

# Dialogové systémy

Luděk Bártek

Fakulta informatiky  
Masarykova univerzita

podzim 2014

# Obsah

- 1 PLS
- 2 CCXML
- 3 State Chart XML
- 4 Simulace dialogu metodou Wizard of Oz
- 5 Multimodální dialogové systémy
- 6 Dialogové systémy a emoce

## Pronunciation Lexicon Specification (PLS)

- Standard W3C
- Aktuální verze 1.0 (říjen 2008)
- Definuje značkování pro specifikaci slovníků výslovnosti pro podporu syntézy a rozpoznávání řeči.
- Specifikace na stránkách W3C Specifikace na stránkách W3C

# Základní elementy PLS

- Kořenový element - lexicon
  - atributy - xmlns - specifikace jmenného prostoru (<http://www.w3.org/2005/01/pronunciation-lexicon>)
  - xml:lang - jazyk dokumentu
  - version - verze dokument (1.0)
  - alphabet - abeceda použitá pro fonetický přepis
- lexeme - obsahuje popis pro jednu lexikální jednotku (slovo, zkratku, ...
  - musí obsahovat aspoň jeden dceřiný element grapheme
- phoneme - obsahuje fonetický přepis dané lexikální jednotky (většinou se používá IPA).

## Ukázka lexikonu

```
<?xml version="1.0" encoding="utf-8"?>
  <lexicon
    version="1.0"
    xmlns="..."
    alphabet="ipa"
    xml:lang="en-US">
    <lexeme>
      <grapheme>color</grapheme>
      <phoneme>
        k' ʌə
      </phonem>
    </lexeme>
  </lexicon>
```

## Ukázka propojení lexikonu a gramatiky

- XML formát SRGS

```
<grammar xmlns="..." xml:lang="en" version="1.0">
  <lexicon
    uri="http://www.example.com/lexicon.file"/>
  <lexicon
    uri="http://www.example.com/strange-city-names.file"
    type="media-type"/>
  ...
</grammar>
```

## Ukázka propojení lexikonu a gramatiky

- ABNF formát SRGS

```
#ABNF V1.0 ISO-8859-1;  
language en-US;  
lexicon <http://www.example.com/lexicon.file>;  
lexicon <http://www.example.com/strange-city-names.file>  
      <media-type>;  
...
```

## Ukázka použití lexikonu v SSML

```
<speak version="1.1" xmlns="..." xml:lang="en-US">  
  <lexicon uri="lexicon.pls" xml:id="pls"/>  
  <lexicon uri="strange-words.file" xml:id="sw"  
    type="media-type"/>  
  <lookup ref="pls"> tokens here are looked up in  
    lexicon.pls  
  <lookup ref="sw"> tokens here are looked up first in  
    strange-words.file and then, if not found, in  
    lexicon.pls  
  </lookup>  
  tokens here are looked up in lexicon.pls  
  </lookup>  
  tokens here are not looked up in lexicon documents ...  
</speak>
```



## Call Control XML

- Slouží k ovládání a řízení telefonních hovorů v průběhu interaktivních hlasových služeb.
- Může být doplňkem dialogových systémů na bázi VoiceXML, který poskytuje pokročilé telekomunikační funkce:
  - konferenční hovory
  - umožňuje přiřazení VoiceXML interpretrů jednotlivým příchozím hovorům,
  - ...
- Může být použit dialogovým systémem pro ovládání spojení uživatele a DS.
- Platformy, které CCXML implementují mohou použít libovolnou z definic ovládání telefonního spojení (např. JAIN Call Control, ECMA CSTA, ...)

# Architektura systému využívajícího CCXML

## CCXML System Architecture

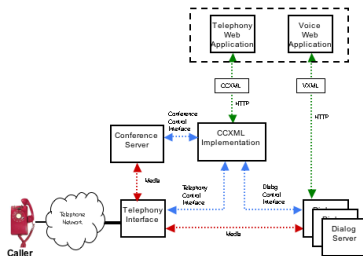


Figure : Architektura CCXML Aplikace (převzato ze specifikace CCXML)

# Struktura CCXML Aplikace

- Aplikace se skládá z dokumentů, které popisují ovládání a správu:
  - CCXML sezení (session) – zpracování CCXML dokumentu, který může být identifikován a odkazován.
  - Spojení – hovor nebo zdroj, který interaguje s hlasovým dialogem.
  - Konferenční objekt – slouží k sestavení a řízení konferenčních hovorů.
  - Dialog – může interagovat s různými spojeními a konferenčními objekty.
- Manipulace s těmito entitami je definována v CCXML, které umožňuje i zpracování asynchronních událostí.

# Životní cyklus CCXML aplikace

- Začátek
  - Příchozí telefonní hovor.
  - Zpracování elementu *createccxml* CCXML aplikací.
  - Požadavek na nové sezení z vnějšku.
- Pro vytvoření sezení je nutné znát:
  - URI počátečního CCXML dokumentu
  - parametry, které mohou ovlivnit získání tohoto dokumentu.

## Ukázka CCXML dokumentu

```
<?xml version="1.0" encoding="utf-8"?>
<ccxml version="1.0"
  xmlns="http://www.w3.org/2002/09/ccxml">
  <var name="state0" expr="'init'"/>
  <eventprocessor statevariable="state0">
    <transition state="init" event="connection.alerting">
<accept/>
    </transition>
    <transition state="init"
      event="connection.connected">
      <log expr="'Prijate volani'"/>
      <dialogstart src="dialog.vxml"/>
      <assign name="state0" expr="'active'"/>
    </transition>
```

# Ukázka CCXML dokumentu

pokračování

```
<transition state="active" event="dialog.exit">  
  <log expr="'Ziskana hodnota: '+event$.values.input+  
  <exit/>  
</transition>  
</eventprocessor>  
</ccxml>
```

## Použitá literatura

- Specifikace CCXML verze 1.0 z července 2011 na stránkách W3C

# SCXML

- State Chart XML – návrh standardu W3C Voice Browser Activity
  - aktuálně ve stádiu Last Call Working Draft (květen 2014)
- Slouží k specifikaci konečných automatů (Mealyho automatů).
- Plánován jako jazyk pro zápis dialogové strategie ve VoiceXML 3.0.
- Principiálně vychází ze CCXML



# Konečný automat

- Konečný automat  $K = (S, \Sigma, \delta, q_0, Q_f)$ 
  - $S$  – konečná množina stavů
  - $\Sigma$  – konečná neprázdná množina vstupních symbolů (abeceda)
  - $\delta$  – přechodová funkce
    - deterministický automat –  $\delta : S \times \Sigma \rightarrow S$
    - nedeterministický automat –  $\delta : S \times \Sigma \rightarrow 2^S$
  - $q_0$  – počáteční stav
  - $Q_f$  – množina koncových stavů
- Mealyho automat  $M = (S, \Sigma, Y, \delta, \psi, q_0, Q_f)$ 
  - $Y$  – neprázdná výstupní abeceda
  - $\psi$  – výstupní funkce:  $S \times \Sigma \rightarrow Y$

# SCXML

## Základní elementy

- *scxml* – kořenový element, slouží jako kontejner pro elementy popisující stavy a přechody mezi nimi.
- *state* – stav konečného automatu
- *parallel* – definice paralelních dceřiných stavů
  - Jsou aktivní současně s rodičem.
- *transition* – definice přechodové funkce
- *initial*, *final* – definice počátečního/koncového stavu
- *onentry*, *onexit* – ošetření události při aktivaci/deaktivaci stavu.
- ...

## Ukázka zápisu konečného automatu v SCXML

```
<?xml version="1.0" encoding="utf-8"?>
  <scxml xmlns="http://www.w3.org/2005/07/scxml"
    xmlns:xi="http://www.w3.org/2001/XInclude"
    version="1.0"
    initial="Main"
    datamodel="ecmascript">
    <state id="main">
      <initial>
        <transition target="Start"/>
      </initial>
    </state>
  </scxml>
```

# Ukázka

pokračování

```
<state id="Start">
  <onentry>
    <log>Vítejte v registraci IS MU</log>
  </onentry>
  <transfer event="AllData" target="process"/>
  <transfer event="course" target="conclusion"/>
  <transfer event="concl" target="courseState"/>
</state>
<final id="process"/>
```

# Ukázka

## pokračování

```
<state id="conclusion">
  <onentry><log>Způsob ukončení</log></onentry>
  <transition target="process"/>
</state>
<state id="courseState">
  <onentry><log>Předmět</log></onentry>
  <transition target="process"/>
</state>
</state>
</scxml>
```

# Metoda Wizard of Oz

- Simulace dialogového rozhraní modelem člověk – člověk.
- Založena na principu popsaném v knize The Wonderful Wizard of Oz (Lyman Frank Baum)
- Princip:
  - Funkce dialogového rozhraní je (skrytě) simulována člověkem.
  - Průběh dialogu je protokolován.
  - Průběh se řídí navrženou dialogovou strategií.
    - Pokud je dostupný prototyp může Wizard pouze modifikovat a předávat komunikaci mezi uživatelem a systémem.

# Metoda Wizard of Oz

- Občas snaha navodit zdání, že uživatel komunikuje s dialogovým systémem – využívají se různé prostředky:
  - vzdálená komunikace kde osoba simulující dialogové rozhraní komunikuje prostřednictvím TTS
  - použití vokodérů, které změní hlas osoby, která provádí testování, aby zněl jako výstup TTS
  - ...

## Tvorba dialogového rozhraní pomocí metody WoZ

- Z korpusu dialogů na dané téma (pro danou doménu) lze vygenerovat dialogové rozhraní následovně:
  - 1 Vytvoříme iniciální korpus metodou WoZ
    - Komunikace pouze čaroděj – uživatel.
  - 2 Odstraní se konflikty a na základě korpusu se vytvoří dialogové rozhraní.
  - 3 Kombinovaně vytvoříme nový korpus.
    - "Čaroděj" se snaží maximálně využívat navržené dialogové rozhraní..
  - 4 Odstranění konfliktů a vygenerování nové verze dialogového rozhraní.
  - 5 Pokud je rozhraní v pořádku, generování končí, jinak se pokračuje krokem 3.



# Multimodální dialogová rozhraní

- Mimo mluvenou řeč umožňuje i další způsoby komunikace člověk – počítač:
  - textová komunikace
  - grafická komunikace
  - emoce
  - ...
- Výhody
  - lepší přístupnost.
    - uživatelé s poruchami sluchu,
    - uživatelé s poruchami řeči,
    - ...
  - lepší pochopení pragmatiky projevu
    - možnost pracovat i s emočním rozpoložením – může mít vliv na pragmatiku projevu.

## Multimodální komunikace počítač – člověk

- Textová:
  - Mimo hlasový výstup je navíc zobrazen i odpovídající textový výstup.
  - Lze využít prostředky pro IM, SMS, ...
- Grafická:
  - Talking Heads – mimo hlasový výstup je navíc zobrazena tvář (hlava, celý člověk, ... jejíž pohyby, zejména úst, odpovídají mluvené řeči.
  - Komunikace znakovou řečí – mluvené slovo je překládáno na znakovou řeč (viz Guimeraes et al. – Structure of the Brazilian Sign Language (Libras) for Computational Tools: Citizenship and Social, in Organizational, Business, and Technological Aspects of the Knowledge Society, CCIS vol. 112, Springer, Heidelberg, 2010, pp. 365 — 370. )
    - Znaková řeč prezentována pomocí rukou nebo avatara.

## Multimodální komunikace člověk – počítač

- Široké spektrum možností zadávání vstupu uživatelem jinak než hlasem:
  - klávesnice (počítač, DTMF, SMS, ...)
  - rukou psaný vstup – dotyková obrazovka + pero, ...
  - ústy ovládaná zařízení
  - ovládání pomocí pohybů očí a víček
  - rozpoznávání řeči pomocí sond detekujících činnost svalů a mozku (viz Schultz, T. – Silent and Weak Speech Based on Elektromyography, in Proceedings of 12th International Conference ICCHP 2010 Part 1, Wien, Springer, Heidelberg, pp. 595 – 604, 2010. )
  - rozpoznávání znakové řeči
  - ...
- Často jako doplněk řečového vstupu.

## Existující nástroje a standardy

- Proprietární řešení:
  - Součást CSLU Toolkitu.
  - Projekt August.
- Otevřená řešení:
  - Návrhy doporučení W3C týkající se multimodálního přístupu – zatím bez implementace.
    - Využívají a propojují i další standardy W3C (CCXML, XHTML, VoiceXML, SVG, SMIL, ...)
  - Výstup W3C Multimodal Interaction WG

# Ukázka multimodální syntetizované řeči

Ukázka avatara

## Co jsou to emoce?

- "This is a very tough question, that has produced significant amounts of headaches to scientists in the past ...", "... many researchers have to opted to study systematically phenomena that most consider emotional." (Laval University of Quebec)
- "Only mathematics is certain, so all must be based on mathematics." (R. Descartes)
- Dělení emocí:
  - Primární (základní) – vyskytují se u všech lidí a u části vyšších živočichů.
  - Sekundární (vyšší) – mohou být intelektuální, morální a estetické. Mohou se lišit mezi jednotlivými kulturami.
- Velkých šest:
  - hněv
  - zklamání
  -

# Základní emoce

- Velkých šest (R. Descartes):
  - hněv
  - zklamání
  - smutek
  - strach
  - překvapení
- Další autoři:
  - Arnold – hněv, averze, odvaha, sklíčenost, touha, zoufalství, strach, nenávisť, láska, smutek.
  - Ekman, Friesen, Ellsworth – hněv, odpor strach, radost, smutek, překvapení.
  - Frijda — touha, štěstí, zájem, překvapení, údiv, zármutek,
  - ...

## Detekce emocí

- Lze provádět pomocí detekce změn různých biometrických vlastností.
  - Změny galvanických vlastností kůže.



- Změny tlaku krve a pulsu.





## Detekce emocí

- Použitelné biometrické charakteristiky:
  - změny dýchání



- změny elektrické aktivity mozku



- změny charakteristik řeči
- výraz tváře (Yale Face Database).

# Ukázky z Yale Face Database

- Radost



- Smutek



# Ukázky z Yale Face Database

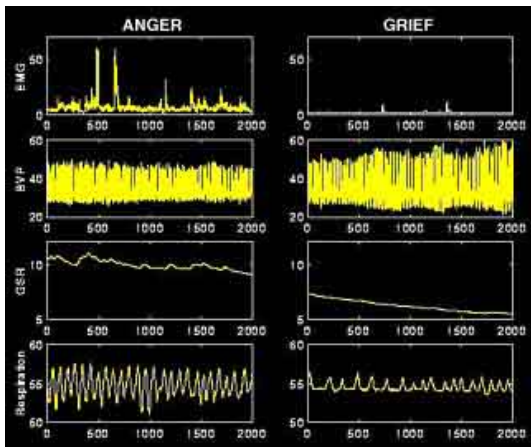
- Ospalost



- Překvapení



## Ukázky průběhů charakteristik pro smutek a hněv



## Využití emocí v dialogovém rozhraní

- Dialogová rozhraní informačních systémů
  - uzpůsobení dialogové strategie emočnímu stavu uživatele (klid, stres, hněv, ...)
  - přepojení uživatele na lidského operátora.
- Výukové DS:
  - uzpůsobení dialogové strategie koncentraci uživatele.
- ...