

PB095 – Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta informatiky
Masarykova univerzita

podzim 2023

- 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
 - ...

- 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 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

- Pro tvorbu dotazovacích systémů 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
 - ...

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

- 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 řízení dané odpovědi,
 - gramatické zpracování – kontrola vstupu resp. gramatické opravy,
 - podmíněčné zpracování – může být implementováno s elementem `<srail>`, nebezpečí nekonečné rekurze,
 - kombinace předchozích.

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