

Řečová interakce a sociální sítě

Luděk Bártek

Fakulta informatiky
Masarykova univerzita Brno

podzim 2013

Obsah

- 1 Způsoby implementace dialogových rozhraní
- 2 Dialogová komunikace
- 3 Standardy W3C Voice Browser Activity
- 4 VoiceXML

Historie

- Prostředky logického programování (Prolog, ...)
 - uchování databáze znalostí
 - průběh dialogu byl řízen interpretem jazyka na základě pravidel pro logické programování.
- Vyšší programovací jazyky
 - C - příklad systém AudiC vyvíjený na FI; dialogový systém na programování v jazyce C
 - ...
- Značkovací jazyky pro popis dialogových rozhraní:
 - 2. polovina 90. let 20. století.
 - VoxML (1999)
 - Standardy W3C VoiceBrowser Activity: - přelom 20. a 21. století

Standardy W3C Voice Browser Activity

- VoiceXML - popis dialogové strategie.
- SRGS - popis gramatiky pro rozpoznávání řeči.
- SISR - doplnění SRGS o sémantickou interpretaci promluvy.
- SSML - popis prozodických vlastností generované promluvy.
- a další viz stránky W3C Voice Browser Working Group

Komunikace uživatel ↔ systém

- PSTN:
 - Uživatel používá ke komunikaci veřejnou telefonní síť.
 - Počítač je připojen k veřejné telefonní síti - závěrečná digitalizace probíhá na straně poskytovatele služby (server(y), telefonní ústředna poskytovatele, ...).
- VoIP:
 - Komunikace digitalizována na straně uživatele:
 - uživatelův počítač
 - VoIP telefon.
 - Komunikace uživatelský počítač ↔ server: prostřednictvím některého VoIP protokolu:
 - SIP
 - H.323
 - proprietární protokoly (Skype, ...)
 - ...

Komunikace uživatel ↔ systém

- Textová komunikace:
 - SMS
 - IM
 - webové rozhraní
 - DTMF
- Vyrovnání rozdílů mezi různými protokoly - standard Call Control XML

Rozpoznávání řeči a sémantická interpretace

- Je vhodné umožnit co nejpřirozenější komunikaci - nutnost rozpoznávání souvislé promluvy a schopnost korektního určení zadaných dat.
 - Relativně nízké úspěšnost rozpoznávání - pro zvýšení se používají gramatiky (JSGF, SRGS, ...).
- Získání relevantních dat z rozpoznané promluvy - sémantická interpretace.
 - Bývá součástí gramatiky pro rozpoznávání řeči - pravidlům je přiřazena sémantická interpretace.
 - atribut
 - dceřiný element.
 - Lze provádět odvozování sémantiky na základě sémantiky částí promluvy - pomocí výrazů v jazyce ECMAScript.
 - JSGF - součást specifikace.
 - SRGS - související standard SISR.

Řízení průběhu dialogu

- Průběh dialogu řízen dialogovou strategií.
 - Jazyky pro zápis dialogové strategie:
 - VoiceXML
 - CallXML
 - Jedná se o značkovací jazyky - nutnost interpretace - VoiceXML platformy:
 - JVoiceXML
 - VoiceGlue
 - Voxeo Prophecy
 - BeVocal Cafe
 - OptimTalk
 - ...

Generování promluvy a syntéza řeči

- Generování promluvy:
 - Rámce výstupních promluv jsou součástí dialogové strategie (VoiceXML).
 - Vlastní promluva zapsána pomocí jazyků pro popis syntetizované řeči:
 - SABLE
 - SSML

Dialogové korpusy

- Dialogový korpus
 - rozsáhlá databáze průběhů dialogů
 - mluvená resp. textová
 - textový může obsahovat buď přepisy hlasových dialogů nebo záznam textové komunikace.
 - Může obsahovat navíc informace:
 - účastníci (věk, pohlaví, obeznámenost s tématem, ...)
 - téma dialogu
 - výsledek dialogu
 - hodnocení průběhu dialogu z pohledu obou účastníků
 - ...

Tvorba dialogového korpusu

- 1 Záznam reálných dialogů živých lidí
 - různé zdroje
 - záznamy rozhovorů mezi lidmi
 - záznamy z linek podpory
 - záznamy komunikace s existujícím dialogovým rozhraním nebo jeho prototypem
 - on-line diskuse/ dialogy na Internetu.
- 2 Simulace průběhu dialogu
 - metoda Wizard of Oz
 - ...

Metoda Wizard of Oz

- Vychází z pohádky L F. Bauma - The Wonderfull Wizard of Oz.
- Dialogové rozhraní je nahrazeno lidským operátorem, který simuluje chování dialogového rozhraní.
 - Měl by se držet navržené dialogové strategie.
- Průběh dialogu je zaznamenáván.
- Lze použít prostředky pro zvýšení důvěryhodnosti komunikace uživatele s počítačem:
 - operátor využívá odpovídající IS - převádí uživatelské dotazy do formy vhodné pro IS
 - operátor nekomunikuje s uživatelem přímo a prostřednictvím TTS a případně ASR

Generování dialogových rozhraní z korpusu

- Korpus - rozsáhlá databáze označovaných dat"
 - řečové korpusy
 - textové korpusy
 - dialogové korpusy.
- Na základě korpusu lze generovat dialogové rozhraní, které korpusu vyhovuje:
 - 1 Vytvoří se korpus metodou Wizard of Oz (komunikuje pouze "čaroděj").
 - 2 Odstranění konfliktů a převod korpusu na dialogové rozhraní.
 - 3 Kombinované vytvoření nového korpusu
 - Čaroděj se snaží v maximální možné míře využívat navržené dialogové rozhraní.
 - 4 Odstranění konfliktů a vytvoření další verze dialogového rozhraní.
 - 5 Pokud je rozhraní v pořádku, proces končí, jinak se pokračuje krokem 3.

W3C Voice Browser Activity

- 1999 - založena W3C Voice Browser Group.
- Cíl - návrh standardů umožňujících přístup k Webu pomocí hlasu a telefonu.
- Členové:
 - Hewlet-Packard
 - Nuance Communication
 - Lucent Technologies
 - Motorola
 - ScanSoft
 - IBM
 - Tellme Networks
 - Vocalocity
 - ...

Standardy W3C Voice Browser Activity

- VoiceXML - jazyk pro popis dialogových strategií.
- Speech Recognition Grammar Specification - jazyk pro zápis gramatik pro podporu rozpoznávání řeči.
- Semantic Interpretation for Speech Recognition - jazyk pro podporu sémantické interpretace rozpoznané promluvy.
- Speech Synthesis Markup Language - jazyk pro popis prozodických charakteristik pro syntézu řeči.
- Pronunciation Lexicon Specification - jazyk pro popis výslovnosti pro rozpoznávání a syntézu řeči.
- Call Control XML - jazyk pro popis řízení telefonního spojení uživatele a systému.
- State Chart XML - jazyk pro popis obecně použitelných stavových automatů.

Implementace dialogových rozhraní - nástroje

- Standardy W3C jsou značkovací jazyky - nutná interpretace.
- Existuje řada platforem:
 - volně dostupné desktopové - JVoiceXML, Public VoiceXML, ...
 - komerční desktopové - OptimTalk
 - dříve existovala volně dostupná verze
 - laboratoř LSD má zakoupenou licenci na laboratorní stroje.
 - Volně dostupné on-line - VoIP ústředna Asterisk + VoiceGlue/OpenVXI, ...
 - Komerční on-line - Voxeo Prophecy, BeVocal Cafe - lze vyzkoušet a omezeně používat on-line (max. 2 paralelní hovory).

VoiceXML

- Značkovací jazyk pro popis dialogových strategií.
- Historie:
 - 1995 - započat vývoj Phone Markup Language (PML, AT&T)
 - 1998 - založeno VoiceXML forum (AT&T, IBM, Lucent, Motorola) - spojení úsilí při návrhu jednotného standardu jazyka pro popis a návrh dialogových rozhraní.
 - 2000 - VoiceXML forum uvolnilo VoiceXML 1.0. Přijato mezi doporučení W3C.
- Cíle:
 - Minimalizace interakce mezi klientem a serverem - více interakcí v jednom dokumentu.
 - Odstínění vývojáře od nízkoúrovňových a platformě závislých detailů.
 - Oddělení UI od aplikační logiky.
 - Přenositelnost služba mezi implementačními platformami.
 - Snadná použitelnost pro jednoduché dialogy a podpora komplexních dialogových rozhraní.

Základní pojmy

- Kořenový element *vxml* - povinné atributy:
 - version - číslo verze - aktuálně 2.1
 - xmlns - deklarace jmenného prostoru dokumentu . Musí obsahovat <http://www.w3.org/2001/vxml>
- Dialogy a dílčí dialogy (subdialogy) - reprezentace pomocí elementů
 - form - skládá se ze vstupních polí, ... (viz dále)
 - menu - obsahuje položky, které se odkazují na další formuláře dialogu.
 - subdialog - odpovídá volání funkce ve strukturovaném programování.
 - grammar - gramatika přiřazená dialogu
 - definice a ošetření událostí.- throw/catch

Způsob interpretace dokumentu - Form Interpretation Algorithm

- 1 Výběr vstupního pole - vybere se první vstupní pole, které vyhovuje (nemá přiřazenu hodnotu a pokud obsahuje vstupní podmínku, tak ta je splněna). Pokud takové pole není, dialog končí.
- 2 Přehraje se jeho výzva.
- 3 Získají se vstupy od uživatele a ošetří se případné události (nomatch, noinput, filled)
- 4 Pokračuje se krokem 1.

Form

- Obsahuje:
 - vstupní pole:
 - field
 - record
 - transfer
 - object
 - subdialog
 - řídicí příkazy:
 - block
 - initial
- Ukázky dialogů s iniciativou systému - dokumenty dostupné na adrese <http://www.fi.muni.cz/~bar/pb095/>