

Lingvistický software

- Morfologický analyzátor – AJKA
- Morfologická databáze – I_PAR
- Desambiguace

Základní pojmy

- Značka (tag, index) – kód přiřazený k jednotlivým tvarům slov nesoucí informaci o jejich gramatických charakteristikách
- Tagset – soubor používaných značek
- Značkování (tagování, tagging, anotace, indexování) – ruční, automatické, poloautomatické
- Desambiguace – zjednoznačnění, výběr správné značky v závislosti na postavení slova v kontextu

Využití morfologického analyzátoru

- Morfologická databáze
- Morfologické značkování korpusů, syntaktická analýza, sémantická analýza
- Využití pro jiné slovanské jazyky – slovenština
- Kontrola pravopisu, překladače, slovníky, webové prohlížeče

Formální popis morfologie

- slovo = řetězec znaků ohraničený z obou stran mezerami

- Segmentace slova

KMZ – IS – T

(kmenový základ, intersegment, koncovka)

- Koncovkové množiny, slovník kmenů a intersegmentů
- Seznam vzorů

Počet vzorů a lemmat v I_PARu

- substantiva 778 lemmat 131 738
- adjektiva 69 lemmat 170 768
- zájmena 105 lemmat 199
- slovesa 757 lemmat 42 716
- adverbia 72 lemmat 41 593
- všech vzorů 1838 celkem lemmat 389 732

Průběh morfologické analýzy

- Rozeznání neohebných slovních druhů
- Rozeznávání slova od začátku
 - záporka –ne
 - superlativní prefix –nej
- Segmentace slova odzadu
 - koncovka
 - intersegment
 - kmenový základ
 - přiřazení ke vzoru

Homografie +

Homonymie -

Desambiguace

- Ruční, automatická
- Nástroj pro desambiguaci CED
- Nelze desambiguovat

Německá firma Tebis v Hannoveru představila kompaktní zařízení pro firemní modelárny.
Technické řešení těsnění nádrží a podlah...

Myrha je přírodní pryskyřice, aloe je vonné dřevo.

V osmi letech měl za sebou účinkování v mnoha televizních show...

Dolní listy jsou obvejčité, čepel se zužuje v ouškatý řapík.

Jak lze z názvu vytušit, jde o nástroje pro zprostředkování databázových transakcí a tvorbu dotazů prostřednictvím standardu SQL.

Jak nám řekl ředitel tohoto závodu, nebyla to jejich chyba...