

# PB095 – Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta informatiky  
Masarykova univerzita

podzim 2022

# Obsah

## 1 Dotazovací systémy

# Dotazovací systémy (Chat Bots)

- Dotazovací systém – konverzační software určený k rozmluvě s uživatelem.
- Druh dialogových systémů.
- Téměř neomezené oblasti použití:
  - automatizované systémy telefonické podpory, dotazování nad strukturovanými daty (db)
  - výukové účely – jazyky, soft skills
  - hry, zábava
  - ...

# Artificial Intelligence Markup Language

- Jazyk na bázi XML
- Autor návrhu Dr. Richard Wallace
- vyvíjen komunitou v letech 1995 — 2002
- stal se základem pro rozšíření chatbota Eliza zvaného A.L.I.C.E.
  - Artificial Linguistic Internet Computer Entity
- popisuje znalostní bázi pro dotazovací systémy.
- Více viz úvod do problematiky na Pandorabots.

# AIML - základní jednotky znalostní databáze

- AIML popisuje třídy objektů dat a částečně popisuje chování programů, které je zpracovávají.
- AIML objekty dat se skládají z jednotek zvaných témata (topics) a kategorie (categories)
  - obsahují buď strukturované nebo nestrukturovaná data
- strukturovaná data obsahují buď textová data nebo AIML elementy
  - AIML elementy obsahují znalost ve formě dvojic stimul - odpověď

# AIML - použití pro tvorbu dotazovacích systémů

- Pro tvorbu dotazovacích systému se využívají klíčová slova *Keywords*
- Struktura:
  - kategorie (category)
  - obsahují vzor (pattern) dotazu a šablonu odpovědi (template)
  - viz ukázky v adresáři data/aiml/
- Jedná se o značkový jazyk – nutnost interpretace.
  - Platforma Pandorabots
  - seznam otevřených interpretů na GitHub.com
  - seznam otevřených interpretů na sourceforge.net
  - Platforma Bot Libre
  - ...

# AIML

## Kategorie

- Základní jednotka uchování znalostí v AIML.
- Každá kategorie se skládá z:
  - vstupního dotazu
  - výstupní odpovědi – nazývá se vzor (template)
  - volitelného kontextu

# Zpracování

- Dokument je zpracováván rekurzivně.
- Využívají se následující způsoby rekurze:
  - symbolická redukce – zjednodušuje komplexní gramatické zápisy,
  - rozděl a panuj – rozdělí vstup do dvou a více částí a zkombinuje odpovědi na ně,
  - synonyma – umožňuje různé způsoby řečení dané odpovědi,
  - gramatické zpracování – kontrola vstupu resp. gramatické opravy,
  - podmíněčné zpracování – může být implementováno s elementem `<srail>`, nebezpečí nekonečné rekurze,
  - kombinace předchozích.



# Ukázky

Viz adresář data/aiml/ ve studijních materiálech z přednášky.