

Zadání anotační úlohy

pro posluchače kurzu JKB003

Digitální komunikace a práce s informacemi

FSV UK - 30. listopadu 2021

Barbora Hladká, Jiří Mírovský, Matyáš Kopp zdroj@ufal.mff.cuni.cz

Ústav formální a aplikované lingvistiky MFF UK

Anotační úloha je organizována v rámci projektu [TL05000057](#)

Signál a šum v éře Žurnalistiky 5.0 - komparativní perspektiva novinářských žánrů automatizovaných obsahů.

iROZHLAS



[DOMOV](#)
[SVĚT](#)
[EKONOMIKA](#)
[SPORT](#)
[KULTURA](#)
[VĚDA](#)
[KOMENTÁŘE](#)
[ŽIVOTNÍ STYL](#)
[VOLBY](#)
[POČASÍ](#)
[VINOHRADSKÁ 12](#)

[VĚDA](#)
[VESMÍR](#)
[PŘÍRODA](#)
[TECHNOLOGIE](#)
[HISTORIE](#)
[TÝDEN VODY](#)

Kde se nacházíte: [iROZHLAS.cz](#) / [Věda a technologie](#) / [Příroda](#) | Související témata: [Aljaška](#) [Severní ledový oceán](#) [oteplování](#) [vysoké teploty](#) [zima](#) [USA](#)

Nejrychleji oteplující se stát USA. Teploty na Aljašce se pohybují až 15 stupňů nad normálem

Mimořádně teplý konec zimy s teplotami i o 15 stupňů Celsia nad sezónními průměry zažívá letos nejsevernější americký stát - Aljaška. O víkendu by mohla vlna teplého počasí ještě vygradovat, napsala agentura AFP. Na severu státu se očekávají teploty převyšující obvyklé hodnoty i o 20 stupňů.

Juneau (USA) 8:35 30. března 2019



ZPRÁVY, KTERÉ JSTE NEČETLI



Prezident se nově spolehne na péči soukromé firmy. Postará se o něj tým zdravotních sester z Penty



Michalík nebude ministrem. Nechce, aby

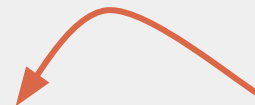
Zdroj: [url](#)

Anotační úloha

V článcích ze serveru [iRozhlas](#)

- označit a klasifikovat citační **zdroje**
- označit citační **fráze**
- spojit **zdroje s frázemi**

O víkendu by mohla vlna teplého počasí ještě vygradovat, **napsala** **agentura AFP**.
oficiální, nepolitický



Motivace

- důvěryhodnost informací
 - viz anotační úloha Verifree
- automatická analýza zdrojů informací
 - dílčí cíl:
 - detekovat zdroje v člancích automaticky
 - anotace = příklady pro strojové učení

„Trapné a ponižující.“ My a Rusko? Šándor: Změna nutná, zůstaneme sami

03.06.2021 14:47

„Osobně jsem přesvědčen, že hlavním ekonomickým zájmem Ruska je prodávat svoje komodity na Západ. Rusko dobře ví, že nemůže zároveň prodávat ropu a hlavně zemní plyn do Německa a západní Evropy a válčit v Pobaltí nebo něco takového,“ objasnil generál v záloze Andor Šándor. „Měli bychom pečlivě sledovat, jaké skutečné zájmy mají jednotlivé, především velké členské země EU a NATO ve vztahu k Rusku, tak abychom se nedostali mimo náš národní zájem. A ten je s každým, kdo o ně usiluje, mít dobré vztahy. Zajistit si ekonomickou, obchodní a jinou spolupráci, což v konečném důsledku povede k bezpečnosti České republiky. Mám pocit, že se tak úplně neděje,“ domníval se.

Zdroj: [url](#)

Citační fráze

je slovní vyjádření odkazu ke zdroji citace. Může být jednoslovné i víceslovné.

Nejrychleji oteplující se stát USA. Teploty na Aljašce se pohybují až 15 stupňů nad normálem

Mimořádně teplý konec zimy s teplotami i o 15 stupňů Celsia nad sezónními průměry zažívá letos nejsevernější americký stát - Aljaška. O víkendu by mohla vlna teplého počasí ještě vygradovat, [napsala agentura AFP](#). Na severu státu se očekávají teploty převyšující obvyklé hodnoty i o 20 stupňů. „Letošní únor i březen byly mimořádně teplé. Spousta míst má našlápnuto k překonání teplotních rekordů pro březen,“ [řekl AFP klimatolog specializující se na Aljašku Rick Thoman](#). Ve městech na severu státu jako Wainwright, Nuiqsut, Kaktovi a Barrow se o víkendu očekávají teploty o 14 až 22 stupňů Celsia nad normálem. „V Barrow jsme tento měsíc měli celkem pět rekordně teplých dní,“ [dodává Thoman](#) ... [Podle Thomana](#) panují tento měsíc podmínky, které jsou obvyklé v dubnu nebo květnu. [Deník The Washington Post](#) [dodává](#), že na některých ...

Citační zdroj

je slovní spojení označující zdroj. Může být jednoslovné i víceslovné.

Nejrychleji oteplující se stát USA. Teploty na Aljašce se pohybují až 15 stupňů nad normálem

Mimořádně teplý konec zimy s teplotami i o 15 stupňů Celsia nad sezónními průměry zažívá letos nejsevernější americký stát - Aljaška. O víkendu by mohla vlna teplého počasí ještě vygradovat, [napsala agentura AFP](#). Na severu státu se očekávají teploty převyšující obvyklé hodnoty i o 20 stupňů. „Letošní únor i březen byly mimořádně teplé. Spousta míst má našlápnuto k překonání teplotních rekordů pro březen,“ řekl AFP [klimatolog specializující se na Aljašku Rick Thoman](#). Ve městech na severu státu jako Wainwright, Nuiqsut, Kaktovi a Barrow se o víkendu očekávají teploty o 14 až 22 stupňů Celsia nad normálem. „V Barrow jsme tento měsíc měli celkem pět rekordně teplých dní,“ [dodává Thoman](#)... Podle [Thomana](#) panují tento měsíc podmínky, které jsou obvyklé v dubnu nebo květnu. [Deník The Washington Post](#) [dodává](#), že na některých ...

Klasifikace zdrojů v anotacích

Kredit: Václav Moravec

- Nepojmenované
 - Anonymní
 - Anonymní částečně
- Pojmenované
 - Oficiální - institucionální příslušnost
 - politický
 - nepolitický
 - Neoficiální - „obyčejní lidé“

Nabízí se klasifikovat zdroje i např. dle pohlaví

Ale to umíme automatickým tvaroslovným rozbořem

Jeho gól byl klíčový i **podle** trenéra hostů Petera Lavioletta.

Z fotbalových fanoušků **podle** Georgije projevili zatím největší zájem ...

České ministerstvo zahraničí nemá zatím **podle** mluvčí Michaely Lagronové ...

... většina z nich ale **podle** Dvořákové již byla proti onemocnění ...

Příklady anonymních zdrojů

Podle informací od důvěryhodných zdrojů se policisté ve čtvrtek zajímali ...

... píše CNN s odkazem na nejmenované zdroje ...

Sunday Times s odvoláním na anonymní zdroj v neděli napsal, že ...

U budovy FSB podle neoficiálních anonymních zdrojů v pátek ráno nejprve ...

Příklady částečně anonymních zdrojů

... **potvrdily** serveru iROZHLAS.cz **dva důvěryhodné zdroje blízké vyšetřování** ...

Nejmenovaný zdroj obeznámený se situací sdělil Reuters

Zdroj blízký premiérovi Babišovi Právu řekl, ...

... klidu, **popsal** mimo záznam **zdroj blízký jednáním**.

Der Tagesspiegel **citoval zdroj blízký německé tajné službě**, ...

Anotace

1. Ve webovém nástroji **Brat** (velmi přátelský)
2. Články ze serveru iRozhlas ze sekcí
ekonomika, komentáře, kultura, sport, věda a technologie, zprávy z domova,
zprávy ze světa, životní styl a společnost
3. Pro každého studenta články ze všech sekcí o celkovém objemu **cca 3 400 slov**

Anotace :: inicializační e-mail

- na adresy studentů uvedené v SIS
- odesílání 30.11.2021 od 10:00

od: zdroj@ufal.mff.cuni.cz

předmět: JKB003 - anotační instrukce pro *jméno studentky/studenta*

Milá studentko/milý studente,

posíláme vám přístupové údaje k anotační úloze řešené v rámci kurzu Digitální komunikace a práce s informacemi (JKB003).
Vaše články k anotaci najdete na adrese <https://quest.ms.mff.cuni.cz/brat/iRozhlas/index.xhtml#/TEST2/>

Vaše přístupové údaje jsou: *studentské číslo/heslo*

V případě nejasností konzultujte prosím webovou stránku <https://ufal.mff.cuni.cz/anotace-citacnich-frazi-v-datech-irozhlas>,
případně se můžete obrátit přímo na nás prostřednictvím e-mailu zdroj@ufal.mff.cuni.cz.

Anotační období končí 7.1.2022. Přejeme vám příjemně strávený čas při anotování.

Za anotační tým,

Jiří Mírovský

**Pokud neobdržíte e-mail
do půlnoci 30.11.,
pište prosím
na zdroj@ufal.mff.cuni.cz.**

Brat :: přihlášení v prohlížeči

quest.ms.mff.cuni.cz/brat/iRozhlas/index.xhtml#/TEST2/

link z e-mailu

Tutorial 1

Welcome to the brat annotation tool!

Below is a mini-tutorial for basic usage. For detailed instructions, please see the [brat user manual](#).

Selecting a document

After closing this tutorial, you will see the collection browser, which allows you to access the different text collections and individual documents in those collections on your brat installation. Simply double-click on a collection name to show its contents or on a document name to open it.

You can later return to the collection browser by pressing the TAB key or by clicking on "Collection" in the menu.

Visualization

When a document is selected, the main window area shows a visualization of the text and annotations of that document. Placing the mouse cursor over an annotation shows further information about that annotation.

Menu

Placing the mouse cursor over the blue bar on top of the window opens the tool menu. This provides access to the following features:

- **Collection:** collection browser
- **Data:** export annotation data
- **Search:** search current document or collection

OK

Open 2

Collection

/TEST2/

Collection Information

Document

Document

Document	Modified	Entities	Rel
../			
doc-8323011-Stredni_trida	12 minutes ago	39	
doc-8514812-Masivni_kseft	9 minutes ago	27	

OK

Cancel

články k anotování, vyberte libovolný z nich

Brat :: přihlášení v prohlížeči

1

1 Střední třída v chudých regionech se vnímá hůř než chudí v bohatším prostředí, varuje Daniel Prokop

2

1 Střední třída v chudých regionech se vnímá hůř než chudí v bohatším prostředí, varuje Daniel Prokop

3 Data ukazují, že počet lidí, kterým hrozí příjmová chudoba, se za loňský rok zvýšil o víc než 60 tisíc.

Brat :: přihlášení

Login
×

Username

Username viz e-mail

Password

Password viz e-mail

OK

Cancel

Brat :: články k anotování

← → /TEST2/doc-8323011-Stredni_trida brat
Collection Data Search Options Logout Hladka Help



Open

Collection

Collection Information

Document

Document	Modified	Entities	Rela
../			
doc-8323011-Stredni_trida	21 minutes ago	39	
doc-8514812-Masivni_kseft	18 minutes ago	27	

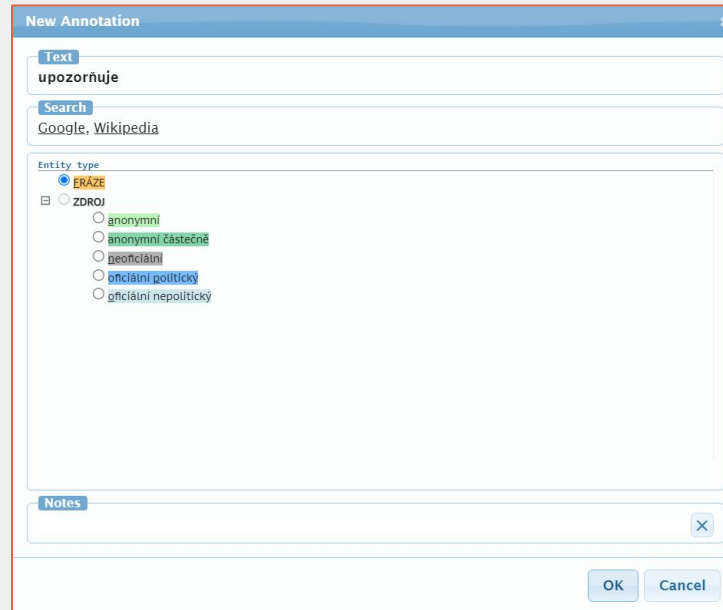
OK

Cancel

články k anotování

Brat :: označení fráze

1. označit frází myší
2. stisknout klávesu f, nebo potvrdit ok



New Annotation

Text
upozorňuje

Search
Google, Wikipedia

Entity_type

- FRÁZE
- zdroj
 - anonymní
 - anonymní částečně
 - neoficiální
 - oficiální politický
 - oficiální nepolitický

Notes

OK Cancel

Brat :: označení a klasifikace zdroje

1. označit zdroj myší

Označujte co nejdelší jmennou frázi, která může obsahovat podstatná jména, přídavná jména, zájmena, číslovky, spojky, předložky

klimatolog specializující se na Aljašku Rick Thoman

nejmenovaný zdroj obeznámený se situací

I zájmeno může označovat zdroj

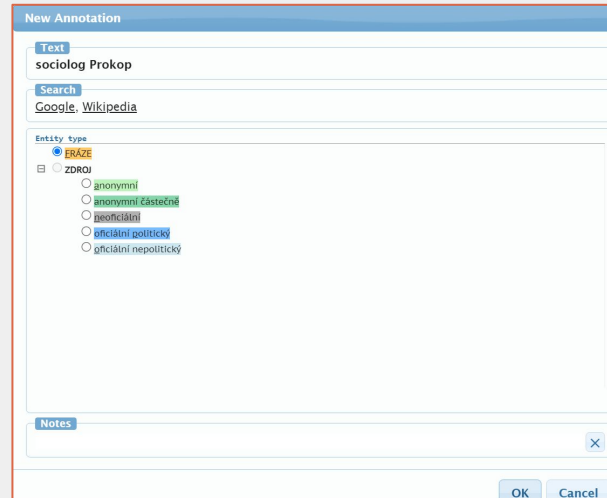
Zranitelných domácností je podle něj asi čtvrtina, v chudších krajích až polovina.

Brat :: označení a klasifikace zdroje

2. vybrat třídu zdroje myší, nebo klávesovou zkratkou

a anonymní, c anonymní částečně, n neoficiální,

p oficiální politický, o oficiální nepolitický



Brat :: spojení zdroje a fráze

1. přetáhnout zdroj přes frázi (opačně to nepůjde)

1 Střední třída v chudých regionech se vnímá hůř než chudí v bohatším prostředí, varuje Daniel Prokop

1 Střední třída v chudých regionech se vnímá hůř než chudí v bohatším prostředí, varuje Daniel Prokop

← → /TEST2/doc-8323011-Stredni_trida

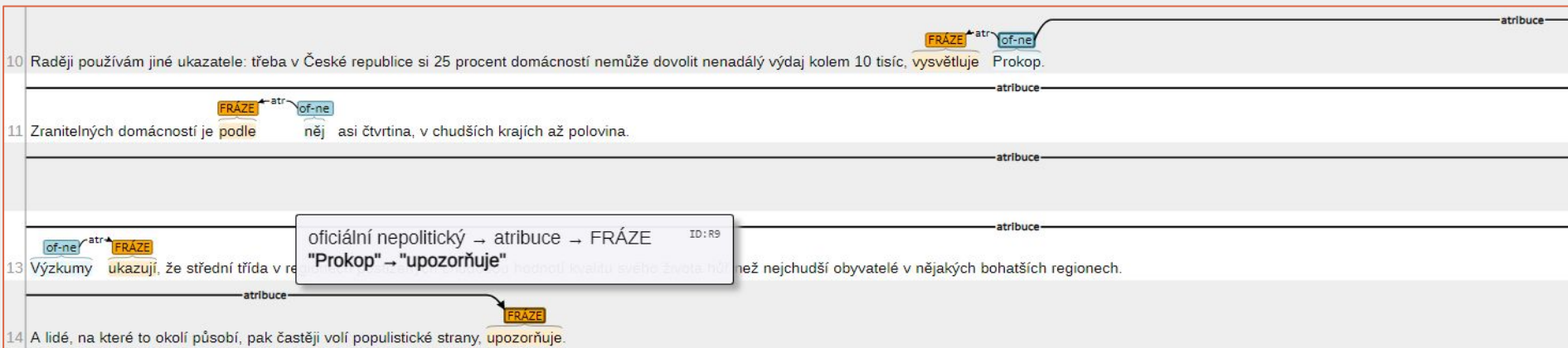
oficiální nepolitický → atribuce → FRÁZE ID:R1
 "Daniel Prokop" → "varuje"

1 Střední třída v chudých regionech se vnímá hůř než chudí v bohatším prostředí, varuje Daniel Prokop

Pokud není zdroj ve větě explicitně uveden

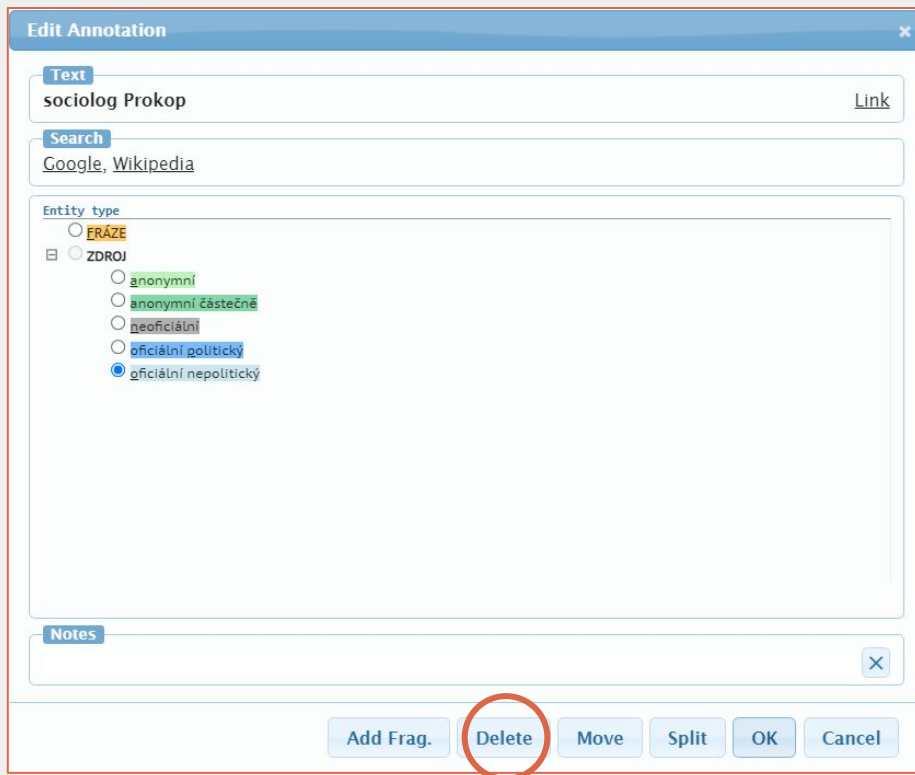
Sociolog a zakladatel platformy PAQ research Daniel Prokop nicméně upozorňuje, že příjmová chudoba není nejlepší ukazatel stavu společnosti. Český chudý je totiž výrazně chudší než německý chudý, vysvětluje.

1. s frází spojit konkrétní označení zdroje z nejbližší předcházející věty



Brat :: mazání anotací

1. dvojklik na anotaci
2. delete



Svět není černobílý, tedy ani anotace nejsou

Některé citační fráze jsou zřejmé a vymysleli bychom je bez anotace stovek článků, ale některé objevíme až po přečtení článků.

"Přestanou nám chodit hosté", **bojí se číšníci.**

Harmonogram anotace

- 30.11.2021 – odeslání inicializačních e-mailů
- 07.01.2022 – ukončení anotace

Helpline

- Pište prosím na zdroj@ufal.mff.cuni.cz
- Sledujte <https://ufal.mff.cuni.cz/anotace-citacnich-frazi-v-datech-irozhlas>
- V dotazech na citace v konkrétních větách nemusíte přepisovat celé věty. Stačí do e-mailu vložit jejich url, které získáte dvojklikem na číslo věty

quest.ms.mff.cuni.cz/brat/iRozhlas/index.xhtml#/TEST2/doc-8323011-Stredni_trida?focus=sent~11

		atribuce
11	Zranitelných domácností je podle FRÁZE ^{atr} of-ne něj asi čtvrtina, v chudších krajích až polovina.	
		atribuce

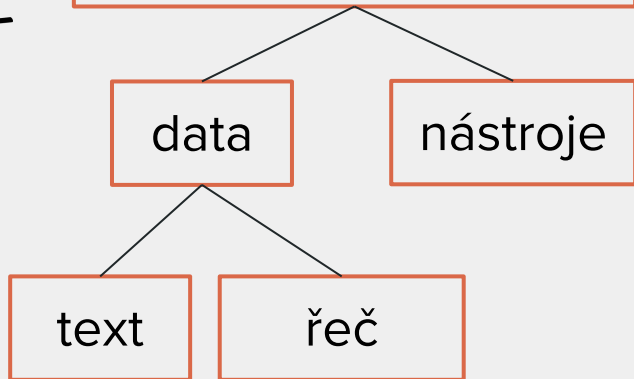
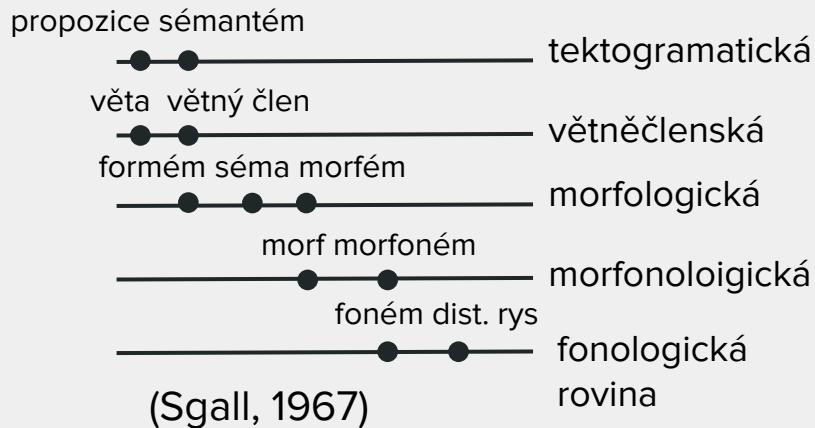
Ústav formální a aplikované lingvistiky MFF UK

<http://ufal.mff.cuni.cz>

formální lingvistika

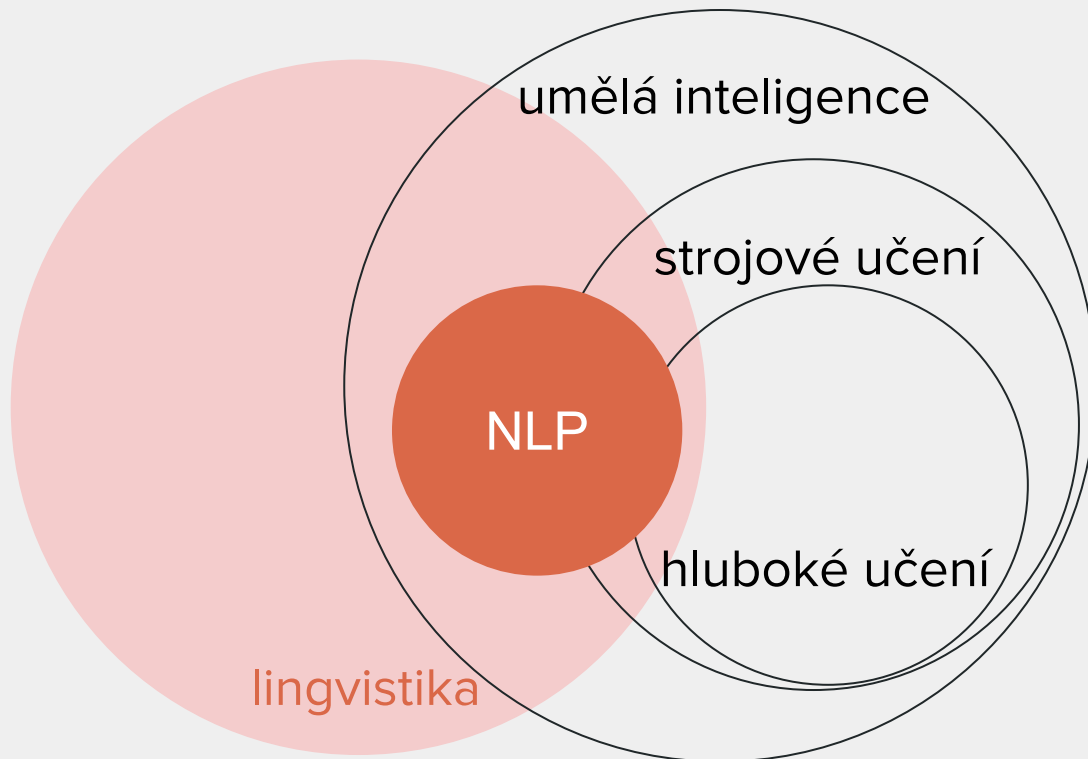


aplikovaná lingvistika*



* počítačová lingvistika nebo komputační lingvistika

Počítačové zpracování přirozeného jazyka (NLP)



Co si pod tím představit?

- překlad z jednoho jazyka do druhého
- titulkování řeči
- sumarizace textu
- extrakce informací z textu
- analýza sentimentu
- ...