

# Chyťal tloušťě na višni

---

Barbora Hladká

Ústav formální a aplikované lingvistiky

Matematicko-fyzikální fakulta

Univerzita Karlova v Praze

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



Jedna noc s informatikou a matematikou  
MFF UK, 26. ledna 2016

# Chytil tlouště na višni

---



**Tloušť ulovený na třešně**

David Hrstka

# Jak vyzrát na vychytralého tlouště. Vyzkoušejte bahýnko a ovoce

---

**Tloušti** se nenechají jen tak napálit. Pro většinu rybářů představují výzvu: "Kdo s koho..." Vyzkoušejte neobvyklé **návnady**, které by vám mohly pomoci. ... Při **lovu na ovoce\*** je velmi důležité vybrat vhodné lovné místo. Šanci na krásné úlovky budete mít v místech, kde jsou přímo nad hladinou větve ovocných stromů (třešeň, **višeň**)...

---

\*A co lov tloušťů na maso?

# Chytal tloušťě na višni **aneb** Učíme počítače (nejen) česky

---



# Učíme počítače (nejen) česky

---

□ např. větné rozbory

# Chceme-li naučit počítač větné rozборы



...

# Chceme-li ho naučit víc ...

---

## Jak přežít (den)2 na matfyzu

Víte, proč nosí matfyzáci kostkované košile? Aby si určili souřadnice, kde je svědí záda.

Hned vedle chrámu sv. Mikuláše na Malé Straně úřaduje Informatická sekce matematicko-fyzikální fakulty. Polovina jejích studentů si na škole vyláme zuby. Soustíčkem, které zaskočí, bývá často matematická analýza, tak obávaná, že je kultovní. Obsah předmětu zní hrůzostrašně: funkce, limity, derivace, integrály, teorie míry, nekonečné řady a analytické funkce. Prý všechno kromě algebry, analyzuje se cokoliv, co se dá. A dá se toho hodně ...

# Chceme-li naučit počítač

---

- např. opravovat gramatické chyby



# Chceme-li naučit počítač opravovat gramatické chyby

## Jak přežít (den)2 na matfyzu

Víte, proč nosí matfyzáci kostkované košile? Aby si určili souřadnice, kde je svědí záda.

Hned vedle chrámu sv. Mikuláše na Malé Straně úřaduje **I**nformatická sekce **m**atematicko-fyzikální fakulty. Polovina jejích studentů si na škole vyláme zuby. Soustíčkem, které zaskočí, bývá často matematická analýza, tak obávaná, že je kultovní. Obsah předmětu zní hrůzostrašně: funkce, limity, derivace, integrály, teorie míry, nekonečné řady a analytické funkce. Prý všechno kromě algebry, analyzuje se cokoliv, co se dá. A dá se toho hodně ...

# Chceme-li naučit počítač

---

- např. sumarizovat text

# Chceme-li naučit počítač sumarizovat text

---

## Jak přežít (den)2 na matfyzu

Víte, proč nosí matfyzáci kostkované košile? Aby si určili souřadnice, kde je svědí záda.

Hned vedle chrámu sv. Mikuláše na Malé Straně úřaduje Informatická sekce matematicko-fyzikální fakulty ...



Přednáška Matematická analýza I začíná v devět ráno v prvním patře. Matfyzák je chytrý a krásný.


# Chceme-li naučit počítač

---

□ např. překládat

# Chceme-li naučit počítač překládat

---

 Translate

Czech to English (large, generic) ▾

Auto-translate

Víte, proč matfyzáci nosí kostkované košile? Aby si určili souřadnice, kde je svědí záda.

You know, why matfyzáci wears plaid shirts? To determine the coordinates of where the itch is back.

# Chceme-li naučit počítač

---

- např. doplnit obrázek k textu

# Chceme-li naučit počítač doplnit obrázek k textu

Stock Photo Finder

1710 words in czech language — 1,395,677 images found

Add tags...

Suggested tags: **kluka** **pedagoga** **minuty** **derivate** **přednáška**


Matfyzákem se člověk nestává, ale rodí. Matfyzáci to tvrdí s pýchou v hlase. Matematiku, fyziku nebo informatiku můžete studovat i na dalších školách. Matfyzákem se ale stanete jen na jediné z nich – na pražské matematicko-fyzikální fakultě. Zkusila jsem zjistit co tihle lidé ve škole vlastně dělají.

Víte, proč nosí matfyzáci kostkované košile? Aby si určili souřadnice, kde je svěděl záda.

Hned vedle chrámu sv. Mikuláše na Malé Straně úřaduje Informatická sekce matematicko-fyzikální fakulty. Polovina jejích studentů si na škole vyláme zuby. Soustičkem, které zaskočí, bývá často matematická analýza, tak obávaná, že je kultovní. Obsah předmětu zní hrůzostrašně: funkce, limity, derivace, integrály, teorie míry, nekonečné řady a analytické funkce. Prý všechno kromě algebry, analyzuje se cokoliv, co se dá. A dá se z toho hodně.

„Například si představ, že stoupáš pod určitým úhlem do kopce. Jedna věc je spočítat, jak rychle na ten kopec vyjdeš. Druhá je zjistit, kolik energie spotřebuješ. Aby se to dalo derivovat a integrovat,“ zkouší hranice mé představivosti informatik a bývalý matfyzák Martin.

Hafo vzorečků a vět je pro větší legraci u zkoušky rozděleno na praktickou a teoretickou část. K teorii se prý lze propočítat. Právě přednáška z analýzy slibuje



college professor teacher front statistical regression equation formula chalkboard academia academic analysis analyze answer blackboard board calculate college chalk chalkboard class classroom equation formula formulae green high instructor math maths mathematical mathematics problem regression school statistical statistics professor

# Chceme-li naučit počítač

---

- např. vyhledat spojení dle hlasové komunikace s uživatelem



# Chceme-li naučit počítač vyhledat spojení

---

- Volejte **800 899 998** a získejte informace o veřejné dopravě v ČR



# Chceme-li učit počítač

---

Jak na to?

# Pozor na ...

---

## Jak přežít (den)2 na matfyzu

Víte, proč nosí matfyzáci kostkované **košile**? Aby si určili **souřadnice**, kde je svědčí záda.

Hned vedle chrámu sv. Mikuláše na Malé **Straně** úřaduje Informatická sekce matematicko-fyzikální fakulty. Polovina jejích studentů si na škole vyláme zuby. Soustíčkem, které zaskočí, bývá často matematická analýza, tak obávaná, že je kultovní. Obsah předmětu zní hrůzostrašně: funkce, limity, derivace, integrály, teorie míry, nekonečné řady a analytické funkce. **Prý všechno kromě algebry, analyzuje se cokoliv, co se dá.** A dá se toho hodně ...



---

Naučme počítač  
ohýbat česká slova

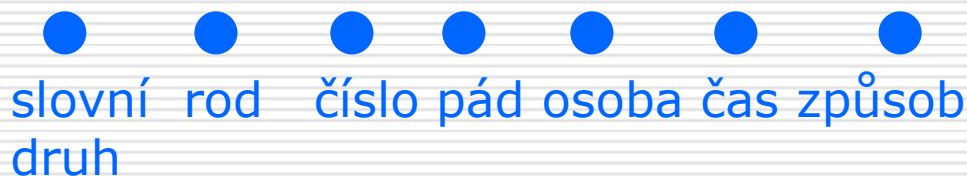
# Učíme počítač ohýbat česká slova

## □ Druhy slov

- 1 ■ Podstatná jména
- 2 ■ Přídavná jména
- 3 ■ Zájmena
- 4 ■ Číslovky
- 5 ■ Slovesa
- 6 ■ Příslovce
- 7 ■ Předložky
- 8 ■ Spojky
- 9 ■ Částice
- 10 ■ Citoslovce

## □ Mluvnické kategorie

- rod M,I,F,N
- číslo S,P
- pád 1,2,3,4,5,6,7
- osoba 1,2,3
- čas R,P,F
- způsob N,I,C
- ...



# Učíme počítač ohýbat česká slova

---

## □ Příprava učebnice tvaroslovných rozborů

slovo	značka
-----	
Matfyzák	1MS1---
je	5MS-3RN
chytrý	2MS1---
a	8-----
krásný	2MS1---

# Učíme počítač ohýbat česká slova

---

## **Příprava učebnice** tvaroslovných rozborů

1. Vybrat věty

2. Udělat rozbory

Ne na papíře!

Např. v editoru Čapek



# Učíme počítač ohýbat česká slova

---

## □ Učebnice

Honza 1MS1--- měl 5MS-3PN  
pak 6----- příležitost 1FS4---  
promluvit 5----- si 3—3--- s 7-----  
ním 3MS7--- mezi 7----- čtyřma 4-----  
očima 1NP7---.

Moje 3XS1--- sestra 1FS1--- teď 6-----  
myje 5FS-3RN talíře 1MP4--- od 7-----  
oběda 1IS2--- ...



# Učíme počítač ohýbat česká slova

## □ Učebnice v tabulkách

slovo se značkou	kolikrát	dvojice značek za sebou	kolikrát
Honza 1MS1---	10		
měl 5MS-3PN	15	1MS1--- 5MS-3PN	10
pak 6-----	30	5MS-3PN 6-----	15
příležitost 1FS4---	10	7 3MS7---	2
příležitost 1FS1---	5	...	
...			
se 7-----	40		
myje 5MS-3RN	1		
myje 5FS-3RN	1		
...			

# Počítač už tvarosloví umí

## učebnice

Honza 1MS1---	10x
měl 5MS-3PN	15x
pak 6-----	30x
příležitost 1FS4---	10x
příležitost 1FS1---	5x
se 7-----	40x
myje 5MS-3RN	1x
myje 5FS-3RN	1x
1MS1--- 5MS-3RN	10x
5MS-3RN 6-----	15x
7 3MS7---	2x

Honza 1MS1---

1 myje 5MS-3RN

2 nádobí 6-----

Honza myje nádobí



učebnice

nádobí a předložka?



# Počítač už tvarosloví umí

---

- ❑ Učebnice (asi) nikdy nebude obsahovat všechny možné věty
- ❑ Vždy bude pouze vzorkem
- ❑ Proto počítač bude dělat chyby, ale
  - čím větší učebnice, tím méně chyb


# Toto byl lehounký úvod do strojového učení s učitelem

---

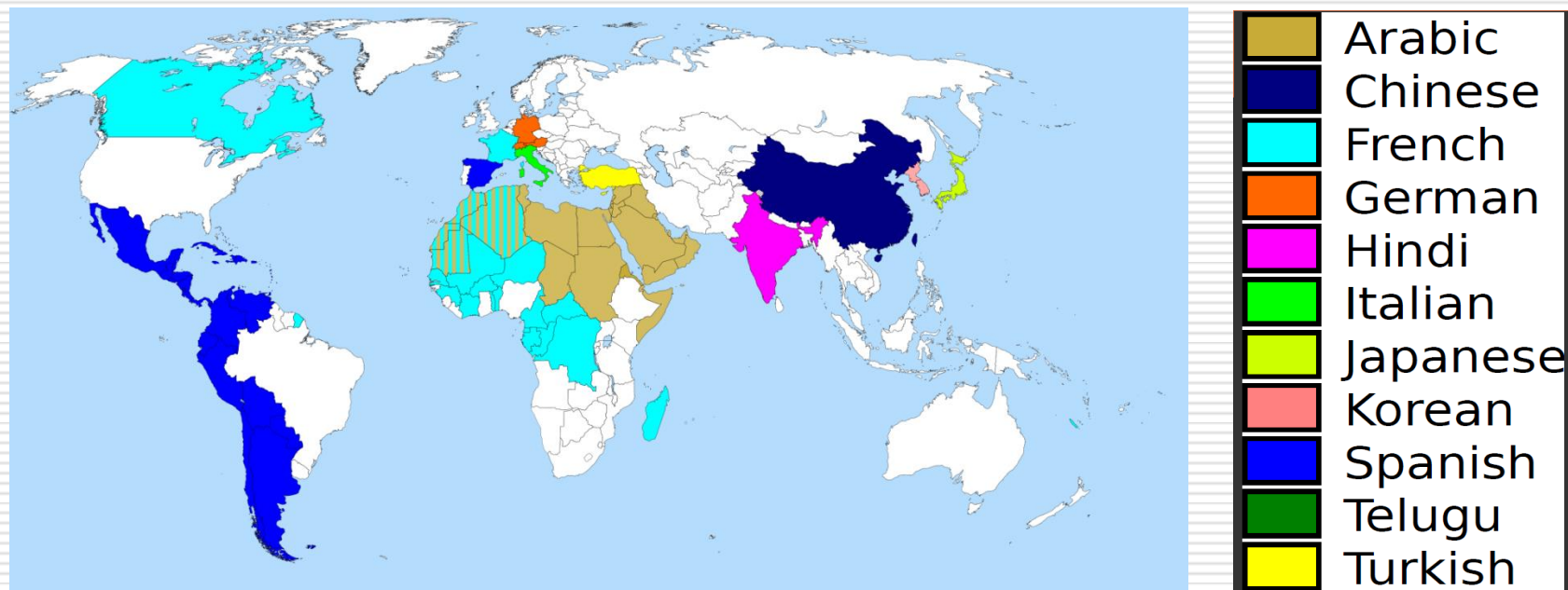
1. Ukázat počítači **příklady** toho, co chceme, aby uměl, tj. připravit učebnici
2. Říct počítači, na co se má v učebnici dívat, tzv. **příznaky**
  - ❑ např. jak jdou tvaroslovné značky za sebou
3. Převést příznaky do **matematických** struktur
  - ❑ např. relativní frekvence

# Učíme počítače (nejen) česky

---

- Učíme počítač ohýbat anglická slova 
- 1. Připravíme učebnici v editoru Čapek
  - anglické věty v Čapkově označujeme
- 2. Dále postupujeme jako pro češtinu

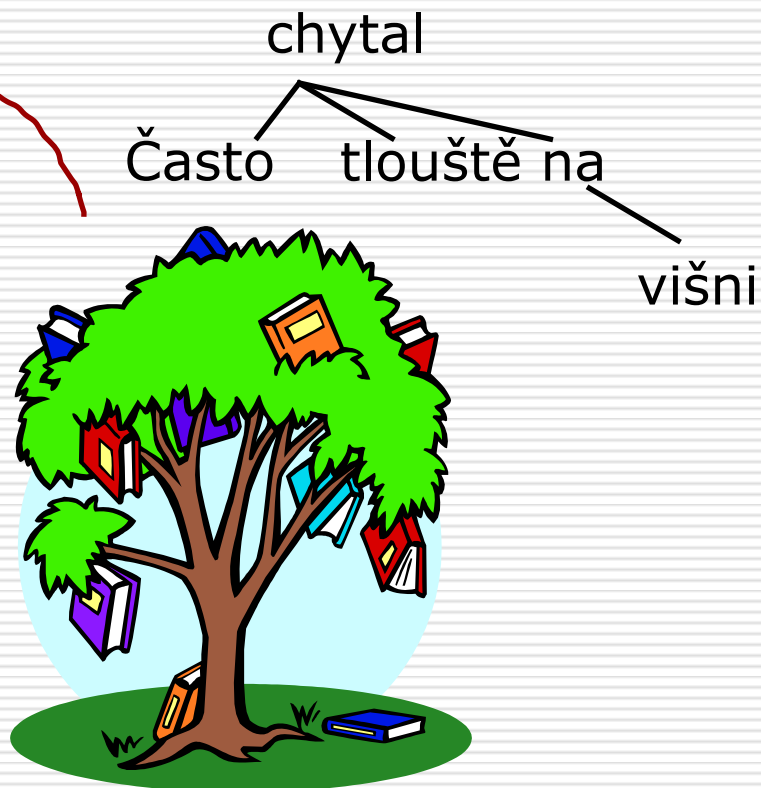
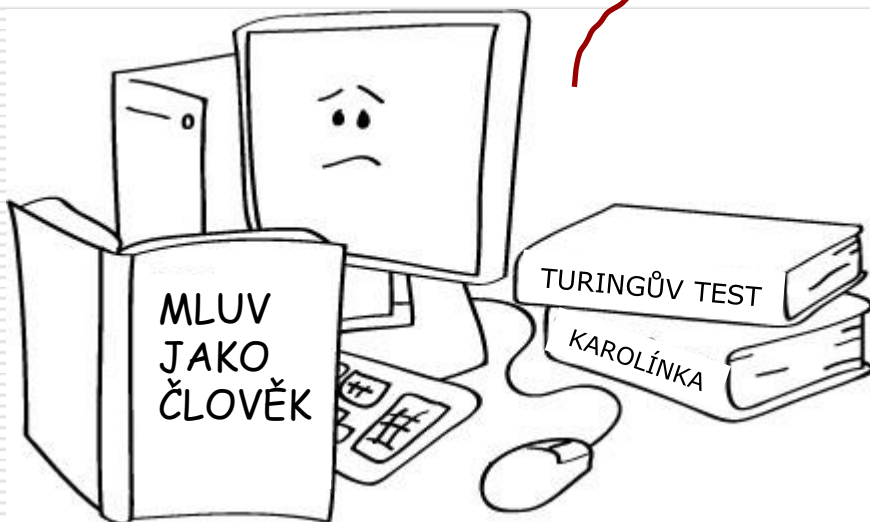
# Učíme počítače (nejen) česky



Skupina studentů, rodilých mluvčích 11 jazyků, píše anglickou esej. Anonymně. Pro každou esej určete rodný jazyk jejího autora.

se věnuje automatické komunikaci člověka s počítačem v přirozeném jazyce.

UČENÍ



UNIVERZITA KARLOVA V PRAZE

matematicko-fyzikální fakulta



Lingvistika  
na matfyzu



I ♥  
50<sub>s</sub>

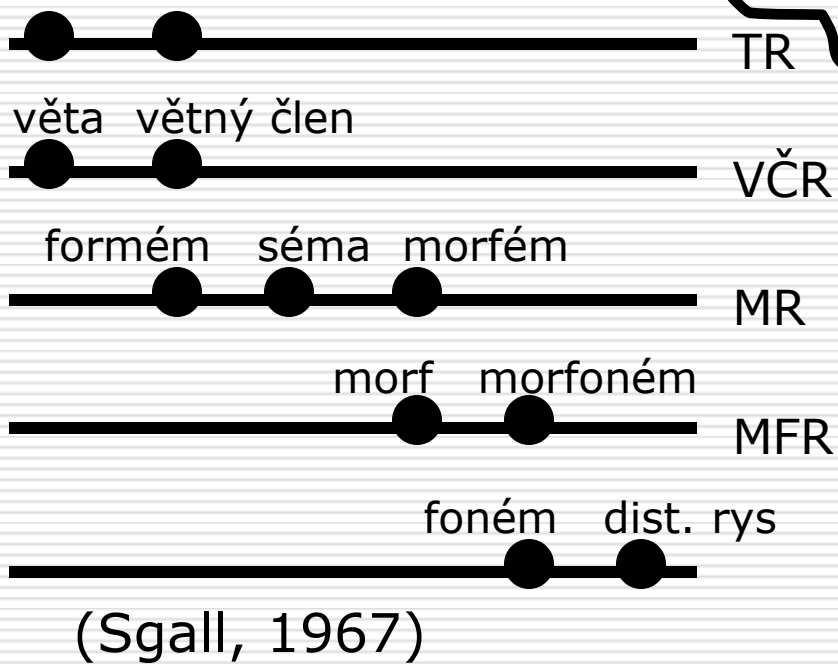
I ♥  
60<sub>s</sub>

I ♥  
90<sub>s</sub>



formální lingvistika

aplikovaná lingvistika\*



data

nástroje

text a řeč



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

# garantuje studijní plán I3

## Matematická lingvistika

### Povinné předměty

kód	Předmět
<a href="#">NPFL067</a>	Statistické metody zpracování přirozených jazyků I
<a href="#">NPFL092</a>	Technologie pro NLP
<a href="#">NPRG027</a>	Zápočet k projektu
<a href="#">NPRG023</a>	Softwarový projekt
<a href="#">NTIN090</a>	Základy složitosti a vyčíslitelnosti
<a href="#">NTIN066</a>	Datové struktury I
<a href="#">NSZZ023</a>	Diplomová práce I
<a href="#">NSZZ024</a>	Diplomová práce II
<a href="#">NSZZ025</a>	Diplomová práce III

- ❑ Mgr.
- ❑ Ph.D.

### Povinně volitelné předměty

kód	Předmět
<a href="#">NPFL068</a>	Statistické metody zpracování přirozených jazyků II
<a href="#">NPFL083</a>	Lingvistická teorie a gramatické formalismy
<a href="#">NPFL070</a>	Zdroje lingvistických dat
<a href="#">NPFL075</a>	Pražský závislostní korpus
<a href="#">NPFL054</a>	Úvod do strojového učení
<a href="#">NPFL093</a>	Aplikace NLP
<a href="#">NPOZ009</a>	Odborné vyjadřování a styl
<a href="#">NPFL087</a>	Statistický strojový překlad
<a href="#">NPFL094</a>	Morfologická a syntaktická analýza
<a href="#">NPFL006</a>	Úvod do formální lingvistiky
<a href="#">NPFL095</a>	Moderní metody v počítačové lingvistice I
<a href="#">NPFL038</a>	Základy rozpoznávání a generování mluvené řeči
<a href="#">NPFL082</a>	Informační struktura věty a výstavba diskurzu
<a href="#">NPFL096</a>	Komputační morfologie
<a href="#">NPFL079</a>	Algoritmy rozpoznávání mluvené řeči
<a href="#">NPFL099</a>	Statistické dialogové systémy
<a href="#">NPFL106</a>	Obecná lingvistika
<a href="#">NPFL103</a>	Vyhledávání informací

# Učíme počítače

---

- metody strojového učení jsou oborově nezávislé
- lékařství, bankovníctví, zábavní průmysl, ...

## Matfyz pomáhá dolovat data

19. 12. 2014 | Fakulta | článek



Ústav formální a aplikované lingvistiky (ÚFAL) MFF UK a společnost DATLOWE, která se zabývá vývojem řešení pro inteligentní zpracování textů a chytrým vyhledáváním v těchto textech, se dohodly na oficiální spolupráci. DATLOWE bude přenášet výsledky výzkumu ÚFAL do komerčních aplikací v různých odvětvích.

# Odkazy

---

- Doležalová Marie. [Přízemní herečka na matfyzu](#), 8. ledna 2015, Reportér
- ÚFAL – [popularizace](#)