Жеребцова Ю.А., *руководитель проектов ООО «Инфорсер Инжиниринг»*
Чижик А.В., *преподаватель кафедры информационных систем в искусстве и гуманитарных науках, СПбГУ, кандидат культурологии*
Клышинский Э.С., *доцент школы лингвистики, НИУ ВШЭ, кандидат технических наук*

Проблемы обработки естественного языка в диалоговых системах

Диалоговые системы (ДС) – на сегодняшний день одна из самых стремительно развивающихся и перспективных областей искусственного интеллекта. Исследования в этой области вызывают интерес как научного сообщества, так и индустрии, где все чаще говорят о внедрении интеллектуальных агентов с разговорным интерфейсом. Одним из наиболее обсуждаемых направлений развития ДС последнего времени является реализация возможности интеллектуального агента поддерживать непринужденный диалог с пользователем на общие темы и при этом точно определять его намерения, выполняя поставленные задачи. Ключевой проблемой при разработке такого диалогового агента является понимание системой разговорного языка при сохранении контекста диалога. В данной статье представлен обзор типов диалоговых систем, особенности их современной архитектуры и перспективы развития. Обозначены проблемы понимания естественного языка в преломлении на специфику ДС, описаны основные методы и библиотеки обработки. Уделено внимание понятию диалога, проанализированы особенности человеко-машинного дискурса и кратко рассмотрены подходы к оценке качества ДС.

1. Введение

Больше чем полвека назад профессором MIT Джозефом Вейценбаумом была создана знаменитая компьютерная программа Элиза (ELIZA) (Weizenbaum, 1966) — виртуальный собеседник с диалоговым интерфейсом, который пародирует разговор психотерапевта с пациентом при первой встрече. Элиза считается одной из первых попыток создать разговорный агент (чат-бот) и является первой программой, попытавшейся пройти известный тест Тьюринга (Turing, 1950).

Диалоговые системы (разговорные интерфейсы, диалоговые агенты) – это компьютерные системы, с которыми пользователь взаимодействует на естественном языке (ЕЯ): письменном, устном или комбинируя оба способа. Сегодня диалоговые системы получили широкое распространение: мы часто можем встретить их в повседневной жизни, когда звоним в банк, обращаемся в службу технической поддержки, указываем маршрут в навигаторе. Диалоговые интерфейсы стали основой современных персональных помощников, которые помогают выполнять повседневные задачи. Среди самых популярных можно назвать Алису от Яндекс, Siri от Apple, Google Assistant, Microsoft Cortana и Amazon Alexa. Однако отметим, что на текущий момент три последних помощника пока еще не имеют поддержки русского языка.

Почему так востребованы и привлекательны диалоговые агенты? Потому что это естественный и интуитивно понятный для любого человека способ взаимодействия с устройством или компьютерной программой: нет необходимости знать ни Python, ни Java, ни какой-либо другой язык программирования, вы просто говорите или пишете, а программа выдает необходимую информацию или выполняет поставленную задачу.

Однако, несмотря на значительный прорыв в последние годы, современные диалоговые системы встречают множество критики (Berg, 2014). Зачастую пользователю хочется вести не только целенаправленный диалог для выполнения разговорным агентом конкретной задачи, как например, запуск любимой песни, - но и воспользоваться поиском ответа на вопрос, попытаться найти информацию в Интернете или просто поддержать неформальную беседу ради развлечения. Создание универсального интеллектуального диалогового агента, сочетающего в себе не только возможности выполнения конкретных повседневных сценариев от пользователя, но и элементы эмоциональности и интерактивности, поддержание контекста беседы, а также возможность ответить на вопрос собеседника в удобном для него виде (Huang, 2019), является интересной, перспективной и при этом очень сложной задачей, которую исследователям и инженерам еще предстоит решить.

В рамках сложившегося подхода к разработке диалоговых систем, неотъемлемой частью ее архитектуры является модуль обработки естественного языка (Рис. 1) (Bellegarda, 2013). Он получает на вход реплику пользователя, затем анализирует его, определяя ее тему (предметную область), после чего определяет цель обращения пользователя (его намерение) и учитывая дополнительную

информацию, формирует некоторое семантическое представление пользовательского запроса для дальнейшего анализа системой. На этом этапе также выполняется ряд дополнительных шагов обработки входного текста: сегментация, токенизация, лемматизация, синтаксический разбор, выделение именованных сущностей, разрешение анафоры и неоднозначности (Masche, 2018). Проблемы обработки ЕЯ в диалоговых системах имеют ряд своих особенностей и ограничений по сравнению с другими областями, где с данными проблемами приходится сталкиваться, к примеру, машинный перевод или анализ тональности текстов. Обработка и понимание ЕЯ в диалоговых системах требует особого подхода к сбору данных, выбору методов анализа и решению возникающих проблем.

В данной статье мы рассмотрим понятие диалога, опишем классификацию существующих диалоговых систем, сфокусируемся на актуальных проблемах и задачах обработки естественного языка, затем представим краткий обзор методов, используемых для их решения и разберемся, в каком направлении будут развиваться современные диалоговые агенты в ближайшем будущем.

2. Диалог

Диалог – это рациональное коммуникативное взаимодействие людей на естественном языке, для которого характерны постоянная смена ролей слушающий-говорящий и наличие строго определенной ролевой модели (Searle, 1997).

2.1. Структура и типы диалога

Структуру диалога можно представить себе в виде ориентированного графа: коммуникативный акт состоит из взаимосвязанных реплик собеседников, которые имеют иерархическую структуру, очередность их появления заложена известными всем агентам, участвующим в коммуникации, правилами разговора. Каждую реплику диалога можно воспринимать как отдельный коммуникативный акт, характер которого закладывается намерением говорящего, а его форма обеспечивает получение релевантного результата. Складываясь в последовательную цепочку обмена информацией, реплики образуют дискурс. При этом процесс двустороннего общения всегда обусловлен конкретной ситуацией, которая по мере развития диалога изменяется по мере обретения общего контекста. Основная черта диалога – целенаправленность, то есть наличие намерения у агента (что принимает определенные речевые формы, например, вопрос или повеление). Для достижения своих целей адресант своей репликой должен побудить адресата к действию (ответу). Реплика-стимул может быть выражена прямо (в форме глагола в повелительном наклонении) или косвенно.

Ролевая стратегия любого диалога предполагает, что инициатива управления диалогом переходит между агентами: произнося свою реплику, адресант передает контроль над диалогом адресату. Методы распределения ходов обычно делятся на два типа сценариев (Sacks, 1974):

- агент, который будет держать инициативу и говорить, назначается текущим оратором (например, человек задает прямой вопрос и этим действием очередь передается для ответа);
- агент, удерживающий инициативу, выделяется путем самоотбора (человек решает прервать текущего оратора).

Итак, диалог как коммуникативная единица включает участников, их цели и текущую ситуацию. Его характеристиками являются: набор намерений собеседников, наличие общего контекста, текущая ситуация и ее понимание сторонами, порядок реплик в беседе и инициативность агентов, а также эмоциональная модальность.

Человеко-машинное взаимодействие копирует структуру классического диалога и ролевые модели взаимодействия агентов внутри него. Держать инициативу может агент, имеющий мотивацию продолжать коммуникативный акт. Поэтому с точки зрения архитектуры диалог, в котором система берет на себя функцию управления взаимодействием, построен на линейном (простом) или древовидном (мультизадачность или мультитематичность) сценарии, заранее прописанном с помощью формальных правил. Мотивация системы строить диалог в данном случае закладывается наличием целеориентированного алгоритма. Взяв на себя инициативу ведения диалога, система последовательно получает от пользователя информацию, необходимую ей для выполнения функционала, который в ней заложен. В этом случае анализ реплик пользователя осуществляется точнее, но диалоги могут казаться длинными, неестественными и даже недружественными.

Рис. 1. Архитектура современного диалогового агента

В настоящее время большое распространение приобретают системы, взаимодействие с которыми строится на основании сценария, выбранного пользователем. Мотивацией поддержания беседы в таком случае является намерение человека (интент), при этом система обеспечивает возможность реализации различных пользовательских сценариев за счет алгоритмов поддержания диалога и распознавания интентов. Очевидно, что количество действий, которые может выполнить диалоговая система конечно, однако поддержать коммуникативный акт и закрыть тем самым пробелы в отсутствие навыков – возможно. При этом диалог обретает черты естественного, но точность понимания пользователя может снижаться.

2.2. Модели диалога

Диалогическую речь в широком смысле принято воспринимать как развитие отношений субъектов, в процессе которого происходит формирование их оценок, эмоций и смыслов. Существует два основных взгляда на диалогические модели: коммуникативный акт можно рассматривать как процесс, ориентированный на реализацию намерений партнеров и достижение цели; с другой стороны – как процесс обмена информацией между участниками (Cohen, 1996). Выделяют два стратегических подхода к моделированию диалога.

- **Использование диалоговой грамматики.** Коммуникативное взаимодействие агентов строится на предположении о том, что диалог как четко структурированная система имеет ряд последовательных регулярностей: приветствие-приветствие, вопрос-ответ, предложение-принятие/отклонение. В таком случае для описания сценария можно использовать диалоговую грамматику, где терминальными элементами являются типы сказанных реплик, а нетерминальными – стадии диалога.
- **Методы, основанные на интентах.** Коммуникативные акты определены целями участников диалога, которые в свою очередь можно разделить на дифференциальные составные части (пункты плана по достижению конечного желания). В таком случае диалог должен учитывать контекст и потенциал изменения намерения агента в процессе диалога. Каждый участник диалога обязан понять своих партнеров (основой может служить уточнение информации). (Nils, 1971).

2.3. Ведение диалога с системой

Человеко-машинный диалог предполагает наличие в нем повелительных, повествовательных и вопросительных предложений. Чтобы взаимодействие для пользователя было комфортным, диалоговая система должна:

- определить, когда пользователь закончил говорить и ждет ответа;
- иметь с пользователем общий контекст диалога (вести, пополнять и не забывать его);
- понимать роли собеседников и порядок реплик;
- вовремя забирать и/или отдавать инициативу.

Основная проблема построения диалога с системой на таком уровне – понимание смысла сказанного человеком внутри контекста. Поэтому необходимым условием успешной коммуникации в рамках диалога является наличие общей картины мира, включающей в себя общую базу знаний, убеждений и предположений, которые устанавливаются в ходе взаимодействия (Lison, 2013). Также для обеспечения устойчивого общего контекста диалога участники должны обмениваться информацией о понимании или непонимании передаваемой информации (Skantze, 2007). Для этого служат фразы-уточнения и фразы, сообщающие, что высказывание понято (например, через уточняющую интерпретацию высказывания). Необходимо учитывать также, что характерной для диалога на естественном языке является возможность понимать реплики, в которых конкретно не указана суть информации, но она подразумевается контекстом (Lison, 2013). Поэтому диалоговые системы должны уметь обрабатывать импликатуру подаваемого на вход высказывания.

Таким образом, блок управления диалогом (диалоговый менеджер) в системе должен выполнять следующие задачи (Kronlid, 2011):

- мониторинг структуры диалога;
- соблюдение логической целостности;
- использование речевых приемов;
- вход и выход из диалога.

А в качестве критериев успешности диалога можно выделить:

- достижение основной цели диалога;
- реализация набора интентов;
- разрешение коммуникативных неудач;
- эмоциональное удовлетворение пользователя.

2.4. Проектирование диалога

Процесс проектирования и реализации диалогов состоит из определения модели, по которой будут строиться взаимодействия системы и пользователя, оценки конечного множества необходимых сценариев, определения типа и формы каждого диалога, а также учет синтаксиса и семантики языка, на котором будет происходить взаимодействие.

Проектирование абстрактного диалога обычно начинается с оценки первичной ситуации для пользователя в моменте начала взаимодействия с ним. Сценарий диалога имеет несколько составляющих: шаги, принимаемые решения, намерения, сущности и переменные. Он определяет последующее событие в зависимости от входящей информации и внутренних правил. Каждое состояние ассоциируется с конечным множеством действий, релевантных для удовлетворения пользователя. Один шаг системы – одно такое сопоставление. Его прохождение реализуется за счет следования логическим направлениям в зависимости от вводимой пользователем информации, с учетом которой диалоговая система принимает решения относительно сценария диалога.

3. Современные диалоговые системы

Значимым событием в разработке диалоговых систем стало появление языка разметки AIML (диалект XML), разработанный в 2001-м году Ричардом Уэлсом и созданный на его основе диалоговый агент A.L.I.C.E (Wallace, 2003). Rule-based подход развивался на протяжении следующих 10 лет. Возможность описать мир формальными правилами базируется в данном случае на двух фундаментальных задачах – выделении семантически значимых элементов фраз и описании сценариев диалогов. Основная проблема такого подхода – масштабируемость до мультизадачности и открытой тематики диалога. Частично эту проблему решили появившиеся в последние годы алгоритмы определения семантической близости текстов и другие технологии машинного обучения, помогающие расширить возможности диалоговых систем.

3.1. Типы современных диалоговых систем

Диалоговые системы (ДС) по их функциональности можно условно разделить на следующие типы:

- **ДС для контакт-центров.** Диалоговые агенты могут быть использованы для ответов на типовые вопросы (по принципу

FAQ) и в режиме «call steering» для маршрутизации пользователя в нужный ему отдел компании.

- **Говорящие устройства (smart speakers).** Термин обычно относится к устройствам, которые помогают пользователю выполнять различные простейшие бытовые задачи внутри «умного» дома; вызываются и управляются с помощью голосовых команд. Примеры: Amazon Echo, Google's Home и Apple HomePod. К этому типу диалоговых систем также относятся роботы, детские игрушки, интеллектуальная бытовая техника.
- **Голосовые ассистенты (intelligent virtual assistant).** Обычно встраивается в конкретное приложение для использования на мобильном устройстве. Голосовые ассистенты определяют интенты (намерения) пользователей и исполняют команды.

3.2. Общая архитектура диалоговой системы

В настоящее время принято разделять диалоговые системы на целеориентированные (task-oriented dialogue agents) и системы общего назначения (general dialogue agents).

3.2.1. Целеориентированные диалоговые системы

Целеориентированные диалоговые системы предназначены для решения заранее определенного набора задач. Для них характерны короткие беседы, по заранее прописанному разработчиками сценарию (получить информацию от пользователя для выполнения поставленной цели). Интерактивные мобильные помощники (Siri, Cortana, Alexa и т.д.) являются примерами такого типа диалоговых систем. Они ориентированы на выполнение интерактивного поиска, управление бытовой техникой или системами умного дома, помощи в совершении звонков или отправки текстовых сообщений. Интерактивная помощь клиентам на различных веб-сайтах также относится к этому типу систем.

Целевые диалоговые системы обычно включают в себя следующие компоненты:

- **анализатор естественного языка,** который преобразует высказывания пользователя во внутреннюю структуру данных, понимаемую системой;
- **диалоговый менеджер** определяет текущее состояние диалога, которое используется для выбора следующего действия;
- **модуль определения действия** выбирает следующее действие диалоговой системы в зависимости от текущего состояния (т.н. политика поведения);
- **генератор естественного языка** преобразует результат действия из внутреннего представления системы в ответ пользователю (речь).

Основной цикл обработки запроса пользователя состоит из следующих событий и действий (Рис. 2):

1. Система получает запрос пользователя в модуль управления диалогом – диалоговый менеджер.

2. Модуль управления диалогом (диалоговый менеджер) загружает контекст диалога из базы данных. Если в этот момент в системе уже есть история диалога (т.е. данный запрос не первый), то задача на этом шаге — сохранить новую информацию и объединить ее с уже имеющимися данными о предыдущих запросах пользователя.

3. Запрос пользователя вместе с контекстом отправляется в модуль обработки естественного языка. Происходит выделение тематики запроса, по которой определяются намерения пользователя (интент). Затем идет выделение сущностей, то есть атрибутов, которые определяют, как именно можно было бы ответить на вопрос пользователя. На выходе формируется некоторое формальное представление о запросе пользователя – семантический шаблон (фрейм).

4. На основе сценария диалога и извлечённых данных, диалоговый менеджер определяет следующее наиболее подходящее состояние/действие (этап политики поведения).

5. Вызов внешних информационных систем или обращение к внутренним базам данных и базам знаний.

Рис. 2. Архитектура диалоговой системы целеориентированного типа

Рис. 3. Входные данные и результат работы блока обработки ЕЯ в диалоговой системе

6. Действие поступает в модуль генерации ответа.

7. Контекст и параметры диалога сохраняются для обработки последующих обращений.

8. Ответа отправляется пользователю.

Каждая целеориентированная диалоговая система создаётся и настраивается под конкретную задачу. Большинство из них основаны на фреймовой архитектуре (frame-based): используется набор слотов, каждый из которых может принимать одно конкретное значение из фиксированного набора. Также для целеориентированных систем характерно большое число правил, созданных вручную. Все это выделяет основную особенность таких систем – наличие сценария: на каждом этапе диалога у системы есть четкая инструкция.

3.2.2. Чат-боты

Для диалоговых систем общего назначения, также называемых чат-ботами, стоит задача поддержать разговор с человеком на произвольную тему. В этом случае всех возможных целей пользователя не предусмотреть, поэтому невозможно и написание достаточного количества сценариев взаимодействия. На данный момент системы этого типа часто имеют развлекательную функцию (в качестве примера может служить переключение в режим «болталки» голосового помощника Алиса от компании Яндекс).

Чат-боты могут быть генеративными (generative) и поисковыми (retrieval-based). В генеративном подходе ответ на сообщение пользователя порождается с помощью некоторой модели. В 2011 г. было предложено не определять для чат-бота сценарий всего диалога, а попытаться обучить систему естественно отвечать на последнюю реплику пользователя (Ritter, 2011). Подход был основан на машинном переводе – реплики пользователя «переводились» в подходящие ответы на них. Первые эксперименты давали большую погрешность, но появившиеся в 2014 г. нейросетевые модели позволили улучшить качество диалога, поэтому модель encoder-decoder (представление предложения как вектора фиксированной длины с последующим декодированием в другое предложение) считается одним из стандартов разработки чат-ботов. В настоящее время seq2seq-модели – один из популярнейших способов построения таких систем (Vinyals, 2015; Miller, 2016).

У такого подхода обнаружился ряд проблем, который стоит на пути осмысленного диалога с пользователем, в частности потеря контекста диалога и тенденция давать довольно общие ответы. Поэтому большинство практических систем сейчас все еще базируется на заранее заготовленных ответах и ранжировании вариантов (с помощью правил и машинного обучения). Так, например, для диалоговых агентов (Siri, Алиса) каждая тематика, которую они понимают, тщательно настраивается вручную (Sarikaya, 2016), что позволяет добиться высокого качества ответов. Google Smart Reply использует ранжирование, реализованное с использованием нейросетей (Henderson, 2017). В большинстве случаев ранжирование эффективно за счет синтаксических преобразований и простых правил (Ameixa, 2014). В таком случае, смысл сказанного пользователем опускается. Однако большие нейросети демонстрируют оперирование общетематическими знаниями (Tarasov, 2017). В поисковом подходе ответ на сообщение пользователя выбирается из большого набора готовых ответов. Как правило, для сообщения и всех ответов строятся векторные представления одной размерности, после чего ответ выбирается в соответствии с некоторой метрикой.

Из приведенных примеров можно увидеть, что диалоговые системы охарактеризуются еще и способностью поддерживать диалог на конкретную заранее определенную тему (closed domain) или произвольную (open domain). Целеориентированность системы означает наличие одного или множества сценариев, которые система должна включать на основании точной информации полученной из модуля обработки ЕЯ, поэтому для таких систем актуально точно классифицировать пользовательские интенты и заполнять значения фреймов. Если пользователь обратился к системе за рамками ее навыков, режим разговора на общую тему (чат-бота) может быть полезным инструментом поддержания диалога.

Обобщая характеристики, можно также предложить разделение диалоговых систем по методам генерации ответов пользователю:

- системы, основанные на правилах (rule-based). На запрос (команду) пользователя ответ формируется на основе заранее заданного шаблона или правила;
- системы, выбирающие для заданного пользователем сообщения ответы из заранее заданного списка (retrieval-based);
- системы, основанные на порождающих (generative) моделях без заранее заданного списка ответов.

3.3. Оценка качества диалоговых систем

Оценка качества диалоговой системы является достаточно сложной и специфической задачей, так как классические для машинного обучения критерии оценки не применимы из-за большой вариативности выдаваемого пользователю ответа. Поэтому достаточно часто работу диалоговой системы оценивают эксперты или конечные пользователи. Целеориентирвоанные диалоговые системы можно оценивать по критериям, соответствующим достижению (или не достижению) конечной цели, и количеству сообщений, которое для этого потребовалось.

Для чат-ботов принято использовать метрики оценки качества машинного перевода: BLEU, ROUGE и модификации (Papineni, 2002). В ряде исследований отмечается, что значения метрик не всегда совпадают с оценками экспертов (Liu, 2016; Al-Rfou, 2016). Также, если целью является определение качества диалогового опыта при использовании системы, то для оценки хорошо

подходит тест Тьюринга (Venkatesh, 2018). В 2017 г. был предложен метод, основанный на обучении с подкреплением, где в качестве вознаграждения для порождающей модели используется заключение бинарного классификатора, играющего роль эксперта в тесте Тьюринга (Li, 2017).

4. Проблемы обработки ЕЯ в диалоговых системах

Для того чтобы лучше интерпретировать намерения пользователя, а затем корректно возвращать запрашиваемую им информацию, в любой диалоговой системе существует блок обработки естественного языка (ЕЯ), который играет важнейшую роль: получая от пользователя разговорную команду, сконвертированную модулем распознавания речи в текстовую строку, блок обработки ЕЯ возвращает ее некоторое внутреннее представление, понимаемое диалоговой системой, которая далее принимает решение, как реагировать на команду (Рис.3).

Часто говорят, что на данном этапе диалоговая система «понимает» естественный язык и называют данный блок – блоком понимания ЕЯ (Natural Language Understanding, NLU). Данный блок может состоять из одного или нескольких модулей, каждый из которых предназначен для решения конкретной задачи обработки входного текста, начиная от корректировки опечаток, затем продолжая обработку текста на уровне синтаксиса, переходя к его семантической интерпретации.

С точки зрения специфики естественного языка, анализируемого системой, диалог человека и компьютера является особенным типом дискурса, в рамках которого структура речи и текста отличается от множества форм текста, которые обычно изучает компьютерная лингвистика (напр., новости, отзывы, научные статьи и т.д.).

Как правило, выделяют следующие особенности диалога, которые значительно осложняют решение задач обработки ЕЯ в диалоговых системах (DeVault, 2013).

- **Наличие ошибок автоматического распознавания речи, опечатки.**

  Например,

  Речь пользователя: *Сколько стоит билет на концерт на завтра?*
  Результат распознавания: *Сколько стоит билет на концерт на зал?*

  Наличие ошибок распознавания речи говорит о том, что модуль обработки ЕЯ не всегда получает на вход исходную корректную реплику от пользователя. В зависимости от типа диалоговой системы, предметной области и многих других особенностей, ошибки модуля распознавания речи могут достигать величины, которые в значительной степени могут влиять на корректность ответа системы.

- **Особенности устной речи: спонтанность, сокращения, слэнг.**

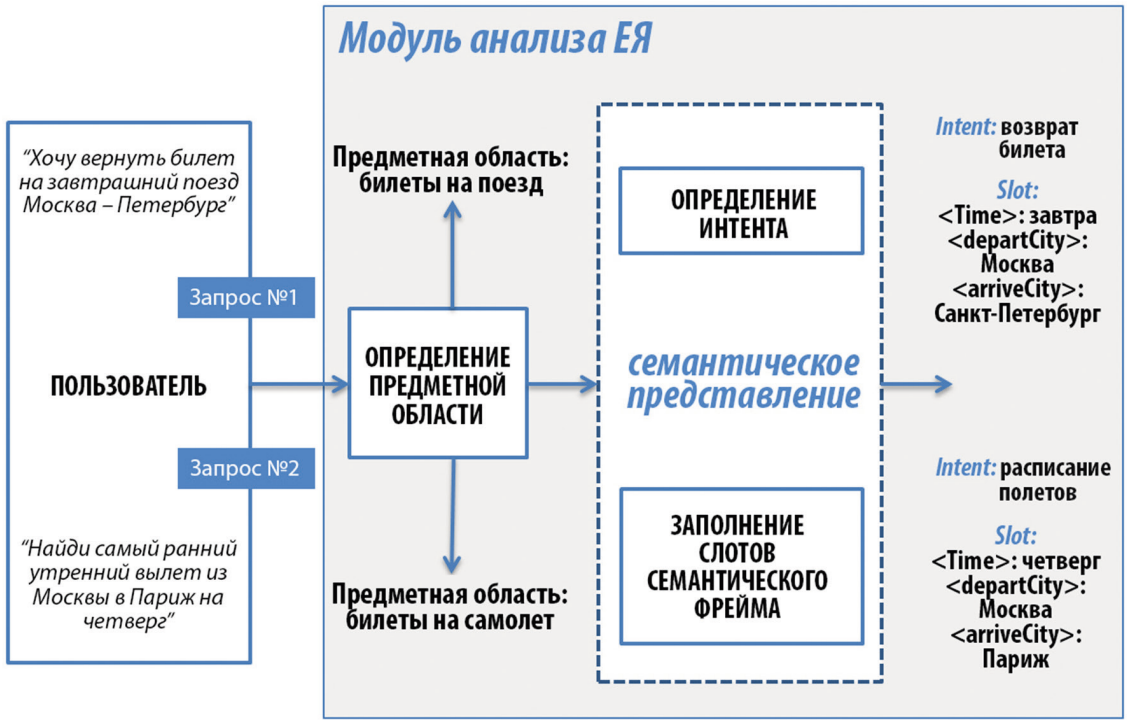
  Часто устная разговорная речь более сложная и «шумная» чем письменная из-за того, что она более быстрая, отрывистая, с большим количеством пропусков. В устной речи часто встречаются паузы, заполненные каким-либо выражением («хм», «мммм»), краткие неформальные реплики («угу»), разговорные и сленговые слова («где потусить?»). Пользователь может спонтанно дополнять только что сказанную фразу или говорить фрагментам по несколько слов, повторить одно и то же или внезапно начать заново, думая, что система не расслышала. Модуль обработки ЕЯ диалоговой системы должен уметь справляться со спецификой устной речи, чтобы минимизировать впоследствии ошибки понимания пользователя.

- **Неоднозначность реплик.**

  Например,

  – *Мне нужно оформить карту (карта – географическая или банковская?)*
  – *Во сколько открывается банк? (какой конкретно банк?)*
  – *А когда там обед? (где?)*
  – *Напомни завтра взять кошелек с собой (кошелек и себя или только кошелек?)*

Рис 4. Пример обработки реплики пользователя диалоговой системой



Проблема наличия лингвистической неоднозначности (лексической, синтаксической, контекстной) пронизывает задачи обработки ЕЯ, решаемые каждый день в различных прикладных областях и диалоговые системы – не исключение.

- **Синонимия, избыточность формулировок.**

  Например,

  – Где я нахожусь?

  – Подскажи, где сейчас мое местоположение?

  – Расскажи о моей геолокации?

  – Мое текущее расположение на карте

  (во всех случаях должны быть возвращены текущие координаты пользователя)

Избыточность информации в репликах пользователя и их синонимичность является проблемой, которую необходимо разрешать, т.к. идентичные по смыслу команды, сформулированные различными способами, должны приводить систему к одинаковому результату.

- **Зависимость фраз от контекста.**

  Например,

  – А есть такое же, но красное и дешевле?

  (в предыдущих репликах пользователь, возможно, выбирал какой-то товар, о котором диалоговая система должна «помнить» контекст разговора)

Диалог на естественном языке состоит из реплик, связанных между собой текущим контекстом. Одинаковые реплики пользователя в различные моменты времени могут иметь различный смысл и должны приводить систему к различным ответам.

- **Сложность выбора оптимального семантического представления.**

Значение, которое пользователь вкладывает в свою речь, может быть очень сложным и включать в себя множество нюансов. Для того чтобы понимать эти нюансы и поддерживать с пользователем эффективный диалог, при создании диалоговой системы важно найти универсальную форму семантического представления поступающей от пользователя информации. Семантическое представление текста внутри современных диалоговых систем может отличаться друг от друга и зависит от специфики задач, для решения которых она создавалась. Кроме того, разработчики диалоговых систем создают собственные уникальные семантические представления для поддержания более свободного диалога с пользователем или диалога в специфической предметной области.

5. Задачи обработки ЕЯ

В ранних диалоговых системах блок обработки ЕЯ обычно выполнял две задачи: определение намерения пользователя (классификация интента) и заполнение переменных семантического представления этого интента, т.е. слотов некоторого шаблона, определенного заранее. С увеличением количества доступной информации о множестве предметных областей человеческой жизни, а также по мере развития технологий и подходов к разработке диалоговых систем, модуль обработки ЕЯ современного диалогового агента получил третью задачу – определение предметной области, к которой относится реплика пользователя Рис.4. (Liu, 2019).

Таким образом, получение семантичекого представления запроса пользователя происходит в следующей последовательности.

- Определение предметной области входной реплики: просит ли пользователь поставить будильник, или показать погоду на завтра, или задает общий вопрос, ответ на который необходимо искать в поисковике или базе знаний? Классификация предметной области входной реплики и ответ пользователю с учетом данного знания – стандарт разработки современного диалогового агента.
- Определение намерения пользователя (классификация интента): каково намерение пользователя при обращении к тому или иному функционалу системы? Он пытается поставить напоминание или удалить его? Спрашивает расписание фильмов на сегодня или на всю неделю?
- Заполнение интента: получение значений переменных (слотов) шаблона, заполнение полей семантического фрейма конкретного интента для дальнейшего поддержания диалога системой.

Например, для реплики пользователя *Установи будильник на 6 утра* система заполнит заранее определенный внутренний семантической фрейм следующими значениями:

ПРЕДМЕТНАЯ ОБЛАСТЬ: *будильник*

ИНТЕНТ: *установить будильник*

ВРЕМЯ: *2019-08-01 06:00:00*

А на вопрос *Какая столица в Исландии?* будет заполнен шаблон для поискового запроса или запроса в базу знаний:

ПРЕДМЕТНАЯ ОБЛАСТЬ: *поисковый запрос*

ЗАПРОС: *столица-страны (Исландия)*

В разных диалоговых системах блок обработки ЕЯ может возвращать различные семантические представления запросов пользователя. Это могут быть как заполненные одиночные слоты, так и набор слотов, объединенный в заранее определенный шаблон. Также блок обработки ЕЯ может возвращать тег с типом устной реплики, к которому относится (Bunt, 2010, Jurafsky 1997, Bach and Harnish, 1979). Например,

Пользователь: *да, прекрасно!*

Тип реплики: *утверждение-мнение*

Рис 5. Этапы предварительного анализа реплики пользователя модулем обработки ЕЯ

Пользователь:*нет, спасибо*
Тип реплики: *отрицание*
Пользователь: *[смех]*
Тип реплики: *невербальная реакция*

Определение предметной области и интента с последующим заполнением его искомых значений, является сложным этапом глубокого семантического анализа естественного языка, которому обычно предшествуют необходимые шаги предобработки текста. На данном этапе в рамках разработки диалогового агента обычно решают следующие классические задачи обработки ЕЯ, изображенные на Рис.5 (Berg, 2014): сегментация (Latorre-Navarro et al., 2015), токенизация и лемматизация (Veselov, 2010; Wilks et al., 2010; Eugene, 2014; Bogatu et al., 2015; Amilon, 2015), определение частей речи (Lasguido et al., 2013; Dingli et al., 2013; Higashinaka et al., 2014; Ravichandran et al., 2015), извлечение именованных сущностей и семантических связей (Wilks et al., 2010; Lasguido et al., 2013).

Также помимо упомянутых задач, в современных диалоговых системах решаются и специфические для данной прикладной области задачи обработки диалога на естественном языке (Jurafsky, 2008): обработка последовательности из нескольких фраз, дополняющих друг друга; поиск ссылок одной фразы на другую; обработка чередования различных типов интентов подряд; обработка запроса пользователя, если во фразе найдено более одного интента или более одной предметной области; генерация ответов и уточнений со стороны системы, их обработка.

6. Методы и библиотеки обработки ЕЯ

Как было показано выше, существует множество задач обработки ЕЯ, которые необходимо иметь возможность решать при создании хорошей диалоговой системы. Ниже мы рассмотрим методы, используемые для решения упомянутых задач.

Традиционными подходами к анализу намерения пользователя (классификация интента, извлечение значений слотов его семантического шаблона) считаются методы, основанные на заранее подготовленном наборе правил (Appelt, 1993) или использующие статистические методы, например, скрытые марковские модели (Pengju, 2000). Данные методы использовались в ранних диалоговых системах и являлись достаточно трудоемкими в поддержке и имели невысокую точность результатов.

Также применяются различные методы машинного обучения для задачи классификации: наивный байесовский классификатор (McCallum, 1998), метод опорных векторов (Haffner, 2003), логистическая регрессия (Genkin, 2007), метод k-ближайших соседей (Lefevre, 2009) или комбинацию этих методов (Schapire, 2000).

С развитием нейронных сетей и алгоритмов глубокого обучения наиболее часто стали использоваться модели векторных представлений слов (Kim, 2015), сверточные нейронные сети (Hashemi, 2016), рекурентные нейронные сети, в особенности архитектуры с долгой краткосрочной памятью (LSTM) или управляемыми рекуррентными блоками (GRU), а также модификации и комбинации этих моделей (Ravuri, 2015; Dey 2017; Liu 2016). По сравнению с традиционными методами машинного обучения, нейросетевые модели требуют больше данных и ресурсов для обучения, но, тем не менее, показывают высокую точность результатов.

Среди существующих библиотек обработки ЕЯ, которые могут быть использованы при создании диалоговой системы, можно выделить следующие (Burtsev, 2018):

- DeepPavlov – открытая библиотека для создания диалоговых агентов, включающая предобученные нейросетевые модели. Разработана лабораторией нейронных систем и глубокого обучения Московского физико-технического института;
- Rasa Stack – платформа для создания целеориентированных диалоговых агентов, включающая средства для создания основных модулей, входящих в современную архитектуру современной диалоговой системы;
- ParlAI – платформа для создания диалоговых систем, включающая множество дополнительных наборов данных для обучения моделей и анализа;
- а также AllenNLP, spaCy и Stanford CoreNLP – библиотеки с мощным инструментарием для решения задач обработки ЕЯ, в т.ч. токенизация, лемматизация, определение частей речи, а также средства для синтаксического и семантического анализа текстов.

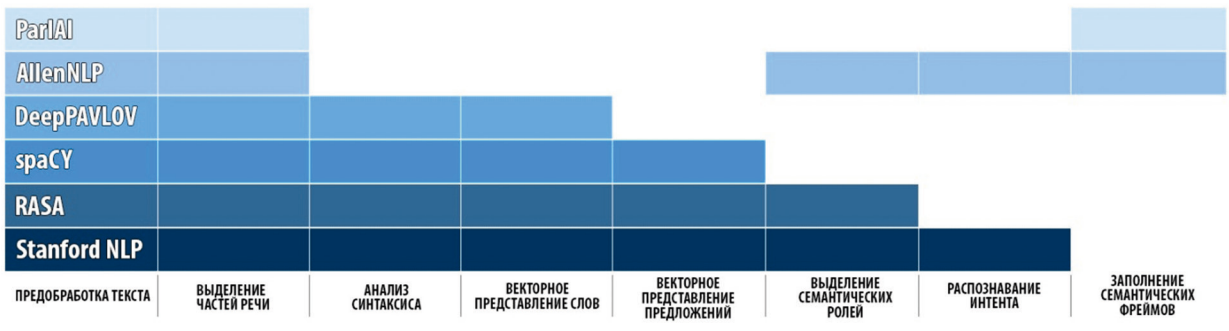
На Рис. 6 представлена сводная информация о возможностях данных библиотек.

7. Заключение

В данной статье рассмотрены актуальные подходы к построению диалоговых систем, выделены их типы, проанализированы особенности, как с точки зрения реализации, так и с точки зрения идеи, которая закладывается на этапе разработки, а также приведена краткая историческая справка, показывающая тенденции исследований в области создания человеко-машинного диалога.

Также мы выделили основные модели диалогов и описали различные ракурсы научного взгляда на этот тип коммуникации. Выделенные характеристики диалога и вариативность его моделей было предложено учитывать при моделировании архитектуры диалоговой системы. В рамках данного обзора коммуникативный акт

Рис. 6. Сравнение библиотек обработки ЕЯ



между пользователем и системой выделен как элемент, имеющий важнейшее значение: именно с его изучения должно начинаться фундаментальное исследование проблемы создания диалогового интерфейса. Нами была предпринята попытка рассмотреть проблемы обработки естественного языка в диалоговых системах именно через призму междисциплинарности. Стратегия ведения диалога со стороны пользователя также предмет для дальнейшего изучения: определение влияния коммуникативной цели высказывания на его текст предполагает учет сложного взаимодействия многих контекстных и социолингвистических факторов, что может являться темой отдельного исследования.

На сегодняшний день не существует какой-либо строгой классификации диалоговых систем по формальным признакам, однако их удобно разделять на ДС общего назначения или целеориентированные, а также по способности говорить на любую тему или только на строго определенную. Вектор разработки диалоговых систем направлен на создание целеориентированных ДС, так как потребности современного рынка заключаются по большей части в наличии агента, который решает конкретный набор повседневных задач для удобства пользователя. Вместе с тем стоит задача обеспечить способность агента поддерживать связный диалог, обладающий признаками коммуникации на естественном языке. Важно, например, «не забывать» контекст беседы и правильно понимать смысл сказанного пользователем.

Стоит отметить, что на данный момент нет методов, которые бы обеспечили высокое качество удовлетворения запросов пользователя диалоговой системы в открытой предметной области. С другой стороны, естественность беседы может дать пользователю эмоциональное удовлетворение, которое является одним из критериев успешности ведения диалога. Поэтому, на наш взгляд, стоит обратить внимание на проектирование «идеальной» целеориентированной системы с навыками разговора на любую тему системы (task-oriented + open domain). При этом такая система должна обладать такими важными свойствами как понимание неявных намерений пользователя, целостность ведения диалога, осведомленность о пользователе, способность планирования, эмоциональность, личность.

Блок обработки естественного языка современной диалоговой системы состоит из модулей, решающих множество задач обработки ЕЯ: от разделения реплики на предложения до глубокого семантического анализа. По сравнению с традиционными методами машинного обучения, нейросетевые модели показывают высокую точность результатов, но требуют достаточно больших ресурсов для обучения моделей, их поддержки, а также результат их работы является трудно интерпретируемым и мало предсказуемым. Поэтому современные подходы к разработке диалоговых агентов сочетают в себе и традиционные методы обработки ЕЯ, и классические алгоритмы машинного обучения, и нейросетевые модели.

Дальнейшие исследования должны проводиться в области изучения особенностей создания сценариев построения человеко-машинного диалога в целеориентированных системах с открытой предметной областью, и проведения экспериментов по комбинированию различных подходов к моделированию диалога в рамках одного диалогового агента. Представляется полезным создание качественной метрики, для оценки успешности диалога пользователя с диалоговой системой, которая бы учитывала как релевантность выполняемых системой действий, так и уровень удовлетворенности пользователя диалогом. Актуальны также междисциплинарные исследования по взаимодействию человека с устройствами, имеющими диалоговый интерфейс и исследования коммуникативного пространства, возникающего на этом фоне, в частности интересен анализ возникающего в этот момент человеко-машинного дискурса.

[1] (Al-Rfou, 2018) Al-Rfou R. et al. Conversational contextual cues: The case of personalization and history for response ranking // arXiv preprint arXiv:1606.00372. 2016.

[2] (Ameixa, 2014) Ameixa, David, et al. Luke, I am your father: dealing with out-of-domain requests by using movies subtitles. International Conference on Intelligent Virtual Agents, 2014.

[3] (Appelt, 1993) Appelt D, Bear J, Cherny L, et al. GEMINI: A natural language system for spoken-language understanding // Proc. Meeting of the Association for Computational Linguistics. 1993:54-61.

[4] (Bach, 1979) Bach, K. and Harnish, R. (1979). Linguistic communication and speech acts. MIT Press.

[5] (Bellegarda, 2013) Jerome R. Bellegarda. Large–Scale Personal Assistant Technology Deployment: the Siri Experience.

[6] (Berg, 2014) Markus Berg, Modelling of Natural Dialogues in the Context of Speech-based Information and Control Systems.

[7] (Bunt, 2010) Bunt, H., Alexandersson, J., Carletta, et al. Towards an ISO standard for dialogue act annotation. In Proceedings 8th International Conference on Language Resources and Evaluation (LREC 2010), Malta, pp. 2548-2558.

[8] (Burtsev, 2018) DeepPavlov: Open-Source Library for Dialogue Systems / Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-System Demonstrations, pages 122–127 Melbourne, Australia, 2018.

[9] (Cohen, 1996) Cohen A.D., Scott K. A synthesis of approaches to Assessing Language Learning Strategies // Language Learning Strategies around the World. 1996. P. 89-109

[10] (DeVault, 2013) David DeVault CSCI 599 Special Topic: Natural Language Dialogue Systems Spring 2013.

[11] (Dey, 2017) Dey R, Salemt F M. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks // IEEE 60th International Midwest Symposium on Circuits and Systems. IEEE, 2017:1597-1600.

[12] (Eugene, 2014). Turing Test success marks milestone in computing history. University of Reading. Retrieved 8 June 2014.

[13] (Genkin, 2007) Genkin A, Lewis D D, Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization. Technometrics, 2007, 49(3):291-304.

[14] (Haffner, 2003) Haffner P, Tur G, Wright J H. Optimizing SVMs for complex call classification // IEEE International Conference on Acoustics. IEEE, 2003:632-635.

[15] (Hashemi, 2016) Hashemi H B., Asiaee A, Kraft R, Query intent detection using convolutional neural networks // International Conference on Web Search and Data Mining, Workshop on Query Understanding , 2016.

[16] (Henderson, 2017) Henderson Matthew, et al. Efficient Natural Language Response Suggestion for Smart Reply. 2017.

[17] (Huang, 2000) Yinfei Huang, Fang Zheng, Mingxing Xu, Pengju Yan, Wenhu Wu. Language Understanding Component for Chinese Dialogue System. Sixth International Conference on Spoken Language Processing (ICSLP 2000).

[18] (Huang, 2019) Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2019. Challenges in building intelligent open-domain dialog systems. arXiv preprint arXiv:1905.05709.

[19] (Jurafsky, 2008) Title Speech and Language Processing; Author(s) Dan Jurafsky, James H. Martin; Publisher: Prentice Hall, 2nd edition. May 16, 2008.

[20] (Jurafsky, 1997) Dan Jurafsky, Liz Shriberg, Debra Biasca. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual. – http://groups.inf.ed.ac.uk/switchboard/dialactmanual.html.

[21] (Kim, 2015) Kim D , Lee Y, Zhang J, et al. Lexical feature embedding for classifying dialogue acts on Korean conversations //Proc. of 42th Winter Conference on Korean Institute of Information Scientists and Engineers, 2015: 575–577.
[22] (Kronlid, 2011) Fredrik Kronlid, Jessica Villing, Alexander Berman, Staffan Larsson, Comparing system-driven and free dialogue in in-vehicle interaction, Proceedings of Interspeech 2011, Conference paper 2011, c. 1.
[23] (Lefevre, 2009) k-Nearest Neighbor Monte-Carlo Control Algorithm for POMDP-based Dialogue Systems F. Lefèvre, et al. Spoken Dialogue Systems Group Cambridge University Engineering Department, 2009.
[24] (Li, 2017) Li J. et al. Adversarial Learning for Neural Dialogue Generation // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. – 2017. – P. 2157-2169.
[25] (Lison, 2013) Pierre Lison, Structured Probabilistic Modelling for Dialogue Management, PhD Thesis, 2013.
[26] (Liu, 2016) Liu B, Lane I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling //17th Annual Conference of the International Speech Communication Association. 2016:685-689.
[27] (Liu, 2016) Liu C.W. et al. How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. – 2016. – P. 2122-2132.
[28] (Liu, 2019) Jiao Liu, Yanling Li and Min Lin. Review of Intent Detection Methods in the Human-Machine Dialogue System.
[29] (Masche, 2018) A Review of Technologies for Conversational Systems Conference Paper in Advances in Intelligent Systems and Computing.
[30] (McCallum, 2000) Andrew McCallum, Kamal Nigam, et al.A comparison of event models for naive bayes text classification// In AAAI-98 workshop on learning for text categorization. 1998: 41–48.
[31] (Miller, 2016) Alexander H. Miller, Key-Value Memory Networks for Directly Reading Documents. Conference on Empirical Methods in Natural Language Processing, pages 1400–1409, Austin, 2016.
[32] (Nilsson, 1971) Nils J. Nilsson, STRIPS: a new approach to the application of theorem proving to problem solving. Proceeding IJCAI'71 Proceedings of the 2nd international joint conference on Artificial intelligence, England, 1971.
[33] (Pengju, 2000) WORD-CLASS STOCHASTIC MODEL IN A SPOKEN LANGUAGE DIALOGUE SYSTEM, YAN Language Processing, pp. 141-144, Oct. 13-15, 2000, Beijing.
[34] (Papineni, 2002) Papineni K. et al. BLEU: a method for automatic evaluation of machine translation // Proceedings of the 40th annual meeting on association for computational linguistics. – Association for Computational Linguistics, 2002. – P. 311-318
[35] (Ravuri, 2015) Ravuri S V, Stolcke A. Recurrent neural network and LSTM models for lexical utterance classification // 16th Annual Conference of the International Speech Communication Association. 2015:135-139.
[36] (Ritter, 2011) A Ritter, Data-Driven Response Generation in Social Media. Conference on Empirical Methods in Natural Language Processing, p 583–593, Edinburgh, 2011.
[37] (Ruhi, 2016) Sarikaya, Ruhi, et al. An overview of end-to-end language understanding and dialog management for personal digital assistants. Spoken Language Technology Workshop (SLT), 2016.
[38] (Sacks, 1974) Harvey Sacks, Emanuel A Schegloff, Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation, Language, 1974., c. 696-735.
[39] (Schapire, 2000) Schapire R E, Singer Y. BoosTexter: a boosting-based system for text categorization. Machine Learning, 2000, 39(2-3):135-168.
[40] (Skantze, 2007) Gabriel Skantze, Error handling in spoken dialogue systems, PhD Thesis, 2007.
[41] (Searle, 1997) Searle J.R. Classification of Illocutionary Acts // Proceedings of the Texas Conference on Performatives, Presuppositions and Implications. - Arlington: Univ. Press, 1977. - P. 27 - 45.
[42] (Tarasov, 2017) Tarasov, D. S., and E. D. Izotova. Common Sense Knowledge in Large Scale Neural Conversational Models. International Conference on Neuroinformatics, 2017.
[43] (Tur, 2011) Tur G. Spoken Language Understanding: Systems for extracting semantic information from speech. New York, NY: John Wiley and Sons.
[44] (Turing, 1950) Turing, A.M. Computing Machinery and Intelligence. Mind 49: 433-460.
[45] (Venkatesh, 2018) Venkatesh A. et al. On Evaluating and Comparing Conversational Agents //arXiv preprint arXiv:1801.03625. – 2018.
[46] (Vinyals, 2015) Vinyals Oriol, and Quoc Le. «A neural conversational model» arXiv preprint arXiv:1506.05869 (2015).
[47] (Wallace, 2003) R. Wallace, The Elements of AIML Style, ALICE A.I Foundation, 2003.
[48] (Weizenbaum, 1966) Weizenbaum, J. ELIZA – A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), P. 36–45.

Ключевые слова: Диалоговая система, интеллектуальный агент, обработка естественного языка, компьютерная лингвистика, целеориентированная ДС, чат-бот, искусственный интеллект.

Natural Language Understanding Challenges in Dialogue Systems

Chizhik Anna, *senior lecturer, Department of Information Systems in Arts and Humanities, SPSU, PhD in culturology*
Zherebtsova Yulia*, Project Manager, Inforser Engineering*
Klyshinsky Eduard*, assoc. prof. of School of Linguistics, PhD in Computer Science*

Abstract: *Nowadays, a field of dialogue systems and conversational agents is one of the rapidly growing research areas in artificial intelligence applications. Business and industry are showing increasing interest in implementing intelligent conversational agents into their products. Many recent studies has tended to focus on possibility of developing task-oriented systems which are able to have long and free social chats that occur naturally in social human interactions. In order to better understand the user's expression, and then feedback the correct information, natural language understanding plays an extremely important role. Despite progress made in solving NLP problems, it remains very challenging today in the field of dialogue systems. In this paper, we review the recent progress in developing dialogue systems, its current architecture features and further prospects. We focus on the natural language understanding tasks which are key for building a good conversational agent, and than we are summarizing NLP methods and frameworks, in order to allow researchers to study the potential improvements of the state-of-the-art dialogue systems. Additionally, we consider the dialogue concept in context of human-machine interaction, and briefly describe dialogue evaluation metrics.*

Keywords: *Dialogue system, conversational agent, natural language interface, natural language processing, natural language understanding, computational linguistics, intelligent personal assistant, chat-bot.*