozehřívací vyhledávačka

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

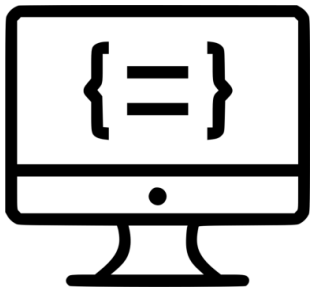
Rozehřívací vyhledávačka

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

Manuscriptorium

Jak hodnotit vyhledávání

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



Co je to relevance?

muni.cz/go/vyhl23

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?

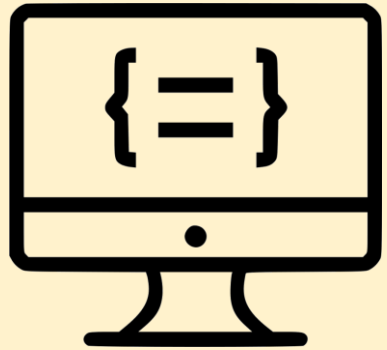
cukr AND tmavý AND bílý AND zdraví AND obezita

relevantní dokument = řeší potřebu

relevantní dokument ≠ obsahuje všechna klíčová slova

„Systémové“ hodnocení

- viděli jsme, že metod jak vystavět IR systém je kupa
- jak zjistíme, že dávají smysl, že jsou efektivní?
- tak třeba 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 = (# relevantní vyhledané d / # vyhledané d)
- 0 až 1 (%)

Úplnost (recall)

- úplnost je množina relevantních vyhledaných *d*, vůči celkovému počtu relevantních *d* v kolekci
- R = (# relevantní vyhledané d / # relevantní d v kolekci)

- *lze to počítat v reálném světě?*

Úplnost + přesnost

	relevantní	*nerelevantní*
vyhledané	v_r	v_n
nevyhledané	n_r	n_n

$P = v_r / (v_r + v_n)$

$R = v_r / (v_r + n_r)$

Úplnost + přesnost

	relevantní	*nerelevantní*
vyhledané	v_r	v_n
nevyhledané	n_r	n_n

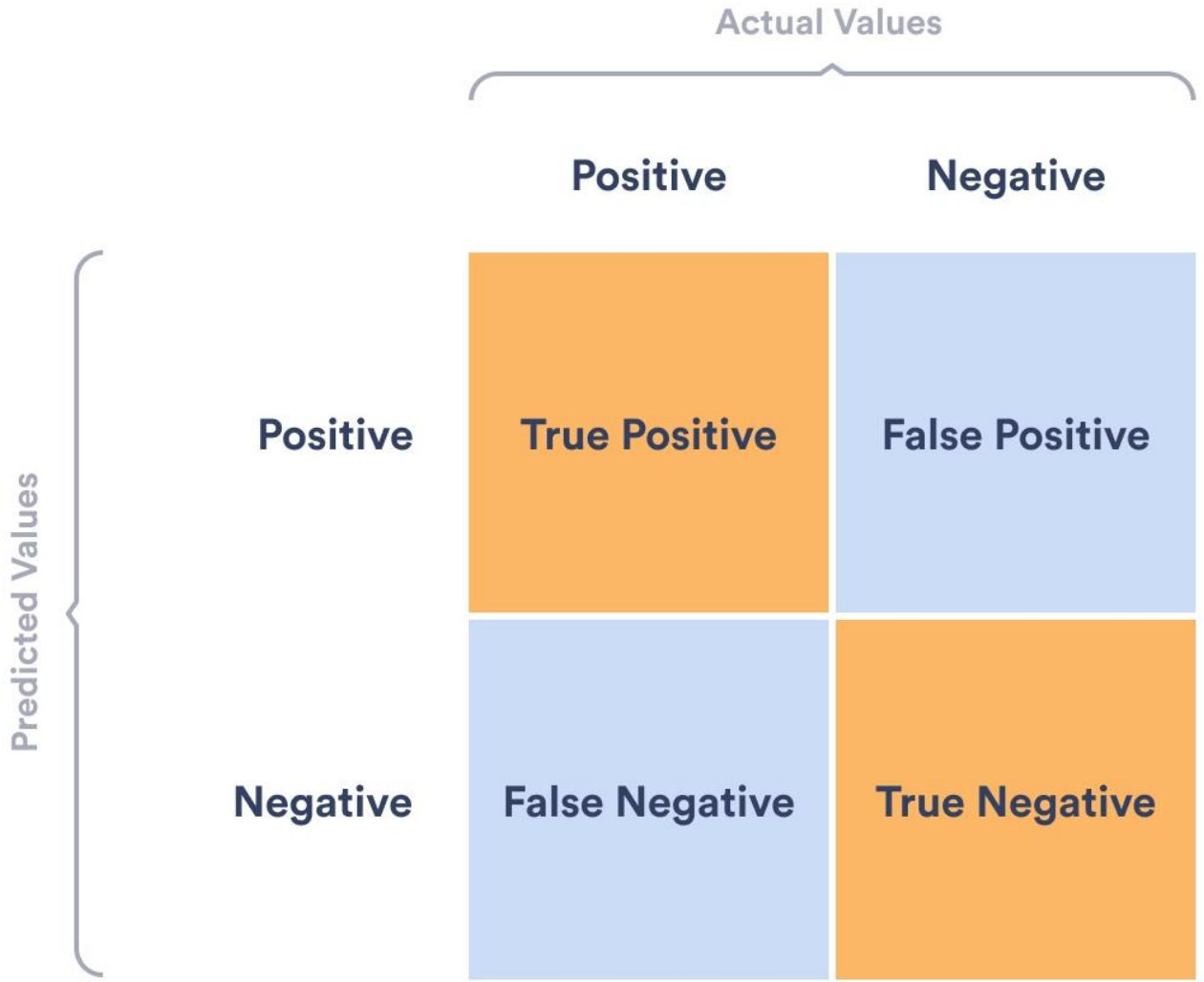
$P = v_r / (v_r + v_n)$

$R = v_r / (v_r + n_r)$

Systém 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?

confusion matrix

v oblasti IR
nás většinou
nezajímá TN



TPR = TP / Actual positive

FNR = FN / Actual Positive

TNR = TN / Actual Negative

FPR = FP / Actual Negative

Úplnost + přesnost

- P a R jsou propojené
- dosáhnout R=1 je jednoduché
- Vytvořte vyhledávací systém, který dosáhne recall = 1

$R = v_r / (v_r + n_r)$

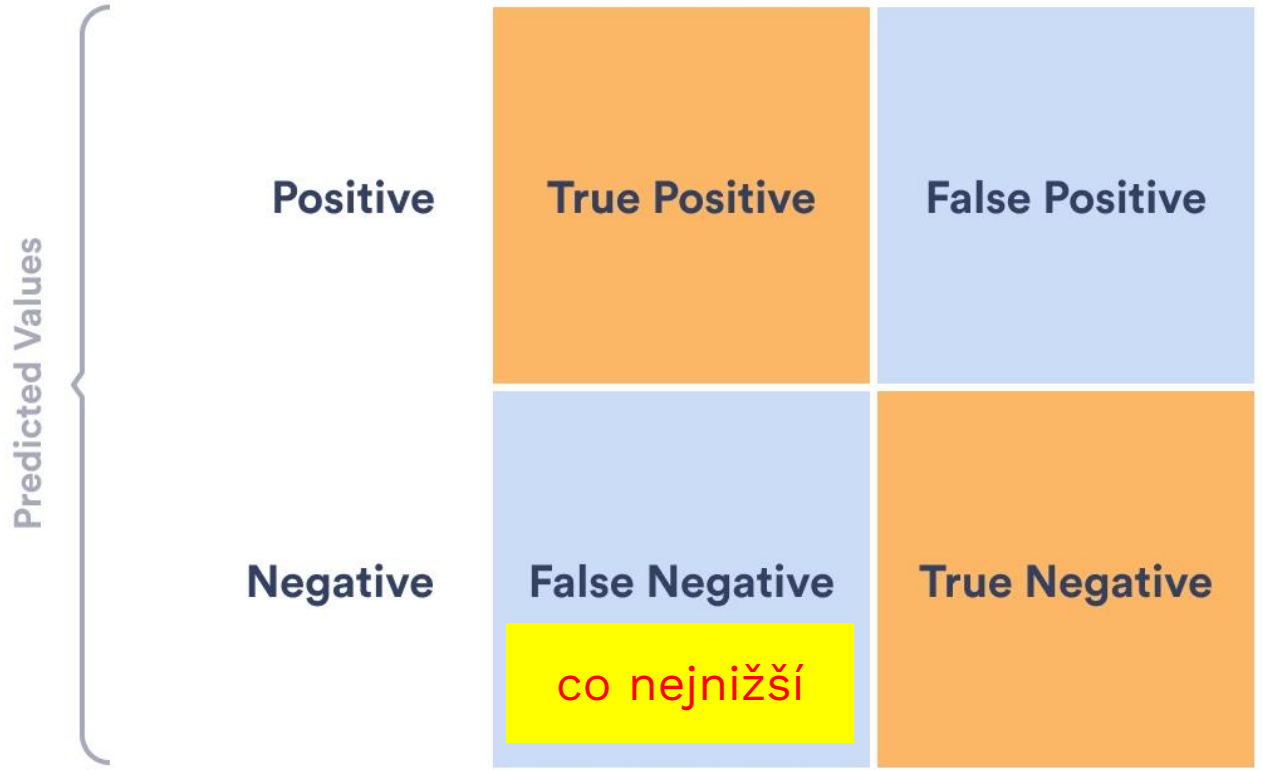
Úplnost + přesnost

- 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

detekce vtipálků

skutečný požár

vtipálek



skutečný požár

vtipálek

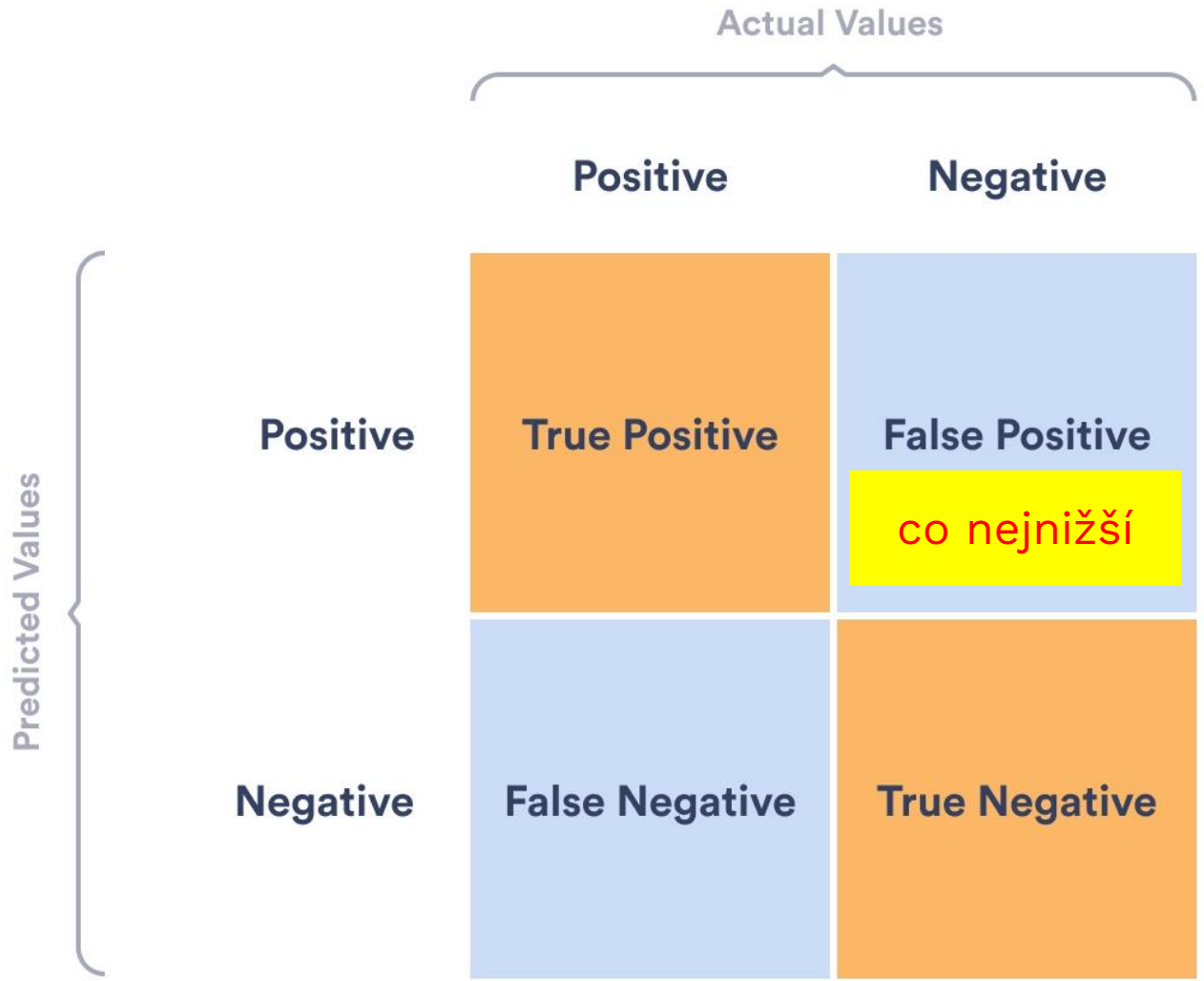
jsme OK s nízkou P

očekáváme vysoký R

detekce SPAMu

je spam

není spam



je spam

není spam

co nejnižší

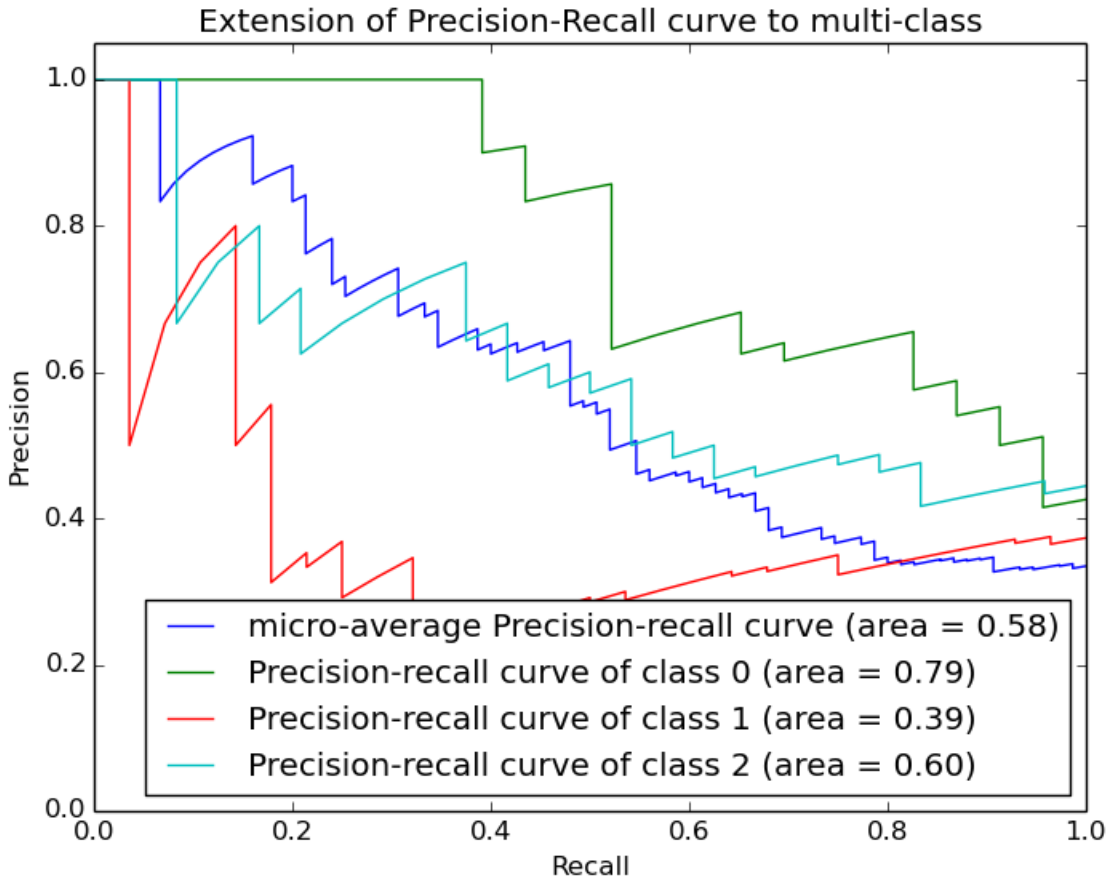
chceme vysokou P

Proč máme dvě míry?

- accuracy=$(v_r+n_n)/(v_r+v_n+n_r+n_n)$ = „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 SCHÜTZE. *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
- …

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,...

Podle čeho posuzujete relevanci výsledku na Google?

Uživatelské hodnocení

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

Seznam.sk

SEZNAM.SK

Všetko Slovensky

Google
google.sk
Vyhľadávač **google** v slovenčine.

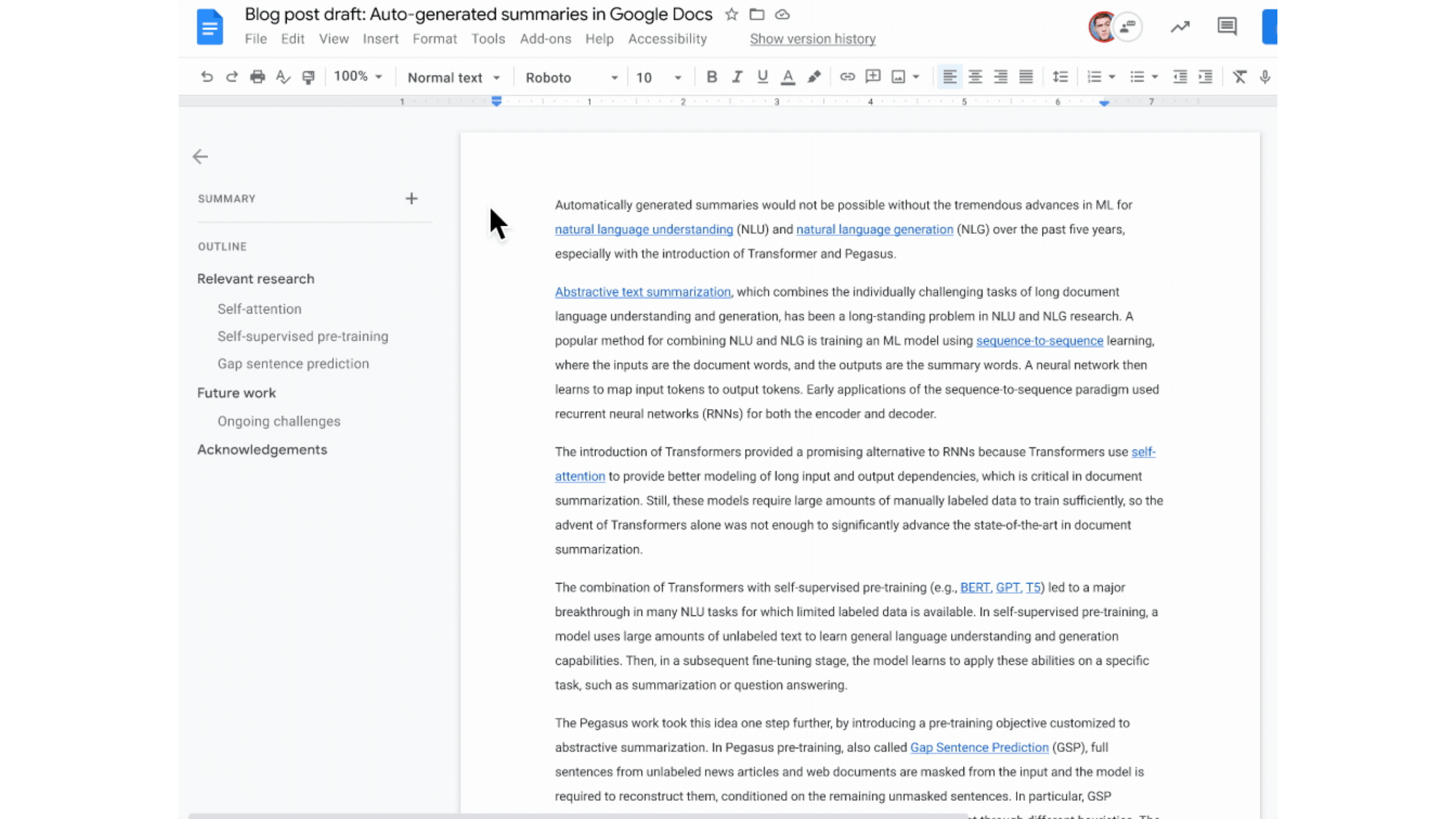
Vitajte na **Facebooku** - zaregistrujte sa, prihláste sa a zistite …
facebook.com
Facebook je nástroj umožnujúci ľuďom spájať sa s priateľmi a ostatnými, ktorí pracujú, študujú a žijú okolo nich.

YouTube
youtube.com
Zdieľajte svoje videá s priateľmi, rodinou a celým svetom.

SME.sk | Najčítanejšie správy na Slovensku
sme.sk
Rýchle a dôveryhodné správy zo Slovenska, sveta i Vášho regiónu. Prihlásenie do Post.sk.

Azet.sk - portál, kde je vždy najviac ľudí
azet.sk
Portál **Azet** zastrešuje služby Katalóg firiem a www stránok, E-mail, Pokec, Fotoalbumy,

Topky.sk - Bleskovky
topky.sk
Topky.sk - Bleskovky | Online spravodajstvo Politika, ekonomika, šport, novinky o


Blog post draft: Auto-generated summaries in Google Docs
File Edit View Insert Format Tools Add-ons Help Accessibility
Show version history
100% Normal text Roboto 10
SUMMARY
OUTLINE
Relevant research
Self-attention
Self-supervised pre-training
Gap sentence prediction
Future work
Ongoing challenges
Acknowledgements
Automatically generated summaries would not be possible without the tremendous advances in ML for natural language understanding (NLU) and natural language generation (NLG) over the past five years, especially with the introduction of Transformer and Pegasus.
Abstractive text summarization, which combines the individually challenging tasks of long document language understanding and generation, has been a long-standing problem in NLU and NLG research. A popular method for combining NLU and NLG is training an ML model using sequence-to-sequence learning, where the inputs are the document words, and the outputs are the summary words. A neural network then learns to map input tokens to output tokens. Early applications of the sequence-to-sequence paradigm used recurrent neural networks (RNNs) for both the encoder and decoder.
The introduction of Transformers provided a promising alternative to RNNs because Transformers use self-attention to provide better modeling of long input and output dependencies, which is critical in document summarization. Still, these models require large amounts of manually labeled data to train sufficiently, so the advent of Transformers alone was not enough to significantly advance the state-of-the-art in document summarization.
The combination of Transformers with self-supervised pre-training (e.g., BERT, GPT, T5) led to a major breakthrough in many NLU tasks for which limited labeled data is available. In self-supervised pre-training, a model uses large amounts of unlabeled text to learn general language understanding and generation capabilities. Then, in a subsequent fine-tuning stage, the model learns to apply these abilities on a specific task, such as summarization or question answering.
The Pegasus work took this idea one step further, by introducing a pre-training objective customized to abstractive summarization. In Pegasus pre-training, also called Gap Sentence Prediction (GSP), full sentences from unlabeled news articles and web documents are masked from the input and the model is required to reconstruct them, conditioned on the remaining unmasked sentences. In particular, GSP

tl;dr: this AI sums up research p

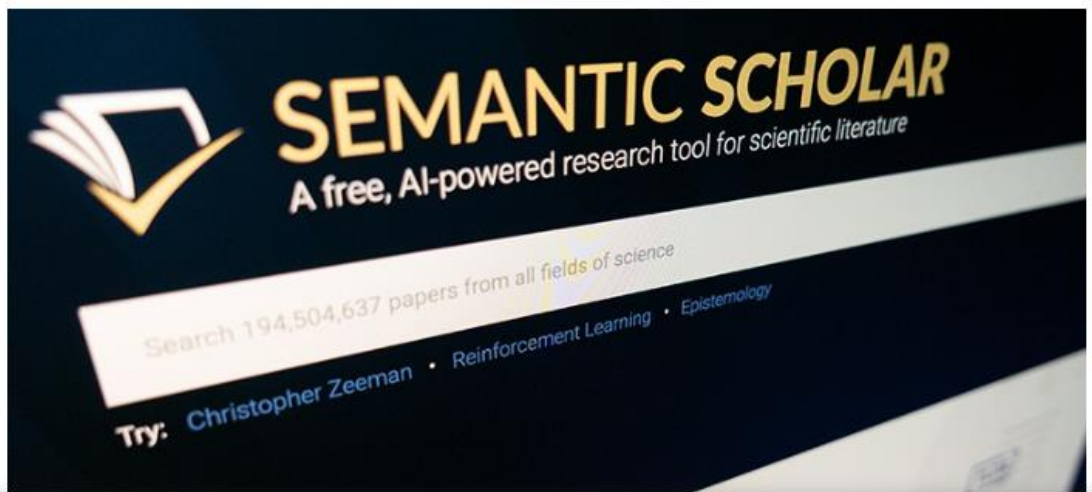
https://www.nature.com/articles/d41586-020-03277-2

NEWS | 23 November 2020

tl;dr: this AI sums up research papers in a sentence

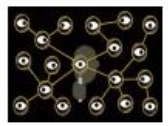
Search engine's tool for summarizing studies promises easier skim-reading.

Jeffrey M. Perkel & Richard Van Noorden



Related Articles

How AI technology can tame the scientific literature



AI science search engines expand their reach

Coronavirus in context: Scite.ai tracks positive and negative citations for COVID-19 literature

Science search engine links papers to grants and patents

Artificial-intelligence tools aim to tame the coronavirus literature

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

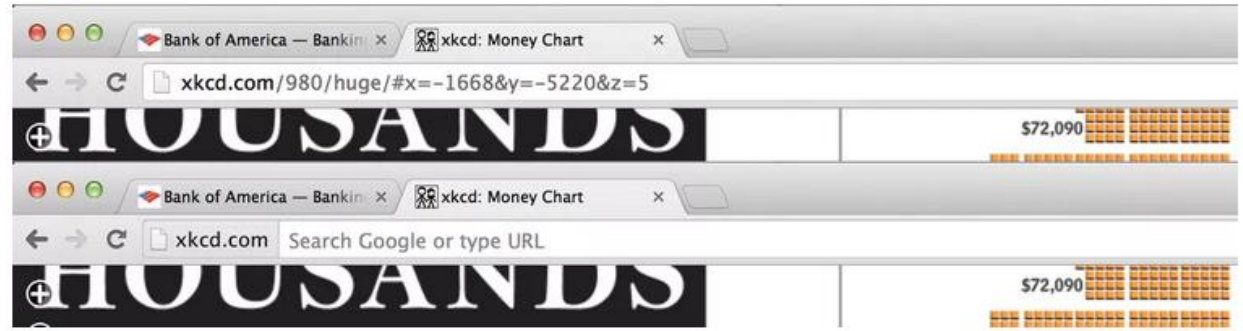
Mortgage calculator | How much mortgage can I afford ...

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 vodítko pro výběr knihovních 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

https://hapgood.us/2018/09/14/a-short-history-of-craap/

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

	Dimension	# of times	Definitions *1[Wang & Strong; 1996]
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

http://inform.nu/Articles/Vol8/v8p159-172Knig.pdf

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?

Adapted from the CSU Chico CRAAP Test for the UC San Diego Library

RADCAB

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

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

Specifická hlediska

- stav dokumentu: *preprinty*
- změny v podobách vědecké publikace
- cOAlition S a nové plány
- *preprint first* a dostupnost recenzních posudků
- přítomnost dat k ověření

Kontrolujte kvalitu *d*
ve svých rešerších!

Uklidňovací vyhledávačka

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