

Z Wikipedie, otevřené encyklopedie
ikona

Tento článek potřebuje 

[vylepšíte](#)

[Vzhled a styl](#) [Encyklopedicky styl](#) [Odkazy](#)

Článek je o některých nejdůležitějších kapitolách zcela nezpracovaný, ale chybí i významné části.

[Přesněji](#) [Zpracování přirozeného jazyka](#)

[vylepšíte](#)

Článek je převeden do

minulého času a případně přesuňte do části článku věnované dějinám.

Počítačové **zpracování přirozeného jazyka** (anglicky **Natural language processing, NLP**) je soubor technik na pomezí ([počítačové](#)) [lingvistiky](#), [informatiky](#) ([umělé inteligence](#)), popř. též [akustiky](#) a dalších. Věnuje se analýze či generování textů nebo mluveného slova, které vyžadují určitou (ne absolutní) míru porozumění [přirozenému jazyku](#) strojem.

Aplikace NLP jsou např. [strojový překlad](#), [odpovídání na otázky](#) ([en:Question answering](#)), [dolnávání z textu](#) (i výťah, z textu, [en:Automatic summarization](#)), a automatická [korektura textu](#) či chatboti.^[1]

Mezi úkoly, které přispívají k řešení těchto problémů, patří mj. [extrakce informací](#), [strojový překlad](#), [generování přirozeného jazyka](#) ([en:Natural language generation](#)) a [rozpoznávání a syntéza řeči](#) (text-to-speech).

Zpracování přirozeného jazyka má tři historické fáze:

1. symbolické NLP (50. až 90. léta 20. stol.)
2. statistické NLP (90. léta 20. stol. až 00. léta 21. stol.)
3. neuromové NLP (počátky 2000, rozvoj po roce 2010 díky [Tomáši Mikolovi](#) a programu [Word2vec](#))

Tradiční (strukturalistický) přístup

Související informace najeznete také v článku [Lingvistická analýza](#).

Počítačové zpracování přirozeného jazyka je interdisciplinární obor. Tento obor mimo jiné zkoumá přirozený jazyk jako matematický systém. Přirozený jazyk jako hlavní nástroj lidské komunikace je za pomocí aktivní účasti uživatele transformován prostřednictvím speciálních technologií ve formální jazyk (*interlingua*; logická reprezentace významu), který je vhodný pro sémantickou prezentaci. Vyjadřuje význam jednotlivých prvků přirozeného jazyka, který je počítačově zpracován. Základem je [algoritmus](#) popis jednotlivých rovin přirozeného jazyka, který je zároveň nezávislý na konkrétním jazyku.[\[zdroj?\]](#) Základem interakce člověka s počítačem je dotazovací jazyk, u kterého je odstraněna víceznačnost jednotlivých prvků na všech úrovních. S ohledem na přesnost a jednoznačnost reprezentace samotného významu je nutná existence samostatné reprezentace pro každý významový prvek přirozeného jazyka. Struktury formálního jazyka jsou na konkrétních jazycích nezávislé.

Při zpracování jazyka bylo nutné vymezit pravidla tzv. jazykové roviny. Každá jazyková rovina je pak určena svým hlavním jazykovým prvkem nebo třídou prvků, které jsou pro konkrétní rovinu typické. Každá rovina má vstupní a [výstupní reprezentaci](#).

- **Fonetická rovina:** Výstupem fonetické roviny je zpracování posloupnosti fónů ve fonetické abecedě.
- **Fonologická rovina:** Výstupem fonologické roviny je posloupnost

[zdroj?](#)

Je to proces přiřazení selekčních obrazů dokumentům nebo dotazům. Selekčním obrazem se rozumí výraz nebo množina výrazů určitého selekčního jazyka, např. všechna podstatná jména, předem daná podstatná jména, výrazy ve tvaru „podstatné jméno – přídavné jméno“ apod.

Klíčovým problémem automatického indexování bývá určení, která slova textu nejlépe charakterizují jeho celkový obsah.

Lingvistické problémy automatického indexování:

*Tvarosloví
Synonymie*

Homonymie

- **Syntaktická analýza**

- Jednotlivé indexační termíny jsou složeny do sousloví.
- Na základě předložek či spojek jsou vyhledávána spojení jmenných frází.
- Úprava vah termínů – sečtení vah jednotlivých výskytů výrazů do vaz výrazů jako takových.
- ~~Termíny mohou být používány i v disciplínách s ustálenou terminologií.~~

Též je možné automaticky indexovat tezaurus.

Automatické referování

- **Referát** (abstrakt) je uváděn jako jeden z možných výstupů intelektuálního procesu nazývaného informační analýza dokumentů.
- Referát je zkrácený výklad obsahu dokumentu (nebo jeho části) s hlavními věcnými údaji a závěry, který zdůrazňuje nové poznatky a umožňuje rozhodnout se o účelnosti studia původního dokumentu. Výklad obsahu musí být stručný a přesný.
- **Automatické referování extrahuje vhodný počet vět, které nejlépe vystihují, co text přináší nového**

- Syntaktické – chyby v použití slov – vynechání slova, přidání nadbytečného slova, nesprávná kombinace tvarů slov, záměna slovního tvaru jiným slovním tvarem a chyby v interpunkci.
 - Stylistické chyby: oprava spočívá v automatické detekci často novějšími algoritmy a metodami, které jsou využívány v moderních aplikacích, využívají nové algoritmy (je možné je upravit tak, aby zlepily výkon) a využívají různé typy dat.

Міжнародні зв'язки України з Європою та світом

© 2018 Pearson Education, Inc. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part.

- = **specific** **method** **for** **multiple** **uses**: **is** **designed** **in** **such** **a** **way** **that** **it** **can** **be** **adapted** **to** **several** **purposes**
= **more** **flexible**, **more** **versatile**, **more** **efficient**, **more** **convenient**, **by** **a** **single** **method** **with** **one** **set** **of** **tools**
 - = **multiple** **uses**, **when** **using** **multiple** **methods**, **provide** **the** **best** **possible** **results** **for** **each** **use** **case**

卷之三

*בְּרֵבָדָה וְבְרֵבָדָה

"*CEPTELLIC*"

Masarykova univerzita, Filozofická fakulta, Ústav jazykovědy a baltistiky.
Vedoucí práce PhDr. Ondřej Šefčík, Ph.D.

- HAJIČOVÁ, Eva, PANEVOVÁ, Jarmila, SGALL, Petr. Úvod do teoretické a počítacové lingvistiky: I. svazek – Teoretická lingvistika. Praha: Karolinum, 2003. 156 s. [ISBN 80-246-0470-1](#).
- STROSSA, Petr. Vybrané kapitoly z počítacového zpracování přirozeného jazyka. 1. vyd. Opava: Slezská univerzita v Opavě, Filozoficko-přírodovědecká fakulta, Ústav informatiky, 1999. 277 s. [ISBN 80-7248-041-3](#).
- HABROVSKÁ, Pavlína. Krátce o zpracování přirozeného jazyka. Inflow: information journal [online]. 2010, roč. 3, č. 9 [cit. 2013-04-24].
Dostupný: [http://www.inflow.cz/kratce-o-zpracovani-prizeneho-jazyka](http://www.inflow.cz/kratce-o-zpracovani-prirozeneho-jazyka)

Archivováno 4. 3. 2016 na [Wayback Machine](#). ISSN 1802-9736



<http://web.archive.org/web/20050520022902/http://www.muni.cz/~habrovska/2003/preservativy.html#text>

Seznamující články

- [Počítacová lingvistika](#)

Externí odkazy

- Obrázky, zvuky či video z tématu [zpracování přirozeného jazyka](#) na [Wikimedia Commons](#)

Podíly: Jazyk

https://commons.wikimedia.org/w/index.php?title=File:Obr%C4%9Bzec_1.jpg&oldid=1500000

Kategorie:

- [Zpracování přirozeného jazyka](#)
- [Počítačová lingvistika](#)
- [Digitální humanitní vědy](#)

Skryté kategorie:

- [Údržba:Články k úpravě](#)
- [Údržba:Články k aktualizaci](#)
- [Údržba:Články obsahující nedoložená tvrzení](#)
- [Údržba:Články obsahující odkazy na nedostupné zdroje](#)
- [Monitoring:Články s identifikátorem NKC](#)
- [Monitoring:Články s identifikátorem PSH](#)
- [Monitoring:Články s identifikátorem LCCN](#)
- [Monitoring:Články s identifikátorem NDL](#)
- [Monitoring:Články s identifikátorem NLI](#)
- [Portál Jazyk/Zapojené články](#)

Hledání

Speciální:Hledání

Hledat

Zpracování přirozeného jazyka

71 jazyků

[Přidat téma](#)