

В.В. Пилипенко, С.А. Биднюк, Р.А. Селюх, А.В. Пилипенко

Построение сценариев формализованного устного диалога на примере заказа билетов на железнодорожные поезда

Представлена экспериментальная система речевого диалога на базе метаязыка задания сценариев. Для создания такого языка проведена формализация речевых диалогов и выделен класс диалогов, в которых ведущую роль выполняет компьютер. Элементы метаязыка реализованы в виде XML-документа. Описана интеграция метаязыка и многодикторной системы распознавания речи на примере заказа билетов на железнодорожные поезда.

The paper presents experimental interactive voice response system based on the metalanguage for dialog scenarios. A formalization of speech conversation was carried out. The dialog scenarios class in which the computer plays the leading role is highlighted. The metalanguage elements are implemented as XML-document. Integration of the meta-language and speech recognition system on the example of a ticket train service is described.

Представлено експериментальну систему мовленнєвого діалогу на базі метамови формування сценаріїв. Для формування мови проведено формалізацію мовленнєвих діалогів і виділено клас діалогів, в яких провідна роль належить комп'ютеру. Елементи метамови реалізовано у вигляді XML-документа. Описано інтеграцію метамови та багатодикторної системи розпізнавання мовлення на прикладі замовлення квитків на залізничні потяги.

Введение. Организация диалога человека с компьютером чрезвычайно актуальна, поскольку широкое использование диалоговых систем делает рутинную работу человека, связанную с получением необходимой информации, более простой и комфортной. Сегодня существует неотложная потребность в построении сценариев речевого диалога человека с компьютеризованными системами. Такими системами могут быть разные информационно-справочные системы, системы заказа билетов на железнодорожные поезда или самолеты.

Для речевых диалоговых систем на базе XML-стандарта предложен язык формирования сценария диалога *VoiceXML* [1, 2]. За время существования этого стандарта он был развит к сложному метаязыку со многими свойствами, но, с самого начала, в этом стандарте был зафиксирован единственный подход к созданию диалога, ограничивающий возможности проведения научных исследований. Было создано несколько промышленных интерпретаторов языка *VoiceXML*, позволивших разрабатывать системы речевого диалога на достаточно формализованном уровне.

Задачей этой работы было создание более гибкой версии метаязыка *VoiceXML*, пригодной для научных исследований речевых диалоговых систем. С одной стороны, этот метаязык должен быть достаточно формализованным и

универсальным для формирования развитых сценариев речевого диалога со следующей целью исследований новейших подходов, а с другой, – достаточно упрощенной, чтобы облегчить усвоение языка.

При разработке формализованного речевого диалога на примере заказа железнодорожных билетов используется подход, максимально приближенный к естественному поведению человека в эргатической системе [3].

Особенности проектирования речевого диалога

При проектировании эргономичных моделей речевого диалога человека с системой необходимо учитывать, что пользователи ожидают:

- достаточно простого, понятного и максимально эффективного комфортного средства общения;
- динамичной реакции со стороны системы, что означает:
 - распознавание команд пользователя в реальном масштабе времени;
 - синтез голосовых ответов от системы, которые не раздражают пользователя и не перегружают длинной синтезированной речевой информацией во времени;
 - максимального приближения к естественному поведению в человеческом обществе.

С учетом этих ожиданий пользователя, определим функции как системы, так и пользова-

теля в модели формализованного речевого диалога.

Функции системы:

- синтез информации и подсказки для действий пользователя;
- загрузка функциональных словарей;
- распознавание команд пользователя в duplexном режиме, т.е. одновременное распознавание и синтез речевой информации для пользователя;
- распознавание команды «Ошибка» и коррекция (под)словарей;
- формирование электронного прототипа билета на железнодорожный поезд.

Функции пользователя:

- четкие и лаконичные ответы на запросы системы;
- дружественное отношение к системе во время речевого диалога.

При проектировании речевого диалога для заказа билетов на железнодорожные поезда используется подход с применением функциональных словарей и динамических (под)словарей.

Определение 1. Под функциональными словарями будем понимать словари, которые загружаются для распознавания на каждом шаге диалога и индивидуальны на данном этапе распознавания.

Определение 2. Под динамическими (под)словарями понимаем такие, которые, в зависимости от правильности распознавания команд словаря, будут переменными, т.е., если какая-то команда словаря распознана неправильно, то на следующих итерациях распознавания она исключается из словаря; или, если указывается станция отправления на первом шаге диалога, то на втором – название этой станции из словаря станций прибытия исключается.

Общая структура систем речевого диалога

Эти системы состоят из нескольких компонентов:

- пользователя, который произносит какие-то утверждения и слушает ответы компьютера;
- компьютера, укомплектованного микрофоном, громкоговорителями и устройством АЦП/ЦАП звукового сигнала (*SoundBlaster*);

- программного обеспечения, что позволяет:
 - распознавать введенный речевой сигнал;
 - анализировать распознанный текст (понимать его);
 - выполнять полезные для человека действия;
 - руководить ходом диалога (поддерживать диалог);
 - формировать вопрос к пользователю;
 - озвучивать вопрос с помощью синтеза речи.

На рис. 1 приведена общая схема системы формализованного речевого диалога, показывающая поток информации во время общения человека с компьютером. В данном случае инициатива общения принадлежит компьютеру – он задает устный вопрос пользователю. Формализованный диалог предусматривает общение пользователя с системой голосового заказа железнодорожных билетов под управлением системы. Это значит, что основной информационный груз ложится на систему. Со стороны пользователя требуются только четкие и лаконичные ответы на вопросы системы. Такое построение диалога максимально упрощает получение необходимой и достоверной информации для пользователя. С другой стороны, такой подход – наиболее эргономичен и привычен для пользователя в речевом диалоге с системой.



Рис. 1

Пользователь отвечает на вопрос компьютера голосом. Ответы пользователя превращаются модулем распознавания речи в текстовый вид.

Управление потоков информации в компьютере осуществляется менеджером диалогов. Диалоговый менеджер следует такому алгоритму:

Шаг 1. Для текущего шага сценария диалога:

- построение фразы вопроса компьютера к пользователю и соответствующего ей звукового файла;
- построение грамматики возможных ответов пользователя;
- ожидание ответа подсистемы распознавания речи;
- анализ ответа подсистемы распознавания речи как последовательности слов; нахождение элемента дерева сценария диалога, связанного со смыслом ответа (этот элемент становится текущим);
- выполнение действия, связанного с этим элементом (при необходимости – остановка выполнения алгоритма);

Шаг 2. Переход к выполнению первого шага.

Менеджер диалога руководствуется сценарием в виде XML-документа, описывающим фразы, воспроизводимые компьютером, возможные ответы пользователя, шаги диалога и действия компьютера. С этой целью разработан язык описания сценария диалога в соответствии с XML-стандартом разметки текстов.

Модель формализованного речевого диалога

На рис. 2 представлена блок-схема речевого диалога при заказе билетов на железнодорожные поезда, так называемый, формализованный диалог.

Метаязык формирования сценария речевого диалога

Язык XML – общеизвестный стандарт создания структурированных текстовых документов, определяющий набор базовых лексических и синтаксических правил для построения языка описания информации путем применения простых тэгов. Этот формат достаточно гибок, поэтому пригоден для применения в разных отраслях. Создана более гибкая версия метаязыка VoiceXML, пригодная для научных исследований речевых диалоговых систем.

Язык состоит из таких XML-элементов: *model*, *dialog*, *command*, *phrase*, *answer*, *set*, *or*, *switch* или *case*. Рассмотрим их:

- *model* – элемент, объединяющий несколько сценариев диалога;
- *dialog* – объединяет элементы одного сценария диалога;

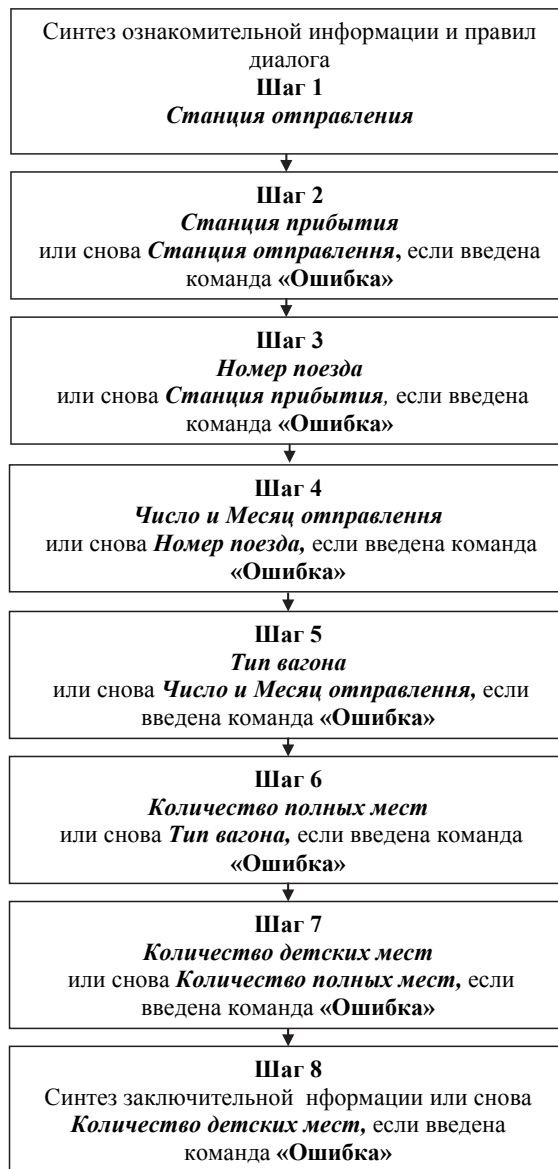


Рис. 2

- *command* – задает ответы пользователя, действительные для всех шагов диалога. Этот элемент складывается из атрибута *key* и фразы. Атрибут *key* задает действие, которое состоится после произнесения пользователем заданной фразы. Например, `<command key=cancel>` вернуться `</command>` задает возврат на один шаг диалога во время произнесения пользователем фразы «Вернуться».

• *phrase* – задает фразу, воспроизводимую компьютером. Этот элемент складывается из атрибута *filename* и текста фразы. Атрибут *filename* задает имя звукового файла с записанной фразой, а текст фразы выводится на экран;

• *answer* – задает фразу, которую может произнести человек в процессе диалога. В самом простом случае этот элемент состоит из одной фразы, которую может произнести человек. Фраза может содержать элемент `<set>`;

• *set*, *or* – задают множество фраз, которые может сказать человек, в виде альтернативных вариантов, разделенных элементом `<or/>`.

Например, `<set> один <or/> два <or/> три </set>` задает возможность произнести одно из слов «один», «два» или «три».

• *switch*, *case* – задают альтернативы ведения диалога. Альтернативы перечисляются элементами `<case>`. Каждый элемент `<case>` состоит из последовательности элементов `<answer>` и возможного продолжения диалога.

Для каждого из элементов можно задать атрибут *name* и в дальнейшем ссылаться на него по имени с помощью атрибута *ref*. Например, `<answer name=«aDestination»> пункт отправления <set> киев <or/> донецк </set> </answer>`, а дальше по тексту: `<answer name=«someDestination» ref=«aDestination» />`.

Современные системы распознавания обрабатывают речь человека, составленную из слов наперед определенного словаря. Обычно такой словарь содержит от одной тысячи до сотен тысяч слов.

Для улучшения надежности распознавания множество допустимых для распознавания фраз задается грамматикой. Это может быть как грамматика, задаваемая правилами, так и вероятностная грамматика. В данной статье используется грамматика, задаваемая двумя правилами – конкатенацией и альтернативой.

Например, фразы для задания любого дня года можно описать такой грамматикой:

```
<set>
Первое <or/>
Второе <or/>
Третье <or/>
...
Тридцать первое
```

```
</set>
<set>
Января <or/>
Февраля <or/>
Марта <or/>
...
Декабря
</set>
```

Здесь элемент *set* задает множество альтернатив слов. Таким образом, можно произнести одно из этих слов. В данном примере заданы последовательно два множества, т.е. можно произнести любое слово из первого множества, а затем – любое слово из второго. С помощью элементов *set* и *or*, которые позволяют задавать альтернативы, и с помощью конкатенации можно задать дату.

Моделирование диалога между человеком и машиной в рамках метаязыка

В процессе диалога формируются разные действия (например, печать бланка заказа), для чего используются данные, сообщаемые человеком, и запоминаемые в некоторых переменных. Когда все переменные для данного действия заполнены, формируется собственно действие.

Например, заказ железнодорожных билетов состоит из выяснения пункта отправления и пункта назначения, даты, типа вагона и количества необходимых мест. Все эти данные выясняются на соответствующих шагах диалога и запоминаются в переменных. Когда все шаги диалога пройдены, тогда формируется бланк заказа билетов по определенному образцу. В этот шаблон входят переменные, которые заполнялись на предыдущих шагах.

При формировании диалога правило конкатенации используется тогда, когда следующий шаг диалога не зависит от ответа пользователя. Например, после задания пункта отправления, всегда уточняется пункт назначения.

Правило альтернативы используется в случае, когда следующий шаг диалога зависит от ответа пользователя. Например, компьютер может спросить: *Заказать билеты на поезд или самолет?* В зависимости от ответа пользователя, диалог будет происходить по ветви заказа билетов на поезд или на самолет.

Система распознавания речи

В данной статье как базовая система используется инструментарий НТК [4] на основе скрытых Марковских моделей (СММ). Инструментарий НТК использовался для построения акустических и лингвистических моделей. Для распознавания речи разработан программный комплекс, совместимый с акустическими и лингвистическими моделями НТК

Речевой сигнал преобразуется в последовательность векторов признаков с интервалом анализа 25 мс и шагом анализа 10 мс. Вначале речевой сигнал фильтруется на высоких частотах с характеристикой. Затем применяется окно Хэмминга и вычисляется быстрое преобразование Фурье. Спектральные коэффициенты усредняются с использованием 26 треугольных окон, расположенных в мел-шкале, и вычисляются 12 кепстральных коэффициентов.

Логарифм энергии добавляется в качестве 13-го коэффициента. Эти 13 коэффициентов расширяются до 39-мерного вектора параметров путем дописывания первой и второй разностей от коэффициентов, соседних по времени. Для учета влияния канала применяется вычитание среднего кепстра.

В качестве акустических моделей используются скрытые Марковские модели. 824 украинских контекстно-зависимых фонем (включая фонему-паузу) моделируются тремя состояниями Марковской цепи без пропусков. Используется диагональный вид Гауссовских функций плотности вероятности. Акустические модели построены на звуковом материале, записанном в условиях офиса на персональном компьютере. Использовалась речь 30 дикторов.

Словарь транскрипций создается автоматически из орфографического словаря с использованием контекстно-зависимых правил.

Исходя из первого этапа разработки речевого диалога при заказе билетов на железнодорожные поезда был разработан детализированный алгоритм формализованного диалога. Этот алгоритм детально описывает, по шагам речевого диалога, функции системы и действия пользователя. Внешний вид экспериментальной диалоговой системы представлен на рис. 3.

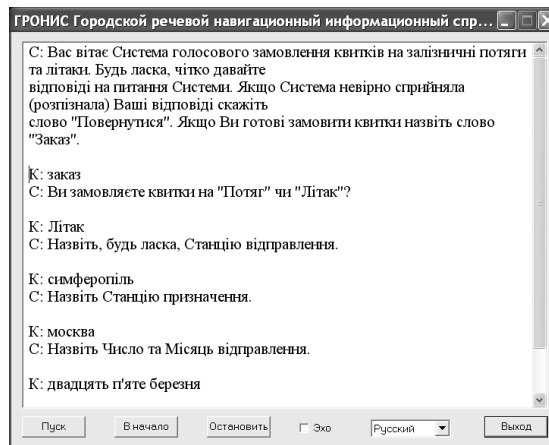


Рис. 3

Заключение. Исходя из первого этапа разработки речевого диалога при заказе билетов на железнодорожные поезда был разработан детализированный алгоритм формализованного диалога. Этот диалог детально описывает, по шагам речевого диалога, функции системы и действия пользователя, что позволило сформулировать метаязык описания сценариев речевого диалога человека с компьютером.

Для проверки предлагаемых подходов разработана экспериментальная система заказа железнодорожных билетов и показана эффективность этих подходов.

Дальнейшие исследования направлены на создание гибких сценариев диалога на базе статистического описания поведения пользователя системы.

1. Шалагинов В.А., Ярлыкова С.М. Особенности разработки инфокоммуникационных услуг на языке VoiceXML // Сети и системы связи. – 2008. – № 1. – С. 44–47.
2. Пилипенко В. Конструювання мовленнєвих діалогових систем на базі XML документів // Тези 12-ї міжнар. конф. «Автоматика 2005». – Харків, 2005. – С. 75–76.
3. Біднюк С. Про один підхід до розроблення моделей діалогу користувачів з ПК // Пр. Сьомої Всеукр. міжнар. конф. УкрОБРАЗ'2004. – Київ, 2004. – С. 119–122.
4. The НТК Book / S. Young, G. Evermann, D. Kershaw et al. – Cambridge University Engineering Department, 2002. – 277 с.

Тел. для справок: +38 044 502-6333 (Київ)
© В.В. Пилипенко, С.А. Биднюк, Р.А. Селюк,
А.В. Пилипенко, 2013