

PLIN037 Sémantika a počítače

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

7. března 2019

Organizace předmětu v jarním semestru 2018

Význam a počítače

Co je význam

Co studuje sémantika?

Repräsentace znalostí

Víceznačnost

Praxe

Co se bude dít v průběhu semestru?

- přednášky
- cvičení
 - práce s algoritmy NLP
 - vytváření datových zdrojů
- prezentace → hodnocení

prezentace 9. 5. 2019

Co byste měli umět?

- pracovat s textovým korpusem
- zhruba porozumět kódu programovacího jazyka
- porozumět anglickým (nebo francouzským) odborným textům

Proč je nauka o významu obtížná

- nedokážeme formálně přesně popsat význam
- jeden a týž znak se používá pro označení více významů (víceznačnost)
- věty v přirozeném jazyce se nevyjadřují úplně (implicitní vs. explicitní informace)
- k významovým posunům dochází častěji, než by se zdálo

Sémantika

- nauka o významu **znaků**
- znak = něco, co něco zastupuje (reprezentant)
- v našem případě nauka o významu morfémů, slov, slovních spojení, vět, promluv

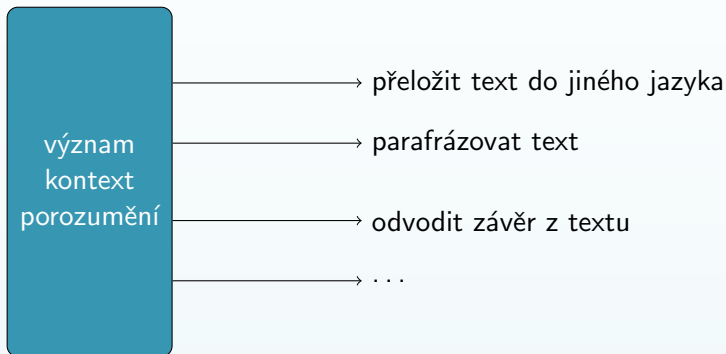
Co není význam

- reference: jednorozec? nikdo? já? vynálezce hračky jo-jo
- (vědecká) znalost: sůl je chlorid sodný
- pravdivostní hodnota: $2 + 2 = 4$

Co je význam

- klasické konceptualistické pojetí: význam je strukturovaná myšlenka (koncept) v mysli
John Locke [in [Goddard, 2011]]:
[W]ords, in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them . . .
- strukturalistické pojetí: význam se skládá ze souhrnu vazeb, které má toto slovo s ostatními
- extensionalismus: rozdíly ve významu slov lze nejlépe zkoumat ve vztahu k extenzím (referencím)
extensionalistické pojetí významu bývá prezentováno „mapami užití“
- moderní konceptualismus: méně filozofie, více praktického pojetí

Jak se projevuje porozumění?



Co studuje sémantika?

- sémantické relace: synonymie, antonymie, hyperonymie, meronymie . . .
- taxonomické a kolektivní nadřazené koncepty
- vyplývání, kontradikce, parafráze
- vágnost – slovo může být užito v různých kontextech
- víceznačnost – slovo má víc významů (a různé gramatické vlastnosti)
- prototypy
- konotace – „subtilní aspekty významu“
- kolokace, ustálená slovní spojení

Reprezentace znalostí

- jazyková znalost
- znalost „světa“

Jazykové znalosti

znaky jazyka, segmentace na věty, segmentace na slova, určení slovních druhů, syntax jazyka, seznam slov jazyka („slovník“)

Znalosti o světě

- je možné mít jazykovou analýzu bez znalosti významů jednotlivých slov
- na určité úrovni tuto znalost potřebujeme alespoň na úrovni vazby mezi slovy

Přečetl jsem ten článek o evoluci člověka za posledních deset tisíc let.

Přečetl jsem ten článek o evoluci člověka za půl hodiny.

Víceznačnost na různých úrovních

- morfémy: koncovka -ík
- slova: stát, kolej, bez, jen, jak
- slovní spojení: velký člověk
- věty: Ta je má.
Karel mluvil o sexu s Britney Spears.
Nájemní smlouva musí obsahovat označení bytu, jeho příslušenství, rozsah jejich užívání a způsob výpočtu nájemného a úhrady za plnění spojená s užíváním bytu nebo jejich výši.
- ...

Sémantika a počítače v praxi

- méně teorie, více „hrubé síly“
- s nástupem velkých korpusů získávají statistické přístupy lepší výsledky než pravidlové
- pro některé úlohy dosahuje strojové učení lepších výsledků než statistické přístupy
- s rozmachem velké výpočetní síly a nových algoritmů je pro některé úlohy výhodné použít neuronové sítě
- potřebná velikost dat záleží na jazyce a typu úlohy
- důležité je znát vstupní data
- hybridní přístup



Goddard, C. (2011).

Semantic Analysis: A Practical Introduction.

Oxford Textbooks in Linguistics. Oxford University Press.