

Zadání anotační úlohy pro posluchače kurzu JKB070 Etika pro novináře

FSV UK - 30. března 2022

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](#)

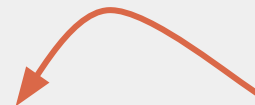
Signál a šum v éře Žurnalistiky 5.0 - komparativní perspektiva novinářských žánrů automatizovaných obsahů.

Anotační úloha

V novinových člancích

- 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, která vás teprve čeká
- 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“

Příklady anonymních zdrojů

... 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**, ...

Anotační úloha

- navazuje na úlohu [Anotace citačních frází a zdrojů 1](#)
 - anotace v člancích ze serveru iRozhlas, vznikla kolekce [SIR 1.0](#)
- anotační instrukce a editor zůstávají
- 30.3.2022 – odeslání inicializačních e-mailů
- 2 anotační etapy
 - 31.3.-7.4.2022: 4 články z iRozhlas/student
 - 13.4.-20.4.2022: iRozhlas + další zdroje
- 12.4.2022 – odeslání informačních e-mailů o druhé etapě

Anotace :: inicializační e-mail

na adresy studentů uvedené v SIS, odesílání 30.3.2022

od: zdroj@ufal.mff.cuni.cz

předmět: JKB070 - anotační instrukce pro *jméno studentky/studenta*

Milá studentko/milý studente,

anotační úloha bude probíhat ve dvou etapách: první 31.3.-7.4.2022 a druhá 13.4.-20.4.2022. Po první etapě vám poskytneme zpětnou vazbu k vašim anotačním výsledkům a odkazy na články pro druhou etapu.

Zde jsou vaše přístupové údaje do nástroje Brat: články k anotaci v první etapě najdete na adrese <https://quest.ms.mff.cuni.cz/brat/iRozhlas2/index.xhtml#/TEST1/>. 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/irozhlas2>, případně se můžete obrátit přímo na nás prostřednictvím e-mailu zdroj@ufal.mff.cuni.cz.

Přejeme vám příjemně strávený čas při anotování.

Za anotační tým Jiří Mírovský

Pokud neobdržíte e-mail do půlnoci 30.3.,
pište prosím
na zdroj@ufal.mff.cuni.cz.

Anotační editor Brat :: přihlášení v prohlížeči

<https://quest.ms.mff.cuni.cz/brat/iRozhlas2/index.xhtml#/TEST1/> ← link z e-mailu

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

2

Open

Collection: /TEST1/

Collection Information:

Document: ../

Document	Modified	Entities	Relations	Events
../				
doc-7982716-Cernobyl	1 day ago	0	0	
pokusny_dokument	1 day ago	0	0	

OK Cancel

články
k anotování,
vyberte
libovolný
z nich

Brat :: přihlášení v prohlížeči

1

← → /TEST2/doc-8323011-Stredni_trida brat

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

FRÁZE ← str → of-ne

2

← → /TEST2/doc-8323011-Stredni_trida brat

Collection Data Search Options Login

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

3 Data of-ne ← str → FRÁZE 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

/TEST1/pokusny_dokument brat
Collection Data Search Options Logout Hladka
Help



Open ✕

Collection
/TEST1/

Collection Information

Document
../

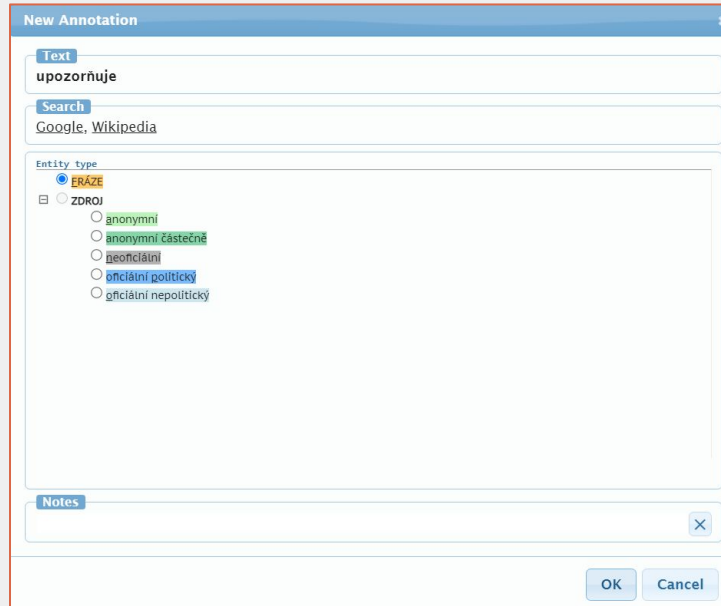
Document	Modified	Entities	Relations	Events
../				
doc-7982716-Cernobyl	1 day ago	0	0	
pokusny_dokument	1 day ago	0	0	

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

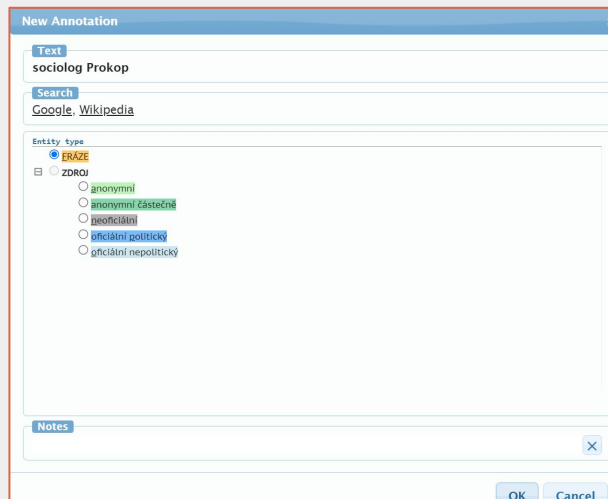
Zranitelných domácností je podle něj asi čtvrtina, v chudších krajích až polovina.

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

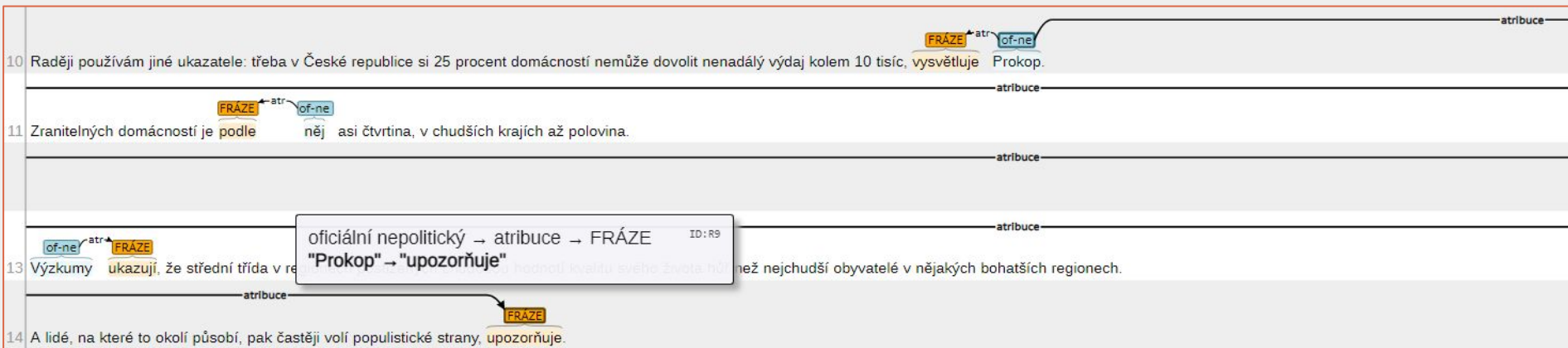
oficiální nepolitický → atribuce → FRÁZE ID:R1
 "Daniel Prokop" → "varuje"

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

The screenshot shows the 'Edit Annotation' window in the Brat interface. It contains the following elements:

- Text:** sociolog Prokop (with a 'Link' button on the right)
- Search:** Google, Wikipedia
- Entity type:**
 - FRÁZE
 - ZDROJ
 - anonymní
 - anonymní částečně
 - neoficiální
 - oficiální politický
 - oficiální nepolitický
- Notes:** (empty text area with a close button)
- Buttons:** Add Frag., Delete (circled in red), Move, Split, OK, Cancel

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.**

Helpline

- citační fráze anotované v první anotační úloze

https://ufal.mff.cuni.cz/~hladka/2021/docs/phrases_2021_12_21_sorted.txt

- zaměřte se na ty s nízkou frekvencí, protože se v nich vyskytují anotace, které nejsou citačními frázemi, např. *Zvědavost ve vědě je hrozně důležitá* (bez fráze), *zopakoval svůj názor* (pouze **zopakoval** je fráze), *vysvětlil pro Radiožurnál* (pouze **vysvětlil** je fráze).

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

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

atribuce

atribuce

Ústav formální a aplikované lingvistiky MFF UK

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

formální lingvistika



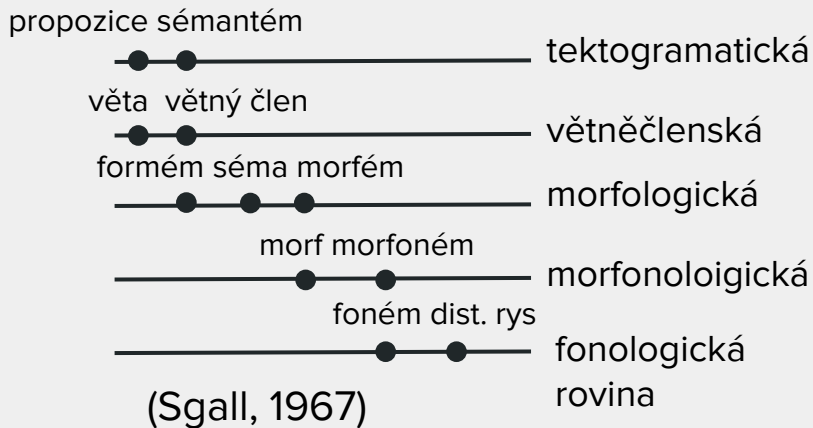
aplikovaná lingvistika*

data

nástroje

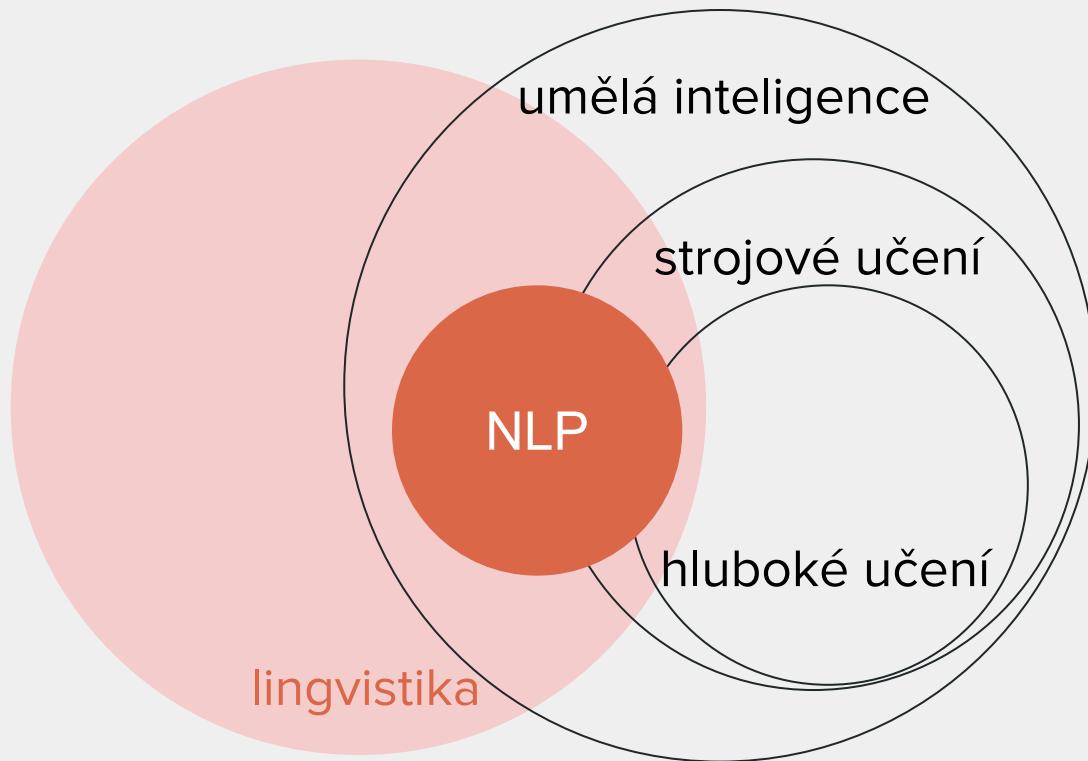
text

řeč



* 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
- ...

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 ...