

Syntaktická analýza přirozeného jazyka

Vojtěch Kovář

Centrum zpracování přirozeného jazyka
Fakulta Informatiky, Masarykova Univerzita
Botanická 68a, 602 00 Brno
xkovar3@fi.muni.cz

PA153 Počítačové zpracování přirozeného jazyka

Outline

- 1 Co? Proč?
- 2 Notace
- 3 Jak?
- 4 Co děláme v NLP centru

Syntaktická analýza přirozeného jazyka

■ Co?

- odhalení povrchové struktury věty
- základ pro analýzu jazyka na vyšších úrovních

■ Proč?

- jakékoli pokročilejší zpracování jazyka
- např. vztahy mezi slovy → logické konstrukce
- identifikace frází v textu
- ...

Notace syntaxe přirozeného jazyka

■ Závislostní formalismus

- strukturní vztahy kódovány závislostmi mezi slovy na vstupu
- pražský korpus závislostních stromů PDT

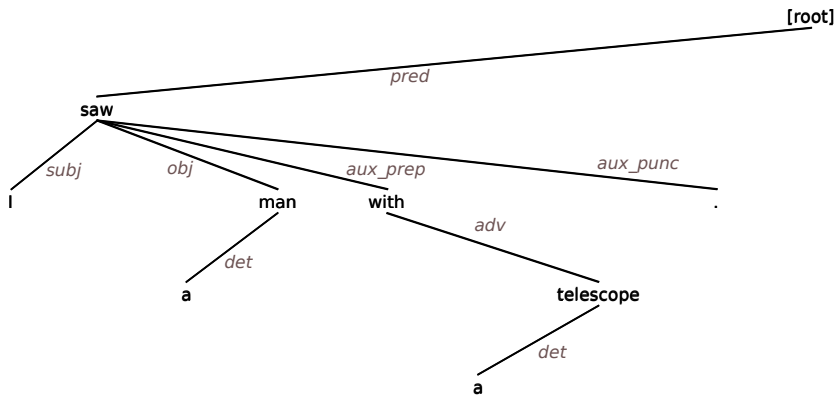
■ Složkový formalismus

- strukturní vztahy kódovány stromem odvození z gramatiky
- brněnský analyzátor synt

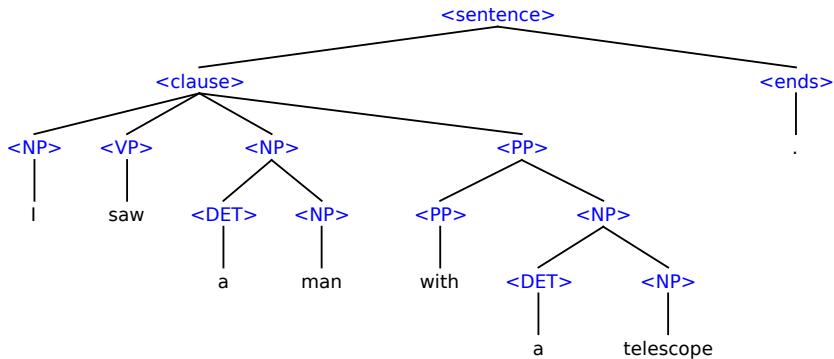
■ Parciální syntaktická analýza

- nezajímá nás kompletní strom, jen některé vztahy
- např. systém VaDis, **Word Sketches**

Příklad: závislostní strom



Příklad: složkový strom



Jak analyzovat syntax automaticky

■ Zásadní problém

- víceznačnost: někdy problém i pro lidi
- „Karel mluvil o sexu s Britney Spears.”
- problémy s vyhodnocením

■ Jak analyzovat syntax jazyka

- manuálně vytvořené gramatiky (bezkontextové, závislostní)
- statisticky naučené gramatiky
- statistické odhadování podoby stromu

Gramatické formalismy

- Bezkontextová gramatika
 - CKY
 - chart parser
- Závislostní syntax
 - aplikace algoritmu hledání nejmenší kostry na statisticky naučené vazby
- Definite clause grammar (Prolog)
- Head-driven phrase structure grammar
- Combinatory categorial grammar
- Tree adjoining grammar

Co děláme v NLP centru

■ Synt

- složkový analyzátor
- základ = bezkontextová gramatika + chart parser
- koncept metagramatiky, kontextové akce

■ SET

- hybridní analyzátor
- založen na detekci konečných vzorků v textu

■ Dis/VaDis

- založen na DCG
- použití při morfologické desambiguaci

■ IOBBER

- statistický rozpoznávač frází (chunker)

Odkazy

www.diotavelli.net/people/void/demos/cky.html
en.wikipedia.org/wiki/Definite_clause_grammar
en.wikipedia.org/wiki/Combinatory_categorical_grammar
en.wikipedia.org/wiki/Head-driven_phrase_structure_grammar
nlp.fi.muni.cz/projekty/wwwsynt
nlp.fi.muni.cz/projekty/wwwsynt/query.cgi
nlp.fi.muni.cz/trac/set
nlp.fi.muni.cz/projekty/set/wwwset.cgi/first_page
ufal.mff.cuni.cz/pdt2.0/index-cz.html