

Dialog o przemiennej inicjatywie w systemie dialogowym języka mówionego

Andrzej M. WIŚNIEWSKI

Zakład Automatyki, Instytut Teleinformatyki i Automatyki WAT,
ul. Kaliskiego 2, 00-908 Warszawa

STRESZCZENIE: Przedstawiono szczegółową architekturę systemu dialogowego języka mówionego. Scharakteryzowano style dialogu i cechy dialogu o przemiennej inicjatywie. Omówiono problemy zarządzania dialogiem z przemiennej inicjatywą i dostosowania systemów dialogowych do użytkownika.

SŁOWA KLUCZOWE: system dialogowy, przemiennej inicjatywa

1. Wstęp

Coraz częściej ludzie są zainteresowani dostępem do informacji w Internecie za pomocą telefonu (stacjonarnego, komórkowego czy internetowego) lub z wykorzystaniem małego, mieszczącego się w dłoni urządzenia (palmtop, iPod, iPhone). Wtedy tradycyjne interfejsy, czyli klawiaturę i myszkę, może zastąpić interfejs głosowy, wykorzystujący najbardziej naturalny i najefektywniejszy sposób komunikacji między ludźmi. Interfejs głosowy jest wygodnym rozwiązaniem, zapewniającym użytkownikowi możliwość mówienia i słyszenia w języku naturalnym.

Obserwuje się rosnącą chęć wykorzystywania interfejsu głosowego także w komputerach przenośnych i stacjonarnych. Przewiduje się, że w szczególności klawiatury i myszki zostaną zastąpione w wielu interfejsach użytkownika przez urządzenia wykorzystujące bardziej intuicyjne techniki, takie jak mowa i dotyk.

Przykładem zastosowania interfejsu głosowego jest portal głosowy, który umożliwia dostęp do internetu za pomocą telefonu. Wybranie odpowiedniego numeru telefonu uruchamia aplikację głosową, będącą realizacją systemu dialogowego. Portal głosowy składa się z dwóch zasadniczych części: serwera

aplikacji, który jest serwerem WWW, oraz serwera VoiceXML. Na serwerze WWW przechowywane są aplikacje głosowe utworzone z wykorzystaniem standardu VoiceXML (sterujące dialogiem poprzez przetwarzanie transakcji) oraz inne aplikacje wykorzystywane w aplikacji głosowej (np. interfejs baz danych, pobieranie danych i treści, logika serwisowa). Na serwer VoiceXML składa się przeglądarka głosowa i usługi obsługujące dialog: automatyczne rozpoznawanie mowy (ASR, *automatic speech recognition*), synteza mowy (TTS, *text-to-speech*), rozpoznawanie telefonicznych sygnałów wybierania tonowego (DTMF, *dual-tone multi-frequency*), rejestracja i odtwarzanie plików dźwiękowych. Głównym zadaniem przeglądarki głosowej jest interpretacja dokumentów VoiceXML składających się na aplikację głosową.

2. System dialogowy języka mówionego

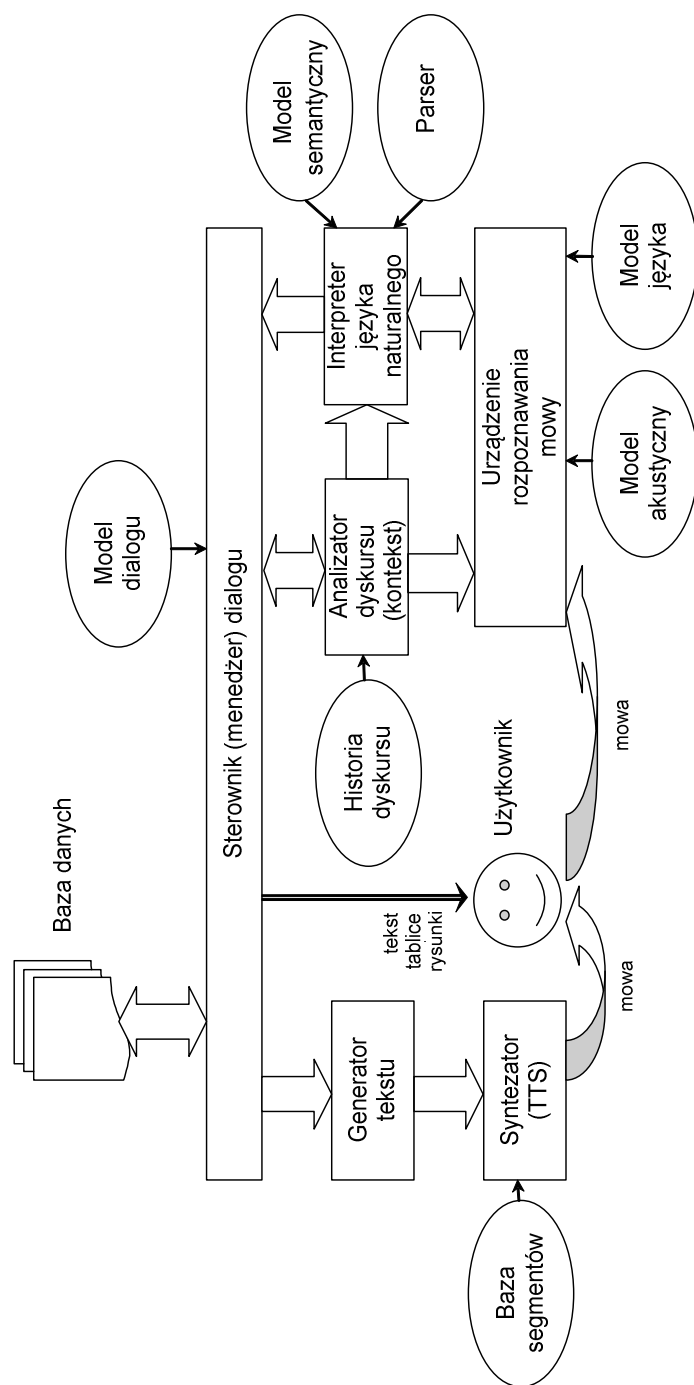
Dialog jest współdziałaniem pomiędzy użytkownikiem i komputerem w osiągnięciu szczególnego celu. Najmniejszą jednostką dialogu jest transakcja, czyli akcja użytkownika i skojarzona z nią odpowiedź komputera (lub na odwrót). Dialog jest serią transakcji.

System dialogowy języka mówionego (SDS, *spoken dialogue system*) jest interfejsem systemu komputerowego, przeznaczonym do konwersacji z człowiekiem i wykorzystującym do tego mowę.

Architekturę systemu dialogowego języka mówionego przedstawia rys. 1. Dialog człowiek – komputer jest traktowany jako proces sekwencyjny, w którym użytkownik-człowiek i sterownik (menedżer) dialogu oddziałują na siebie poprzez podsystemy przetwarzania mowy i języka, takie jak:

- automatyczne rozpoznawanie mowy, realizowane przez urządzenie rozpoznawania mowy, wykorzystujące modele akustyczne rozpoznawanych jednostek fonetycznych i model języka;
- rozumienie języka naturalnego wykonywane przez interpreter języka naturalnego, który może wykorzystywać, w zależności od stopnia złożoności, parser do analizy syntaktycznej oraz model semantyczny w celu określenia znaczenia wypowiedzi;
- generowanie języka naturalnego, które wykonuje generator tekstu;
- synteza sygnału mowy za pomocą syntezy TTS, wykorzystującego bazę segmentów do konkatencji sygnału akustycznego.

Sterownik dialogu ma zwykle dostęp do zewnętrznego źródła informacji – bazy danych.



Rys. 1. Architektura systemu dialogowego języka mówionego

Komunikację werbalną między ludźmi, która jest procesem dwukierunkowym dotyczącym aktywnych uczestników, nazywa się dyskursem. Zdolność analizowania dyskursu (prowadzonego przez analizator dyskursu, który używa do tego formalnie zapisanej historii dyskursu) umożliwia systemowi dialogowemu zrozumienie wypowiedzi w kontekście poprzednich interakcji. Jest to ważne zarówno dla zrozumienia języka naturalnego, jak również dla rozpoznawania mowy. Możliwość dziedziczenia informacji z poprzednich wypowiedzi jest szczególnie pomocna w obliczu błędów rozpoznawania. Na przykład, użytkownik może zadać złożone, wymagające kilku atrybutów (wyróżników), pytanie – urządzenie rozpoznawania może nie rozpoznać pojedynczego słowa w tej wypowiedzi, np. numeru lotu lub czasu przylotu w pytaniu o cenę biletu lotniczego. Jeśli istnieje model kontekstowy, system najpierw wygeneruje pytanie uzupełniające, użytkownik wypowie krótką frazę korekcyjną, a następnie system będzie potrafił wstawić do wypowiedzi tylko źle zrozumiane słowo, zapobiegając konieczności powtarzania całej wypowiedzi i zmniejszając w ten sposób ryzyko kolejnych błędów rozpoznawania i rozumienia.

3. Style dialogu

Na przebieg dialogu największy wpływ ma przyjęty model dialogu, najczęściej traktowany jako maszyna o skończonej liczbie stanów i realizowany w postaci scenariusza dialogu.

W zależności od stopnia aktywności systemu dialogowego oraz swobody użytkownika (o co, w jaki sposób i kiedy zapyta lub wyrazi życzenie) w dialogu wyróżnia się:

- dialog kierowany przez system;
- dialog kierowany przez użytkownika;
- dialog o przemiennej inicjatywie.

W dialogu kierowanym przez system (system-initiative, system-directed) system dialogowy zachowuje całkowitą kontrolę nad przebiegiem konwersacji. Wymaga się od rozmówcy odpowiedzi na wcześniej sformułowane pytania, udziela instrukcji, czeka na odpowiedź użytkownika, który odpowiada ściśle na pytania, mówiąc lub wybierając przycisk w telefonie. Taki dialog cechuje łatwość modelowania i kodowania. Ponieważ reakcje rozmówcy są ograniczane przez system, konwersacja zwykle kończy się sukcesem. Ten typ dialogu został zastosowany w pierwszych udanych implementacjach systemów dialogowych [1, 2, 3]. Od użytkownika nie wymaga się pamiętania komend i opcji, jednak

dialog jest postrzegany jako zbyt złożony (zabiera za dużo czasu) i nieelastyczny.

W dialogu kierowanym przez użytkownika (user-initiative, user-directed) użytkownik ma całkowitą swobodę wypowiedzi, wydaje systemowi polecenia wykonania kolejnych czynności, system pozostaje pasywny, wykonuje odpowiednią czynność, gdy potrzeba prosi o wyjaśnienia, potwierdza wynik, czeka na następne życzenie rozmówcy. Tutaj użytkownik prowadzi dialog przez inicjację każdej jego części bez wyraźnej zachęty ze strony systemu, nie musi wysłuchiwać długich podpowiedzi. Ten styl dialogu wymaga jednak od rozmówcy zapamiętania nazw komend i parametrów (jest to trudność dla początkującego użytkownika – można ją zminimalizować za pomocą uważnie zaprojektowanej pomocy lub listy komend zapisanej na kartce). Rozmówca nie odczuwa dominacji systemu, ale może się czuć niepewnie, gdy nie posiada pełnej wiedzy o przeznaczeniu oraz możliwościach i ograniczeniach systemu dialogowego.

Istotą dialogu o przemiennej inicjatywie (mixed-initiative, goal-oriented) jest aktywny udział zarówno użytkownika, jak i systemu dialogowego w osiągnięciu celu (rozwiązaniu problemu) – strategia dialogu musi być bardziej elastyczna, pytania są bardziej złożone niż w systemach z dialogiem kierowanym przez system oraz użytkownika. Użytkownik przejmuje inicjatywę, aby przekazać więcej informacji, niż w tym momencie wymaga system – może zastosować skróty zwiększające tempo dialogu (znając strukturę i komendy aplikacji może udzielać odpowiedzi na pytania bez czekania na ich zadanie). W dialogu o przemiennej inicjatywie początkujący użytkownicy mają trudności z wykorzystaniem systemu (co robić?, kiedy przerwać?). Projektowanie systemu wymaga większego wysiłku i staranności, a kod aplikacji jest złożony.

4. Cechy dialogu o przemiennej inicjatywie

Dialog o przemiennej inicjatywie jest naturalnym i efektywnym sposobem komunikacji, ponieważ cechuje go podobieństwo do dialogu człowiek – człowiek.

Poniższa transkrypcja przykładowej rozmowy telefonicznej między klientem (U) i pracownikiem kina (A) umożliwi określenie cech charakterystycznych dialogu człowiek – człowiek.

A: Dzień Dobry, witamy...

U: OK, [mmm] interesuje mnie kino Ochota.

A: OK, chcesz wiedzieć, co gramy, czy...

U: Tak właśnie.

A: Szukasz konkretnego filmu?

U: [mmm] co gracie?

A: OK, chwileczkę.

A: Gramy Sztuczki.

U: Nie.

A: Rok 1920.

U: O której godzinie to leci?

A: O siedemnastej.

U: [Mhmm] a inne?

A: Zastługujesz na więcej.

U: Nie.

.....

.....

A: To wszystko.

U: OK, dziękuję.

A: Przyjemnego wieczoru. Do usłyszenia.

Dialog człowiek – człowiek prowadzony w celu uzyskania informacji cechuje spontaniczność, czyli:

- brak płynności, przerwy, nakładające się wypowiedzi;
- niedokończone, niejasne lub zawierające niespójne fragmenty zdania;
- występowanie potwierżeń, wyjaśnień, elips (opuszczenie w zdaniu wyrazu lub wyrazów domyślnych w szerszym kontekście) i anafor (zaimek wskazujący, zapobiegający powtarzaniu podmiotu z poprzedniego zdania);
- zmiany tematu dialogu w trakcie jego trwania;
- trudności we właściwym rozumieniu niektórych wypowiedzi bez wiedzy o kontekście, w którym się pojawiły.

Na pytanie, czy przy tworzeniu interfejsu konwersacyjnego powinno się naśladować interakcję człowiek – człowiek, trudno udzielić jednoznacznej odpowiedzi. Można zauważyć dwa podejścia:

- zaleca się użytkownikom systemu zmianę zachowania i sformalizowanie interakcji (dialog ma przejrzystą strukturę), ponieważ niektóre z powyższych zjawisk mogą nie służyć bezpośrednio osiągnięciu celu (goal-directed problem solving);
- nie formułuje się ograniczeń dotyczących przebiegu dialogu –

użytkownicy czują się bardziej komfortowo z interfejsem, w którym komputer posiada cechy agenta-człowieka.

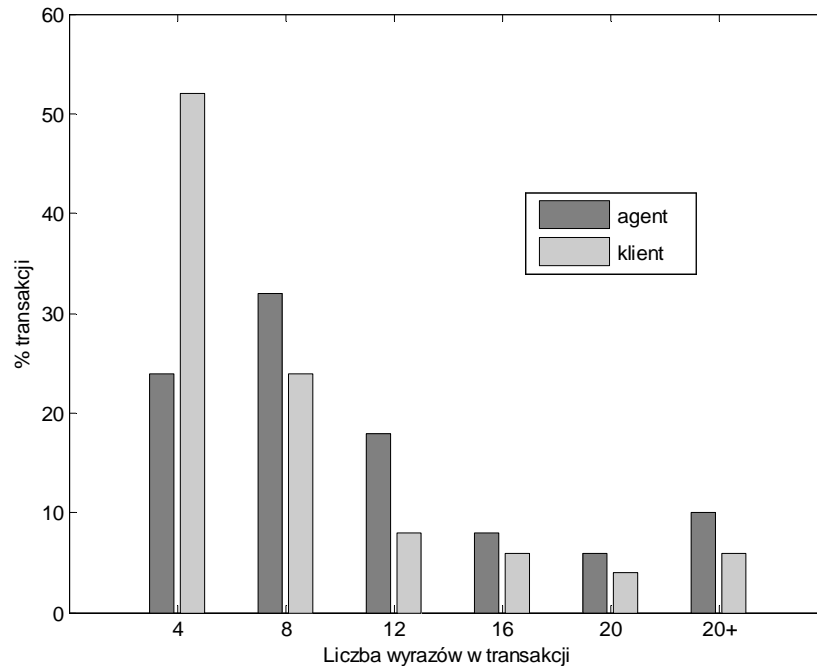
Obecnie pracujące systemy nie radzą sobie ze zjawiskami spontanicznego dialogu, niektóre robią to w ograniczonym zakresie.

Niektóre zjawiska spontanicznej rozmowy pełnią pożyteczną rolę w komunikacji między ludźmi i dlatego powinny być włączone do interfejsu człowiek – komputer. Na przykład: początkowy brak płynności w dialogu może pełnić rolę skupiania (zwracania) uwagi; wypełniana przerwa i zwrotne potwierdzenia są zapewnieniem, że wypowiedź jest zrozumiana lub partner dialogu wciąż pracuje nad problemem.

Zawsze jednak pożyteczne jest studiowanie dialogów między ludźmi i porównywanie ich do dialogów człowiek – komputer. Poniżej przedstawiono niektóre statystyki dialogów prowadzonych przez ludzi w zadaniach dostępu do różnych informacji telefonicznych (rys. 2 przedstawia histogram długości transakcji w takich dialogach [35, 4]). Przeciętny dialog składa się z 28 transakcji. Tabela 1 pokazuje statystykę treści dialogów prowadzonych w telefonicznym dostępie do informacji o filmach [35, 4].

Tab. 1. Statystyka treści dialogów informacyjnych

Typ treści	Użytkownik-klient		Agent	
	Częstość [%]	Liczba słów	Częstość [%]	Liczba słów
potwierdzenia	47,9	2,3	30,8	3,1
życzenia	29,5	9,0	15,0	12,3
zatwierdzenia	13,1	5,3	11,3	6,4
informacje	5,9	7,9	27,8	12,7
wypowiedzi	3,4	6,9	15,0	6,7



Rys. 2. Histogram długości transakcji w dialogach informacji telefonicznej

5. Zarządzanie dialogiem z przemienną inicjatywą

Ogólny cel zarządzania (sterowania) dialogiem można sformułować następująco: brać aktywny udział w kierowaniu konwersacją, tak aby zakończyła się pomyślnie dla użytkownika.

U podstaw sterowania dialogiem leży modelowanie dialogu. Technologia wytwarzania modelu jest jedną z najmniej dopracowanych w zakresie technologii języka naturalnego, szczególnie dla dialogów z przemienną inicjatywą. Modelowanie dialogu może oznaczać:

1. Planowanie działań i rozwiązywanie problemów interakcji.
2. Przygotowanie systemu do każdej transakcji (odpowiedź słowna, graficzna i za pomocą tablicy, jak również żądanie wyjaśnień).

Rola zarządzania dialogiem zależy od stanu dialogu. W początkowym stadium sterowanie gromadzi informacje i dane od użytkownika oraz, gdy potrzeba, wyjaśnia niejasności, aby mogło być wygenerowane zapytanie bazodanowe. Jednocześnie sterownik dialogu musi umieć rozwiązać niejasności

spowodowane błędami rozpoznawania (*Powiedziałeś Warszawa czy Warka?*) lub niepełną informacją (*W jakim dniu chcesz polecieć?*).

W późniejszym stadium konwersacji, gdy informacja z bazy danych została udostępniona, sterownik dialogu może uczestniczyć w negocjacjach z użytkownikiem. Na przykład, gdy informacja z bazy danych jest obszerna, system może sugerować dodatkowe ograniczenia, aby pomóc zawęzić liczbę opcji. System musi umieć zainicjować propozycję redukcji informacji (*Znalazłem 10 lotów, czy ma Pan ulubioną linię?*). Oprócz tego sterownik musi informować i prowadzić użytkownika poprzez:

- sugerowanie kolejnych celów pośrednich (*Czy może Pan określić pułap cenowy?*);
- oferowanie pomocy na prośbę;
- pomoc w łągodzeniu ograniczeń;
- dostarczenie prawdopodobnych alternatyw, gdy pożądana informacja jest niedostępna (*Nie mam danych o wschodzie słońca w Grodzisku, ale w Warszawie...*);
- inicjowanie dodatkowych dialogów wyjaśniających dla potwierdzenia pozyskanych danych.

Oprócz wyżej wymienionych zadań, menedżer dialogu wpływa na pozostałe elementy systemu dialogowego, na przykład poprzez dynamiczne modyfikowanie modelu języka lub historii dyskursu. Gdy wykryje fakt, że przedmiot dialogu dotyczy określonej poddziedziny (np. pogody, planowania wakacji, podróży samolotem, nawigacji w mieście), może dostosowywać każdorazowo model języka (w szczególności poprzez zawężanie słownika) do aktualnych potrzeb. Na przykład w dialogu o planowaniu podróży, początkowa wypowiedź *Planuję wakacje we Francji* pozwoli systemowi znacznie polepszyć rozpoznawanie nazw miejscowości we Francji poprzez zmianę modelu języka. Podobnie wygenerowanie pytań przez system może wpływać na zmianę modelu języka spodziewanej odpowiedzi (po pytaniu o datę, system może oczekiwać wyrażen z datami w odpowiedzi).

Trudność w projektowaniu dialogu o przemiennej inicjatywie polega na pogodzeniu potrzeby swobody użytkownika (elastyczności w reakcji na pytanie lub odpowiedź systemu) i wierności rozumienia przez system (drogą ograniczeń na dopuszczalne wejścia użytkownika). Można to osiągnąć następująco:

- umożliwić użytkownikowi zadawanie dowolnych pytań w dowolnym miejscu dialogu, czyli cały słownik jest aktywny przez cały czas;

- zmieniać aktywny słownik lub ograniczać możliwość rozumienia dialogu w zależności od jego stanu.

Nie ma również ogólnie przyjętego standardu implementacji sterownika dialogu, stosowane są albo języki skryptowe jako ogólny mechanizm, albo graf obiektów lub modułów dialogowych.

W tabeli 2 przedstawione zostały podstawowe charakterystyki praktycznie udostępnionych (laboratoryjnych i komercyjnych) aplikacji systemów dialogowych języka mówionego (dialogi człowiek – komputer) [4]. Ostatni wiersz tablicy dotyczy przytaczanych wcześniej dialogów człowiek – człowiek, prowadzonych w zadaniach dostępu do różnych informacji.

Tab. 2. Charakterystyki systemów dialogowych języka mówionego

Dziedzina	Język	Rozmiar słownika	Średnia liczba	
			słów/trans	trans/dialog
CSELT Train Timetable Info	włoski	700	1,6	6,6
Speech Works Air Ravel Reservation	ang.	1000	1,9	10,6
Philips Train Timetable Info	niem.	1850	2,7	7,0
CMU Movie Information	ang.	757	3,5	9,2
CMU Air Travel Reservation	ang.	2851	3,6	12,0
LIMSI Train Timetable Info	fran.	1800	4,4	14,6
MIT Weather Information	ang.	1963	5,2	5,6
MIT Air Travel Reservation	ang.	1100	5,3	14,1
AT&T Operator Assistance	ang.	4000	7,0	3,0
Air Travel Reservation (ludzie)	ang.	?	8,0	27,5

6. Aktualne problemy rozwojowe

Projektowanie systemów dialogowych powinno uwzględniać założenie, że użytkownik może być (komputerowym) analfabetą, zatem system musi uwzględniać jego zachowanie i potrzeby.

Budowa systemu dialogowego jest procesem nieustającym. Kluczowym krokiem w procesie tworzenia jest udostępnienie tego systemu rzeczywistym użytkownikom: efektywny rozwój technologii następuje w warunkach rzeczywistych (a nie dla sztucznych scenariuszy). Rzeczywiste aplikacje (zapewne początkowo uproszczone) są źródłem pożytecznych danych, daleko lepszych niż zgromadzone w laboratoriach – umożliwiają tworzenie coraz lepszych słowników oraz modeli akustycznych i modeli języka. Istnieje bezpośredni związek między ilością zgromadzonych danych, pozyskiwanych

w pracującym systemie, a liczbą błędnie rozpoznanych słów. Miary jakości systemu – niezbędne do śledzenia postępu – powinny zawierać oceny składników systemu, ale również oceniać działanie całego systemu (ważna jest także satysfakcja użytkownika).

Najważniejsze problemy w zakresie:

- modelowania dialogu – automatyzacja procesu tworzenia modelu dialogu poprzez wykorzystanie technik uczenia maszynowego (w większości obecnie powstających systemów projektowanie przebiegu dialogu jest rękodzielne, bazuje na intuicji, jest to czasochłonny proces szczególnie dla dialogów z przemiennej inicjatywą, którego wyniki nie mogą być łatwo uogólniane dla różnych dziedzin i aplikacji);
- automatycznego rozpoznawania mowy:
 - wykrywanie i przyswajanie nowych słów (nie można przewidzieć wszystkich słów, pojawiają się nowe, np. w słowniku nazw własnych);
 - zapewnienie odporności na zakłócenia spowodowane zmiennością kanałów dostępu;
 - zapewnienie niezależności od mówcy – dotyczy to osób posługujących się dialektami, posiadających obcy akcent, a także dzieci (adaptacja systemu rozpoznawania mowy do takich mówców powinna być albo krótkookresowa dla użytkowników jednorazowych, albo wykorzystywać utworzony w trakcie współpracy profil użytkownika);
- rozumienia wypowiedzi: opracowanie metod zaawansowanej analizy i automatyzacji wykrywania faktów językowych, gdy interakcja staje się bardziej złożona (dla prostych zadań do zrozumienia może wystarczyć rozpoznanie).

Aktualny stan rozwoju systemów dialogowych wymaga:

- 1) opracowania niezbędnej do badań nad dialogiem infrastruktury, na którą składają się zbiory danych o dialogach (człowiek – człowiek i człowiek – komputer) oraz narzędzia i konwencje anotowania (opatrywania przypisami, indeksowania pozyskanych danych);
- 2) zaawansowanych badań nad dialogiem, obejmujących między innymi przenośność, ocenianie, wspólne zasoby i standardy, a wymagających zespolenia wysiłków badaczy (powstają organizacje prowadzące takie badania);
- 3) zapewnienia elastyczności i swobody użytkownikowi – jest to szczególnie ważne dla systemów dialogowych z przemiennej inicjatywą, w których zwiększa się zagrożenie opuszczenia zakresu

- kompetencji systemu (na przykład w systemie z informacjami o prognozach krótkoterminowych pogody zdarzają się pytania związane z pogodą, ale odległe od możliwości systemu: *Jakie są średnie opady w Indiach we wrześniu?, Kiedy jest przyptyw?*). Należy dopasować ograniczone możliwości współczesnych systemów dialogowych do potrzeb użytkownika, w szczególności potrzebna jest pomoc kontekstowa oraz, zależne od stanu dialogu, wskazówki dla użytkownika;
- 4) przewidywania reakcji na nieuniknione błędy systemu (trudno jest wykryć błąd, określić jego przyczynę oraz sposób naprawy sytuacji) – doświadczenie uczy, że błąd pociąga następne błędy;
 - 5) identyfikacji niepewnych, w sensie rozpoznania, wejść (wypowiedzi) użytkownika – wtedy nawet ich częściowe zrozumienie może umożliwić obliczanie miar zaufania do rozpoznawania i rozumienia;
 - 6) przenośności (portability), która dotyczy środowisk akustycznych, baz danych, dziedzin wiedzy, języków:
 - wykorzystuje się fakt, że dialog może być często podzielony na mniejsze poddialogi (dotyczące dat, adresów, ...);
 - technologia powinna umożliwić budowę systemu z modułów;
 - tworzenie systemów nie powinno wymagać wiedzy specjalistycznej;
 - trzeba dążyć do rozdzielenia algorytmicznych aspektów systemu dialogowego od specyfiki aplikacji – potrzeba (pół)automatycznych metod pozyskiwania modeli akustycznych, języka, gramatyk, struktur semantycznych do rozumienia języka, modeli dialogów dla nowych aplikacji.

7. Elastyczne i dostosowane do użytkownika systemy dialogowe

Użytkownik SDS o przemiennej inicjatywie będzie miał możliwość dynamicznej konfiguracji systemu na podstawie nowych danych. Na przykład baza (dziedzina) hotelowa dla Polski mogłaby ograniczyć początkowo rozpoznawanie nazw do głównych sieci, jak sieć hoteli Orbis, ponieważ dla wszystkich hoteli w Polsce byłoby to niewykonalne. Gdy użytkownik określi rejon geograficzny, system może poprawić poziom szczegółowości słownika roboczego, w szczególności na zawężające żądanie użytkownika, umożliwiając rozpoznawanie nazw hoteli w wyznaczonym regionie.

Użytkownik powinien mieć możliwość dalszej personalizacji systemu przez dodawanie na bieżąco nowych słów, wykorzystując do tego dialog,

np. może to obejmować klasy semantyczne słów: „chcę dodać nazwę Jan Kowalski do mojej książki adresowej”.

Aby wizje stały się rzeczywistością:

- 1) należy opracować narzędzia umożliwiające szybką konfigurację systemu dialogowego w nowej dziedzinie wiedzy, sterowaną głównie przez źródła informacji zależne od dziedziny;
- 2) potrzeba ciągłego (przyrostowego) uaktualniania słownika i modelu językowego dla rozpoznawania i rozumienia mowy; to umożliwi dużą elastyczność w pojedynczym dialogu, gdzie użytkownik może pytać o nazwy własne, które nie są jeszcze (na początku dialogu) znane systemowi;
- 3) jeżeli chcielibyśmy wykorzystać duże zasoby leksykalne do modelowania wymowy, musiałyby być dostępna możliwość wykonania transkrypcji fonetycznej (transformacji liter na dźwięki), integrującej wielorakie źródła wiedzy;
- 4) potrzeba inteligentnych systemów akwizycji wiedzy, zdolnych zapełnić bazy danych, wykorzystując źródła sieciowe.

Powyższe pomysły można zilustrować przykładowymi scenariuszami.

Przykład 1.

U: Jestem zainteresowany restauracjami w Warszawie na Woli.

S: Znam 38 restauracji na Woli. Z tego 15 serwuje dania polskie, pozostałe dania włoskie, meksykańskie, żydowskie oraz owoce morza.

U: Jestem zainteresowany owocami morza.

S: Restauracja Muszelka znajduje się na Wolskiej pod numerem 68. Telefon 022 640 32 45. Ceny potraw od 14 do 45 złotych.

Użytkownik zaczyna od prośby o restaurację na Woli. System szybko samodzielnie się konfiguruje, aby być gotowym do wykorzystywania odpowiedniego podjęzyka i przedstawia listę restauracji spełniających kolejne życzenia użytkownika, prowadząc do konkretnego wyboru.

Przykład 2.

U: Proszę o telefon do restauracji Thessa w Warszawie.

S: Chyba nie znam tej restauracji. Proszę przeliterować jej nazwę.

U: t-h-e-s-s-a.

S: Telefon do Thessy jest 022 665 34 23.

Użytkownik prosi o numer telefonu znanej sobie restauracji. System po przeszukaniu w swoim słowniku określa słowo jako nieznanne w zbiorze nazw restauracji. Może w tej sytuacji pozyskać z sieci zbiór kandydatów nazw restauracji w Warszawie, aby sformułować dodatkowe ograniczenia na

dostarczone literowanie. Integracja literowania, wymowy i listy z sieci powinna dostarczyć wystarczających ograniczeń do rozwiązania tego problemu z wysoką dokładnością. System może wyszukać pożądaną informację.

8. Podsumowanie

Coraz więcej centrali telefonicznych czy centrów kontaktowych (Call Center, Contact Center) dużych i średnich firm zastępuje operatorów portalami głosowymi (Voice Portal). Zadaniem portali głosowych jest umożliwienie interakcji głosowej z użytkownikiem. Portale głosowe są wyposażone w mechanizmy interakcji, których podstawą jest rozpoznawanie i rozumienie mowy oraz konwersja pobranej z bazy danych informacji tekstowej do postaci dźwiękowej.

Portal głosowy jest nie tylko systemem do prowadzenia konwersacji z komputerem, lecz przede wszystkim stanowi bazę danych z informacjami dla potencjalnych klientów serwisu. Informacje te przechowywane są w postaci tekstowej na serwerach baz danych, skąd pobierane są przez skrypty, zlokalizowane na serwerach WWW, obsługujące zapytania, np. SQL. Wyselekcjonowane wiadomości konwertowane są do postaci dźwiękowej przez przeglądarkę głosową za pomocą syntezatora TTS.

Technologia IVP (Internet Voice Portal), mimo że jest jeszcze bardzo młoda, przeżywa swój rozkwit. Pojawiło się wiele bogatych serwisów informacyjnych zarówno udostępniających własne zasoby, jak i korzystających z zasobów Internetu. Część z nich umożliwia także realizację podstawowej usługi internetowej, czyli dostępu do poczty elektronicznej. Portale te są powszechnie dostępne na terenie całych Stanów Zjednoczonych, a korzystanie z nich jest bezpłatne (BeVocal, Tellme, HeyAnita, VoiceGenie, AOL).

Popularny staje się stale rozwijany język (standard) VoiceXML umożliwiający realizację systemów dialogowych, zapewniających dostęp do treści i informacji w Internecie poprzez telefon.

W załączonym spisie literatury umieszczone zostały liczne pozycje, dostarczające opisów praktycznych realizacji systemów dialogowych języka mówionego.

Ostatnia dekada przyniosła realizację kilku systemów dialogowych o coraz bardziej zaawansowanych, lecz ograniczonych możliwościach. Mimo umiarkowanych sukcesów opracowanie takich interfejsów będzie wymagało ciągłych ulepszeń podstawowych technologii języka naturalnego i wielu wysiłków badawczych. Jednym z celów jest technologia języka mówionego niewrażliwa na narodowość użytkownika.

Technologie języka naturalnego będą odgrywały główną rolę w zapewnieniu

interfejsu, który zasadniczo zmieni paradygmat komunikacji człowiek – komputer w kierunku konwersacji (dyskusji).

Literatura

- [1] ALLEN J., et al., *The TRAINS project: A case study in defining a conversational planning agent*, J. of Experimental and Theoretical AI, vol. 7, 1995, pp. 7–48.
- [2] BAPTIST L., SENEFF S., *A versatile system for language generation in conversational system applications*, Proc. ICSLP, Beijing, 2000.
- [3] BARNARD E., et al., *A consistent approach to designing spoken-dialog systems*, Proc. ASRU Workshop, Keystone, CO, 1999.
- [4] BAZZI I., GLASS J., *Modeling out-of-vocabulary words for robust speech recognition*, Proc. ICSLP, Beijing, China, 2000, pp. 401–404.
- [5] BILLI R., et al., *Automation of Telecom Italia Directory Assistance Service: Field Trial Results*, Proc. IVTTA, 1998, pp. 11–16.
- [6] BLACK A., LENZO K., PAGEL V., *Issues in building general letter to sound rules*, Proc. ESCA Speech Synthesis Workshop, 1998.
- [7] BLOMBERG M., et al., *An experimental dialogue system: Waxholm*, Proc. Eurospeech, 1993, pp. 1867-1870.
- [8] DERTHICK M., et al., *Example-Based Generation of Custom Data Analysis Applications*, Proc. IUI, Santa Fe, 2001, pp. 57–64.
- [9] FILISKO E., SENEFF S., *A context resolution server for the conversational systems*, Proc. Eurospeech, Geneva, 2003, pp. 197–290.
- [10] FLAMMIA G., *Discourse Segmentation of Spoken Dialogue: An Empirical Approach*, Ph.D. Thesis, MIT, 1998.
- [11] GLASS J., *A probabilistic framework for segment-based speech recognition*, Computer, Speech, and Language, 17, 2003, pp. 137-152.
- [12] GLASS J., WEINSTEIN E., *Facilitating spoken dialogue systems development*, Proc. Eurospeech, Aalborg, 2001, pp. 1335–1338.
- [13] GORIN A., RICCARDI G., WRIGHT J., *How may I help you?*, Speech Communication, 23, 1997, pp. 113–127.
- [14] MIT Project Oxygen web site, <http://oxygen.lcs.mit.edu>.
- [15] NAKANO M., HAZEN T., *Using untranscribed user utterances for improving language models based on confidence scoring*, Proc. Eurospeech, Geneva, 2003, pp. 417–420.
- [16] Nuance Communications, <http://www.nuance.com>.

- [17] POLIFRONI J., et al., *Towards the automatic generation of mixed-initiative dialogue systems from web content*, Proc. Eurospeech, Geneva, 2003.
- [18] POLIFRONI J., CHUNG G., *Promoting portability in dialogue management*, Proc. ICSLP, Denver, CO, 2002, pp. 2721–2724.
- [19] POPESCU A.M., ETZIONI O., KAUTZ H., *Towards a Theory of Natural Language Interfaces to Databases*, Proc. IUI, Miami, 2003.
- [20] ROSSET S., et al., *Design strategies for spoken language dialog systems*, Proc. Eurospeech, 1999, pp. 1535–1538.
- [21] RUDNICKY A., et al., *Creating natural dialogs in the Carnegie Mellon Communicator system*, Proc. Eurospeech, 1999, pp. 1531–1534.
- [22] SCHALKWYK J., et al., *Speech recognition with dynamic grammars using finite-state transducers*, Proc. Eurospeech, Geneva, 2003, pp. 1969–1972.
- [23] SENEFF S., *A natural language system for spoken language applications*, Computational Linguistics, 18(1), 1992.
- [24] SENEFF S., et al., *Automatic induction of gram language models from a natural language grammar*, Proc. Eurospeech, Geneva, 2003, pp. 641–644.
- [25] SENEFF S., et al., *A reference architecture for conversational system development*, Proc. ICSLP, 1998, pp. 931–934.
- [26] SENEFF S., POLIFRONI J., *Dialogue management in the flight reservation system*, Proc. ANLP-NAACL Sat. Workshop, Seattle, 2000.
- [27] SOUVIGNIER V., et al., *The thoughtful elephant: Strategies for spoken dialogue systems*, IEEE Trans. SAP, 8(1), 2000, pp. 51–62.
- [28] STARKIE B., et al., *Lyrebird: Developing spoken dialog systems using examples*, Proc. ICGI, 2002, pp. 309–311.
- [29] STEELE K., WATERMAN J., WEINSTEIN E., *The Oxygen H21 handheld*, MIT Lab. for Computer Science Research Summary, March 2003.
- [30] SUTTON S., et al., *Universal speech tools: The CSLU toolkit*, Proc. ICSLP, Sydney, 1998, pp. 3221–3224.
- [31] TOTH A., et al., *Towards every-citizen's speech interface: An application generator for speech interfaces to databases*, Proc. ICSLP, Denver, CO, 2002, pp.1497–1500.
- [32] ZUE V. et al., *A telephone-based conversational interface for weather information*, IEEE Trans. SAP, 8(1), 2000, pp. 85–96.
- [33] ZUE V., GLASS J., *Conversational interfaces: Advances and challenges*, Proceedings of the IEEE, 88(8), 2000, pp. 1166–1180.
- [34] <http://www.microsoft.com/sql/evaluation/features/english.asp>
<http://www.w3.org/TR/voicexml/>.

Mixed initiative dialogue in spoken language dialogue system

ABSTRACT: In this paper, the detailed structure of spoken language dialogue system was described. Dialogue styles and features of mixed initiative dialogue were characterized. Problems of mixed initiative dialogue management and spoken dialogue systems personalization were described.

KEYWORDS: dialogue system, mixed initiative

Praca wpłynęła do redakcji: 22.12.2009.