

ÚFALÍ mapa 2006

kdo co dělá / chystá se / chtěl by dělat
v tomto a následujícím roce
(příp. i v letech dalších)

Kvilda 22.-29.1.2006

Jan Hajič

- Vedení ÚFAL -> člen kolegia sekce informatiky
- Výuka (ved. učitel I-3, Mgr. studium)
 - PFL067/8 Statistika
 - Vedení doktorandů
- Projekty / vedení
 - VZ ÚFAL (lexikální věci, sémantika), CKL (mluvený jazyk), IS „Sémantika“, Malach, GAČR „angličtina“ (<->PIRE), GAUK 3x, Kontakt (PropBank->EngVallex), Kontakt Malach, EU Euromatrix (MT), EU Companions (?!)
- Mimo fakultu:
 - posudky & hodnocení (granty GAČR, GAUK, projekty NSF, oponentury projektů), komise pro SZZ, RZ
- Údržba morfologie

GAČR pro PIRE

- PIRE
 - Partnership for International Research and Education:
 - „Collaborative Investigation of Meaning Representation in Language Understanding“, 4 roky
 - JHU, Brown <-> UK, Saarland Univ.
- Speech Reconstruction
 - Význam, ne doslovný přepis
- Machine Translation
 - Struktura a sémantika věty, ne (jen) překlad po frázích
- ÚFAL:
 - Výuka (návštěvy studentů na min. 4 měsíce), workshopy (VMC)
 - Kontakt, podáno (výsledek 19.2.), 4.8 mil. Kč na 5 let
 - Naši studenti na 90 dnů ročně v USA (platíme jen cestu)
 - Výzkum: TR angličtiny, paralelní texty, transfer
 - GAČR, 3 r., 3.9 mil. Kč

Euromatrix

- Projekt EU, typ STREP
 - „Statistical and Hybrid Machine Translation Between All European Languages“, 30 měsíců (pouze!)
 - Partneři (7): DFKI (A. Eisele), Edinburgh (P. Koehn), CELCT Itálie, Globalware Německo, Morphologic Maďarsko, Linear-B Ltd. V. Británie
- Strojový překlad se strukturálním transferem
- ÚFAL (kEUR 190(??) – 30.3.2006 se rozhodne):
 - Definice hloubkové struktury, statistický transfer
- Inzerát: hledá se 1-1 ½ člověka ☺

Eva Hajičová

Projekty využívající TFA anotace

- Automatické členění věty na Topic a Focus (Lešnerová, Havelka)
 - Implementace algoritmu
 - Evaluace na vzorku PDT
- Ověření hypotézy o systémovém uspořádání (Lešnerová)

Jarmila Panevová

- vlastní náplň práce
 - úpravy a dotahování tektogramatické roviny (TGTS v PDT 2,5 i v teorii)
 - ve spoluautorství (zatím nechci jmenovat) připravit publikaci pro nakl. Karolinum: nové poznatky v syntaxi na základě vývoje PDT a dat v něm uložených
- „veřejně prospěšná“ činnost
 - zástupce ředitele ÚFAL
 - předsedkyně Rady doktorského studijního oboru (RDSO) I3
 - předsedkyně rigorózní komise oboru I3
 - členka státnicové zkušební komise
 - školitelka 5 doktorandů
 - (2 skončí přímo hned, 1 přerušené studium, zbývají 2 lehké případy)
 - členka 3 oborových komisí PGDS na FF UK
 - (matematická lingvistika, obecná lingvistika, současný český jazyk)
 - předsedkyně podoborové komise (POK 405) Lingvistika, literární věda GAČR
 - místopředsedkyně OK 4 (Společenské vědy) GAČR
- výuka
 - Vybrané kapitoly ze současné syntaxe češtiny
 - Gramatická cvičení pro doktorandy
 - Úvod do obecné lingvistiky
 - Seminář z úvodu do obecné lingvistiky
 - Odborné vviadřování a stvl

Bára Vidová-Hladká

- **Projekty**
 - *Data a nástroje pro informační systémy*, IS, 2004-2008
 - *cvičebnice STYX*, návaznost na dp: podaný návrh GAUK 2006
- **Vlastní bádání**
 - struktura diskurzu: aktivovanost objektů
- **Výuka**
 - přednáška Úvod do strojového učení (společně s Kirilem)
- **Zkušenosti**
 - podaný návrh projektu v rámci Marie Curie Training Programme (VMC, nepřijato)
 - podaná návrh projektu v rámci Central European Initiative (VMC, přijato se zkrácenými prostředky)
- **‘Černota’**
 - správa stránky o českém tagování
 - správa stránky s přehledem popularizačních článků, rozhovorů, ... ÚFALu

Alla Bémová

- substantiva v PDT-VALLEXu
 - dokumentace k substantivním doplněním
- projekt IS: Data a nástroje pro informační systémy
 - analytická rovina Českého akademického korpusu
- „veřejně prospěšná“ činnost
 - knihovna

Lucie Kučová

- Rozšíření anotace koreferenčních vztahů
 - substantiva
 - asociativní vztahy (bridging anafora)
 - koreferenční řetězce
 - nový anotační prostředek
- GAUK
 - propojení s TFA
- Stupně aktivovanosti textu
 - projekt Báry Hladké

Šárka Zikánová

- „Faktory koherence textu“
kontrolní anotace AČV
systémové uspořádání
- Anotace pojmenovaných entit
- Tektogramatická anotace

Jiří Havelka

- **Projekt značkování základu a ohniska v textu**
 - Prostředí pro anotaci
 - Koordinace anotace na technické úrovni
 - Cíl projektu: vyhodnocení automatické extrakce základu a ohniska z anotace aktuálního členění v PDT
- **Výzkum**
 - Vlastností závislostních stromů (navazující na práci na anotaci aktuálního členění v PDT)
 - Alternativní formulace podmínky projektivity
 - Typy neprojektivních hran
 - Pojem kanonické projektivizace a jeho zobecnění
 - Algoritmy
 - Snaha o vytvoření statistického modelu pro přiřazování hloubkového slovosledu a TFA
- **Výuka**
 - Komputační sémantika
- **Veřejně prospěšné práce**
 - Tajemník pro výuku za ÚFAL

Marie Mikulová

- **Manuál pro tektogramatickou rovinu PDT 2.0**
(dlouhá a krátká verze, organizace překladu)
- „Školení“ nových anotátorů pro anotaci
na tektogramatické rovině
- Disertace: elipsy
(doplňování uzlů na tektogramatické rovině
vzhledem k rovině analytické)

Magda Razímová

- gramatémy
- doktorský projekt „Collegium Informaticum“
- GA UK 2005 (?a 2006) „PZK: Analýza vybraných jevů z české funkční onomatologie a syntaxe“
- infromatická společnost „Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů“ - anotace pojmenovaných entit
- zastupování V. Kolářové v organizaci pondělních seminářů

Petr Pajas

- Projekt informační společnosti - Jazz
 - **“Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů“**, pod vedením Jarky Hlaváčové
 - Naše skupina (Pajas+Štěpánek+Zeman) vyvíjí PML, obecný datový formát pro jazykové zdroje založený na XML a použitý v PDT 2.0 a s ním související nástroje (zejm. ntred).
- Vývoj programových nástrojů
 - Editor stromů – TrEd + btred/ntred;anotační makra
 - XSH – jazyk/shell pro práci s XML
- Odborné zájmy a výzkum
 - Převážně nelingvistické: teorie modelů, logika, teorie množin, nestandardní metody
- Výuka
 - Logika a TM (pro učitelské obory, okolo 50 studentů); Nestandardní seminář
- Veřejně prospěšné práce
 - Správa dat z PDT 1.0 a PDT 2.0
 - Správa repozitáře Perlových modulů v /net/su/h/...
 - Poradenská činnost :-)

Jan Štěpánek

- **Informační společnost JAZZ** (Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů)
 - „žlutá skupina“ – XML formát
- **PDT 2.0**
 - kontroly konzistence dat
 - sample data
- **Anotace**
 - tvorba maker pro anotátory, rozdělování a sbírání dat k anotaci
- **Výuka**
 - PDT 2.0
 - Komputační sémantika
- **Veřejně prospěšné práce**
 - správa databáze pracovních cest

Jiří Mírovský

- **Minulost**
 - Morfologická disambiguace (tagování) češtiny pomocí HMM taggeru
 - Vytváření webových korpusů a selekce textů pro projekt MALACH
- **Učení založené na paměti**
 - Aplikace na tagování češtiny, v budoucnu na parsování do tektogramatické roviny
 - Vývoj metod pro určování vah atributů
- **Netgraph**
 - Vývoj vyhledávacího nástroje (Java, C)
 - Školení v- a pomoc při práci s Netgraphem
 - Správa databází stromů přístupných Netgraphem na serveru quest.ms.mff.cuni.cz
- **(Další) pracovní-vědecké zájmy**
 - tagování, parsování, tektogramatická rovina, valence, TFA, thajština
- **Veřejně prospěšné práce**
 - Nákup a půjčování software pro ÚFAL (kromě MS Windows a MS Office)
 - Nákup a správa dat vydávaných LDC; správa dat v /f/data
 - Pomocná správa databáze Amoeba
 - Půjčování diapojektoru

Otakar Smrž

Prague Arabic Dependency Treebank

- Vedení projektu – GAUK, publikace, správa dat
- Nové směry – TFA, valence, TG anotace (et alia)
- Bakalářská práce – XML reprezentace dat PADT
- Aktuálně – CoNLL-X, CCISSA @ CASL, PADT ++

Functional Arabic Morphology

- “F. A. M. – Formal System and Implementation”
- Functional Morphology, MorphoTrees, F-B Tagger
- Haskell a otázky funkcionálního programování

UFAL mě už ostatních povinností zprostil :)_7

Markéta Lopatková

- **VALLEX**
 - teorie, koordinace anotací, anotace
 - promítání alternací do dat (spolu s Karolinou Skwarskou)
 - GAČR 2004-2006
- **IS – Metody inteligentních systémů a jejich aplikace při dobývání znalostí a zpracování přirozeného jazyka**
 - redukční analýza, modelování syntaktické analýzy (s Martinem Plátkem)
 - koordinace „úfalí“ části
- **LCT ~ Language and Communication Technologies**
 - integrovaný magisterský program
- **výuka**
 - Vybrané problémy z lingvistiky I,II
 - Matematické metody v lingvistice I,II
- **další „veřejně prospěšná“ činnost**
 - program na hory

Zdeněk Žabokrtský

- **Valence sloves**
 - valenční slovník českých sloves VALLEX (2003 v.1.0, 2006 v.2.0)
 - alternační model valence (disertace 2005)
- **Treebanking**
 - PDT 2.0 (koreference, gramatémy, opravy, netred)
 - převody složkových stromů na závislostní (PTB, Negra/Tiger)
 - závislostní parsing (čeština, slovinština, němčina, rumunština, arabština)
 - pojmenované entity
- **Strojový překlad přes tektogramatickou rovinu**
 - vývoj prototypu od podzimu 2005
- Výuka: PFL070 – Zdroje lingvistických dat

Veronika Kolářová

- Výzkumný záměr, Informační společnost (J. Šíma), GAUK PZK (J. Panevová):
 - PDT 2.0 (podíl na anotacích a psaní textů do manuálu)
 - valence deverbativních substantiv
 - valence složených predikátů
- Momentálně: mateřská dovolená
- Veřejně prospěšná činnost:
 - příprava pozvánek na pondělní seminář (v současnosti předáno Magdě Razímové)

Zdeňka Urešová

- 1. Valence sloves: (dosud, dále???)
 - *anotace PDT – analytická i tektogramatická rovina*
 - *valenční slovník PDT-Vallex*
- 2. Aplikace tektogramatické roviny na mluvený jazyk (nově)
 - *anotace (manuálního) přepisu, data z projektu MALACH*
 - *postavení tektogram. stromů nad standardizovaným přepisem*
 - *analýza vztahu mluvený jazyk - TR*
- Odpovědnost:
 - *publikační databáze (ve spolupráci s Petrem Homolou)*

Václava Benešová

- Informační společnost,
- Centrum základního výzkumu,
- The Valency Lexicon of Czech Verbs with Complex Syntactic-Semantic Annotation,
- ???podaný GAUK: Problematické jevy ve valenci českých sloves a jejich zpracování v rámci Funkčního generativního popisu (M. Lopatková, Z. Urešová: valence sloves v mluveném projevu, M. Hučínová: zpracování zejm. analytických predikátů, V. Benešová: třídy sloves)
 - valence sloves
 - revize VALLEXu 1.0 po experimentu VALEVAL (O. Bojar, J. Semecký)
 - další anotace valenčního slovníku VALLEX (nová slovesa)
 - anotace slovesných tříd + jejich další rozpracování
(propracování tříd sloves communication tradičně označovaných jako verba dicendi a cogitandi) + návrh exp. s těmito slovesy (O. Bojar)
- „Veřejně prospěšné práce“
 - nástěnk. ufalí stránkv

Silvie Cinková: TR popis angličtiny

- **Úkoly:**
 - Vypracování anotačního manuálu podle české TR
 - Vytvoření valenčního slovníku anglických sloves (**EngValLex**) (s J.Vejvodovou a J. Semeckým)
 - Anotace korpusu PTB-WSJ (s J. Vejvodovou a J.Semeckým)
- **Účast v projektech:**
 - IS - Data a nástroje pro informační systémy (2004-2008)
 - IS - Od jazyka ke znalostem a sémantickému webu (2005-2009)
 - GAUK – Tektogramatická reprezentace angličtiny (2004-2006)
 - GAČR – Tektogramatický popis jazyka pro strojový překlad a rozpoznávání mluveného jazyka (2006-2008)
- **Veřejně prospěšné práce:**
 - od podzimu 2004 do února 2005 práce na první verzi překladu TR manuálu pro češtinu

Jiří Semecký

- **Disambiguace slovesných rámců**
 - disambiguace korpusu ValEval podle VALLEXu
 - disambiguace PDT podle PDT-VALLEXu
- **ValEval**
 - Správa webu projektu (<http://ufal.ms.mff.cuni.cz/~semecky/vallex/>)
 - Nástroje pro kontrolu anotačních textů
- **Prague English Dependency Treebank**
 - Editor valenčních rámců pro angličtinu
 - Generování a údržba valenčního slovníku pro angličtinu
 - Překlad anglických korpusů do PML (Prague English Dependency Treebank)
- **Veřejně prospěšné práce**
 - Překlad manuálů PDT z XML do PDF, PS, HTML

Ondřej Bojar

- **Slovesné rámce – experimenty kolem VALLEXu**
 - Zejména VALEVAL, 100x100 výskytů sloves označených rámcem, WSD
 - Souhrnně moje disertace: jak se učit VALLEXová hesla/součástky z korpusu (reflexivita, obligatornost, rámce, odlišení významů)
- **Česko-anglický slovník pro strojový překlad (GAUK)**
 - Ruční alignment testovacích dat PCEDT po slovech
 - Poloautomatické čištění hesel z elektronických slovníků
 - Doplnování chybějících údajů na základě výskytů v korpusu (morfologie, syntax, frekvence)
- **Experiment se statistickým strojovým překladem (DAAD)**
 - Více na jednom z pondělních seminářů
 - Zřejmě se budu podílet na čištění Czengu (Zdeňkův paralelní korpus)
- **Okrajově loni též Informační společnost**
 - Úvahy o učení se redukční analýzy z korpusu
- **Veřejně prospěšné**
 - PDT: instalace, pomoc s překladem manuálu
 - Konferenční stránka
 - Zde na Kvildě: unixové nástroje pro lingvisty, remcání

Pavel Straňák

- Lexikálně semantická anotace PDT
 - Anotace lexikálních významů
 - Moje práce: co anotovat (na které rovině, hranice jednotek), čím anotovat (konstrukce slovníku lex. významů), jak anotovat (metodologie, anotační nástroj)
 - Vlastní anotace: Marie Hučínová, Pavel Šidák
- GAUK: Možnosti extrakce morfosémantických tříd ve více jazycích
 - které jazyky jsou pro extrakci morfosémantických tříd vhodné,
 - jaké sémantické třídy je možné identifikovat v jednotlivých jazycích,
 - Jak informace z různých jazyků kumulovat v jednom slovníku
 - nakolik je možno celý proces automatizovat
- Veřejně prospěšné práce
 - Příprava PBML do tisku a na web

Lexikálně-sémantické anotace PDT

Petra Möllerová

- 1. kolo:

2 anotátoři nezávisle na sobě anotovali data z PDT 1.0 → přiřazení významů z CWN ==> u neshody data dále analyzována 3. anotátorem

V případě, že byla nalezena nesrovnalost mezi významy slov v anotacích a v CWN, byla data „poslána“ do 2. kola.

- 2.kolo:

Na počátku podrobeny analýze jednotlivé synsety v CWN → v případě, že byly nekonzistentní, následovala oprava synsetu přímo v CWN ==> anotace dat „novým“ významy

- Problémy:

- převod synsetů z WordNetu do CzechWordNetu → nekryjí se významy
- některé významy slov v CWN naprosto chybějí, nebo jsou nejasně definované
- nejasnosti v použití literálů a multiliterálů při anotaci

Pavel Pecina

- disertace: **Lexikální asociační míry**
 - *extrakce kolokací, word alignment*
- projekt: **Malach**
 - *testovací kolekce pro Information Retrieval*
 - *automatický překlad tezauru*
- výuka: **PFL 067/068, NLP Reading Group**
 - *asistent doc. Hajiče při výuce statistických metod v NLP*
 - *pravidelná týdenní setkání nad články z oboru NLP*
- vedení: **softwarové projekty, bakalářské a diplomové práce**
 - *web corpus, basic text processing, near-duplicate detection*

Pavel Schlesinger

- Výzkum na téma „Adaptace statistických metod na problémy počítačové lingvistiky“
 - disertace ve stavu řešerše
 - snad trénování vah pro parser pomocí klasifikace, změna možná
 - kolokace spolu s Pavlem Pecinou (já: hlavně stat. analýza dat)
 - vylepšování implementace spolu s J. D. Tebbensem (AV ČR) lineární diskriminační analýzy
 - použití také v IR, WSD
 - k promyšlení: co takhle v týmu zkusit příští Senseval-4?
- Statistický servis
 - vedení pravidelného setkávání z kraje týdne (po. 9:00)
 - kombinace rad pro aplikaci statistiky a její implementaci
 - bylo: první krůčky v **R** (výpočetní prostředí pro práci s daty), obrázky a grafy, vizualizace vícerozměrných dat, řešení typických úloh při začátku experimentu ...
 - bude: metody pro klasifikaci, clustering..., ale i podle **poptávky**
- Veřejně prospěšné práce
 - v nejbližší době převzetí břemene od Václava Klimeše s časovou koordinací a přípravou časopisu PBML

Jarka Hlaváčová

- Aktualizace morfologického slovníku češtiny
 - opravy, nová slova, kontroly, morfologický servis
- Výuka - Perl pro lingvist(k)y – letos ne
- Projekty Informační společnosti
 - REST (Data a nástroje pro jazykové systémy)
 - JAZZ (Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů)
- Evropský projekt It4el (vk)
- EUROVOC – spolupráce s parlamentní knihovnou

Miroslav Spousta

Lingvistika:

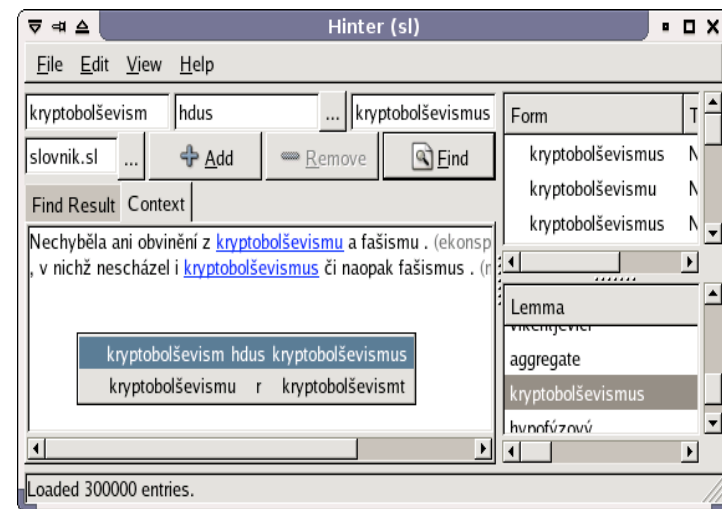
- DP: přiřazení tvaroslovných vzorů (v češtině)
 - přidal jsem grafický interface pro přidávání slov do morf. slovníku
 - knihovnu jsem používal i pro jiné úkoly
 - přidávání vidu do korpusu
- projekt LT4eL
 - e-learningové materiály, programování
- získávání textů z webu
 - úprava nástroje Holmes

Něco společného s lingvistikou

- Bara, systém pro správu lístečkové databáze na UJC AV

Veřejně prospěšné práce

- neprovozují
- učení: seminář TCP/IP






Dan Zeman

- **Informační společnost se ZČÚ Plzeň**
 - Strojový překlad z češtiny do znakové řeči (3000 pojmů)
 - Na nás (Zeman+Klimeš+Semecký) je ta část, kterou stejně děláme: analýza psané češtiny až po t. rovinu
- **Parsing češtiny**
 - Kombinace parserů (loni se Zdeňkem)
 - Snaha pustit parsery na PDT 2.0
- **Momentálně na University of Maryland**
 - Použití parseru natrénovaného na PDT pro příbuzné jazyky (teprve počáteční pokusy)
- **Veřejně prospěšné práce**
 - Správa databáze adres lingvistů a jiných lidí, jejichž adresy může někdo na ÚFALu potřebovat. Webové rozhraní, tisk štítků aj. Telefonní seznam pracovníků ÚFALu.
 - Správa webových formulářů pro objednávky a registrace korpusů z rodiny PDT. Odpovědi na dotazy na pdtd@ufal.mff.cuni.cz.
 - Zatím neoficiální webový formulář pro hlášení publikací Zdeně Uřešové.

Kiril Ribarov

- Parsing češtiny a projekt informační společnosti
 - Testování rysů
 - Rozšíření MST závislostního parsru
- Stromečky v českém akademickém korpusu (projekt Data a nástroje pro inf. systémy)
 - Převod syntaktické anotace na analytické PZK stromy
- Moje plánované působení
 - Duben až červenec – Bulharska akademie věd, Sofie, Transfer of Knowledge projekt: Digital Cultural and Scientific Heritage
 - Od října 2006, zatím nepotvrzené, Itálie
- Zodpovědný za
 - Erasmus/Socrates výměna lektorů a studentů s Univerzitou v Pise

Václav Klimeš

- Dizertace (bude hotova snad v létě)
 - analytický, a hlavně tektogramatický parsing; učící se metoda (TBL)
- Grant AV ČR „Informační společnost“, spolu se ZČU
 - téma: strojový překlad čeština – znaková řeč
 - my (Zeman+Klimeš+Semecký) děláme analýzu psané češtiny až po TR
 - nyní za MFF vedení projektu
- Příprava CD s PDT 2.0 k vydání
 - správa průvodce CDčkem a nástrojů pro strojovou anotaci dat, CSS, ...
- Vedení ročníkových projektů a bakalářských prací
- Správce české mutace Collinsova parseru  Jiří Mírovský
- Redaktor PBML odpovědný za příspěvky  Pavel Schlesinger
- Správce interních webových stránek  Martin Cetkovský

Johanka

- Zařazení: Zástupce ÚTKL v CKL -> formálně elementem ÚFALu, efektivně ÚTKL
- Pracovní náplň: všelike příspěvky k pravidlovému zpracování češtiny (morfologická disambiguace, syntax)
 - Programovací, lingvistická a testovací činnost pro disambiguační projekt
 - Vývoj pravidlového parseru postaveného na disambiguačních pravidlech
 - Vymýšlení a testování různých metod kombinace pravidel a statistiky, a to jak v taggování (ve spolupráci s Annou K. alias Honzou Votrubcem), tak i v parsingu
 - Momentálně i koordinace značkování nového SYN2005
- Otročina: pouze pro ÚTKL :) (správa serveru, dat apod.)
- Fyzicky se vyskytují na obou pracovištích – lze po mně poslat libovolnou poštu, vzkazy, peníze ;) ³⁵

Přehled aktivit v roce 2005 – Vlád'a Kuboň

Projekty

- LATER – Language Technology Erasmus Mundus – propagace CL mimo EU
- LCT – příprava magisterského studia v rámci programu Erasmus Mundus
- LT4EL – Language Technology For E-Learning – STREP v 6.rámcovém programu EU
schválen v červenci, zahájen 1.12.2005, trvá 30 měsíců
- GAČR – Překlad ekonomických textů (skončil 31.12.2005)
- GAUK – formální vedení projektu Ondřeje Bojara a Magdy Prokopové
- Collegium Informaticum – organizační zajištění společně s dr.Bartákem

Výuka

- přednáška Úvod do počítačové lingvistiky (podzim)
- semináře Syntaktická analýza češtiny (jaro)
- seminář Nástroje automatického překladu (podzim)
- Praktikum z informatiky (jaro)
- studentský experiment s EBMT v Hamburku v květnu
- příprava studentských softwarových projektů Fotosoutěž a MOAP (astrologie)

Aplikace

- spolupráce při vývoji českého grammar checkeru pro Microsoft 36

Petr Homola

- Informační společnost s AV
 - formalismus pro parsing (FS)
- *Česílko* — MT mezi blízkými jazyky
 - čeština - polština, litevština
- pro ÚFAL
 - webová aplikace pro publikace

Petr Němec

- strojový překlad (ČJ \Rightarrow AJ)
 - transfer skrze tektogramatickou rovinu
 - generativní komponenta
 - třeba dokončit samotný transfer
 - ve spolupráci s K.Hallem a CLSP
- automatická analýza časových vztahů
 - systém (ne)rovností mezi počátky a konci dějů vyjádřených v diskurzu
 - základem je informace daná gramaticky (J.Panevová)
 - další informace (sémantika časových adverbíí, další obsahové informace)

Václav Novák

- **Informační společnost**
 - Návrh 4. roviny – znalostní báze
 - Porovnání sémantických formalismů (DRT, TIL, MultiNet, KRL)
 - Prof. Hermann Helbig z Hagenu promluví o **MultiNetu** 20. února na semináři (PFL004)
- **Parsing češtiny**
 - Spolu s Keithem Hallem
 - Opravy výstupu parseru (statistické modelování oprav)
 - Jako reranking, jenže varianty nemáme od parseru, ale tvoříme je sami
 - Zneprojektivnění u projektivních parserů
 - Snaha pustit MST parser na PDT 2.0 (zatím 84.24%, produkuje i funktoxy)
 - Generování n-best listu pomocí lokálních změn rodičů
- **Výuka**
 - Internet (SWI096), 2 skupiny

Petr Podveský

- Dizertace o rozpoznávání řeči
- Studenti
 - Bc. HMM tagging
 - Mgr. automatická detekce chyb ve výstupu z rozpoznávače řeči
- Editor M-roviny
 - rozšíření PDT o Speech Reconstruction
- Veřejně prospěšné práce
 - technické zprávy

Nino Peterek

- Doktorandská práce zaměřená na modelování prosodie mluvené češtiny
 - sběr a anotování dat mluveného prosodického korpusu,
 - tvorba softwarových nástrojů umožňujících vyhledávání a analýzy v prosodických korpusech,
 - trénování modelů intonace češtiny použitelných pro rozpoznávání a syntézu mluvené češtiny i prosodických jevů.
- Spolupráce s Ústavem pro jazyk český na korpusu Dialog
 - konverze audio-visuálních nahrávek a prosodických přepisů televizních pořadů do korpusové podoby,
 - generování morfologie a pomoc s postupnou ruční disambiguací přepisů,
 - generování time-alignmentu přepisů a audio-video nahrávek,
 - plný přístup ÚFALu k celému korpusu.
- Výuka
 - seminář Základy rozpoznávání mluvené češtiny (PFL038).
- Veřejně prospěšné práce
 - správa síťové verze softwaru Matlab.

Jan Votrubec

- **„Strojové učení prozodie“
aneb „Co bych měl dělat“**
 - Shromáždění co největšího množství mluvených a označovaných korpusů (jakékoliv zdroje informací vítány!)
- **Morfologické značkování – MORČE
aneb „Co doopravdy
dělám“**
 - Založeno na „Průměrovaném perceptronu“
 - Pro PDT 2.0 zatím nejlepší výsledek – 94,7%
 - Pokusy zkombinovat statistiku a pravidla (s Johankou)
 - Úpravy pro použití na arabská data

David Klusáček

- Informační společnost
 - *Generování formálního důkazu na základě jeho popisu v přirozeném jazyce (pokus o reprezentaci sémantiky textu a její automatické trénování s využitím „teoretické dopočitatelnosti“ důkazu i bez znalosti textu).*
- SW, který by mohl zajímat ostatní
 - *Jako vedlejší produkt při řešení předchozího problému vznikla knihovna (v jazyce C) pro shlukování MMI metodou.*
- Dříve
 - *Rozpoznávání řečníka podle hlasu (speaker detection).*

Milan Fučík

- Střídavý rozvoj a likvidace počítačových systémů UFALu
 - Realizace investic a plánování rozvoje sítě
 - Řešení nových problémů
 - Analýza požadavků na používání a instalování nového SW z hlediska zabezpečení a kompatibility se současným vybavením
 - Bezpečnostní management sítě a systémů
 - Administrace přístupových oprávnění, rozložení služeb a jejich optimalizace
 - Andrej v pozici asistenta (během nepřítomnosti, rady, úkoluji já)
 - Andrej řeší obvykle známé problémy, o kterých jsme se dohodli, jak k nim přistupovat
 - Udržování přehledu o dostupných technologiích a jejich testování z hlediska použitelnosti pro potřeby katedry
 - Nasazování toho co projde jako užitečné
 - Učit se, učit se, učit se a občas taky nezbláznit se 😊