

## Ročníkový projekt LS 2007/08, ZS 2008/09

### Zadání ročníkového projektu

**Téma:** Konverzní procedura vnitřních formátů pražských korpusů

**Anotace:** Cílem rp je implementace automatické procedury pro převod mezi formáty, které se používají pro vnitřní reprezentaci tzv. pražských korpusů (banky textů obohacené o jazykovědné informace). Konkrétně se jedná o formát založený na XML, tzv. PML a o formát založený na SGML, tzv. CSTS.

**Student:** Michal Vachna, *e-mail:* michal.vachna@condornet.sk

**Vedoucí:** Barbora Vidová Hladká, ÚFAL MFF UK, *e-mail:* hladka@ufal.mff.cuni.cz

---

### ŘEŠENÍ PROJEKTU

#### Časový harmonogram

##### LS 2007/2008

1. nastudovat formáty PML a CSTS,
2. rámcově se seznámit s nástrojem TrEd
3. seznámit se se současnou převodní procedurou (modul nástroje bTrEd),
4. seznámit se se s nástrojem tool\_chain,
5. navrhnout převodní proceduru z formátu PML do formátu CSTS a naopak pro [amw]-soubory,
6. sepsat podrobnou specifikaci,
7. implementovat pilotní verzi, tj. navrženou proceduru pro w-soubory,
8. testovat převodní proceduru jejím začleněním do skriptu tool\_chain,

##### ZS 2008/2009

9. implementovat navrženou převodní proceduru z formátu PML do formátu CSTS a naopak pro m-soubory a a-soubory,
10. testovat převodní proceduru jejím začleněním do skriptu tool\_chain,
11. sepsat závěrečnou dokumentaci
  - a. instalační příručka - popis instalace a spuštění programu
  - b. uživatelská příručka - popis ovládání programu
  - c. programátorská dokumentace - postup překladu programu, popis implementace (co, v čem, jak), popis netriviálních algoritmů

#### Studijní materiály

##### Formát PML

- stručný popis - <http://ufal.mff.cuni.cz/morce/cac/?chapter=3#data-format>
- podrobná dokumentace - <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch03.html#a-data-formats-pml>
- video tutoriál - <http://lectures.ms.mff.cuni.cz/video/recordshow/index/17/34>

## Formát CSTS

- stručný popis - <http://ufal.mff.cuni.cz/morce/cac/?chapter=3#data-format>
- podrobná dokumentace - <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch03.html#a-data-formats-csts>

Ukázka souborů v reprezentaci PML a CSTS – viz příložené CD-ROM

## nástroj TrEd

- domovská stránka - <http://ufal.mff.cuni.cz/~pajas/tred/index.html>
- video tutoriál <http://lectures.ms.mff.cuni.cz/video/recordshow/index/2/23>

## nástroj bTrEd

- manuálová stránka - <http://ufal.mff.cuni.cz/pdt2.0/doc/tools/tred/btred.html>
- textový tutoriál k - <http://ufal.mff.cuni.cz/pdt2.0/doc/tools/tred/bn-tutorial.html>

## nástroj tool\_chain

- stručný popis - <http://ufal.mff.cuni.cz/morce/cac/?chapter=3#nastroje-zprac>
- dále viz příložené CD-ROM

## Obsah podrobné specifikace

1. Detailní popis problematiky, podle něhož by jiný programátor napsal "tentýž" program.
2. Analýza existujícího modulu, jeho popis – zdůraznit, čím se bude nově vzniklý nástroj lišit od již existujícího (v čem bude lepší, co bude mít jiného, apod.).
3. Návrh struktury programu (moduly, knihovny, vzájemná provázanost).
4. OS, jazyk, vývojové prostředí

## Obsah pilotní verze

Převodní procedura PML<->CSTS pro [mw]-soubory.

## Obsah závěrečné dokumentace

1. Instalační příručka = popis instalace a spuštění programu.
2. Uživatelská příručka = popis ovládání programu.
3. Programátorská dokumentace = postup překladu programu, popis implementace (co, v čem, jak), popis netriviálních algoritmů.

## ORGANIZAČNÍ ZÁLEŽITOSTI

### Termíny

30. června 2008 – odevzdání pilotní verze

Projektová wiki stránka – <https://wiki.ufal.ms.mff.cuni.cz/user:hladka:michal-vachna>