

Лекция 10:

Общуване с компютрите на ограничен естествен език

Същност на общуването. Процес, при който един агент (*говорещ*) се стреми да направи така, че друг агент (*слушач*) да промени своите знания и/или цели.

Етапи на общуването за говорещия:

- *намерение* – вземане на решение относно информацията, която ще бъде включена в съответното съобщение. В общия случай предполага планиране и разсъждения за знанията и целите на агентите;
- *пораждане* – превръщане на избраната информация в съобщение, т.е. низ от думи;
- *синтез* – извеждане на съобщението във вид, определен от условията на средата, в която се извършва общуването.

Етапи на общуването за слушащия:

- *възприемане* – превръщане на съобщението в низ от думи (разпознаване на символи, на говорима реч и т.н.);
- *анализ* – установяване на информационното съдържание на съобщението на ниво
 - синтаксис: определяне на функционалната структура на съобщението. На това ниво на анализ обикновено се извършва групиране на думите в йерархични структури, наречени *фрази*;
 - семантика: установяване на *значението* на отделните думи и фрази;
 - прагматика: поставяне на значението в контекста на общуването (извличане на възможните “полезни” изводи от значението в контекста на знанията на слушащия или особеностите на ситуацията);
- *усвояване* – евентуално модифициране на базата от знания или целите в зависимост от отношенията между агентите.

Многозначност

Характерно свойство на естественоезиковото общуване. Отстраняването на многозначността е един от основните проблеми на обработката на естествен език.

Нива на многозначността:

- фонетично (“Петко рита” и “пет корита”);
- синтактично (“Ям макарони със сирене” и “Ям макарони с вилица”: предложната фраза “с със сирене” пояснява съществителното “макарони”, докато предложната фраза “с вилица” пояснява глагола “ям”);
- семантично (омоними);
- прагматично (“Знаете ли колко е часът?” – буквален въпрос относно знанието, с което разполага слушащият, в контекст обикновено еквивалентен на “Колко е часът?” или “Смятам, че вече е твърде късно.”).

Същност на понятието “ограничен естествен език” (ОЕЕ)

Като правило под общуване с компютрите (с програмни системи) на естествен език се разбира общуване на т. нар. *ограничен естествен език* (псевдоестествен език; език, близък до естествения). Ограничаването на естествените езици се извършва най-често в следните два аспекта:

- ограничаване на структурата на допустимите фрази (опростяване на синтаксиса на езика);

- ограничаване на предметната област (създаване на езици за общуване в дадена тясна предметна област – така се ограничава речникът на системата и се ограничават възможните интерпретации на отделните фрази, например се избягва опасността от *омонимия*).

Тук ще разгледаме само някои аспекти на *анализа на писмен текст на ОЕЕ* (т. нар. *обработка на ОЕЕ* или Natural Language Processing, NLP).

СИСТЕМИ ЗА ОБРАБОТКА НА ОЕЕ, КОИТО НЕ ИЗПОЛЗВАТ ЗНАНИЯ ЗА СЕМАНТИКАТА НА ЕЗИКА

В тази група попадат голям брой ранни системи от тип “въпрос – отговор”, които работят на базата на механизми за съпоставяне по образец. Примери: ELIZA, DOCTOR, SIR, STUDENT и др.

СИСТЕМИ ЗА ОБРАБОТКА НА ОЕЕ, КОИТО ИЗПОЛЗВАТ ЗНАНИЯ ЗА СЕМАНТИКАТА НА ЕЗИКА

Най-често това са системи за решаване на задачи от определена предметна област, които са снабдени със съответен естественоезиков интерфейс. При тези системи заявката на потребителя се формулира като фраза или поредица от фрази на ОЕЕ. Интерфейсният (лингвистичният) модул на системата анализира заявката и на основата на знанията на системата, които позволяват еднозначна интерпретация на фразите на ОЕЕ в рамките на конкретната предметна област, той трансформира тази заявка в съответна команда на вътрешния език на системата.

Основни етапи на обработката на фрази на ОЕЕ

- Лексичен (морфологичен) анализ;
- Синтактичен анализ;
- Семантичен анализ;
- Свързване с контекста;
- Прагматичен анализ.

ФОРМАЛИЗМИ ЗА ОПИСАНИЕ НА СИНТАКСИСА И СЕМАНТИКАТА НА ОЕЕ

- Разширени мрежи на преходите (Augmented Transition Networks, ATN);
- Граматики, определени от клаузи (Definite Clause Grammars, DCG);
- Семантични граматки (Semantic Grammars) и др.

Значението на съобщението (фразата) обикновено се представя чрез израз(и) в някаква формална система (например чрез ППФ от езика на предикатното смятане от първи ред). Това означава, че за семантичния анализ е необходимо да се знае

- какво е значението на отделните думи (какви логически изрази – формули или части от формули – им съответстват);
- как се свързват значенията на компонентите на синтактично коректните конструкции.