

УДК 004.522+004.584

*Д.А. Суранова***Разработка модели диалоговых человеко-машинных интерфейсов в биллинговых системах***D.A. Suranova***Development of the Dialog Human-Machine Interface Model in Billing Systems**

Взаимодействие сложной вычислительной системы и человека происходит с помощью специальных инструментов, но не всегда эти инструменты удобны и понятны пользователю. Работа может стать более комфортной, если общение человека с системой будет похоже на общение двух людей, говорящих на одном языке и обладающих некоторым набором знаний в обсуждаемой предметной области. В статье описаны способы построения модели диалоговых человеко-машинных интерфейсов с использованием голосовых возможностей для ввода данных и их последующего анализа.

Ключевые слова: человеко-машинный, биллинговая система, модель.

Введение. Диалог – специфическая форма и организация общения, коммуникации. В диалоге два объекта обмениваются информацией, выстраивая логическую цепочку сообщений.

С появлением автоматизированных систем диалог можно рассматривать как способ общения человека и машины. С изобретением систем, заменяющих некоторые функции человека, общение приобрело более сложные формы, и взаимодействовать с механизмами стали специально обученные люди. На современном этапе управление такими механизмами стараются упростить, адаптировать под обычного человека, свести управление сложной системой к простому и понятному диалогу ее с пользователем [1].

Персональные компьютеры в настоящее время активно используются как связующее звено между сложной вычислительной системой и человеком. Оператор – это человек, обладающий навыками работы с системой и занимающийся вводом и обработкой данных. Посредством программного интерфейса оператор отправляет запросы и получает результат. В процессе такого взаимодействия важными факторами являются:

- время, необходимое на формирование запроса к системе и получение ответа;
- количество ошибок, допускаемых при вводе данных;
- удобство использования системы, т.е. доступность и интуитивная понятность совершаемых действий;

User interacts with computing system by special instruments, but these instruments aren't always convenient and popular for him. His work may be more comfortable if communication with system will be like communication between people, who speak one language and have some knowledge in sphere interested. The article describes ways for building models of human-machine interfaces using voice speech and recognition and data analysis.

Key words: human-machine, billing system, model.

– получение ожидаемого результата в наглядном виде.

Для большинства современных автоматизированных систем расчета главной задачей является получение правильного результата за разумное время, и далеко не все приложения, используемые для работы с такими системами, обладают вышеперечисленными свойствами. В связи с этим операторы, работающие с такими приложениями, испытывают неудобства и иногда допускают ошибки. Проблема актуальна в современных биллинговых системах, используемых для ежемесячного расчета платы за жилищно-коммунальные ресурсы.

В биллинговых системах взаимодействие операторов с вычислительным центром происходит через прикладное приложение, содержащее формы для ввода данных, экранные и печатные формы в виде отчетности, инструментарий для выполнения расчетных операций. Работа оператора с приложением в процессе ввода данных происходит в одностороннем режиме, без обратной связи, что является для человека неестественным и увеличивает вероятность ошибки. Объем данных, как правило, большой, и процесс ввода можно оптимизировать, адаптировав к более привычным средствам коммуникации. Такая оптимизация позволит не только улучшить качество работы операторов, но и может стать инструментом для людей с ограниченными возможностями.

Прогресс в области речевых технологий позволяет использовать голосовые решения как в управлении сложными интеллектуальными системами, так и для простого поиска информации.

В научно-фантастическом исследовании [2], сделанном «The University of Nottingham» по заказу компании Microsoft, говорится о революционных переменах в области человеко-компьютерного взаимодействия, связанных с отказом к 2020 г. от стандартных способов ввода-вывода – клавиатур, мышей, чтения текста с экрана монитора и переходом к тачскринам и системам синтеза и распознавания речи. Один из наиболее известных проектов в области человеко-машинного взаимодействия и синтеза и распознавания речи на сегодняшний день – персональный помощник «Siri» [3] (разработка Международного центра искусственного интеллекта SRI). На момент написания данной статьи русский язык «Siri» не поддерживает. Также огромных успехов в области распознавания и синтеза речи добилась компания Google [4], у которой существует поддержка русского языка. Сервисы компании доступны любому пользователю сети и на сегодняшний день бесплатны. В России наиболее интересны исследования «Центра речевых технологий», например, приложение «Voice Navigator», а также решения компании Speereo, такие как «Умный дом», но в настоящее время данные приложения весьма дорогостоящи.

Цель данной статьи – разработка и описание модели программного интерфейса на примере биллинговой системы, которая в режиме диалога сможет взаимодействовать с пользователем, анализировать его действия, задавать ему вопросы и получать ответы. Такое «живое» общение человека и системы должно минимизировать неудобства, возникающие при работе с системой неопытных пользователей, быть интуитивно понятным, уметь анализировать возможные ошибки и в целом минимизировать время работы с системой.

Описание алгоритма и структуры данных. Диалог компьютера с пользователем можно осуществить как минимум двумя способами: традиционным – читать текст с экрана монитора и вводить соответствующие данные в поля формы при помощи клавиатуры и мыши, а также с развитием новых технологий использовать возможности синтеза и распознавания речи: прослушивать текст, задавать команды и вводить текст голосом. Идеальным вариантом могла бы стать система, которая в зависимости от желаний пользователя может работать в обоих режимах.

В рамках исследования была создана модель программы в виде интернет-сайта, а в качестве системы синтеза и распознавания речи использованы сервисы Google. Клиент-серверная технология распознавания и синтеза речи Google была выбрана ввиду следующих преимуществ перед конкурентами:

- приемлемое качество распознавания русской речи;
- отсутствие необходимости обучения системы (база для обучения сформирована на основе большого числа пользователей Интернета);
- высокая скорость распознавания за счет использования ресурсов компании;
- простота внедрения в web-формы.

Недостатком является необходимость доступа к Интернету.

В качестве форм, подлежащих «озвучиванию», были выбраны наиболее часто используемые формы ввода адреса (как пример формы с последовательным распознаванием слов) и форма ввода показаний по приборам учета (форма с периодически повторяющимися данными для каждого прибора учета). Для каждой формы выработан формат ввода голосовых данных – строгая последовательность слов, состоящая из ключевых слов и значений справочников.

Форма адреса. Адрес в простейшем случае состоит из названия города, улицы, номера дома и квартиры.

Для формы ввода адреса обязательная последовательность выглядит так:

КС«адрес»_КС«город»_ЗС«города»_КС«улица»_ЗС«улицы»_КС«дом»_ЗС«дома»_КС«квартира»_ЗС«квартиры»,

где КС – ключевое слово; ЗС – значение соответствующего справочника; _ – символ-разделитель, например, пробел.

Пример последовательности, которую можно корректно проанализировать

адрес_город_барнаул_улица_попова_дом_95_квартира_3,

при условии, что в соответствующих адресных справочниках есть данные по такому адресу.

Процесс распознавания должен выделять каждую из этих составляющих в строгой последовательности, так как иначе цепочка нарушится, так, например, без города улицу однозначно определить невозможно.

Форма ввода показаний по приборам учета. Для занесения показаний необходимо указать номер лицевого счета абонента, тип показания (начальное, конечное или разность) и величину показания. Показаний для каждого лицевого счета может быть одно, два или три. Последовательность для формы ввода показаний выглядит так:

КС«лицевой»_ЗС«лицевой счет»_КСТ_3[_КСТ_3] [_КСТ_3],

где КС – ключевое слово; ЗС – значение соответствующего справочника; _ – символ-разделитель, например, пробел; КСТ – одно из ключевых слов «начальное», «конечное» или «разность»; 3 – значение показания (положительное число).

В квадратные скобки заключены необязательные выражения.

Для формы ввода показаний последовательность может быть такой:

лицевой_1234_начальное_5_конечное_6

или такой:

лицевой_2345_разность_3.

Для работы с системой анализа распознанного текста понадобилось несколько структур в базе данных. Основные структуры: «Данные», «Ключевые слова», «Верификация». Вспомогательные структуры: «Настройки», блок навигации по страницам («Страницы», «Переходы по страницам», «Формы») и набор справочников. В структуру «Данные» в процессе распознавания последовательно помещаются отдельные слова, разделенные специальным символом. Разбор текста осуществляется с использованием регулярных выражений. При заполнении этой структуры происходит анализ слов, и по каждому из них определяется набор признаков. Ключевые слова используются как вспомогательные и в структуру не заносятся. Для каждого слова из справочника записывается его значение и определяется ключевое слово, к которому оно относится, а для некоторых происходит проверка наличия в соответствующем справочнике такого значения и ставится признак (найден/не найден). Необходимость сверки со справочником определяется признаком ключевого слова.

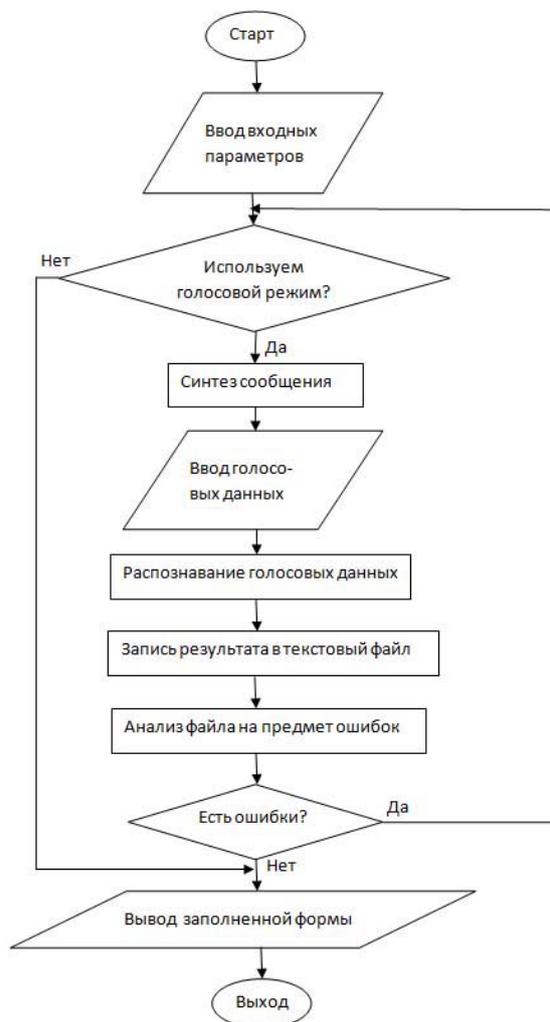
Структура «Ключевые слова» включает:

- название ключевого слова;
- название элемента формы, к которому данное ключевое слово прикрепляется;
- ссылка на родителя – ключевое слово;
- признак сверки со справочниками.

Данная структура позволяет определить принадлежность ключевого слова и его значения к элементу формы. Ссылка на родительское ключевое слово дает возможность работать с массивом вводимых значений, причем уровень иерархии в такой цепочке определяет размерность массива. Так, например, для формы ввода адреса достаточно одномерного массива значений, здесь родительское ключевое слово будет «адрес», тогда как в форме ввода показаний по приборам учета цепочка может выглядеть так: «показания» – «лицевой» – «начальное показание». Правильное написание некоторых значений ключевых слов необходимо проверить – за это отвечает признак сверки со справочниками. В том случае, если он установлен, будет производиться поиск введенного наименования в справочнике. Если запись не будет найдена, то слово введено с ошибкой и система предложит повторить ввод, в качестве подсказки определив наиболее похожие значения справочника.

Структура «Верификация» содержит текст запроса с параметрами и их количеством для поиска зна-

чений по заданным параметрам в соответствующем справочнике. С ее помощью можно одним действием определить корректность введенного слова. Функциональная схема модели приведена на рисунке.



Функциональная схема модели

Необходимые требования к техническому обеспечению и работе оператора. Для того чтобы работа оператора с программой происходила наиболее эффективно, необходимо учитывать следующие факторы:

1. Аппаратное обеспечение. Качество микрофона и его чувствительность оказывают влияние на скорость передачи голосовых данных на сервер и качество распознавания речи. Опытным путем было выяснено, что для Google оптимальными являются настройки в 8–16 КГц, так как при меньшей частоте страдает качество, а при частоте 44 КГц и более выдается ошибка на сервере. При этом расстояние диктора до микрофона должно быть от 15 до 20 см.

2. Программное обеспечение. Технология синтеза и распознавания речи Google предполагает наличие серверной и клиентской частей приложения.

Работа оператора происходит с клиентской частью, на сервер отправляются данные, полученные через микрофон, и результат возвращается обратно в клиентское приложение. Эксперименты показали, что оптимальный промежуток между словами составляет от 0,5 до 1 секунды, а после паузы в 2 секунды данные отправляются на сервер для распознавания. Для форм, предусматривающих ввод периодических данных (форма ввода показаний по приборам учета), оптимальным вариантом будет ввод не сразу всего большого объема, а блоками по 5–8 записей. Это позволит облегчить проверку их после процедуры распознавания, а также не создавать большую нагрузку при передаче их на сервер. Так как выбранный алгоритм распознавания использует большую базу данных [4], основанную на запросах пользователей сети Интернет, все ключевые слова распознаются верно в большинстве случаев. Вероятность правильного распознавания удалось повысить еще значительно путем составления для каждого ключевого слова на основе проведенных тестов словаря синонимов.

3. Человеческий фактор. От оператора, его голоса и дикции в значительной степени зависит качество распознавания озвученного текста. Вне зависимости от алгоритма распознавания четкая дикция и отсутствие посторонних шумов повышают вероятность правильного результата. Голос диктора должен быть спокойным, необходимо говорить негромко, но и не тихо, произносить слова

равномерно, с одинаковыми промежутками между ними.

Как правило, оператор, долгое время работавший с системой в режиме ручного ввода текста, с предубеждениями относится к кардинальным переменам в своей работе. Поэтому для облегчения адаптации предусмотрено переключение между методами ввода в любой удобный для оператора момент. Это позволит снизить возможную негативную реакцию на период перехода. Если человек или система распознавания ошиблись при вводе значения справочника, то предусмотрен повторный ввод значения и для этого используются интерактивные подсказки.

Заключение. В результате работы создана и описана диалоговая модель программного интерфейса на примере биллинговой системы, которая взаимодействует с пользователем, предоставляя возможность выбора способа ввода данных. Приведены требования, повышающие эффективность работы с новыми возможностями системы, сформированные на основе анализа и тестирования. В дальнейшем планируется проверить модель на альтернативных голосовых системах, например, на тех, которые не требуют подключения к сети Интернет, усовершенствовать логику работы с пользователем, а также определить параметры, по которым данная модель превосходит традиционный способ ввода или ему уступает с помощью дальнейшего тестирования.

Библиографический список

1. Лоу Аверилл М., Кельтон В.Д. Имитационное моделирование. – СПб., 2004.
2. Harper R., Rodden T., Rojers Y., Sellen A. Being Human: Human-Computer Interaction in the Year 2020 [Электронный ресурс]. – URL: http://research.microsoft.com/en-us/um/cambridge/projects/hci2020/downloads/BeingHuman_A4.pdf.
3. Roemmele Brian. Why Siri is Important? [Электронный ресурс]. – URL: <http://www.quora.com/Siri-product/Why-is-Siri-important?q=Why+sir>.
4. Strickland Jonathan. Expert Stuff: Google's Mike Cohen [Электронный ресурс]. – URL: <http://computer.howstuffworks.com/mike-cohen-google.htm>.