Comentarios sobre el Software de Código abierto “Sphinx”

Nota de divulgación

Lic. Gilberto Anduaga Márquez (1) Héctor Caudel García (2)
(1) Miembro del Cuerpo Académico de Sistemas Distribuidos del Instituto Tecnológico de Aguascalientes (2) Alumno de la carrera de Licenciatura en Informática del Instituto Tecnológico de Aguascalientes
Departamento de Sistemas y Computación del Instituto Tecnológico de Aguascalientes, Av. A. López Mateos No 1801, Fracc: Bona Gens Aguascalientes, Aguascalientes C.P 20256 Tel (449)91050002, Fax: (449)970423
ganduaga@hotmail.com, kaudel@gmail.con

Resumen

Este artículo presenta el software de código abierto Sphinx para Unix, mismo que es utilizado como un motor de reconocimiento y síntesis de voz, y puede ser implementado para cambiar el enfoque de las nuevas arquitecturas de los programas para no depender del teclado y el mouse sino sólo de la voz

Palabras clave

Unix, código abierto, voz, reconocimiento, software

Introducción

Muchas veces se ha hablado sobre el reemplazo del teclado y del mouse como las interfaces predilectas para conectarse con las computadoras y las aplicaciones que ejecutan.

La gran mayoría de las aplicaciones modernas que requieren un alto nivel de interacción con un usuario están desarrolladas utilizando una interfaz gráfica (GUI), tanto para escritorio como para Web; en ambos casos el ratón y el teclado son los encargados de activar determinados eventos, a partir de los cuales se centra la programación de acciones a desarrollar y los cambios en la presentación gráfica.

Uno de los errores más graves al implementar el manejo de voz como interfaz con el usuario (SUI, Speech User Interface) en una aplicación es utilizar la interpretación de la voz para completar cuadros de texto o listas en cajas de diálogo, comandar click en botones y opciones del menú, etc. Lamentablemente, ese ha sido el enfoque del software de reconocimiento de voz, en el cual no se modifica en el enfoque de las aplicaciones.

Desarrollo

El desarrollo de aplicaciones que maneje una interfaz con el usuario basada en voz requiere un paradigma en el diseño y un nuevo enfoque pensado desde el diálogo entre el usuario y el sistema informático. En el caso de programar una agenda en el que se utilizará el motor de reconocimiento de voz para reemplazar las acciones realizadas por el mouse y el teclado sería solo un laberinto de palabras sueltas para indicar en qué control de caja se desea posicionar. Por otra parte, si se cambia el paradigma y se transforma esta acción en un diálogo entre un usuario y un prestador de servicios, se encontrarán ordenes, preguntas y respuestas que permitirán ir completando los parámetros necesarios para agendar una cita, como sucedería en un diálogo entre dos personas. Los datos necesarios para completar una orden determinada se van preguntando y se va recopilando la información; cuando se acumula toda, se confirma y lleva a cabo la acción solicitada. Así funcionan los seres humanos, por lo cual es una forma más conveniente se hacer las cosas.

A diferencia de lo que sucede en otros campos, en los motores de voz para GNU/Linux existe una brecha muy marcada entre los productos con licencias comerciales y los de código abierto o licencias libres. Una de las razones por las que aparece esta notable diferencia es que muchos desarrolladores que estaban participando en proyectos de código abierto fueron contratados por empresas para que trabajasen con ellos en la realización de productos comerciales (como sucedió con Cvoice Control). Es por esto que encontramos que muchos proyectos que tenían un gran futuro aparecen sin nuevas actualizaciones y quedaron descontinuados. Por otra parte hay varios que utilizan llamadas al SDk de ViaVoice de IBM, el cual es un producto bajo licencia.

Una de las principales desventajas que se encuentran en los motores de voz de código abierto o licencias libres es que suelen ser muy complejos de configurar y de utilizar. La otra es que es muy común que no exista documentación clara.

Uno de los motores para reconocimiento de voz de código abierto más interesante, adaptable al español, actualizado y moderno es Sphinx, desarrollado por las universidades de Carneige Mellon y con colaboraciones de Sun Microsystems y de los laboratorios de investigación de Mitsubishi Electric.

Es de código abierto y la licencia requiere que se mantengan las menciones de copyright y el nombre de los autores. El nombre completo del proyecto es "CMU Sphinx Group Open Source Speech Recognition Engines".

Lo interesante de Sphinx es que es un proyecto que está vigente, tiene continuidad, una documentación bastante buena incluyendo varios ejemplos y ofrece versiones en lenguaje C y Java, lo cual fue bueno al momento de integrarlo a aplicaciones existentes.

El motor de reconocimiento de voz Sphinx utiliza Modelos Ocultos de Markov (HMM, en inglés) y funciones de densidad probabilísticas con salidas PDF. Sphinx viene en varias versiones las cuales se pueden descargar de manera gratuita del sitio del proyecto. A continuación se mencionarán sus características:

Sphinx-2:

Está orientado a brindar el mejor rendimiento posible en tiempo real, sacrificando un poco la precisión. Está escrito en lenguaje de programación C.

Sphinx-3:

Esta orientado a brindar la mayor precisión posible con vocabularios extensos. Inicialmente fue pensado para el procesamiento por lotes. Luego, con la aparición de mejores procesadores y memorias más económicas, se utilizá para ofrecer un reconocimiento de voz con vocabulario extenso en tiempo real con un rendimiento muy aceptable. Está escrito en lenguaje de programación C.

Sphinx-4:

Tiene las mismas características que Sphinx-3, pero esá totalmente escrito en lenguaje de programación Java. Se debe tener en cuenta que es una versión beta, por lo cual aunque se mantiene estable aun tiene varios inconvenientes.

Para probarlo o integrarlo a alguna aplicación, la versión a utilizar dependerá de la conveniencia pues si la aplicación está escrita en C o C++ es mucho más conveniente acoplarle un motor en C. Si está escrita en Java por ende se usará Sphinx-4.

Las versiones de Sphinx escritas en C ofrecen un rendimiento muy superior a las escritas en Java. Ahora bien la integración de Sphinx-4 es un poco más sencilla, pues como esta construido a base de clases es más simple de comprender y aprovechar. Esta última requiere de mucha mayor capacidad de procesamiento y aproximadamente 1 Gb de memoria para ofrecer un rendimiento conveniente para realizar un reconocimiento de voz en tiempo real. Para máquinas no tan poderosas es conveniente la versión escrita en C.

Sphinx puede trabaja en modo cliente servidor, por lo cual un cliente puede solicitar servicios de reconocimiento de voz a un servidor que se esta ejecutando en otro host. Esto es muy apropiado para el desarrollo de aplicaciones. De manera muy similar a un motor de bases de datos ejecutándose en un servidor independiente del cliente, así también se puede hacer con el motor de reconocimiento de voz de Sphinx.

Por otro lado se tiene el módulo de entrenamiento SphinxTrain para generar los modelos que luego se pueden cargar en diferentes versiones de Sphinx

Conclusiones

El software Sphinx puede ser una herramienta bastante poderosa para desarrollar software con una interfaz de voz. Además de ser gratis se incorpora a dos de los lenguajes muy usados como lo son Java y C, con lo cual su implementación en un sistema puede ser más fácil; asimismo también se puede modificar por ser "open source", que son las tendencias de las nuevas tecnologías de software

Referencias

[1] Revista Mundo Linux No. 88
http://cmusphinx.sourceforge.net/html/cmusphinx.php