

MANAŽMENT REČOVÝCH DIALÓGOVÝCH SYSTÉMOV (MANAGEMENT OF SPOKEN DIALOGUE SYSTEMS)

Ján PAPAJ

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky Technickej univerzity v Košiciach, Letná 9, 042 00 Košice, Slovenská republika, tel.: 055/602 2223, E-mail: jan.papaj@tuke.sk

SUMMARY

Nowadays Human-Computer Interfaces (HCI) are widely studied and became one of the major interests among scientific community. Due to last decades progresses in the field of speech technologies, like Automatic Speech Recognition (ASR) and Text-To-Speech synthesis (TTS) and in the field of Natural Language Processing (NLP), voice enabled interfaces and spoken dialogue systems are hopefully going to become more and more common.

A dialogue management system is a system for managing the interactions between a person and a computer in a conversational interface. Spoken dialogue systems can be classified in to three main types, according to the methods used to control the dialogue with user: finite state-based systems, frame-based systems and agent-based systems. This paper describes basic principles of these methods and their application in dialogue systems as well as dialogue management techniques which have been developed for spoken dialogue systems and methods that have been adopted in some well-know dialogue systems.

Keywords: dialogue management, dialogue control, finite state-based systems, frame-based systems, agent-based systems

1. ÚVOD

Modul manažovania dialógu (MD) je často vnímaný ako centrálny prvok automatizovaných (interaktívnych) rečových komunikačných systémov a v niektorých prípadoch riadi nielen interakciu medzi systémom a používateľom, ale aj interakciu jednotlivých modulov systému. Jeho najdôležitejšou úlohou je však riadenie dialógu – v abstraktnom vyjadrení – riadenie prechodu z jedného stavu dialógu do iného. Mnoho dialógových manažérov iniciuje alebo dokonca vykonáva interpretáciu vstupných správ, generuje výstupné správy a realizuje rôzne špecifické úlohy. Vo všeobecnosti možno konštatovať, že systém manažmentu dialógu sa využíva na simulovanie dialógového procesu [1].

Techniky MD sú nevyhnutné pre systémy využívajúce rozpoznávanie reči. Dialógy s hovoreným jazykom vyžadujú komplikované techniky modelovania, ktoré ale môžu následne zvýšiť účinnosť techník rozpoznávania reči [2].

2. PRÍSTUPY K MANAŽMENTU DIALÓGU

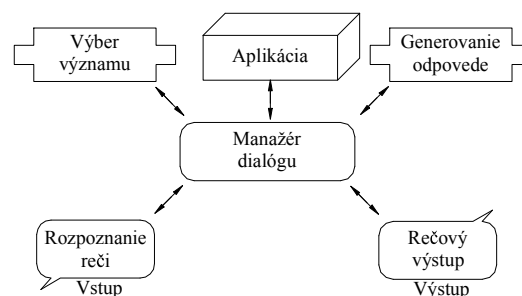
Rečový dialógový systém v podstate pozostáva z komponentov zobrazených na obr. 1. Architektúru dialógového systému znázorňuje obr. 2.

Vstupom do rečového dialógového systému je reč, ktorá je analyzovaná modulom na rozpoznávanie reči. Výstup z tohto modulu je text alebo prepísaná reč, ktorá je ďalej analyzovaná. Na základe analýzy je odvodená sémantická informácia. Manažér dialógu následne koordinuje ostatné komponenty za účelom dokonalého pochopenia všetkých implikácií vstupu v aktuálnom kontexte. Modul generovania odpovede generuje odpoveď

najčastejšie vo forme prirodzeného textu, ktorý je následne transformovaný rečovým syntetizátorom do rečovej podoby (*Text-To-Speech -TTS*).



Obr. 1 Komponenty rečového dialógového systému
Fig. 1 Components of spoken dialogue system



Obr. 2 Architektúra rečového dialógového systému
Fig. 2 Architecture for spoken dialogue system

Hlavnou úlohou MD je riadenie postupnosti dialógu a zároveň musí byť schopný zabezpečiť tieto činnosti [3]:

- rozhodnúť, či bola od používateľa získaná požadovaná informácia pre umožnenie komunikácie s externou aplikáciou,
- komunikácia s externou aplikáciou,
- spätné odovzdanie informácie používateľovi.

V jednoduchých aplikáciách sú tieto činnosti prezentované vo forme procesov, ktoré sú zoradené do presných skupín činností, t. j. systém zistí, resp. overí, čo chce používateľ vedieť alebo robiť, poradí sa s externou aplikáciou a zobrazí späťne výsledky používateľovi. Častejšie však nastáva komplikovanejší proces určovania z dôvodu získania nedostatočného množstva informácií od používateľa potrebných na komunikáciu systému s externými aplikáciami.

3. RIADENIE DIALÓGU

Riadenie dialógu môžeme rozdeliť do nasledujúcich skupín: *systémovo riadené*, *používateľsky riadené* a *ich kombinácia*.

V dialógu riadenom systémom kladie systém sériu otázok potrebných na zistenie požadovaných parametrov úloh od používateľa.

V používateľsky riadenom dialógu používateľ riadi dialóg a kladie systému otázky tak, aby získal potrebné informácie.

V zmiešanom riadení dialógu je riadenie rozložené medzi používateľa a systém. Používateľ môže kedykoľvek kľásť otázky, pričom systém môže prevziať riadenie získavania požadovaných informácií alebo objasniť nejasné informácie.

V systémove a zmiešanom riadení dialógu je nutné riadenie za účelom stanovenia, ktoré otázky a príkazy môže systém požadovať a kedy.

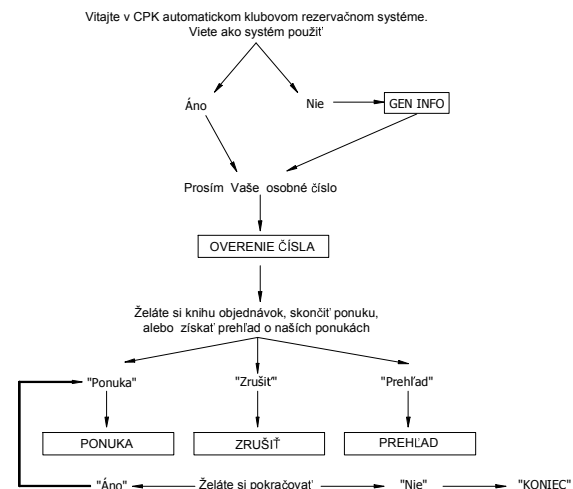
Pre riadenie dialógu sú určené tri hlavné stratégie:

- **Systémy s konečným počtom stavov** (*Finite State-Based Systems* - FSBS) podporujú systémove dialóg, v ktorom sú všetky otázky, na ktoré sa systém pýta, vopred stanovené.
- **Formulovo riadené systémy** (*Frame-Based Systems* - FBS) sú predovšetkým systémove vedené, hoci umožňujú používateľovi určitý obmedzený stupeň prevzatia iniciatívy.
- **Agentové systémy** (*Agent-Based Systems* - ABS) sú charakteristické zmiešanou iniciatívou.

3.1. Systémy s konečným počtom stavov

Štruktúra dialógu v systéme s konečným počtom stavov (FSBS) je zobrazená vo forme stavu prechodovej siete. Prechodová sieť je tvorená uzlami, ktoré predstavujú systémove otázky. Prenosy medzi jednotlivými uzlami determinujú všetky možné cesty dialógu. Tým sú teda špecifikované všetky možné dialógy. Každý stav reprezentuje určitú stavovú informáciu, v ktorej je informácia získaná alebo potvrdzovaná používateľom. Na obr. 3 je znázornený príklad použitia FSBS, ktorý bol aplikovaný v projekte Danish Dialogue System [4]. Predstavuje použitie FSBS na modelovanie postupného dialógu určeného pre automatické rezervačné služby. V tomto dialógu postupuje systém cez sériu prenosových stavov, ktoré determinujú odpovede používateľa. Vyskytujú sa tu rozličné možnosti voľby uzlov a slučiek, ktoré sú znázornené pomocou poddialógov (napr. overenie

čísla, ponuka, zrušenie, prehľad). Okrem toho môže používateľ v tejto konkrétnej architektúre používať kľúčové slová ako: „opakovať“ a „zmeniť“ na vyžiadanie opakovania systémovej výstupu a na zmenu získaného predchádzajúceho parametra.



Obr. 3 Príklad pre dialógový graf určený pre rezervačné automatickú službu

Fig. 3 Example of dialogue graph for an automatic book service

Hlavná výhoda FSBS spočíva v jeho jednoduchosti. Siete prenášajúce stavy sú určené na modelovanie dialógov v dobre štrukturovateľných úlohách. Získavané informácie sa menia v presne stanovenej postupnosti, pričom sa zachováva systémove riadenie dialógu, ktoré definuje, ktorá otázka bude nasledovať.

Ďalšími vhodnými príkladmi FSBS sú dotazníky, cestovné dotazy alebo podadresárová spolupráca za predpokladu, že dialóg je vedený základnými otázkami kladenými systémom za účelom zistenia presného počtu definovaných odpovedí.

Modely dialógov s konečným počtom stavov nie sú vhodné pre modelovanie neštrukturovaných úloh, ktoré sú charakterizované podúlohami. Tieto podúlohy sú náročné na predikciu či už pomocou informácií modelovaných na rozličnej úrovni abstrakcie alebo komplexnými závislosťami medzi informačnými položkami [6].

3.2. Formulovo riadené systémy

Formulovo riadené systémy (FBS) sú založené na analógii s vyplňovaním formuláru (form filling), kde je preddefinovaná skupina informácií zaradená skôr, než sa vytvorí dialóg z preddefinovaných postupností otázok. Napr. v projekte Philips automatic train timetable information system sú otázky, ktoré by sa mohol systém opýtať, zahrnuté spoločne s predpokladmi, t. j. za akých podmienok môžu byť položené [7]. Komponent na manažovanie dialógu (viď obr. 1) rozhodne o otázke, ktorá bude nasledovať na základe splnenia podmienok, pričom však informácie nemusia byť získavané v chronologickom poradí. Na výber požadovaných

otázok sa v určitej etape dialógu môžu vybrať rôzne faktory. V projekte Philips SpeechMania je každá dialógová akcia (zahrňujúca dialógy) kódovaná kľúčovým slovom, ktoré určuje prioritu dialógovej činnosti a nasledovne postupnosti dialógu [8].

Existujú aj iné formy, ktoré môžu byť využité na riadenie FBS Sú to: *E-forma, schéma, grafická štruktúra úloh a typová hierarchia* [9].

E-forma (elektronická forma) je využívaná ako rozhranie medzi hovorenou rečou a databázou. Rozlišuje popísané typy foriem a formúl. Formy a formuly sú usporiadané do slotov, ktoré môžu mať rôznu prioritu pre používateľov (napr. pre niektorého používateľa môže byť rozhodujúca skôr farba auta než typ alebo funkčnosť). *E-forma* umožňuje preskúmať mnohonásobné kombinácie na nájdenie najvhodnejšej odpovede, ktorá spĺňa používateľove požiadavky. Selekcija vhodnej odpovede je uvedená ako optimalizačná úloha, ktorá vyžaduje viac, než len výber záznamov z databázy.

Schémy sú použité v systéme Carnegie Mellon Communication na modelovanie komplexnejších úloh, než len na získavanie základných informácií, ktoré využívajú formy [10]. *Schéma* je stratégia na doplnenie cieľov v úlohovo vedenom dialógu (task-based dialogue) ako je napr. určenie cestovného plánu.

Grafická štruktúra úloh predstavuje podobnú schématickú štruktúru ako *E-forma*. Je používaná na určenie správania sa modulu určeného na riadenie dialógu ako aj modulu rozpoznávania reči [11]. Graf znázorňuje vzťahy medzi elementami služieb zákaznických aplikácií a je použitý na poskytnutie kontextovej interpretácie hovorenej výpovede získanej v dialógu.

Typ hierarchie môže byť podobne využívaný na modelovanie sféry dialógu, ako aj základ pre objasnenie nejasných otázok [12]. Za predpokladu, že informácie v hierarchii môžu chýbať alebo sú nedefinované, sú generované žiadosti za účelom dosiahnutia splnenia a objasnenia nejasných otázok v dialógu.

3.3. Agentové systémy

Pojem agentové systémy (ABS) našiel širší základ v odbore známom pod pojmom umelá inteligencia.

ABS sú navrhované na komplexnú komunikáciu medzi systémom, používateľom a základnou aplikáciou tak, aby riešila nejakú úlohu alebo problém. ABS sa zameriavajú na modelovanie dialógov ako vzájomnej spolupráce medzi inteligentnými agentami.

Z pohľadu riadenia dialógu môžeme ABS začleniť do nasledujúcich skupín [3]:

- *Systémy využívajúce teóriu overovania,*
- *Systémy využívajúce plánovanie,*
- *Systémy s distribuovanou architektúrou,*
- *Systémy na báze racionálnej interakcie.*

4. PRÍKLAD DIALÓGU

V tomto odstavci popíšeme príklad dialógu, ktorý je vedený medzi užívateľom a systémom vo formulovo riadenom systéme (FBS). Dialóg znázorňuje, ako je systém schopný reagovať na jednotlivé vyjadrenia získané od používateľa.

Systém: Dobrý deň. Vitajte v informačnom systéme. Ako vám môžem pomôcť?

Používateľ: Chcel by som letieť z Košíc do Paríža.

Systém: Kedy chcete letieť z Košíc do Prahy?

Používateľ: Nie do Prahy ale do Paríža..

Systém: Prepáčte, kedy chcete letieť z Košíc do Paríža?

Používateľ: Vo štvrtok 20.08.2004 o 8:00.

Systém: Takže vo štvrtok 20.08.2004 letí lietadlo z Košíc do Paríža o 8:30 s príchodom 12:00. Želáte si rezervovať letenku?

Používateľ: Nie ďakujem

Systém: Ďakujem, že ste využili naše služby a prajeme Vám šťastnú cestu a príjemný deň!

V dialógu nie je preddefinovaná presná postupnosť krokov, ktoré budú nasledovať. Postupnosť sa špecifikuje až na základe kontextu získaného z používateľových odpovedí. Takže z danej postupnosti výmien medzi používateľom a systémom si systém získava potrebné údaje, ktoré sú nevyhnutné na splnenie danej úlohy. V dialógu si systém explicitne overuje získané údaje od používateľa a zároveň hľadá príslušné riešenie na nájdenie riešenia, v našom prípade nájdenie leteckého spojenia medzi Košicami a Parížom.

FBS môže nájsť široké uplatnenie v systémoch určených na informovanie cestujúcich, ako to znázorňuje táto aplikácia určená pre asistenčné služby leteckých spoločností.

5. ZÁVER

Pri návrhu rečového dialógového systému je nutné na základe týchto prístupov na MD uvažovať, ktorý prístup je pre danú aplikáciu najvhodnejší.

Siete s konečným počtom stavov (FSBS) sú určené pre jednoduché dobre štrukturované úlohy, ktoré môžu byť rozložené do presne definovaných podúloh. Sú taktiež určené pre úlohy vyžadujúce postupné vyplňovanie vo forme dotazníka (form filling). FSBS je systémovo vedený cez preddefinovanú dialógovú cestu, od ktorej sa užívateľ nemôže odchyliť. Ako odpovede používateľa by boli ideálne krátke odpovede vo forme jednoduchých slov a fráz. Hlavná úloha pri návrhu takéhoto systému vyžaduje systémovú jednoznačnosť na formuláciu a na obmedzenie používateľových odpovedí. Tieto systémy využívajú stavovú prechodovú sieť, ktorá má presnú štruktúru Nevýhodou systému s konečným počtom stavov je, že nie je flexibilný a nemôže byť jednoducho navrhnutý na potvrdenie používateľa vykonať

korekcie predchádzajúcich získaných hodnôt, alebo jednoducho zmeniť jednu podúlohu na inú [13].

Formulovo riadené systémy (FBS) umožňujú vyšší stupeň flexibility, pretože používateľ môže poskytnúť viac informácií než vyžadujú systémové otázky a na záver ich nepotrebuje dopĺňať v presnom čase (one-time), ak to nie je možné. Systém si zachováva záznam o tom, ktorá informácia je nevyhnutná a na základe toho sa opýta adekvátnu otázku. FBS sú obmedzené získavať základné informácie s úlohami bez možnosti sprostredkovania informácií. V systéme založenom na konverzačnej agentúre (conversational agency) môžu byť tieto správania flexibilnejšie a môžu byť založené na procese jasného uváženia systému. FBS predstavuje poskytnuté riešenie s väčšou flexibilitou oproti systému FSBS.

V agentových systémoch je komunikácia znázornená ako interakcia medzi dvomi agentmi, pričom každý z nich je schopný uvažovať o svojich činnostiach a poznatkoch, ktoré nadobudol počas dialógu. Tieto systémy predstavujú zmiešanú iniciatívu na riadenie dialógu.

LITERATÚRA

- [1] G.E.Churcher, E.S Atwell, C.Souter-Dialogue Management Systems: A Survey and Overview, School of computer studies, Research report series, February 1997
- [2] G.E.Churcher, E.S Atwell, C.Souter-Dialogues in Air Traffic Control, in Luperfoy et al.1996
- [3] M.F. McTear: Spoken Dialogue Technologie: Enabling the Conversational User Interface, ACM Computing Surveys, Vol.34, No.1, March 2002, pp.90-169
- [4] L.B.Larsen, A. Baeekgaard: Rapid prototyping of a dialogue system using a generic dialogue development platform. In Proceedings of ICSLP'94(Yokohama, Japan), ICSLP, 919-922
- [5] J. Potjer, A.Russel, L. Boves and E.D. Os: Subjective and objective evaluation of two types of dialogues in a call assistance services. In IVTTA. IEEE, Basking Ridge, NJ, pp.121-124, 1996
- [6] C. Kamm, User interfaces for voice applications. In Voice Communication Between Humas and Machines, D. Roe and J. Wilpon, EDS. National Academy Press, Washington, DC, pp.34-75, 1995
- [7] H. Aust, M.Oerder, F. Seide, V. Steinnbis: The Philips automatic train timetable information system. Speech communication, 1996
- [8] SpeechMania™ Philips: www. Speech. de. philips. com
- [9] D. Goddeau, H. Meng, J. Polifroni, S. Seneff and Busayapongchai: A form-based dialogue manager for spoken language applications. In Proceedings of 4th International Conference on Spoken Language Processing (ICSLP'96, Pittsburg, PA). ICSLP, pp.701-704
- [10] P. Constantinides, S. Hansma, C. Tchou and A. Rudnický: A schema based approach to dialogue control. In Proceedings of 5th International Conference on Spoken Language Processing (ICSLP'98, Sydney, Australia). Vol.2, ICSLP, pp.409-412, 1998
- [11] J. Wright, A. Gorin and A. Abella: Spoken language understanding within dialogs using a graphical model of task structure. In Proceedings of 5th International Conference on Spoken Language Processing (ICSLP'98, Sydney, Australia). Vol.5, ECSLP, 1998
- [12] M. Denecke and A. Waibel: Dialogue strategies guiding users to their communicative goals. In Proceedings of 5th European Conference on Speech communication and Technologies (Eurospeech'97, Rhodes, Greece) ESCA, 1997
- [13] M. McTear, S.Allen, L. Clatworthy, N. Ellison, C.Lavelle and H. McCaffery: Integrating flexibility into a structured dialogue model:Some design considerations. In Proceedings of the 6th International Conferences on Spoken Language Processing (ICSLP'2000, Beijing, China), Vol.1, ICSLP, pp.110-113, 2000

BIOGRAPHY

Ján Papaj was born in Liptovský Miluláš, Slovakia in 1977. In 2001 he graduated at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Since 2001 he is a external PhD. student at the faculty. His research interests are in the fields of dialogue and multimodal systems.