

ВОЗМОЖНОСТИ СИСТЕМ ГЕНЕРАЦИИ РЕЧИ ДЛЯ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА И РАЗВИТИЯ ВЗАИМОДЕЙСТВИЯ «ЧЕЛОВЕК – МАШИНА»*

Д.В. Галкин, Е.В. Кондрушина, К.С. Шилияев

Представлен обзор современного состояния проблемы компьютерной генерации речи в теоретическом и прикладном аспектах. На основе концепции вопросно-ответных систем рассмотрены пути развития взаимодействия «человек – компьютер», предложена модель компьютерной системы, предоставляющей человеку наиболее полный опыт интеракции.

Ключевые слова: речь, язык, взаимодействие «человек – компьютер».

SPEECH GENERATION SYSTEMS FOR THE NATURAL LANGUAGE PROCESSING: SOLUTIONS FOR HUMAN – MACHINE INTERACTION

D.V. Galkin, E.V. Kondrushina, K.S. Shilyaev

The article presents an overview of the present state of the problem of Natural Language Generation in its theoretical and applied aspects. Based on the concept of Question Answering Systems, the ways of developing human-computer interaction are review and a model of a maximally interactive computer system.

Keywords: speech, language, human-computer interaction.

Проблема компьютерной генерации речи

Проблема генерации речи находится в сфере исследований искусственного интеллекта и компьютерной лингвистики и сосредоточивается на компьютерных системах, способных производить понятные тексты на английском или других человеческих языках. Обычно такие системы основываются на нелингвистической репрезентации информации в качестве ввода и используют знание о языке и определенной предметной области для автоматического производства документов, докладов, объяснений, вспомогательных сообщений и других видов текста.

Технология генерации речи, кроме теоретической значимости, имеет множество практических применений. В качестве поля исследований генерация речи позволяет объединить фундаментальные проблемы в области искусственного интеллекта, когнитивной науки и взаимодействия человек-компьютер. Сюда относятся вопросы о том, как следует представлять и обрабатывать лингвистическое и энциклопедическое знание, каковы параметры корректного текста.

* Издание подготовлено при финансовой поддержке проекта № 6.4832.2011 «Организационно-правовые и философско-антропологические основания инновационных социокультурных практик (федеральный и региональный уровень)».

С практической точки зрения компьютерная генерация речи имеет огромное значение для развития моделей взаимодействия человека и компьютера. Все более широкое распространение получают так называемые голосовые интерфейсы для смартфонов, персональных компьютеров, видеоигр, мобильных систем навигации, систем управления транспортом, банкоматов, информационных сервисов и социальных роботов. В перспективе технологии генерации речи помогут создать гораздо более широкое и удобное взаимодействие человека с машиной. Отчасти это связано с естественной перцептивной ограниченностью экранных интерфейсов и большим удобством получать важную информацию от технических систем акустическим путем. Очевидно, что генерация речи может сыграть важную роль в автоматизации создания документов или функции предоставлять и объяснять сложную информацию людям, не имеющим достаточных фоновых знаний.

С исследовательской точки зрения генерация речи выступает в рамках более широкой проблемы обработки естественного языка, которая, в свою очередь, может рассматриваться как пересечение информатики и когнитивной науки. Вопросы, стоящие перед исследователями в области генерации речи, соприкасаются, таким образом, сразу с несколькими дисциплинами:

1. Информатика (взаимодействие человек-машина): Как должно происходить взаимодействие человека и машины? Какой способ лучше подходит для передачи информации компьютером человеку? Какое лингвистическое поведение ожидает человек от компьютера, с которым взаимодействует человек, и как его осуществить?

2. Лингвистика: Что составляет «подходящий» язык в определенной коммуникативной ситуации? Как можно формализовать соответствующие прагматические, семантические, синтаксические и психолингвистические ограничения? Какую роль играет контекст?

3. Исследования в области искусственного интеллекта (ИИ): Как типично компьютерное представление информации преобразовать в подходящее для человека, т. е. большие массивы числовых данных превратить в небольшое количество высокоуровневых понятий? Какие для этого необходимы модели мира и знания [2, с. 2]? Наиболее близко и глубоко исследования ИИ включаются в проблематику генерации речи с точки зрения семантической обработки естественного языка.

Взаимодействие человека и машины на основе естественного языка было и остается краеугольным камнем развития систем искусственного интеллекта. Достаточно упомянуть знаменитый тест Тьюринга, в котором наличие у компьютера интеллекта проверяется в естественном диалоге между человеком и машиной, подобно обычному разговору между

людьми. История развития систем искусственного интеллекта показывает, что эта проблематика не только не потеряла, но продолжает приобретать все большее значение.

Производство речи естественным образом связано с компьютерной обработкой речи: эти две области исследований разделяют многие теоретические основания и в прикладном аспекте часто функционируют вместе. На высоком уровне абстракции можно представить процесс генерации речи как процесс ее понимания наоборот: перевод машинного представления данных на человеческий язык и перевод человеческого языка на язык машинных представлений.

Однако принципиальное отличие генерации речи состоит в центральной роли принятия решений, от содержательного уровня до поверхностного текстового (как выбрать из многих возможных наиболее подходящей способ языковой реализации определенного содержания), в то время как система восприятия и обработки речи в основном занимается менеджментом гипотез (что бы могло значить данное выражение). По емкой формулировке одного из авторитетных исследователей, «там, где исследование интерпретации языка может описывать ограничение возможных опций, с целью наиболее точно определить выбор конкретной опции, исследование генерации должно определять, почему одна опция лучше другой в различных ситуациях» [6, с. 3].

С точки зрения практического применения большинство существующих систем генерации речи используются либо для того, чтобы представить пользователям информацию в более понятном виде или (частично) автоматизировать производство рутинных документов. Предоставление информации в более понятном виде важно потому, что форма внутренних представлений, используемая в компьютерных системах, часто требует значительных знаний для интерпретации. Такие представления, как база данных вылетов и прилетов самолетов, бухгалтерские таблицы, базы знаний экспертных систем, легко обрабатываются компьютером, но зачастую сложны для неподготовленного человека. Отсюда возникает необходимость в системах, которые могут представить данную информацию в понятной форме для не-эксперта. В случае, когда это представление должно осуществляться на естественном языке, используется технология генерации речи. Важно учитывать также тот аспект, будет ли система генерации речи функционировать полностью автономно, или ее задача будет состоять в том, чтобы производить черновики текстов для последующего редактирования автором-человеком, так как в некоторых контекстах невозможно создать тексты надлежащего качества или содержания без вмешательства человека.

В практическом применении современная технология генерации речи и информационные ограничения базисных систем далеко не всегда позволяют создавать конечный продукт. Вместо этого система генерации речи производит черновик документа, который может быть расширен или исправлен автором-человеком. Вариантом такого подхода является сосредоточение системы генерации речи на производстве разделов с фактическими данными (написание которых автору-человеку представляется монотонным и утомительным). При этом их анализ и интерпретация выполняются человеком. Примерами таких систем являются SumTime, FOG (метеорология, составление прогнозов), PlanDoc (приложение для инженеров телефонных сетей), Drafter (для написания технической документации программных продуктов) [2, с. 7–16].

Кроме того, технологии генерации речи нашли свое применение в обучении (система ICICLE помогает глухим людям овладеть английским синтаксисом), маркетинге (генерация описания музыкального альбома, нацеленного на увеличение продаж, система DYD), программах психологической помощи, мотивации и убеждения (генерация текстов, направленных на изменение аффективного состояния), развлекательных целях (генерация шуток и историй), предоставлении справки по работе с UNIX-системами (INTERIX [4, с. 184]).

Во всех перечисленных примерах машина осуществляет работу с формальной или структурной стороной языка – грамматикой и синтаксисом, оставляя семантическую обработку за пределами своих функций.

Машинная генерация речи: стадиальный подход

Генерация речи часто разделяется на три стадии: планирование документа, микропланирование и реализация [5, с. 127–128]. На стадии планирования документа решается вопрос о том, *что* следует передать в тексте (определение контента, информации) и о том, *как* организовать документ (структурирование документа). На выходе данной стадии обычно получается древо сообщений, которые затем могут быть преобразованы в высказывания. Уже на этом уровне задействована интеллектуальная обработка данных на экстралингвистическом уровне. Так, например, система STOP генерирует письма с просьбой прекратить курить на основе психологической модели о том, что следует говорить курильщикам в соответствии с особенностями их отношения к отказу от курения [2, с. 16].

На стадии микропланирования решается задача выражения информации на естественном языке. Интеллектуальной системе приходится принимать целый спектр решений: выбор лексики, референции, синтаксиса

и агрегации (выбор того, сколько сообщений следует выразить в каждом предложении). Входом на уровне микропланирования становится план документа, а выходом – текстовая спецификация: разделение на абзацы и предложения, а также глубокая синтаксическая структура последних.

На стадии реализации создается текст (поверхностная форма), основанный на решениях, которые интеллектуальная система принимает на предыдущих двух стадиях. Большинство программ-реализаторов основываются на формальных грамматиках. Существуют также проекты, основанные на анализе корпусов текста и вычленения из них грамматических правил, но проблема контроля за качеством реализации в таком случае не решена [9, с. 579–586].

Актуальными проблемами для систем генерации речи в настоящее время являются соотношение языка и реальности (например, при генерации языкового описания графического изображения, какой цвет назвать красным и будет ли такое название верным для конкретного контекста), влияние необходимости языкового выражения на до-вербальный анализ данных на входе системы генерации речи, интеграция лингвистического и нелингвистического знания, проблема соотношения текстового и графического выражения на выходе системы, интерактивность систем генерации речи, моделирование пользователя системы с целью подстройки определенных параметров производства речи [9, с. 593–595], моделирование персоналии генератора, основанное на «корреляции значительного числа языковых переменных (многословность, повторение, заполнение пауз, заикание)» [5, с. 139].

Вопросно-ответные системы как поле развития интерактивности

Мы сосредоточимся на проблеме интерактивности и перспектив ее развития в системах генерации речи, поскольку именно в интерактивной диалоговой форме сложнее всего реализовать на практике основные аспекты генерации речи. Наиболее актуально эта проблема встает в вопросно-ответных системах (Question Answering Systems), созданных для предоставления ответов на вопросы пользователя в режиме реального времени. Важно подчеркнуть, что в данном случае речь рассматривается не как монологическое высказывание, а как живая текучая ткань человеческого разговора – интерактивный процесс, основную роль в котором играет формирование разделяемой участниками семантики.

Изначально вопросно-ответные системы были разработаны, чтобы позволить пользователям задавать интересующие их вопросы, основанные на хорошо структурированных наборах данных, таких как бейсбольная статистика, данные персонала или химические исследования лунной

поверхности и образцов почвы [10]. Ответы на эти вопросы «хранились» в чётко структурированных базах данных, и ответная система требовала постановки вопроса, строго соответствующей формулировке, содержащимся в этих базах данных. Для успешного поиска ответа пользователь уже должен был знать, какая информация заложена в систему и как она структурирована компьютером. В этом и была основная проблема, на которой сосредоточились исследователи [1, с. 631].

Однако целью разработки вопросно-ответной системы было не только создание механизма непосредственно ответа на заданный вопрос, но и создание способа общения пользователей с компьютером на естественном языке.

Одной из самых ранних диалоговых систем является SHRDLU. В этой диалоговой системе пользователь мог общаться с роботом, при этом пользователь мог спросить SHRDLU о многом: от прошлых событий робота до его планов на будущее. Общение с SHRDLU и с другими подобными диалоговыми системами осуществлялось при помощи печатного текста. Спустя некоторое время также стали появляться системы с ограниченным речевым взаимодействием.

В настоящее время базовая вопросно-ответная система функционирует в рамках процесса, начало которого идёт от заданного вопроса и завершается выводом ответа или списка наиболее подходящих ответов (рис. 1). При введении вопроса большое значение имеет содержание вопроса и типизация вопроса. Это и есть два существенных пути развития процесса поиска ответа. При типизации вопроса идёт поиск определённого типа информации, а содержание вопроса помогает найти отрывок текста, в котором содержится данная информация, поскольку поиск ответа осуществляется на основе какого-либо текста. При этом типизация вопроса может помочь системе с выбором подходящего ответа при помощи некоторых правил, которые задают структуру определённого типа ответа. Например,

- если вопрос начинается с вопросительных слов *Кто?*, *Кого?* – спрашивается о ЧЕЛОВЕКЕ (типовой ответ – человек);
- если вопрос начинается с вопросительного слова *Где?* – спрашивается о МЕСТОПОЛОЖЕНИИ (типовой ответ – местоположение) [1, с. 635].

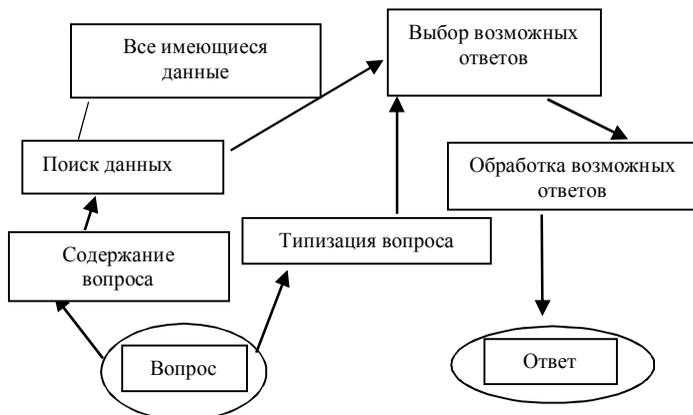


Рис. 1. Базовая вопросно-ответная схема

Вопросно-ответные технологии движутся в нескольких направлениях, три из которых являются основными:

- 1) расширение отношений между вопросом и корпусом текстов;
- 2) увеличение ряда вопросов, на которые можно дать ответ;
- 3) установление более тонкого взаимодействия между пользователем и системой.

На последнем направлении остановимся подробнее.

Несмотря на то, что вопросно-ответные системы разрабатывались в аспекте поиска единственного правильного ответа на поставленный вопрос, следует учитывать, что для разных пользователей разные вопросы могут считаться «правильными». Наглядным примером возможной неоднозначности, где не будет единственного правильного ответа, могут послужить вопросы типа:

Где находится мост Веррзано-Нарроус?

Где находится дамба Три Ущелья?

Житель Северной Америки мог бы ответить на первый вопрос: «Между Бруклином и Статен-Айленд», а на второй: «Где-то в Китае». А житель Азии может дать ответы на те же вопросы: «В Нью-Йорке» и «В западной провинции Хубэй». При этом и первый, и второй человек ответят верно.

Учитывая этот фактор, ряд исследователей предложили не выводить самый вероятный ответ из общего числа возможных, а дать все самые

подходящие из них, позволяя пользователю самому выбрать верный для себя ответ (Quarteroni, Manandhar, 2009) (цит. по [1, с. 646–648]).

Другие авторы продемонстрировали, что подобный подход может быть использован в случае двусмысленных вопросов, например: *Где находится Тадж-Махал?* (захоронение? казино? ресторан? и т.д.) (Dalmas, Webber, 2007) (цит. по [1, с. 646–648]).

Однако самым главным способом обогащения взаимодействия пользователя и вопросно-ответной системы является использование интеракции.

Интерактивная вопросно-ответная система может быть обозначена как процесс, в котором пользователь является значимой частью механизма создания информации: пользователь создаёт вопрос, отслеживает релевантность информации, получает конечный продукт.

Интерактивные вопросно-ответные системы заимствуют от диалоговых систем способ взаимодействия пользователя с компьютером, акцент на завершение пользовательской задачи, их обработку неполных данных. Вместо того чтобы пытаться решить сложную двусмысленность вопроса самостоятельно, независимо от того, вызван он пользователем или контекстом, система может предложить пользователю вопросы, которые служат существенным инструментом ввода и уточнения информации для процесса поиска. Для достижения этого необходимы «понимание» программой контекста вопроса пользователя и некоторые знания из области знаний, на которую ориентирован вопрос.

В рассмотрении интерактивности остается открытым вопрос о возможности лингвистического обучения компьютеров в процессе вопросно-ответного взаимодействия. Интерактивность между людьми – если брать за образец именно ее – предполагает возможность обогащения семантического поля, синтаксического набора и трансформации исходных установок (знаний) участников взаимодействия (включая эмоциональную динамику).

Модель идеальной интерактивной системы генерации речи

Диалоговая составляющая вопросно-ответных систем все еще несовершенна: по специальным методикам измерения, их эффективность не превышает 75% в лучшем случае [1, с. 644], существуют и проблемы в машинной интерпретации естественного языка, и в представлении знаний. Однако уже сегодня, основываясь на таких проектах, как Prosthetic Head [7], можно попытаться выработать требования к системе, способной к подлинно интерактивному процессу генерации речи:

1. Машина должна быть способна воспринимать звучащую человеческую речь (технология Speech Recognition), корректно интерпретировать высказывание исходя как из узкого контекста, так и из широкого (модель мира) (Natural Language Processing) и адекватно реагировать на реплики собеседника-человека. Сюда же относится восприятие просодии высказывания. Кроме того, предполагается возможность восприятия текстовых данных при опосредованной коммуникации человека – машины (в рамках чата, например).

2. Машина должна быть способна воспринимать и интерпретировать паралингвистические параметры коммуникации – жесты, мимику, позу собеседника, направление взгляда, внешний вид (одежда, физическое состояние собеседника).

3. Машина должна быть способна обнаруживать свои реакции вовне путем как речевого синтеза (Speech Synthesis), способного, в идеале, производить неотличимый от естественного языка фонетический продукт (в том числе при необходимости могут имитироваться региональные акценты и индивидуальные особенности произношения), так и текстовой или графической реакции.

4. Машина должна уметь выражать те паралингвистические параметры коммуникации, которые она способна воспринимать. Для этого, очевидно, она должна обладать телом, подобным человеческому (возможно в виде изображения на экране).

Перечисленные требования представляются нам своего рода *идеалом* машины-собеседника, нацеленной на *имитацию* собеседника-человека. Практическое же приложение интерактивности в процессах генерации речи видится нам в практике построения и функционирования вопросно-ответных систем. Выгодно отличаясь от интернет-поисковиков точностью ответа, такие системы могли бы стать новым витком в технологии взаимодействия *человек – компьютер*.

ЛИТЕРАТУРА

1. Clark A., Fox C., Lappin S. The Handbook of Computational Linguistics and Natural Language Processing. USA: Wiley-Blackwell, 2010. 801 p.

2. Dale R., Reiter E. Building Natural Language Generation Systems UK: Cambridge University Press, 2000. 270 p.

3. Dalmás, Tiphaine, & Bonnie Webber. Answer comparison in automated question answering // Journal of Applied Logic 2007. 5(1). P. 104–120.

4. Danlos L. The Linguistic Basis of Text Generation. UK: Cambridge University Press, 1987. 238 p.

5. McDonald D. Natural Language Generation // Handbook of Natural Language Processing / ed. N. Indurkha, F. Damerou. 2nd ed. USA: Chapman & Hall, 2010. 676 p.

6. *McKeown K.* Text Generation. NY: Cambridge University Press, 1992. 246 p.
7. *Prosthetic Head* [Электронный ресурс] : Stelarc. – Офиц. сайт. – Australia, 2012. – Режим доступа: <http://stelarc.org/?catID=20241> (дата обращения: 13.12.2012).
8. *Quarteroni Sylvia, Suresh Manandhar.* Designing an interactive open-domain question answering system // Journal of Natural Language Engineering: Special Issue on Interactive Question Answering. 2009. 15(1). P. 73–95.
9. *Reiter E.* Natural Language Generation // Handbook of Natural Language Processing / ed. A. Clark et al. UK. : Blackwell Publishing, 2010. 801 p.
10. *Simmons, R. F.* Answering English questions by computer: a survey // Communications of the ACM. 1965. 8(1). P. 53–70.