

Učíme počítače (nejen) česky: ÚFAL

Barbora Hladká

Ústav formální a aplikované lingvistiky

Matematicko-fyzikální fakulta

Univerzita Karlova v Praze

<http://ufal.mff.cuni.cz>



Ústav formální a aplikované lingvistiky (ÚFAL MFF UK)

se věnuje automatické komunikaci
člověka s počítačem v přirozeném
jazyce.



Učíme počítače (nejen) česky

- Sgall Petr, Hajičová Eva, Piřha Petr.
Učíme stroje česky, Panorama, 1982.



ÚFAL (MFF UK)

□ Dlouhá tradice

UNIVERZITA KARLOVA V PRAZE

matematicko-fyzikální fakulta



Lingvistika
na matfyzu

60

50



Ústav formální a aplikované lingvistiky

formální lingvistika

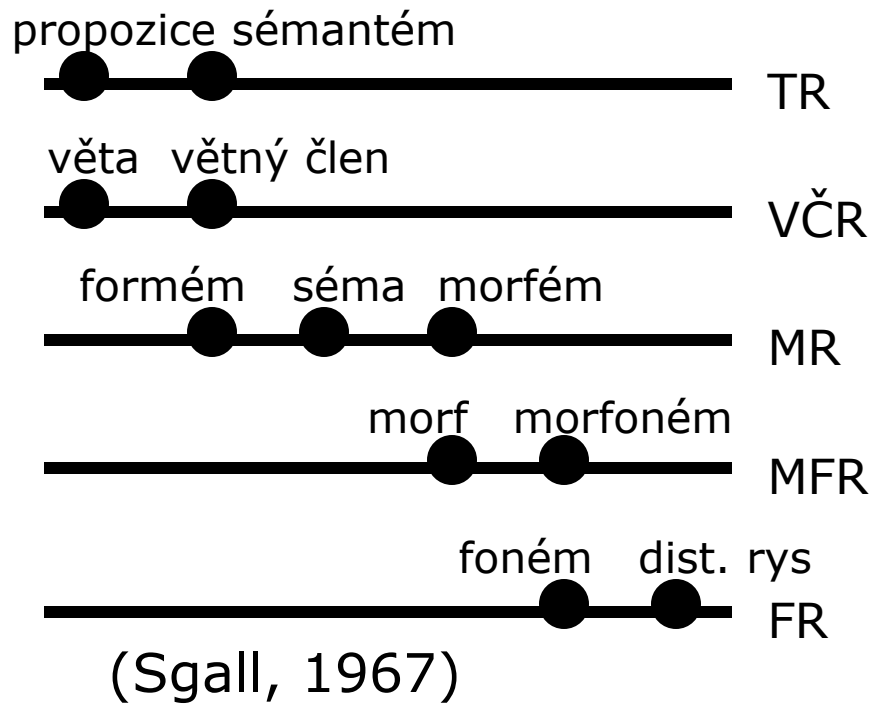
aplikovaná lingvistika*

* Počítačová lingvistika nebo počítačová lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

formální lingvistika

aplikovaná lingvistika*

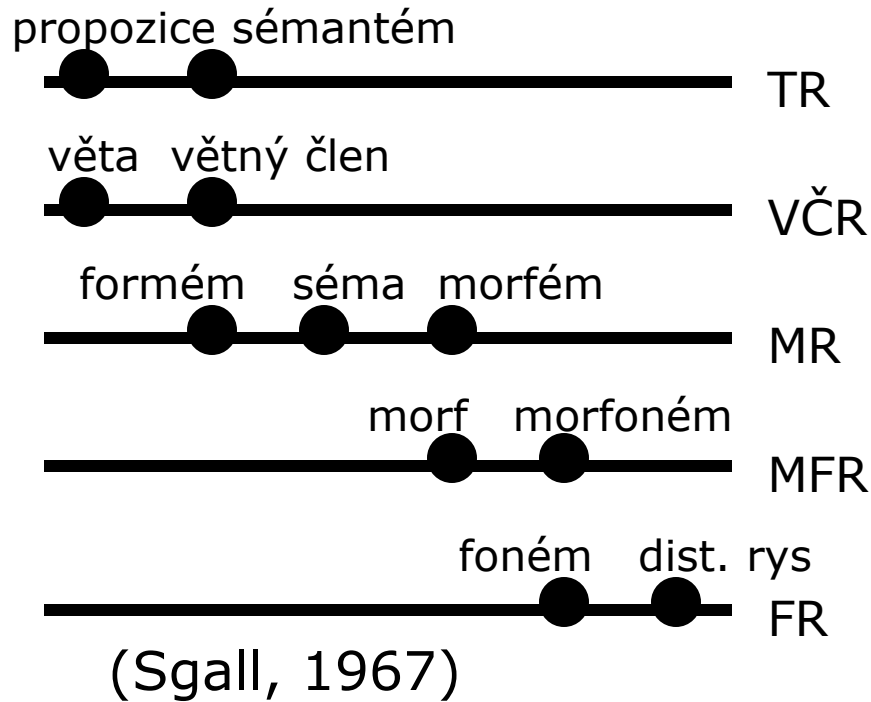


* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

formální lingvistika

aplikovaná lingvistika*



data

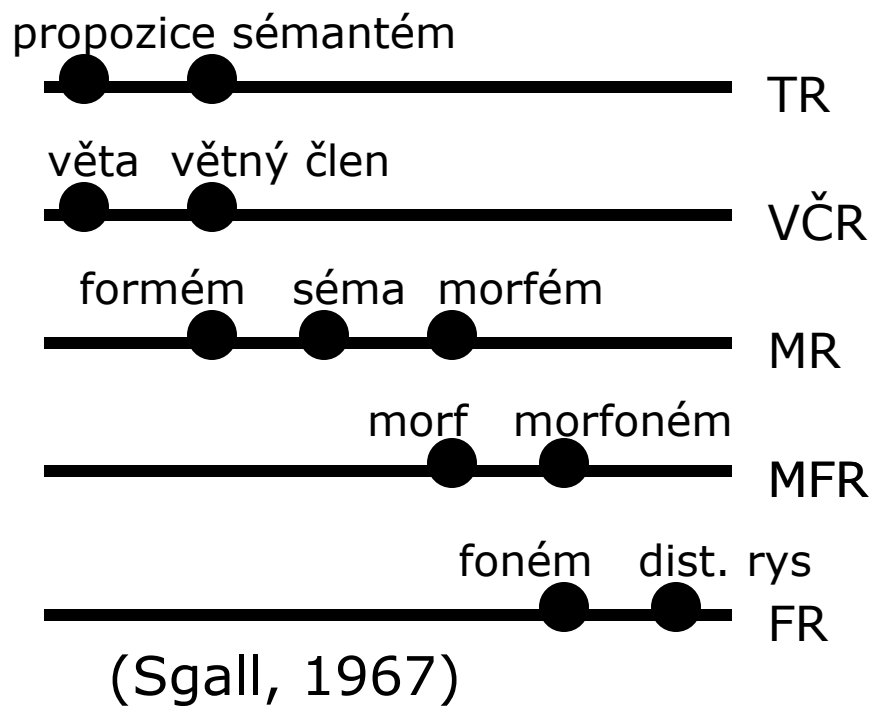
nástroje

* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

formální lingvistika

aplikovaná lingvistika*



data

nástroje

text a řeč



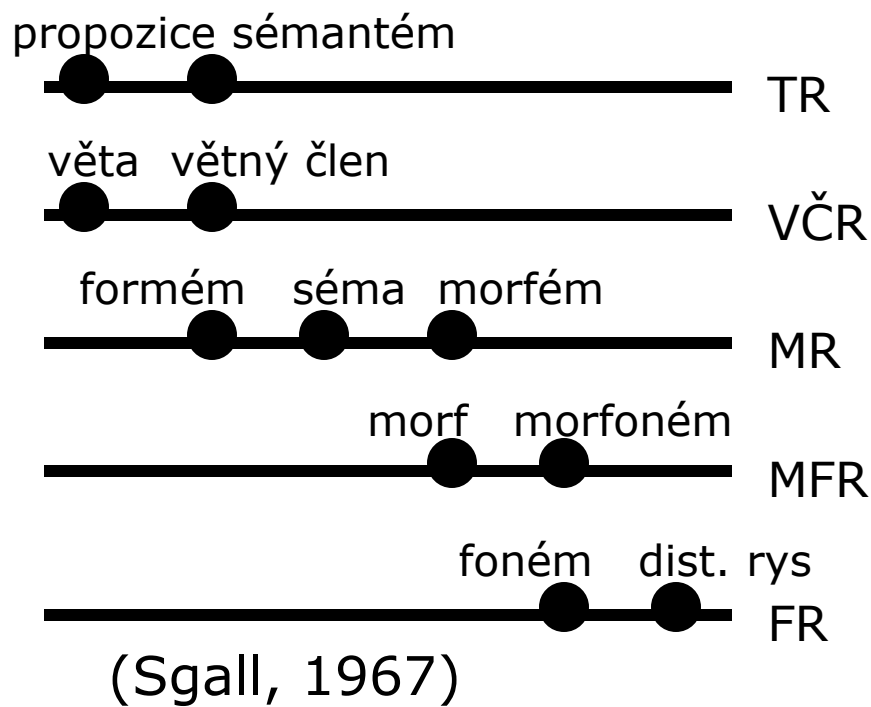
* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

formální lingvistika



aplikovaná lingvistika*



data

nástroje

text a řeč



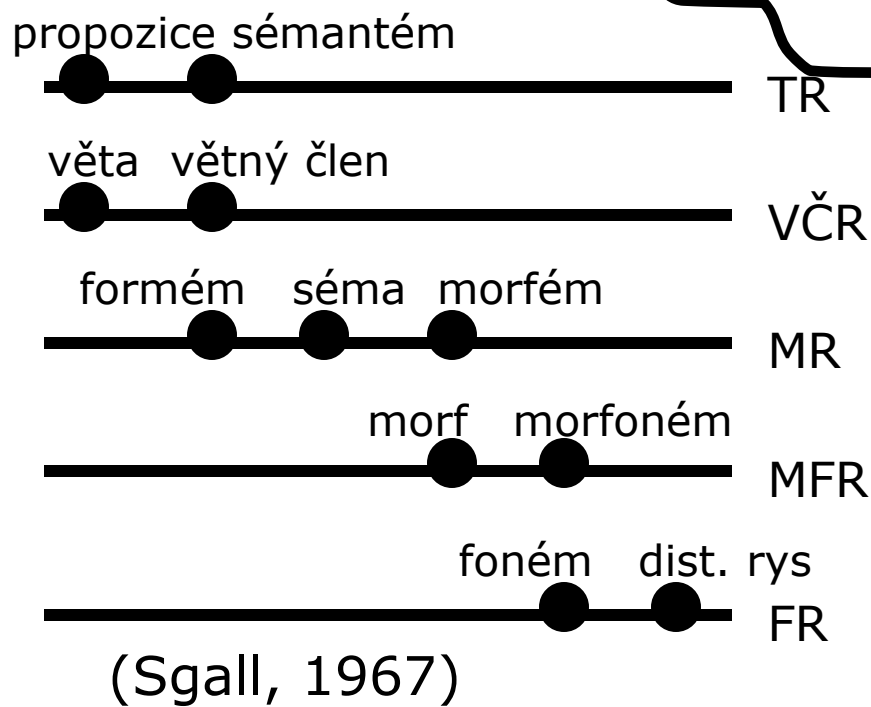
* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

formální lingvistika



aplikovaná lingvistika*



data

nástroje

text a řeč



* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Ústav formální a aplikované lingvistiky

□ Výuka

- Magisterský a doktorský program
Matematická lingvistika

□ Studijní pobyty

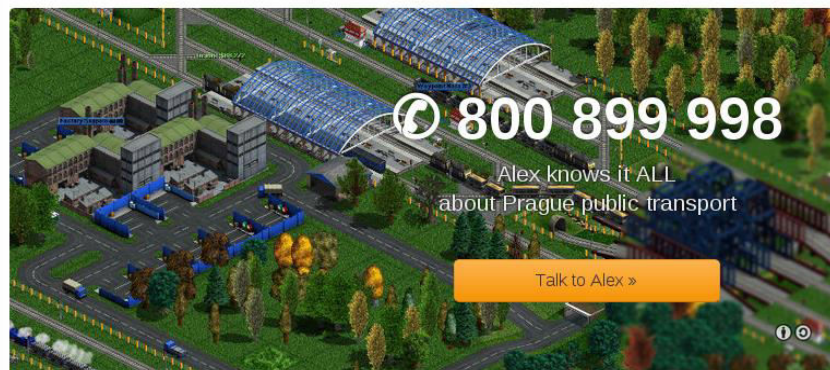
□ Videonahrávky z akcí



Institute of Formal and Applied Linguistics



ABOUT EVENTS RESEARCH PEOPLE TEACHING WIKI



Monday

17

March

Upcoming Event

How to evaluate a corpus

Adam Kilgariff will introduce a framework for corpora evaluation as part of our Fred Jelinek Seminar series.

[View details >](#)

or [open full ÚFAL calendar](#)

PhD position available

Join our efforts in the area of Spoken Dialogue Systems research!

You will investigate and implement statistical models, increase the efficiency of the learning process and reduce the need for large corpora.

[View job details >](#)



Vše na <http://ufal.mff.cuni.cz>

I 
90s

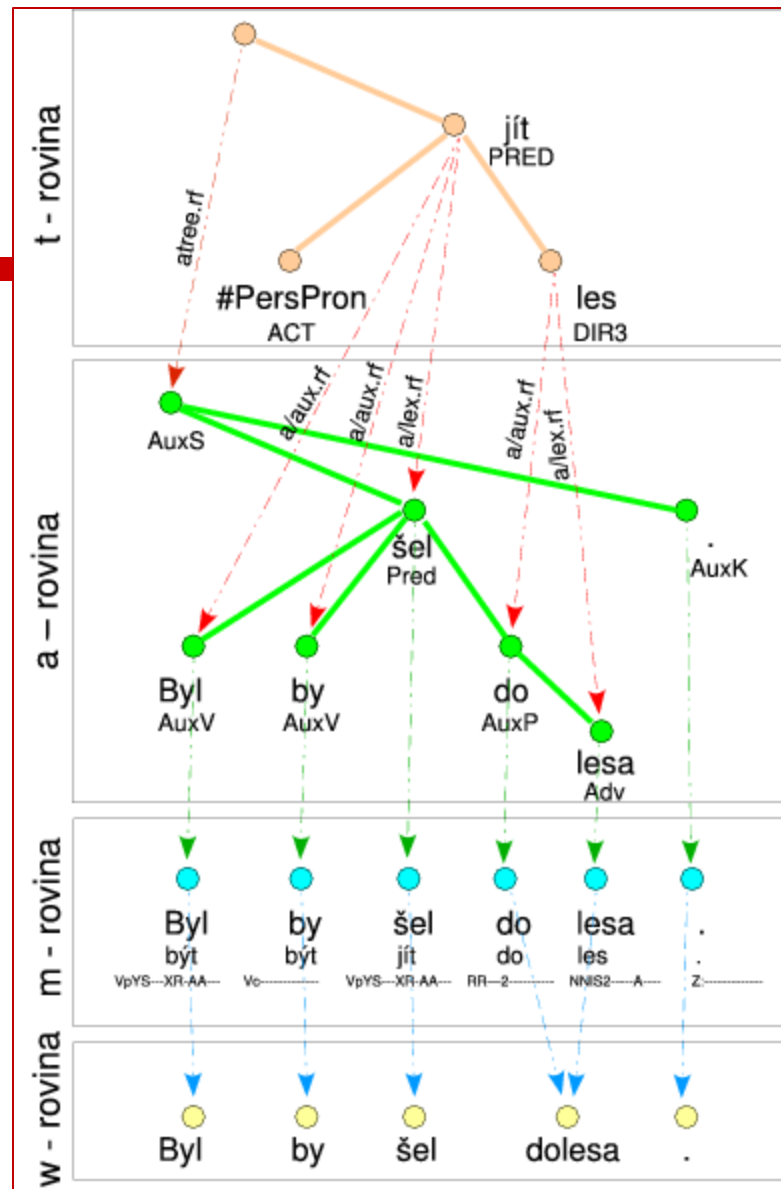
aplikovaná lingvistika*

data

nástroje

Banky dat korpusy textů

- Pražský závislostní korpus
- Pražský anglický závislostní korpus
- Pražský anglicko-český závislostní korpus
- Pražský závislostní korpus mluvené češtiny
- Pražský arabský závislostní korpus
- Český akademický korpus
- ...



Banky dat slovníky

- Vallex 2.6 (valenční slovník sloves)
- Czech SubLex 1.0 (slovník evaluativních výrazů)
- ...

VALLEX 2.6

alphabet class functors forms aspect control reflex. recipr. complexity VALEVAL miscel. home help ?

absolvovat^{blasp}

1 = zakončit, końčit

-frame: ACT₁^{obl} PAT₄^{obl}

SubLex1.0

A	báječný	polarity=positive
V	balamutit_:T	polarity=negative
N	banalita	polarity=negative
A	banální	polarity=negative
N	bankrot	polarity=negative
N	barbar	polarity=negative
A	barbarský	polarity=negative
D	barbarsky_^(*1ý)	polarity=negative
N	barbarství	polarity=negative
N	bastard	polarity=negative
V	bát	polarity=negative
V	bavit_:T	polarity=positive
N	bázeň	polarity=negative
D	báživě_^(*1ý)	polarity=negative
N	bažina	polarity=negative
N	bdělost_^(*3ý)	polarity=positive
I	běda	polarity=negative
V	bědovat_:T	polarity=negative
V	belhat_:T	polarity=negative
V	běsnit_:T	polarity=negative
N	bestie	polarity=negative
N	bezduchost_^(*3ý)	polarity=negative
A	bezduchý	polarity=negative
D	bezděvodně_^(*1ý)	polarity=negative
N	bezmoc	polarity=negative
D	bezmocně_^(*1ý)	polarity=negative
N	bezmocnost_^(*3ý)	polarity=negative

Banky dat

Čím více dat, tím lépe.

Nástroje

Naučme počítač určovat pád, číslo a rod podstatných jmen.

Kupujete motorovou pilu? Deset vychytávek pro kvalitní stroj
25. září 2012

Motorové pily vypadají na první pohled všechny téměř stejně, ne nadarmo se ale podstatně liší cenou. Jde nejen o sílu a objem motoru, ale i o kvalitu válců či robustnost provedení ...

(Zdroj: <http://hobby.idnes.cz>)

Naučme počítač určovat **pád, číslo** a **rod podstatných jmen.**

Kupujete motorovou **pilu**(4JŽ)? Deset **vychytávek**(2MŽ) pro kvalitní **stroj**(4JM)
25. **září**(2JS) 2012

Motorové **pily**(1MŽ) vypadají na první **pohled**(4JM)
všechny téměř stejně, ne nadarmo se ale podstatně
liší **cenou**(7JŽ). Jde nejen o **sílu**(4JŽ) a **objem**(4JM)
motoru(2JM), ale i o **kvalitu**(4JŽ) **válců**(2MM) či
robustnost(4JŽ) **provedení**(2JS).

Naučme počítač určovat pád, číslo a rod podstatných jmen.

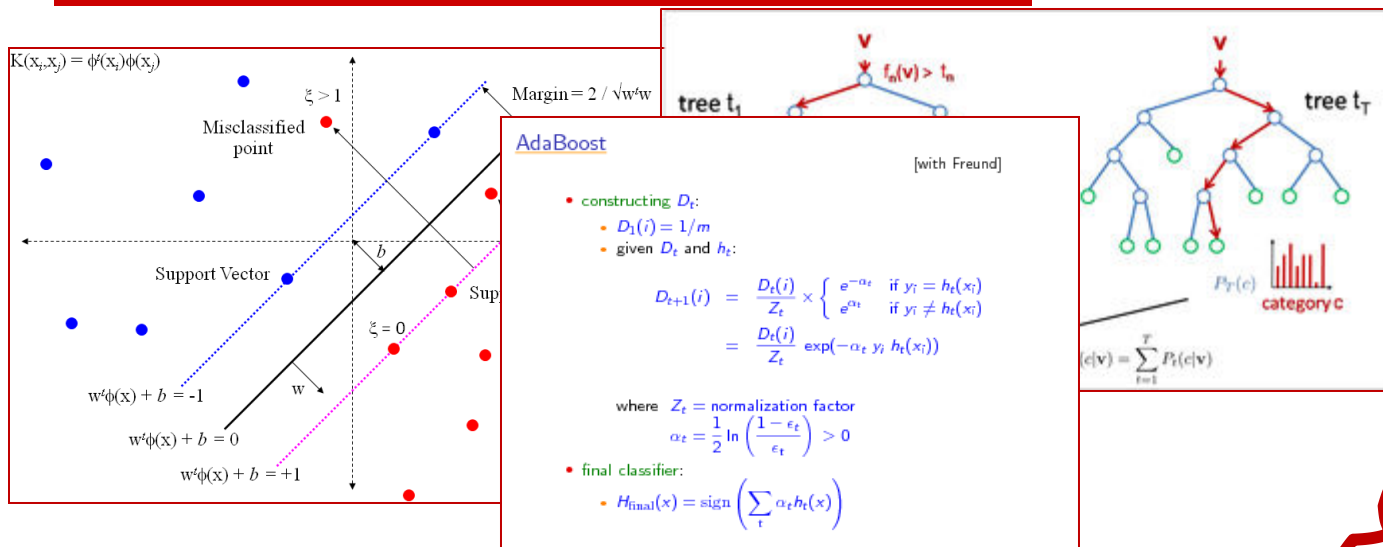
Kupujete motorovou pilu(4JŽ)? Deset vychytávek(2MŽ) pro kvalitní stroj(4JM)
25. září(2JS) 2012

Motorové pily(1MŽ) vypadají na první pohled(4JM) všechny téměř stejně, ne nadarmo se ale podstatně liší cenou(7JŽ). Jde nejen o sílu(4JŽ) a objem(4JM) motoru(2JM), ale i o kvalitu(4JŽ) válců(2MM) či robustnost(4JŽ) provedení(2JS).



anotace

Strojové učení s učitelem



Vzduchový filtr patří mezi opomíjené součásti motoru.

Vzduchový **filtr**(1JM) patří mezi opomíjené **součásti**(4MF) **motoru**(2JM).

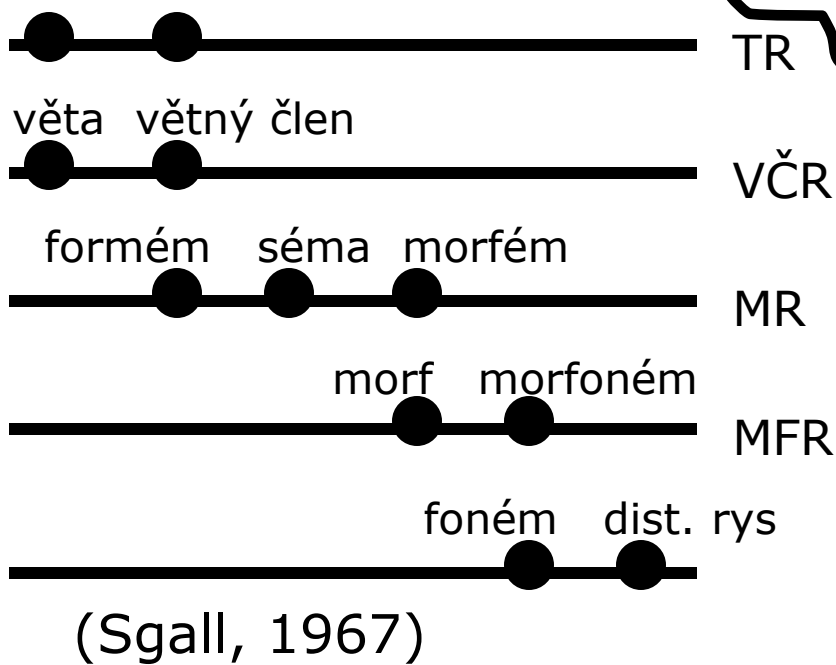
Kupujete motorovou **pilu**(4JŽ)? Deset **vychytávek**(2MŽ) pro kvalitní **stroj**(4JM)



Ústav formální a aplikované lingvistiky

formální lingvistika

aplikovaná lingvistika*



data

nástroje

text a řeč



* Počítačová lingvistika nebo počítačnická lingvistika nebo počítačové zpracování přirozeného jazyka nebo jazykové technologie.

Klasifikace lingvistických nástrojů

Správa dat

Automatické zpracování dat

Klasifikace lingvistických nástrojů

Správa dat

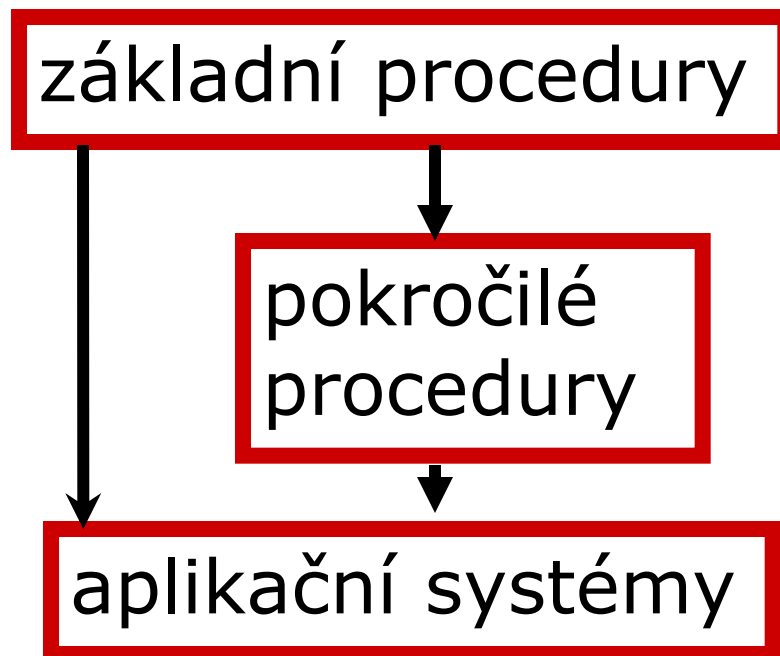
- Vnitřní reprezentace dat
- Anotace/úprava dat (editory, on-line hry)
- Prohlížení dat (prohlížeče)

Automatické zpracování dat

Klasifikace lingvistických nástrojů

- Správa dat
 - Vnitřní reprezentace dat
 - Anotace/úprava dat (editory, on-line hry)
 - Prohlížení dat (prohlížeče)

- Automatické zpracování dat



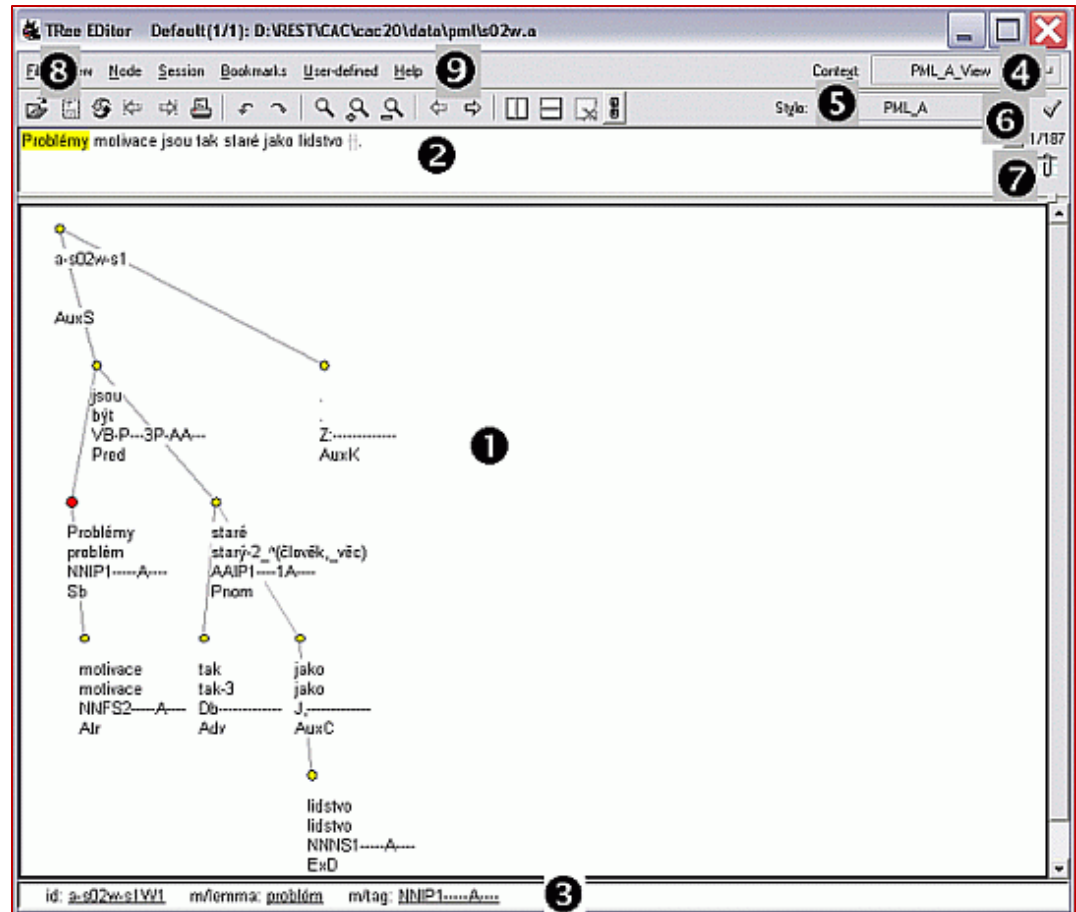
Nástroje z ÚFAL



- Nástroje z projektů, na jejichž řešení se ÚFAL podílí/podílel.
 - Studentské práce a projekty, které vedou/vedli lidé z ÚFALu.
-

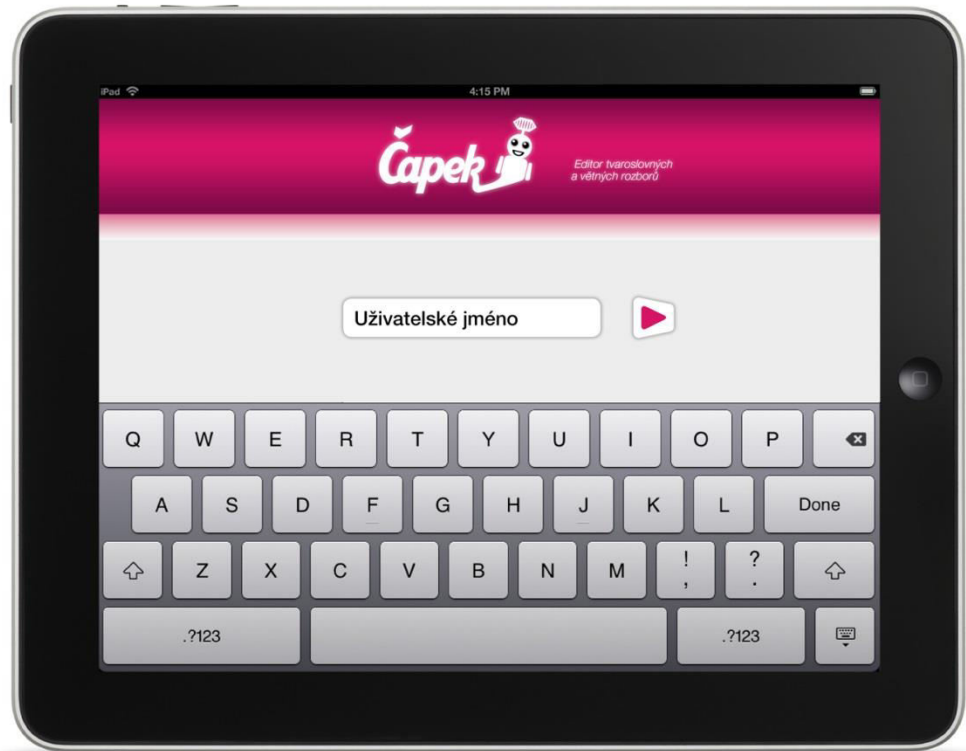
Správa dat - anotační editory

□ např. TrEd



Správa dat - anotační editory

□ např. Čapek



Základní procedury



Základní procedury



Hobby pila Oleo-Mac s výkonem 1,8 kW, objemem motoru 39 ccm a lištou 41 cm má boční napínání řetězu. Vzduchový filtr patří mezi opomíjené součásti motoru.

Základní procedury



Hobby pila Oleo-Mac s výkonem 1,8 kW, objemem motoru 39 ccm a lištou 41 cm má boční napínání řetězu. Vzduchový filtr patří mezi opomíjené součásti motoru.

Základní procedury



Hobby pila Oleo-Mac
s výkonem 1,8 kW ,
objemem motoru 39
ccm a lištou 41 cm
má boční napínání
řetězu . Vzduchový
filtr patří mezi
opomíjené součásti
motoru .

Základní procedury



Hobby pila Oleo-Mac
s výkonem 1,8 kW ,
objemem motoru 39
ccm a lištou 41 cm
má boční napínání
řetězu . Vzduchový
filtr patří mezi
opomíjené součásti
motoru .

Hobby pila Oleo-Mac
s výkonem 1,8 kW ,
objemem motoru 39
ccm a lištou 41 cm
má boční napínání
řetězu . Vzduchový
filtr patří mezi
opomíjené součásti
motoru .

Základní procedury

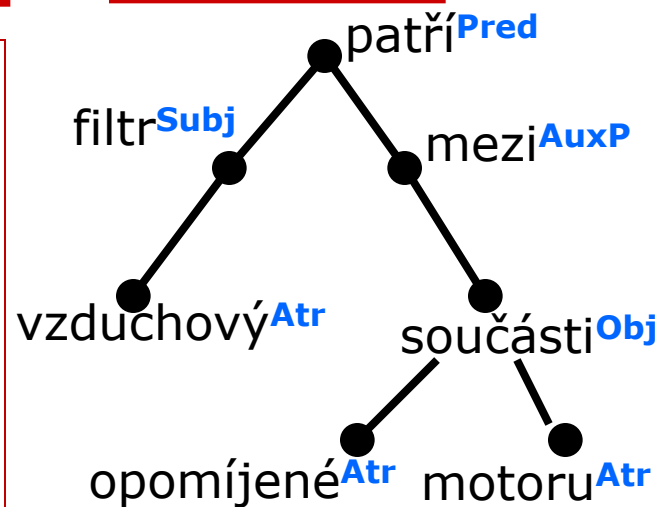
segmentace

tvarosloví

syntax

Hobby pila Oleo-Mac
s výkonem 1,8 kW ,
objemem motoru 39
ccm a lištou 41 cm
má boční napínání
řetězu . Vzduchový
filtr patří mezi
opomíjené součásti
motoru .

Hobby pila Oleo-Mac
s výkonem 1,8 kW ,
objemem motoru 39
ccm a lištou 41 cm
má boční napínání
řetězu . Vzduchový
filtr patří mezi
opomíjené součásti
motoru .



Základní procedury



Demo pro češtinu

- ❑ MorphoDita

<http://ufal.mff.cuni.cz/morphodita/demo>

- ❑ Treex::Web

<https://lindat.mff.cuni.cz/services/treex-web/run>

Pokročilé procedury

- např. určování koreference

Vzduchový filtr patří mezi opomíjené součásti motoru. Je ideální, pokud se k němu můžete dostat bez nářadí a pokud ho lze jednoduše vyčistit vymytím.

Pokročilé procedury

- např. určování koreference

Vzduchový **filtr** patří mezi opomíjené součásti motoru. Je ideální, pokud se k **němu** můžete dostat bez nářadí a pokud **ho** lze jednoduše vyčistit vymytím.

Pokročilé procedury

- např. určování koreference

Vzduchový **filtr** patří mezi opomíjené součásti motoru. Je ideální, pokud se k **němu** můžete dostat bez nářadí a pokud **ho** lze jednoduše vyčistit vymytím.

- např. detekce a klasifikace jmenných entit

Od loňského září bylo v kinech uvedeno 19 dokumentů. Vedle staršího snímku Matěje Mináče Nickyho rodina patřilyk nejnavštěvovanějším Pod sluncem tma Martina Marečka (11 487diváků), Soukromý vesmír Heleny Třeštíkové (6 634 diváků) a portrét olympioničky Věry Čáslavské Věra 68 od Olgy Sommerové (6 305 diváků). (Zdroj:<http://kultura.idnes.cz>)

Pokročilé procedury

- např. určování koreference

Vzduchový **filtr** patří mezi opomíjené součásti motoru. Je ideální, pokud se k **němu** můžete dostat bez nářadí a pokud **ho** lze jednoduše vyčistit vymytím.

- např. detekce a klasifikace jmenných entit

Vedle staršího snímku **Matěje Mináče Nickyho rodina** patřilyk nejnavštěvovanějším **Pod sluncem tma Martina Marečka** (11487diváků), **Soukromý vesmír Heleny Třeštíkové** (6 634 diváků) a portrét olympioničky **Věry Čáslavské Věra 68** od **Olgy Sommerové** (6 305 diváků).

Kulturní artefakt, příjmení, křestní jméno

Pokročilé procedury

- např. detekce a klasifikace jmenných entit

Demo pro češtinu

- NameTag

<http://ufal.mff.cuni.cz/nametag>

- Czech Named Entity Corpus
-

Aplikační systémy

- např. strojový překlad
 - TectoMT
 - Moses
 - 2013
 - překlad angličtina -> čeština
 - TectoMT + Moses + Depfix* beat Google Translator!
 - Dle automatického i ručního hodnocení

* Pravidlový systém pro oprav chyb v překladu.

Aplikační procedury

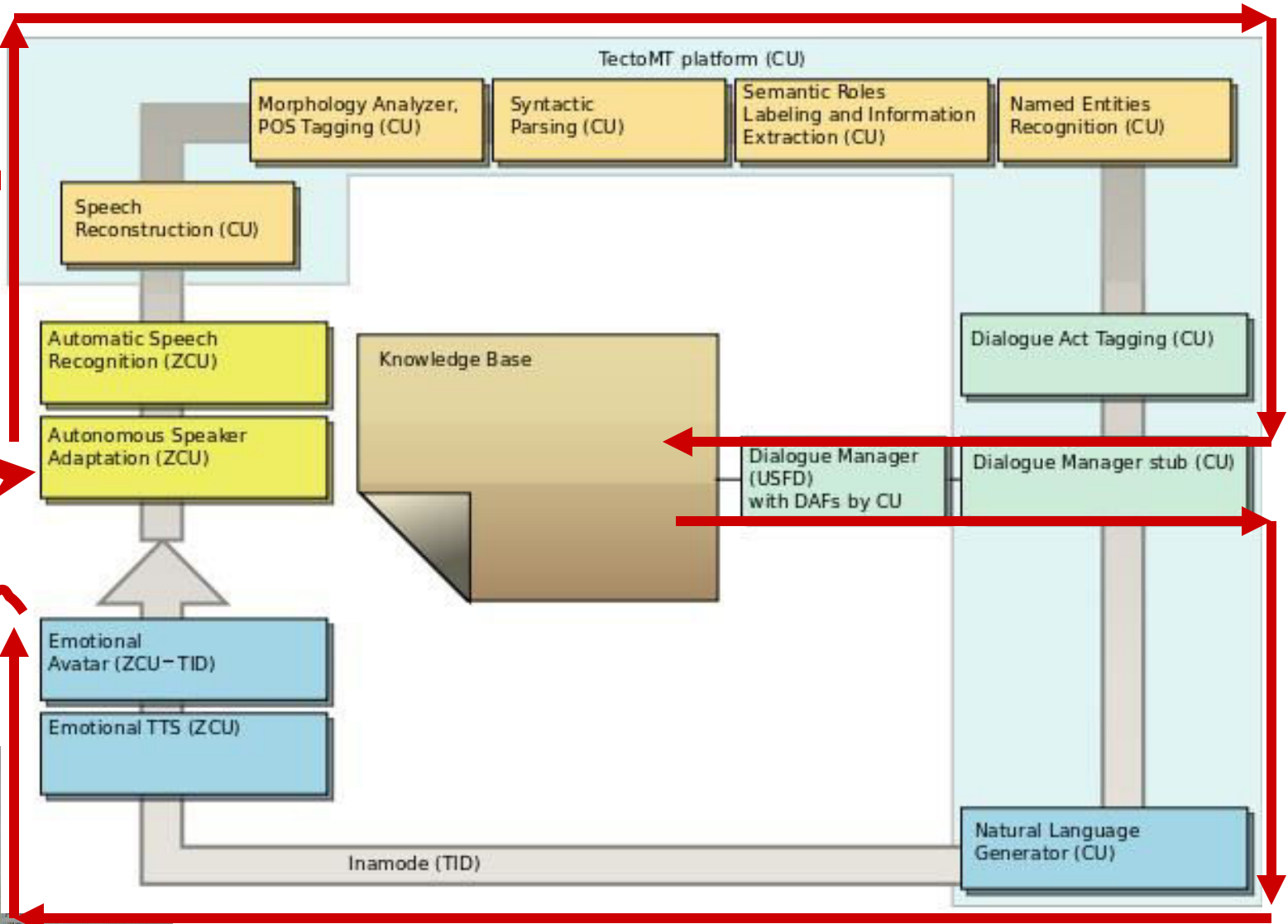
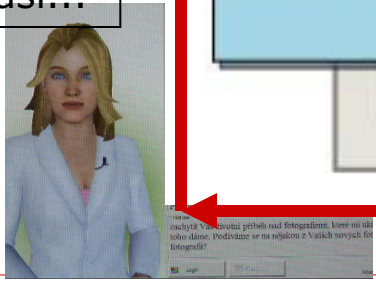
- např. dialogový systém



(Zdroj: <http://www.companions-project.org/>)

To je můj manžel Standa.

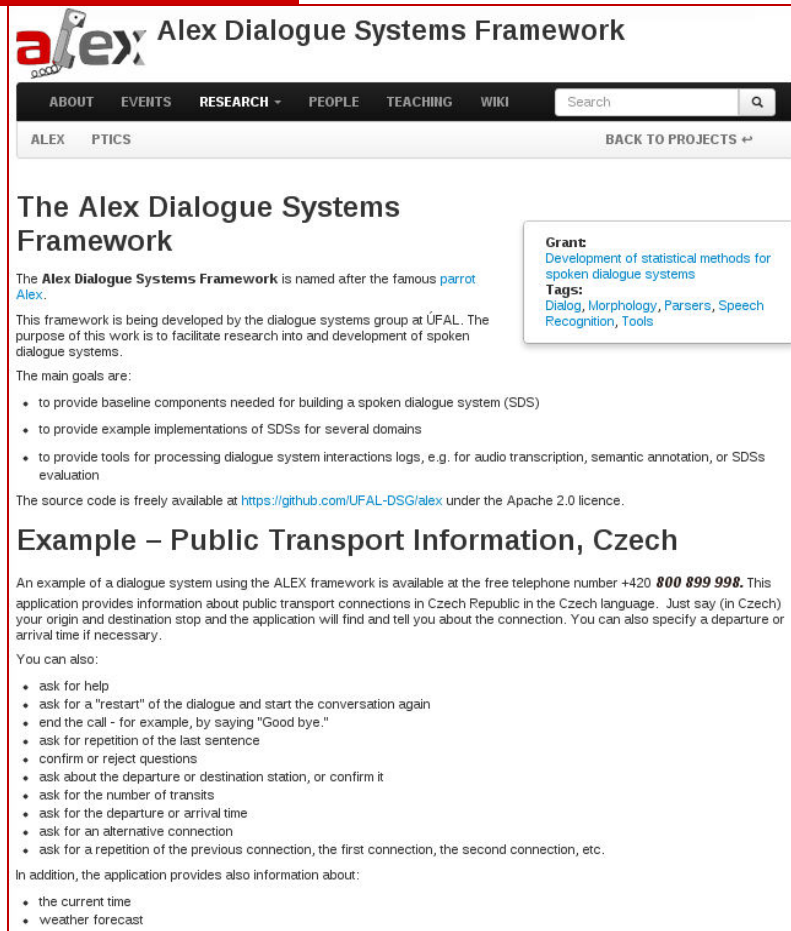
Pěkný obrázek. Moc mu to sluší...



(Zdroj: Hajič, Ircing, 2011)

Aplikační procedury

- např. dialogový systém



The screenshot shows the website for the Alex Dialogue Systems Framework. The header includes the 'alex' logo and the text 'Alex Dialogue Systems Framework'. A navigation menu contains links for 'ABOUT', 'EVENTS', 'RESEARCH', 'PEOPLE', 'TEACHING', and 'WIKI'. A search bar is located on the right side of the menu. Below the menu, there are links for 'ALEX' and 'PTICS', and a 'BACK TO PROJECTS' button. The main content area features the title 'The Alex Dialogue Systems Framework' and a brief description: 'The Alex Dialogue Systems Framework is named after the famous parrot Alex.' It also mentions that the framework is being developed by the dialogue systems group at ÚFAL. A list of main goals is provided, including providing baseline components, example implementations, and tools for processing logs. The source code is noted as being available on GitHub. An example application for public transport information in Czech is described, along with a list of user actions like asking for help, restarting the dialogue, and asking for specific information. The footer mentions additional information provided by the application, such as current time and weather forecast.

alex Alex Dialogue Systems Framework

ABOUT EVENTS RESEARCH PEOPLE TEACHING WIKI Search

ALEX PTICS BACK TO PROJECTS

The Alex Dialogue Systems Framework

The **Alex Dialogue Systems Framework** is named after the famous [parrot Alex](#).

This framework is being developed by the dialogue systems group at ÚFAL. The purpose of this work is to facilitate research into and development of spoken dialogue systems.

The main goals are:

- to provide baseline components needed for building a spoken dialogue system (SDS)
- to provide example implementations of SDSs for several domains
- to provide tools for processing dialogue system interactions logs, e.g. for audio transcription, semantic annotation, or SDSs evaluation

The source code is freely available at <https://github.com/UFAL-DSG/alex> under the Apache 2.0 licence.

Example – Public Transport Information, Czech

An example of a dialogue system using the ALEX framework is available at the free telephone number +420 **800 899 998**. This application provides information about public transport connections in Czech Republic in the Czech language. Just say (in Czech) your origin and destination stop and the application will find and tell you about the connection. You can also specify a departure arrival time if necessary.

You can also:

- ask for help
- ask for a "restart" of the dialogue and start the conversation again
- end the call - for example, by saying "Good bye."
- ask for repetition of the last sentence
- confirm or reject questions
- ask about the departure or destination station, or confirm it
- ask for the number of transits
- ask for the departure or arrival time
- ask for an alternative connection
- ask for a repetition of the previous connection, the first connection, the second connection, etc.

In addition, the application provides also information about:

- the current time
- weather forecast

Grant: Development of statistical methods for spoken dialogue systems
Tags: Dialog, Morphology, Parsers, Speech Recognition, Tools

Nástroje technicky

- Napříč operačními systémy
 - Napříč programovacími jazyky
 - Nástroje samostatně, viz repozitář
 - <https://ufal-point.mff.cuni.cz/xmlui/>
 - Nástroje společně, viz Treex
 - <http://ufal.mff.cuni.cz/treex>
 - platforma pro spouštění nástrojů
 - webového rozhraní Treex::Web
-

Nástroje i data ošetřené licenci

Distribution License

Select license(s):

Specify one or more licenses which will be required for every file in the submission.

- Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0) ▼
- Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)
- Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)
- Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)
- Attribution-NoDerivs 3.0 Unported (CC BY-ND 3.0)
- Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)
- PDT 2.0 License
- Creative Commons - Attribution 3.0 Unported (CC BY 3.0)

(Nejen) česky

□ jazykově nezávislé nástroje

□ jazykově závislá data

The diagram illustrates the AdaBoost algorithm. On the right, a decision tree starts with the root node 'Outlook'. It branches into 'Sunny', 'Rain', and 'Overcast'. 'Sunny' leads to 'Yes'. 'Rain' leads to 'Wind', which further branches into 'Strong' (No) and 'Weak' (Yes). 'Overcast' leads to 'Yes'. On the left, a scatter plot shows data points being separated by a line. A vertical line is labeled 'Margin = 2 / (w1 + w2)'. Below the plot, the text 'AdaBoost [with Freund]' is followed by a list of steps: 'constructing D_t ', ' $D_t(i) = 1/m$ ', and 'given D_t and h_t '. The central formula is
$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } y_i = h_t(x_i) \\ e^{\alpha_t} & \text{if } y_i \neq h_t(x_i) \end{cases}$$

$$= \frac{D_t(i)}{Z_t} \exp(-\alpha_t y_i h_t(x_i))$$
 Below this, it states 'where Z_t = normalization factor' and ' $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) > 0$ '. The final classifier is given as ' $H_{\text{final}}(x) = \text{sign} \left(\sum_t \alpha_t h_t(x) \right)$ '.

Kupujete motorovou pilu(4JŽ)? Deset vychytávek(2MŽ) pro kvalitní stroj(4JM) 25. září(2JS) 2012

Motorové pily(1MŽ) vypadají na první pohled(4JM) všechny téměř stejně, ne nadarmo se ale podstatně liší cenou(7JŽ). Jde nejen o sílu(4JŽ) a objem(4JM) motoru(2JM), ale i o kvalitu(4JŽ) válců(2MM) či robustnost(4JŽ) provedení(2JS).

Former Wimbledon champion (J) Petra (J) Kvitova became the first big name seed(J) to exit the star-studded WTA tournament(J) in Tokyo(J) ...

HamleDT

HamleDT

HamleDT is a compilation of existing dependency treebanks (or dependency conversions of other treebanks), transformed so that they all conform to the same annotation style.

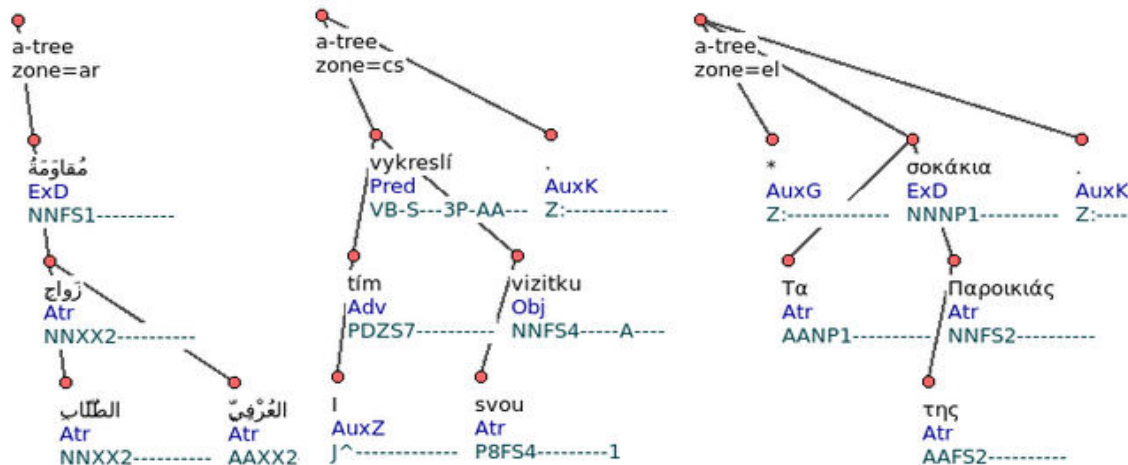
There are as many as 29 treebanks integrated in HamleDT at this moment. A subset of the treebanks whose license terms permit redistribution is available directly for [download from us](#). For the rest, we provide the software that normalizes tree structures in the data obtained by the user from their original providers. All the conversion tools are located in the publicly available repository of the [TreeX](#) (formerly TectoMT) framework in the directory [treeX/devel/hamleDT/](#) (see the [Installation Guide](#) for details about TreeX installation from the SVN repository).

The initial version of HamleDT was described in [Zeman et al.: HamleDT: To Parse or Not to Parse?](#)

Please direct any questions or comments to the following e-mail address: zeman@ufal.mff.cuni.cz.

Tags:

[Annotations](#), [Corpora](#), [Data](#),
[Multilingual](#), [Parsers](#)



O čem

- Představení ÚFAL ✓
 - Banky dat ✓
 - Klasifikace lingvistických nástrojů ✓
 - Nástroje z ÚFAL ✓
 - (Nejen) česky ✓
 - Motivační projekty
 - Odkazy
 - Závěr
-

Motivační projekty: akademické

- Novelizátor zákonů, bakalářská práce, David Mareček, (vedoucí Daniel Zeman), 2006.
 - Úplná znění zákonů z novel zákonů.
 - Novely: omezený (nízký) počet větných konstrukcí, které se opakují. Přesto mohou vzniknout velmi rozmanité věty.
-

Motivační projekty: akademické

- Predikce vývoje ceny ropy na základě textových zpravodajských informací, bakalářská práce, Jan Skalický, (vedoucí Ondřej Bojar), 2010.
 - Předpovídat ceny ropy výhradně z článků, které se jí věnují.
-

Motivační projekty: akademické

- Systém českých číslovek a jejich automatické rozpoznání v textu, diplomová práce, Jan Bureš, (vedoucí Jaroslava Hlaváčová), 2007.
 - Systematické třídění českých číslovek a číselných výrazů.
 - Detekce číslovek v textu a jejich klasifikace dle bodu výše.
-

Motivační projekty: akademické

- Machine learning for Android, bakalářská práce, Juraj Figura (vedoucí Ondřej Bojar), 2011.
 - Knihovna strojového učení (dva algoritmy).
 - Aplikace pro výběr kulturních akcí, které uživatele zajímají.
-

Motivační projekty: akademické

- Sledování témat v elektronickém zpravodajství, Karel Bílek (vedoucí Ondřej Bojar), 2011.
 - Definice témat
 - Jejich automatická detekce (strojové učení)
 - Jak to dělá člověk?
 - Měřitelnost kvality detekce
-

Motivační projekty: akademické

- Interface k online internetovým přenosům v přirozeném jazyce , diplomová práce, Jan Macošek, (vedoucí Jan Hajič), 2007.
 - Hledání živých textových internetových přenosů
 - Jejich převod na přenosy mluvené, které předává elektronický králík, společnost Petra ještě neexistovala
 - Ovládání systému hlasem
-

Motivační projekty: akademické

- Hlasové ovládání televize, bakalářská práce, Patrik Černý, (vedoucí Pavel Straňák), 2012.
 - Hlasové ovládání televize pro jedince se sníženou hybností a s poruchou řeči
-

Motivační projekty: akademické

□ Native Language Identification

- Rozpoznání rodného jazyka autora textu
 - NLI 2013 Shared Task (Hladká, Holub, Kríž, 2013)
-

Motivační projekty: akademické s přesahem do průmyslu

- Sentiment analysis, Kateřina Veselovská a kol.
 - Ruční kategorizace článků ze zdrojů Aktualne.cz, Mall.cz, CSFD.cz, následně učení a automatická kategorizace
 - Rozšíření analýzy sentimentu pro IBM Content Analytics
 - Spolupráce se společnostmi Buzzboot, CaptchaWorks, Wundermann, Zoom International
-

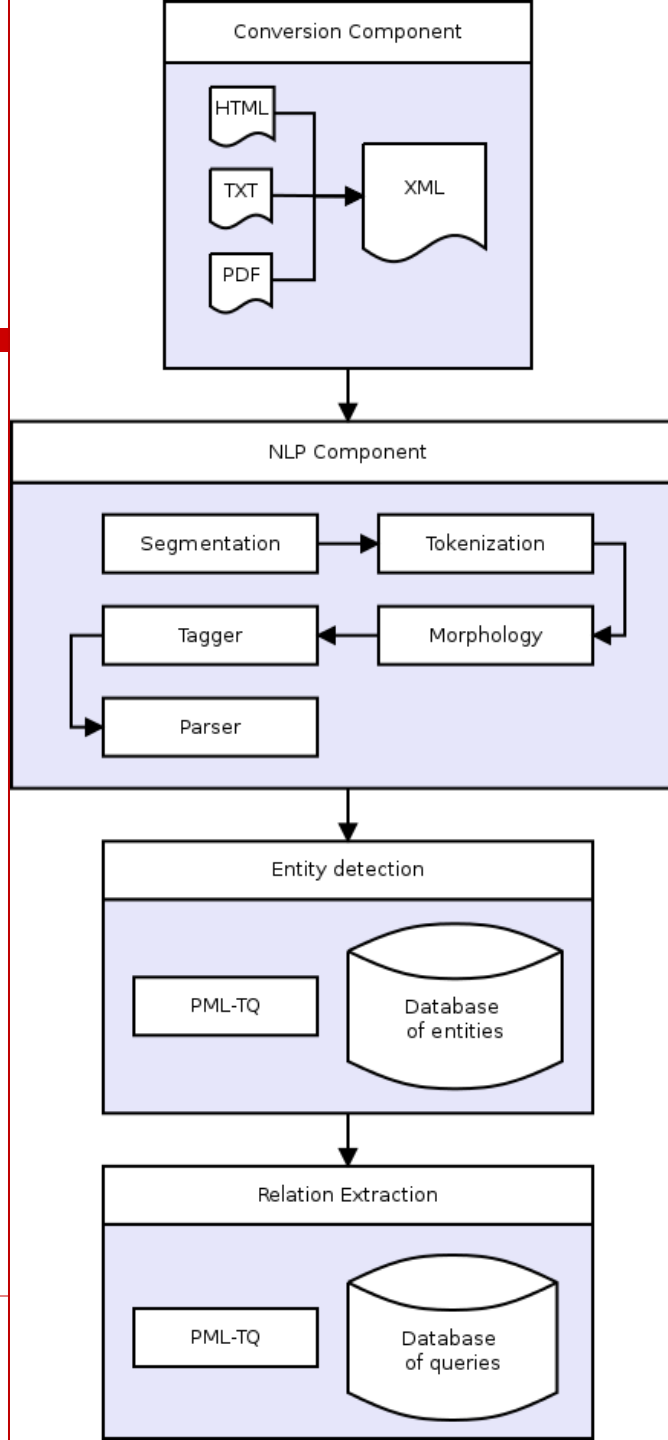
Motivační projekty: akademické s přesahem do průmyslu

- Překladové minislovníky jazykových učebnic nakladatelství Akropolis, Zdeněk Žabokrtský.
 - Autor konkrétní lekce v textu označí slova, která považuje za podstatná/specifická/nová a systém automaticky
 1. pro každé slovo vyrobí slovníkový tvar (v zásadě lemma),
 2. pokud je to sloveso, přidá informaci o vidu,
 3. pokud je to sloveso a má konjugaci se změnou v kmeni, vyrobí i tvar 1os.sg.pres (číst-čtu),
 4. pokud je to substantivum, přidá informaci o rodu,
 5. vyloučí hesla, která se už vyskytla v předcházejících lekcích, a ze zbývajících hesel předpřipraví pro každou stránku učebnice abecedně seřazený minislovníček.
-

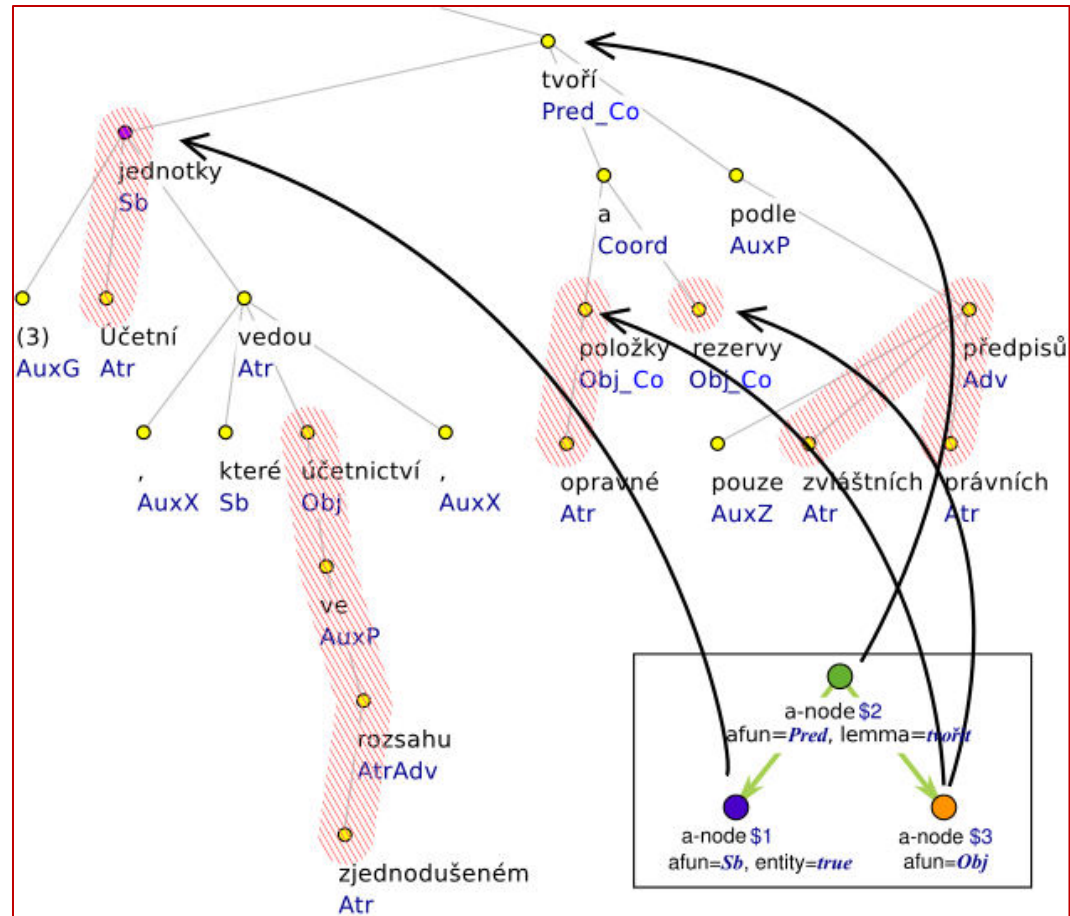
Motivační projekty: akademické s přesahem do průmyslu

- Inteligentní knihovna (INTLIB)
 - Extrakce entit a vztahů mezi nimi
 - Reprezentace ve formě Propojených otevřených dat (Linked Open Data)
-

INTLIB



INTLIB



Odkazy

- Bojar Ondřej. Strojový překlad, *Vesmír 91*, str. 488-490, 2012.
 - Bojar Ondřej. *Čeština a strojový překlad: Strojový překlad našincům, našinci strojovému překladu*. Studies in Computational and Theoretical Linguistics. 2012.
<http://ufal.mff.cuni.cz/books/2012-bojar>
 - Čapek. <http://ufal.mff.cuni.cz/capek/>.
 - Hajič Jan, Hajičová Eva. Some of our friends are statisticians. Ve sborníku Text, Speech and Dialogue Conference (editoři: Václav Matoušek a Pavel Mautner), Plzeň, str. 2-10, 2007.
 - Hajič Jan, Hladká Barbora, Panevová Jarmila. Lingvistika na Matematicko-fyzikální fakultě?, *Vesmír 91*, str. 523-526, 2012.
-

Odkazy

- Hladká Barbora, Martin Holub, Vincent Kríž. Feature Engineering in the NLI Shared Task 2013: Charles University Submission Report. In: *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, Atlanta, Georgia, USA, pp. 232-241, 2013.
 - Jelinek Fred. Some of my friends are linguists. (<http://www.lrec-conf.org/lrec2004/doc/jelinek.pdf>), 2004.
 - Moses. <http://www.statmt.org/moses/>
 - Panevová a kol. Čeština v digitálním věku/The Czech Language in the Digital Age, *Série Bílé knihy/White Paper Series*, Springer, 2012.
-

Odkazy

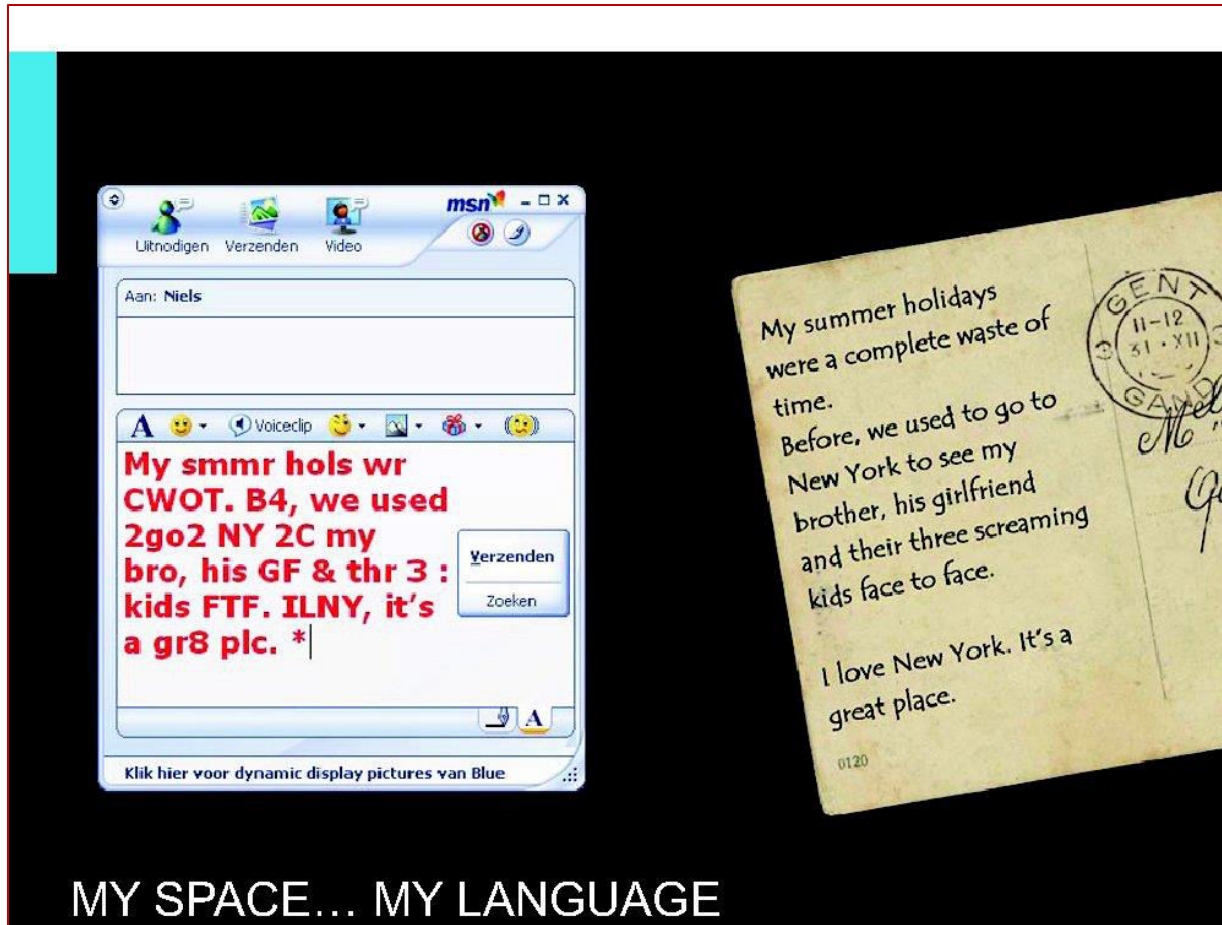
- ❑ Rosa Rudolf. Depfix. <http://ufal.mff.cuni.cz/depfix/>
 - ❑ Sgall Petr. Generativní popis jazyka a česká deklinace. Praha:Academia, 1967.
 - ❑ TrEd. <http://ufal.mff.cuni.cz/tred>
-

Závěr

- Umíme lingvistiku.
- Umíme strojové učení.
- Máme data.
- Máme software.
- Máme zkušenosti.

Ještě něco ...

... neumíme. Zatím?



MY SPACE... MY LANGUAGE