

ZPRACOVÁNÍ PŘIROZENÉHO JAZYKA — TÉMATICKÉ OKRUHY

stínované: k zápočtu/zkoušce za ZS, resp. k 1. otázce zkoušky za LS,
ostatní: k 2. otázce zkoušky za LS

1. Obecné základy a souvislosti [kn.: úvod]:

- a) slovník, gramatika, sémantika, pragmatika (vymezení a vzájemné souvislosti pojmů);
- b) přehled hlavních aplikačních oblastí (automatické indexování, generování tezurů, automatické referování, komunikace s databází / robotem / expertním systémem ap., strojový a počítačem podporovaný překlad, naplňování bází dat/znalostí z textů, automatizovaná korektura textů);
- c) spojitosti s jinými infromatickými obory (teorie formálních jazyků a automatů, mat. logika, umělá inteligence, datové modelování, teorie paralelních procesů ap.).

2. Základní lingvistické problémy automatického indexování textů [kn.: kap. 1]:

- a) rozpoznání slov / termínů;
- b) relevance slov / termínů (možnosti konstrukce negativního slovníku [též kn.: 2.1.2] a váhových funkcí);
- c) tvarosloví (řešení operátorem rozšíření v dotazovacím jazyce, derivátorem, lematizátorem, vzájemné výhody a nevýhody; základní specifika české morfologie);
- d) sémantické vztahy mezi termíny (možnosti použití tezauru, metody *SÉMAN*);
- e) homonymie různých typů a možnosti jejich dílčího řešení.

3. Metoda automatického indexování *MOZAIKA* [kn.: 1.4.1 a příl. A]:

- a) hlavní cíle, myšlenky a funkční prvky, které je realizují (datové struktury — stačí abstraktně! — a algoritmy, které je používají);
- b) co z výše uvedených obecných problémů tato metoda řeší, jakými svými prvky; jaká jsou její omezení.

4. Automatické referování [kn.: kap. 3]:

- a) vztah mezi automatickým indexováním a automatickým referováním (shody a rozdíly);
- b) metoda založená na měření obsahových souvislostí mezi větami.

5. Strojový překlad [kn.: kap. 4]:

- a) přístup 1. generace (metoda *slovo za slovo*, resp. *fráze za frázi*, *věta za větu*), jeho zákonitá omezení a další vývoj systémů tzv. *přímého překladu* (př.: *SYSTRAN*) [kn.: 4.2.1–3];
- b) obecně převládající schéma 2. generace (*analýza — transfer — syntéza*):
 - ba) jaké „dimenze“ a „úrovně“ *jazykového významu* je třeba analyzovat (lexikální a morfologické složky významu a jejich modifikace gramatickými neboli pomocnými slovy; syntaktické složky významu, vč. *aktuálního členění*; různé úlohy slovosledu) [kn.: 4.3.1–2];
 - bb) možné podoby datových struktur pro transfer (složkový a závislostní syntaktický popis, případně formule predikátové logiky ap. — shrnutí argumentů) [kn.: 4.3.3];
 - bc) náročnost syntaktické analýzy přirozeného jazyka (přirozené jazyky ve světle teorie formálních jazyků a automatů) [kn.: příl. B];
 - bd) nástroje prakticky používané, resp. použitelné k analýze a důvody pro jejich použití (*slovník* a *gramatika* jako doplňující se znalostní struktury a různé možnosti rozdělení

jazykových znalostí mezi ně, typy stálých a pracovních slovníků;
rámce jako datové, příp. i řídicí struktury;
datové struktury a pravidla pomáhající vyřešení *koreferencí* a *odkazů* všeho druhu;
princip *nejistoty* veškerých jazykových znalostí) [kn.: 4.4];

- c) *interlingvální přístup* (argumenty pro a proti) [kn.: 4.3.1];
- d) různá chápání pojmu „strojový překlad 3. generace“ [kn.: 4.2.4];
- e) porozumění přirozenému jazyku jako univerzální proces zacházení se znalostmi (potřeba věcných znalostí pro jednoznačné určení některých gramatických struktur; Dikův model interpretace a klasifikace znalostí používaných při interpretaci textu) [kn.: odd. 4.5 a kap. 5];
- f) možnosti *částečné automatizace* překladu (systémy s *postredakcí*, příp. *preredakcí*, *HAMT*, *MAHT*, vývojové třídy podle Boiteta — „pro pozorovatele“, „pro postredaktora“, „pro překladatele“, „pro autora“ — a smysl takového třídění) [kn.: 4.6.2];
- g) princip tzv. *překladové paměti* [kn.: 4.6.3].

6. Automatizovaná korektura textů [kn.: kap. 6]:

- a) 3 teoreticky možné úrovně cíle (zjistit chybu, navrhnout opravu, samočinně opravit);
- b) lexikální (pravopisná) korektura:
 - ba) hlavní typy chyb;
 - bb) rozdíl mezi jazyky jako angličtina a čeština;
 - bc) strategie a reálné možnosti splnění výše uvedených 3 úrovní cíle;
- c) korektura „mechanických chyb“ (překlepy v interpunkci ap.), rozšíření pravopisné korektury na víceslovná spojení, elementární stylová korektura (možnosti, příklady);
- d) gramatická korektura:
 - da) hlavní typy chyb;
 - db) rozdíl mezi jazyky jako angličtina a čeština;
 - dc) strategie a reálné možnosti splnění výše uvedených 3 úrovní cíle.

Pozn.:

„kn.“ = STROSSA, P.: *Počítačové zpracování přirozeného jazyka*, Praha, Oeconomica 2011.