

<b>Текст доклада:</b>	Обзор систем для работы в среде <i>GNU/Linux</i> без зрительного контроля
<b>Автор:</b>	Михаил Пожидаев
<b>Конференция:</b>	Вторая международная конференция разработчиков свободных программ “На Протве”
<b>Место проведения:</b>	г. Обнинск
<b>Дата:</b>	26 июля 2005 г.

Средствами, позволяющими работать без зрительного контроля в некоторой среде, называются программы, анализирующие содержимое экрана и пытающиеся представить графическую информацию в текстовом виде. Пользователь может получать результат в виде синтезированной человеческой речи и через брайлевский дисплей. Брайлевские дисплеи очень дороги, среди пользователей встречаются редко, и поэтому мы не будем их рассматривать в рамках этого доклада. Вся работа рассматриваемого программного обеспечения разбивается на 3 группы:

1. Снятие содержимого экрана и представление его в текстовом виде.
2. Синтезирование человеческой речи.
3. Организация взаимодействия программ из предыдущих двух пунктов и решение проблем национальных языков.

Программы, выполняющие снятие графической информации, обычно называются *screen reader*'ами. В общем случае можно выделить два подхода в их работе:

1. Перехват вывода информации на экран.
2. Анализ внутреннего состояния среды, где работает пользователь.

Как это обычно бывает, каждый из подходов имеет свои достоинства и недостатки. Второй подход на практике оказывается более эффективным, его использование позволяет, например, отделить редактируемый текст от оформления текстового редактора. Пользователь получает только ту информацию, которая ему нужна. Кроме того, он является единственно возможным при использовании графической среды. Его применение требует наличия специальных фильтров для каждого приложения, которые использует пользователь.

Одной из первых программ, созданных для работы пользователей с ослабленным зрением, является *emacspeak*. Она же является сегодня самой развитой. Авторы называют свой продукт не *screen reader*'ом, а *audio desktop*'ом, подчёркивая его развитость. По своей природе она является набором дополнений к среде *GNU Emacs*, написанных на языке *Lisp*. Популярность её объясняется в некоторой мере многофункциональностью среды *GNU Emacs*, в некоторой мере его расширяемостью. При использовании *emacspeak* пользователь может вполне уверенно работать с текстом, программировать, пользоваться электронной почтой и новостями, запускать в ограниченном объёме консольные приложения, читать *man* и *info*. В общем, перечень возможностей определяется возможностями самой среды *GNU Emacs*. К настоящему моменту в *emacspeak* реализована возможность хорошо взаимодействовать с *World Wide Web*. Синтезом речи *emacspeak* не занимается. Предполагается наличие внешней программы, которая способна принять текст для синтеза от оболочки и воспроизвести его для пользователя.

Единственным *screen reader*'ом, позволяющим работать пользователю в графической среде, является программа *Gnopernicus*, предназначенная для работы в среде *Gnome*. Подобно тому, как *emacspeak* встраивается в *GNU Emacs*, *Gnopernicus* встраивается в *Gnome* и, анализируя события, происходящие внутри среды, и ввод пользователя, пытается представить всю необходимую информацию в форме, доступной для пользователя. Нужно отметить, что *Gnopernicus* не ограничивается только выводом синтезированной речи, но и способен выдавать текст на брайлевские дисплеи. Разработчики *Gnopernicus*'а предусмотрели возможность взаимодействия с существующими *OpenSource* речевыми синтезаторами, о которых будет рассказываться далее.

Теперь остановимся на рассмотрении программ для перехвата информации, выводимой на экран в терминальном режиме. Использовать такие программы очень удобно, если пользователю приходится выполнять большое количество действий из командной строки.

Проект *YASR* задумывался как маленькое очень лёгкое, переносимое приложение, порождающее виртуальный терминал и обрабатывающее всё, что в нём будет появляться. На данный момент этот проект находится на стадии разработки, не было выпущено ни одной стабильной версии, но даже отладочные варианты вполне пригодны для использования. При использовании программных речевых синтезаторов *YASR* представляет речевую информацию в том же виде, что и *emacspeak*. Таким образом, можно использовать одно и то же приложение для обработки вывода от *emacspeak* и от *YASR*.

Другой программой, похожей на *YASR* по возможностям, является проект *Speak Up*. Этот пакет очень похож на *YASR* по своему поведению, но выполнен в виде патча к ядру операционной системы. Такая форма реализации позволяет озвучить всю работу операционной системы от начала загрузки до выключения компьютера. Но это возможно только при использовании аппаратных речевых синтезаторов. Общая проблема использования средств экранного доступа осложняется тем, что пользователь должен выбрать что-нибудь одно. На практике воспользоваться какой-либо комбинацией программ для работы в терминальном режиме не получается.

Сейчас рассмотрим, чем можно синтезировать речь в среде *GNU/Linux*. Самым крупным проектом речевого синтезатора является пакет *ViaVoice* компании *IBM*. Долгое время этот пакет разрешалось распространять свободно, но два года назад руководство компании сделало этот пакет недорогим, но всё же коммерческим продуктом. Ограничимся только заметкой о существовании такого варианта.

Исторически первым и самым крупным открытым проектом речевого синтезатора является проект Эдинбургского университета *Festival*. Помимо конкретно самого синтезатора командой разработчиков был создан отдельный пакет инструментов для обработки речи — *speech tools*. Сам по себе он написан на *C++*, но имеет отдельный интерфейс из языка *Schema*. В числе прочих возможностей нужно отметить возможность работать в варианте клиент/сервер. Но на практике он оказывается довольно громоздким и неповоротливым, хотя и обладает приятным качеством звучания синтезированной речи. После длительного периода простоя разработка продолжается, готовится к выпуску версия 2.0. На данный момент уже доступны бета версии.

В технологическом институте Карнеги-Меллон в Питсбурге создали на основе *Festival* отдельный синтезатор речи — *Flite* (*Festival lite*). В результате получился очень лёгкий и простой в использовании речевой англоязычный синтезатор. Но, к сожалению, качество его звучания оставляет желать лучшего. Впоследствии он был переписан на языке *Java*, что привело к созданию синтезатора, известного под именем *FreeTTS*.

Очень приятным качеством звучания обладает речевой синтезатор *Mbrola*. Точнее говоря, утилита *Mbrola* занимается генерацией звукового вывода, используя для этого речь, разложенную на фонетические составляющие. Такого рода разложением должна зани-

маться отдельная программа-препроцессор. Для этого можно использовать свободно распространяемую утилиту *FreeSpeech*. *Mbrola* может быть свободно скачана с сайта разработчика, но её исходные тексты закрыты.

В общем случае, открытого синтезатора, хотя бы англоязычной речи, подходящего во всех отношениях, нет. И такое положение вещей имеет свои причины. Дело в том, что в западных странах и Соединённых Штатах Америки широко распространены аппаратные синтезаторы речи, которые представляют из себя отдельное внешнее устройство, подключенное к компьютеру, и в которое только посылается текст для воспроизведения. Некоторые проекты, такие как *Speak Up*, изначально вообще не поддерживали возможность работы с программными речевыми синтезаторами.

С речевыми синтезаторами, способными генерировать русскую речь, ситуация более чем печальная. Единственным примером такого синтезатора является утилита *ru\_tts*, распространяемая без исходных текстов и каких-либо сформулированных условий распространения. Эта утилита генерирует хорошо различимую речь и очень проста в использовании. В 2002 году был распространён диск *SlackSpeak*, который представлял односторонний вариант дистрибутива *Slackware 7.0*, с подготовленным для использования *Emacspeak*. Речь синтезировалась в нём при помощи *Mbrola* в случае английского языка и при помощи *ru\_tts* в случае русской речи. Интересной особенностью этого диска была возможность полностью установить *Slackware* на жёсткий диск без зрительного контроля, используя консольный вариант программы установки.

Мысль о создании отдельного речевого сервера появилась при изучении существующих речевых серверов для *Emacspeak*. Программа, обрабатывающая вывод от *Emacspeak*, запускается им один раз при старте и получает информацию для озвучивания в отформатированном по специальным правилам виде через поток стандартного ввода. Главная цель такой программы — организовать совместную работу нескольких одновременно запущенных копий *Emacs*'а. Если выясняется, что программа запускается в первый раз, то она порождает новый процесс. Сервер делает небольшую задержку, чтобы подождать, пока основной процесс запустится, а затем подключается к нему как клиент и посылает всё, что он получил от *Emacspeak*. Если программа запускается во второй и более раз, то она просто подключается к уже существующему серверу и посылает текст для воспроизведения.

Что в таком виде неудобно? Сегодня у каждого синтезатора речи есть свой собственный сервер для *emacspeak*. Так *Flite* обладает *eflite*'ом, а у *FreeTTS* речевой сервер есть среди программ демонстрации.

Предпринимались попытки создания сервера для *Festival*, но автор этой идеи бросил её. Было бы приятно иметь один раз написанный сервер и просто реализовывать в нём интерфейсы для разных синтезаторов. Помимо экономии усилий по разработке появляется интересная возможность проговаривать, например, знаки препинания синтезатором, отличным от того, который используется для проговаривания обычного текста. Тем более, сегодня перечень средств для работы без зрительного контроля уже не ограничивается *Emacspeak*'ом, есть *Gnopernicus*, *YASR* и пр. Разработчики *Gnopernicus* явно определили на сайте перечень синтезаторов, к которым реализован интерфейс. Среди них не оказалось *mbrola*, обладающего лучшим качеством звучания, и если *mbrola* не полностью *OpenSource*, то доступен для скачивания с сайта, и пользователь вполне может использовать именно его. Логично было бы переложить всю работу по обработке речи на отдельную программу. И остаётся самый сложный вопрос о поддержке национальных языков. Некоторые синтезаторы обладают возможностью воспроизводить речь на языках, отличных от английского. Так, например, *Festival* способен разговаривать на испанском языке, а *Flite* говорит по-немецки, но в общем случае этого явно недостаточно. Скорее всего, пользователь не

сможет ограничиваться только одним синтезатором, и ему придётся использовать некоторую их комбинацию. Поэтому одной из главных отличий *VoiceMan* является способность анализировать поступающую информацию и принимать решение, в какой синтезатор её направить. Такая возможность есть в речевом сервере на диске *Slackspeak*, но его возможности ограничиваются поддержкой только синтезаторов *ru\_tts* и *Mbrola*.