

Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,
Brno

jaro 2021

Speech Synthesis Markup Language

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Značkovací jazyk, určený pro zvýšení kvality syntézy řeči.
- Standard W3C.
- Aktuální verze 1.1 (září 2010).
- Vychází ze specifikací JSGF a JSML (Sun Microsystems).
- Vychází z něj jazyk SABLE.

- Vytvořit standard pro značkování prozodických jevů mluvené řeči.
- Jazyk by měl být podporován různými TTS.
- Zvýšení kvality syntézy řeči pomocí ovládnání:
 - výslovnosti
 - hlasitosti
 - průběhu základního hlasivkového tónu
 - rychlosti
 - ...

- Kořenový element – *speak*.
- Atributy:
 - *version* – použitá verze SSML (aktuálně 1.0, 1.1)
 - *xml:lang* – přirozený jazyk použitý obsahem tohoto elementu.
- Může obsahovat elementy:
 - výslovnost – *lexicon*, *phoneme*, *say-as*
 - struktura – *p*, *s*
 - prozódie – *emphasis*, *prosody*, *voice*, *break*
 - ostatní – *audio*, *meta*, *metadata*, ...

- Element *p*:
 - Ohraničuje odstavec.
 - Atribut – *xml:lang* – přirozený jazyk tohoto odstavce.
 - Může obsahovat elementy:
 - audio, break, emphasis, mark, phoneme, prosody, say-as, sub, s, voice.
- Element *s*:
 - Ohraničuje větu.
 - Atribut – *xml:lang*.
 - Může obsahovat elementy:
 - audio, break, emphasis, mark, phoneme, prosody, say-as, sub, voice.

- Element *lexicon*:
 - Vkládá odkaz na lexikon výslovnosti (více viz).
 - Atributy:
 - *uri* – URI odkazující na soubor s lexikonem výslovnosti.
 - *type* – mime typ odpovídající typu lexikonu.
- Element *phoneme*:
 - Obsahuje fonetický přepis textu.
 - Atributy:
 - *alphabet* – použitá fonetická abeceda (IPA, případně ještě x-JEITA, x-JEITA-2000 – japonské fonetické abecedy, většinou znaková využívá znakovou sadu UNICODE).
 - *ph* – fonetický přepis textu uzavřeného do tohoto elementu.

- Element *say-as*
 - Popisuje jakým způsobem se má daný text vyslovovat (datum, množství peněz, ...).
 - Atributy:
 - *interpret-as* – o jaký typ dat se jedná (currency, date, ...)
- Element *sub*:
 - Umožňuje definovat aliasy pro daný text (např. přepis zkratek, ...).
 - Atributy:
 - *alias* – alias pro text, který je obsahem daného elementu.

- Umožňuje popsat prozodické vlastnosti promluvy počítače.
- Do jaké míry budou obsaženy ve výsledné řeči závisí na podpoře v konkrétním TTS.
- *voice* – umožňuje ovlivňovat některé charakteristiky použitého hlasu:
 - pohlaví – atribut *gender* – povolené hodnoty male, female, neutral
 - věk – atribut *age* – kladné celé číslo udávající věk mluvčího.
 - variantu – atribut *variant* – kladné celé číslo, které značí která varianta daného hlasu se má použít – musí být podpora v TTS
 - jazyk – atribut *xml:lang* – pokud je dostupný použije se tento jazyk, jinak by se měl použít jiný, co nejbližší jazyk.

- Element *emphasis*
 - daný text by se měl říct s důrazem – pomocí přízvuku, hlasitosti, ...
 - míra důrazu popsána atributem *level* – hodnoty jsou none, reduced, moderate, strong.
- Element *break*
 - výsledkem by měla být pauza v řeči
 - její síla (výraznost) je ovlivněna atributem *strength* – jedna z hodnot none, x-weak, weak, medium, strong, x-strong
 - doba trvání atributem *length* – čas ve formátu shodným s formátem použitým ve specifikaci CSS2.

- Element *prosody* – umožňuje ovlivňovat prozodické charakteristiky promluvy, která je jeho obsahem. Je nutná podpora na straně TTS::
 - F_0 (atribut *pitch*) – hodnota může udávat výšku v Hz, relativní změnu a nebo některou z hodnot x-low, low, medium, high, x-high a nebo default.
 - Průběh F_0 (atribut *contour*) – hodnotou jsou mezerou oddělené uspořádané dvojice (time, pitch), kde time je vyjádřen pomocí procentuálně a výška stejným způsobem jako u atributu pitch.
 - Rozsah F_0 na daném úseku (atribut *range*) – hodnota buď rozsah v Hz, nebo relativní rozsah a nebo jedna z hodnot x-low, low, medium, high, x-high a default.
 - Doba trvání (atribut *duration*) – jak dlouho se má daný text číst (ms resp. s).
 - Hlasitost (atribut *volume*) – hlasitost proslovu – hodnoty v intervalu 0.0 – 100.0 nebo jedna z silent (=0.0), x-soft, soft, medium, loud, x-loud a nebo default (=100.0).

Pronunciation Lexicon Specification

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Standard W3C VoiceBrowser Activity.
- Aktuální verze 1.0 (říjen 2008).
- Popisuje jazyk pro tvorbu lexikonů výslovnosti použitelných při syntéze a rozpoznávání řeči.
 - výslovnost cizích slov
 - výslovnost zkratek
 - ...

Pronunciation Lexicon Specification

Struktura slovníku

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Kořenový element *lexicon*:
 - Atributy:
 - *xml:lang* – přirozený jazyk dokumentu
 - *version* – aktuální verze 1.0
 - *xmlns* – musí být propojen se jmenným prostorem <http://www.w3.org/2005/01/pronunciation-lexicon>
 - *alphabet* – použitá fonetická abeceda.
 - Obsah:
 - Element *metadata* – informace o dokumentu.
 - Element(y) *lexeme* – jednotlivé položky slovníku.
- Element *lexeme*
 - Atribut *role* – popisuje mluvnické kategorie slova, tak aby bylo možné zvolit nejvhodnější výslovnost (např. sloveso vs. podstatné jméno – red vs. red)
 - Obsah:
 - Element(y) *grapheme* – psaná podoba slova.
 - Element(y) *phonemes* – výslovnost(i) slova.
 - Element(y) *alias* – v případě, že *grapheme* obsahuje zkratku, tak její plný tvar (např. ČR – Česká republika).

- Element *phoneme*
 - Atribut *preferred* – pokud je u pojmu uvedeno více různých výslovností, tato je preferovaná.
 - Obsah – fonetický zápis výslovnosti pojmu.
- Element *alias*
 - Atribut *preferred* – pokud je u pojmu uvedeno více různých výkladů, toto je preferovaný.
 - Obsah – plný zápis zkratky.
- Více viz specifikace.

State Chart XML

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Standardu W3C (září 2015)
- Značkovací jazyk pro popis konečných automatů používaných v dialogových rozhraních.
- Standard pro řídicí jazyk ve:
 - VoiceXML 3.0 (aktuálně ve vývoji)
 - budoucích verzích CCXML
 - jazyce pro popis multimodálních rozhraní.

- Konečný automat $(S, \Sigma, \phi, q_0, Q)$:
 - S – konečná neprázdná množina stavů
 - Σ – vstupní abeceda
 - ϕ – přechodová funkce $S \times \Sigma \rightarrow S$
 - q_0 – počáteční stav
 - Q – množina koncových stavů.
- Zápis pomocí SCXML:
 - stav – element *state*:
 - povinný atribut *id* – název stavu
 - počáteční stav – obsahuje dceřiný element *initial*
 - koncový stav – obsahuje dceřiný element *final*
 - přechod(y) – pomocí elementu/ů *transition*:
 - atribut *event* – událost, která vyvolá přechod (nepovinný)
 - atribut *target* – identifikátor cílového stavu
- Příklady a podrobnosti viz specifikace.

SCXML ukázka

Odpovídající SCXML dokument

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

```
<scxml xmlns="http://www.w3.org/2005/07/scxml"
  version="1.0" initialstate="Start">
  <state id="Start">
    <transition event="Credits" target="Credits" >
    </transition>
    <transition event="Course,_credits" target="Course,_credits" >
    </transition>
    <transition event="All" target="All" >
    </transition>
    <transition event="Credits,_finishing" target="Credits,_finishing" >
    </transition>
    <transition event="Course,_finishing" target="Course,_finishing" >
    </transition>
  </state>
```

SCXML ukázka

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

```
<state id="Course">
  <transition event="Finishing" target="Course,_fin
</transition>
  <transition event="Credits" target="Course,_Credi
</transition>
</state>
<state id="Credits">
  <transition event="Course" target="Course,_Credit
</transition>
  <transition event="Finishing" target="Finishing,_
</transition>
</state>
```

```
<state id="Finishing">
  <transition event="Course" target="Course,_finish" >
  </transition>
  <transition event="Credits" target="Finishing,_cr" >
  </transition>
</state>
<state id="Course,_Credits">
  <transition event="Finishing" target="All" >
  </transition>
</state>
<state id="All">
  <transition event="No" target="Start" >
  </transition>
  <transition event="Yes" target="Submit" >
  </transition>
</state>
```

```
<state id="Finishing,_credits">
  <transition event="Finishing" target="All" >
  </transition>
</state>
<state id="Course,_finishing">
  <transition event="Credits" target="All" >
  </transition>
</state>
<final id="Submit">
</final>
</scxml>
```

Call Control XML

Dialogové
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- CCXML je navrženo, aby umožnilo ovládat telefonní hovory z dialogových rozhraní popsaných např. pomocí VoiceXML,
- Umožňuje ovládat hovory na úrovni, která je mimo možnosti VoiceXML:
 - konferenční hovory
 - přiřadit každému hovoru vlastní VoiceXML interpret
 - ovládání odchozích hovorů
 - . . .
- Aktuální verze 1.0 (červenec 2011)