Как работает Siri?

Тут все очень интересно. Если другие голосовые помощники до этого работали просто с поисковой системой, то Siri работает с множеством сервисами, что позволяет точно отвечать на самые разные вопросы, в том числе и очень сложные.   
  
Задав вопрос, он отравляется на сервера Apple (Siri), где обрабатывается и направляется к соответствующему сервису. А тут не только поисковые гиганты *Google и Bing*… Например, для деловых вопросов используются OpenTable, André Gayot, Citysearch, BooRah, Yelp Inc, Yahoo Local, ReserveTravel и Localeze. Для поиска информации о мероприятиях Siri обращается к Eventful, StubHub и LiveKick. Если спрашивать Siri о фильмах, то она отвечает, используя информацию с MovieTickets.com, Rotten Tomatoes и The New York Times... Таким образом голосовой ассистент от Apple справится с большинством повседневных вопросов, но ключевой особенностью является то, что Siri работает с WolframAlpha.   
  
WolframAlpha позволяет*Siri* давать ответы на самые трудные вопросы, так как это не поисковая система. *WolframAlpha*позиционирует себя, как *computational knowledge engine* (перевод: база знаний и набор вычислительных алгоритмов).   
  
Благодаря всему вышеперечисленному, Siri удается понимать речь человека и его вопросы, которые он задает в достаточно свободной форме, а не конкретные команды. На презентации в пример приводили вопрос: «*Стоит ли мне сегодня взять зонтик*». **Siri**анализирует вопрос и понимает, что ей нужно ответить - какая погода ожидается в этой местности.   
  
Конечно же, Siri еще не совершенна, но сам факт того, что такое решение появилось в мобильных телефонах не может не говорить о перспективном будущем технологии распознавания речи и компьютерных голосовых ассистентов.   
  
И напоследок предлагаю Вам посмотреть небольшую подборочку [смешных ответов Siri](http://app-s.ru/news/u_siri_est_chuvstvo_jumora/2011-10-14-540).

Если вы не нашли ответа на свой вопрос или у вас что-то не получилось, а в комментаряих ниже нет подходящего решения, задайте вопрос через нашу [форму помощи](http://app-s.ru/blog/0-0-0-2-1). Это быстро, просто, удобно и не требует регистрации. Ответы на свой и другие вопросы вы найдете в [данном](http://app-s.ru/blog/help/1-0-2)разделе.

# **История развития систем распознавания речи: как мы пришли к Siri**

Оглядываясь назад, мы видим, что развитие технологии распознавания речи похоже на наблюдения процесса взросления у ребенка — прогрессируя от определения отдельных слов, затем все больших словарей и наконец до быстрых ответов на вопросы, как это делает Siri.  
  
Слушая Siri с ее слегка элегантным чувством юмора, мы восхищаемся тем, как далеко продвинулись за эти годы в отрасли распознавания речи. Давайте взглянем в прошлые десятилетия, которые позволили человеку управлять девайсами, используя только голос. 

#### **1950 и 1960: Детский лепет**

Первые системы распознавания речи могли понимать только цифры (учитывая сложность языка, это правильно, что инженеры сначала сфокусировались на цифрах). Bell Laboratories разработали систему «Audrey», которая распознавала цифры, сказанные одним голосом. Через 10 лет, в 1962 году, IBM продемонстрировала их детище — систему "[Shoebox](http://www.ibm.com/ibm100/us/en/icons/speechreco/breakthroughs/)", которая понимала 16 слов на английском.  
  
Лаборатории в США, Японии, Англии и СССР разработали еще несколько аппаратов, которые распознавали отдельные произнесенные звуки, расширив технологию распознавания речи поддержкой четырех гласных и девяти согласных звуков. Звучали они не очень хорошо, но эти первые попытки дали впечатляющий старт, особенно если учитывать, насколько примитивными были компьютеры того времени.

#### **1970-е: Системы постепенно приобретают популярность**

Системы распознавания речи сделали большие шаги в семидесятых благодаря интересу и спонсированию от министерства обороны США. Их программа DARPA Speech Understanding Research (SUR) с 1971 по 1976 год была одной из самой большой в истории распознавания речи, и помимо всего остального она отвечала за систему «Harpy» Университета Карнеги Меллона. «Harpy» понимала 1011 слов, что является средним словарным запасом трехлетнего ребенка.  
  
«Harpy» была значительной вехой, так как она представила более эффективный подход к поиску, называемый [Beam search](http://en.wikipedia.org/wiki/Beam_search), «демонстрируя сеть возможных предложений с конечным числом состояний» ( [Readings in Speech Recognition](http://books.google.com/books?id=iDHgboYRzmgC&pg=PA551&lpg=PA551&dq=darpa+harpy&source=bl&ots=ja6SEQTlgF&sig=YkBDHyTTK-8_0FnR_94sbsoS6IQ&hl=en&ei=1oupTpekCanZ0QHP6vy_Dg&sa=X&oi=book_result&ct=result&resnum=4&ved=0CDUQ6AEwAw" \l "v=onepage&q=darpa%20harpy&f=false)).  
  
70-е годы также отмечены еще несколькими вехами в данной технологии, например основанием первой коммерческой компании Threshold Technology, которая представила систему, которая могла интерпретировать различные голоса.

#### **1980-е: Распознавание речи оправдывает прогнозы**

В следующей декаде благодаря новым подходам и технологиям словарный запас подобных систем вырос с нескольких сотен до нескольких тысяч слов и имел потенциал распознавания неограниченного количества слов. Одной из причин был новый статистический метод, больше известный как [скрытая марковская модель](http://ru.wikipedia.org/wiki/%D0%A1%D0%BA%D1%80%D1%8B%D1%82%D0%B0%D1%8F_%D0%BC%D0%B0%D1%80%D0%BA%D0%BE%D0%B2%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C).   
  
Используя шаблоны для слов и звуковые паттерны, она рассматривала вероятность того, что неизвестные звуки могли быть словами. Эта база использовалась другими системами еще на протяжении двадцати лет ([Automatic Speech RecognitionA Brief History of the Technology Development](http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf)).   
  
С расширенным словарным запасом распознавание речи начало протаптывать себе дорожку в коммерческие приложения для бизнеса и специализированных отраслей, таких как медицина. Она даже вошла в дома обычных людей в 1987 году в виде куклы [Worlds of Wonder's Julie doll](http://www.squidoo.com/julie-talking-doll), которые дети могли натренировать, чтобы она распознавала их голос («Наконец-то кукла, которая тебя понимает»).  
  
  
  
Хоть ПО по распознаванию могло распознавать до 5000 слов, как, например, программа Kurzweil text-to-speech, в них был огромный недостаток — эти программы поддерживали дискретную надиктовку, то есть вы должны были останавливаться после каждого слова, чтобы программа его обработала.

#### **1990-е: Автоматическое распознавание речи идет в массы**

В девяностых компьютеры наконец-то получили быстрые процессоры, и программы по распознаванию речи стали жизнеспособными.  
  
в 1990 году появилась первая общедоступная программа Dragon Dictate c ошеломляющей ценой 9000 долларов. Спустя семь лет вышла улучшенная версия — [Dragon NaturallySpeaking](http://www.pcworld.com/article/4823/naturallyspeaking_voice_recognition_leaps_forward.html). Приложение распознавало нормальную речь, поэтому вы могли говорить в обычном темпе около 100 слов в минуту. Но все равно, вы должны были тренировать программу в течении 45 минут перед использованием, и она имела все еще высокую цену в 695 долларов.  
  
Появление первого голосового портала VAL от BellSouth было в 1996 году. Это была первая интерактивная система распознавания речи, которая давала информацию, основываясь на том, что вы сказали в трубку телефона. VAL вымостила дорогу для всех неточных на то время голосовых меню, которые надоедали звонящим в следующие 15 лет.

#### **2000-е: Застой в распознавании речи — пока не появился Google**

К 2001 году распознавание речи поднялось до 80-процентной точности, и прогресс технологии остановился. Системы распознавали работали отлично, когда языковая вселенная была ограниченной, но они до сих пор «догадывались» при помощи статистических моделей среди похожих слов, языковая вселенная росла вместе с ростом Интернета.   
  
Знали ли вы о том, что распознавание голоса и голосовые команды был встроены в Windows Vista и Mac Os? Большинство пользователей даже не догадывалось, что такая функциональность существует. Windows Speech Recognition и голосовые команды OS X были интересными, но [недостаточно точными](http://www.pcworld.com/article/126613/vistas_voice_recognition_stammers.html)и удобными, как клавиатура и мышь.  
  
Технология распознавания речи получила второе дыхание после одного важного события: появления приложения Google Voice Search для iPhone. Влияние этого приложения было значительным по двум причинам. Во-первых, телефоны и другие мобильные девайсы — это идеальные объекты для распознавания речи, и желание заменить крошечные экранные клавиатуры альтернативными методами ввода было очень велико. Во-вторых, у Google была возможность разгрузить этот процесс, используя свои облачные дата-центры, направив всю их мощь для крупномасштабного анализа данных для поиска совпадений между словами пользователей и огромного числа образцов голосовых запросов, которые они получали.   
  
Вкратце, узким местом распознавания речи всегда было доступность данных и возможность эффективной их обработки. Приложение же добавило к анализу данные миллиардов поисковых запросов, чтобы лучше предугадывать, что же вы сказали.  
  
В 2010 году Google добавил персональное распознавание в голосовой поиск телефонов под управлением Android. Программное обеспечение могло записывать голосовые запросы пользователей для построения более точной голосовой модели. Также компания добавила распознавание голоса в свой браузер Chrome в середине 2011 года. Помните как мы начали с 10 слов и продвинулись до нескольких тысяч? Так вот система Google теперь позволяет распознать 230 миллиардов слов.  
  
Потом появилась Siri. Так же, как и система Goggle Voice Search, она полагается на облачные вычисления. Она использует те данные, которые ей известны о тебе, чтобы сгенерировать вытекающий из контекста ответ и отвечает на твой запрос, как некая личность. Распознавание речи превратилось из инструмента в развлечение. 

#### **Будущее: Точная и повсеместная речь**

Бум приложений по распознаванию речи указывает на то, что время распознавания речи пришло, и мы можем ожидать огромного количества их и в будущем. Эти приложения не только позволят контролировать компьютер с помощью голоса или конвертировать голос в текст — они будут также способны различать разные языки, позволят выбирать голос помощника из различных вариантов.  
  
Вполне вероятно, что технология распознавания речи перейдет и на другие типы девайсов. Нетрудно представить, как в будущем мы будем управлять кофеварками, разговаривать с принтерами и говорить освещению, чтобы оно выключалось