

Apéndice B

Documentación básica de *Sphinx 3*

Los documentos que encontré y que considero relevantes en cuanto a información técnica acerca de *Sphinx* se enlistan a continuación:

Speech at CMU Web Page Es la página de la Universidad de Carnegie Mellon dedicada a la investigación, desarrollo y 'despliegue' de tecnologías de voz. Es el punto de entrada a las tecnologías de voz que maneja esta Universidad, entre ellas *Sphinx* .

The CMU Sphinx Group Open Source Speech Recognition Engines Página principal del *Sphinx Group Project*. Contiene la documentación general de *Sphinx*, los componentes de *Sphinx*, las distintas versiones de los motores de reconocimiento de *Sphinx* que existen y algunas ligas externas a sitios de procesamiento de voz. Las secciones de documentación que destacan son el tutorial de *Sphinx*, el manual de *Sphinx* y la sección de wiki de la documentación.

Robust group's Open Source Tutorial Página que contiene un pequeño tutorial que muestra el proceso de configuración básica de los componentes de *Sphinx*. Además, contiene una breve introducción para aprender a utilizar *Sphinx* y llevar a cabo un entrenamiento y un reconocimiento preliminares. Como apéndices contiene referencias básicas de trifonemas, HMMs, ligado de estados y el modelo del lenguaje.

Manual for the Sphinx-III recognition system Reune 6 documentos escritos por Rita Singh. Contiene una breve explicación del método de estimación por máxima similitud, la diferencia entre HMM discretos, continuos y semicontinuos, el set de instrucciones para el entrenamiento, una sección de preguntas más frecuentemente realizadas (FAQ), una sección de solución de posibles problemas y una sección sobre algunos tópicos de

decodificación. Debido a su relevancia, los últimos 4 documentos también se incluirán en esta lista.

Instruction set for training (Sphinx-3 trainer/decoder manual)

Contiene la serie de procedimientos necesarios que se deben de llevar a cabo tanto para el entrenamiento de modelos continuos como el de modelos semicontinuos con *SphinxTrain*. Describe en forma detallada las aplicaciones que se emplean en cada etapa así como sus parámetros. También incluye una sección de los requerimientos necesarios y consideraciones que se deben de tomar en cuenta antes de comenzar el entrenamiento.

Sphinx-3 FAQ Documento de preguntas más frecuentemente realizadas (FAQ). Contiene una serie de preguntas, junto con sus respuestas, que los usuarios de *Sphinx 3* podrían hacerse antes, durante y después del entrenamiento o de la decodificación.

Troubleshooting: tools and logfiles Documento acerca de la solución de posibles problemas. Muestra la ejecución correcta de cada herramienta que se utiliza tanto en el entrenamiento de modelos continuos como en el de modelos semicontinuos. Asimismo, presenta los posibles problemas, junto con sus soluciones, así como los errores y advertencias que pueden ocurrir durante la ejecución de dichas herramientas. Sin embargo, esta documentación está incompleta.

Sphinx-3 miscellaneous (Decoding) Contiene algunos tópicos acerca de la decodificación. Explica brevemente temas como la generación de celosías y de la lista N-mejor, así como de los resultados de las corridas de las aplicaciones asociadas con este procedimiento. También incluye la explicación del formato en el que se encuentra el archivo del modelo del lenguaje después de ser obtenido con SLM. Otros temas son la generación de hipótesis y la explicación de algunas banderas del decodificador de *Sphinx 2*.

CMU Sphinx documentation Wiki Contiene una wiki dividida en dos secciones principales: documentación para los usuarios y documentación para desarrolladores e investigadores. La sección de documentación de usuario contiene una subwiki de *Sphinx*, una sección cómo hacer un modelado del lenguaje sencillo con *Sphinx*, una sección de cómo adaptar los modelos acústicos por omisión a la voz del usuario y como adaptar el decodificador para que sea más rápido. En la sección de documentación para desarrolladores e investigadores describe aspectos particulares del código fuente de *Sphinx* y también aspectos que sirven de guía a los desarrolladores para llevar a cabo tareas de reconocimiento de voz que deseen.

Sphinx subwiki Documentación acerca de información del sistema de reconocimiento *Sphinx* haciendo énfasis en *Sphinx 3*. Describe con mayor precisión cada aspecto relevante de *Sphinx*. Tiene información concreta acerca de la utilización del decodificador *Sphinx 3*. Muestra solamente los

pasos generales que se deben seguir desde la grabación de las señales de audio, pasando por la descarga y configuración de *SphinxBase* hasta llegar a la instalación y configuración de los decodificadores **sphinx3_livedecode**, **sphinx3_livepretend** y **PocketSphinx**. También contiene tutoriales y documentación para desarrolladores, que cubre los aspectos de modelado acústico, modelado de lenguaje, elaboración de gramáticas de estados finitos, el diccionario de pronunciación que se usa en *CMU Sphinx* y bases de datos de audio, entre otros.

Hello World Decoder QuickStart Guide Una guía completa que muestra, paso a paso, cómo utilizar el decodificador de *Sphinx*. Una vez realizada la configuración del decodificador, se detallan los pasos necesarios para llevar a cabo la ejecución de los decodificadores; es equivalente al manual del usuario para la decodificación. Esta guía me fue de mucha utilidad ya que en otra documentación no existe tal descripción.

Sphinx 3 Internals Documentation Contiene las funciones, los archivos y las variables internas de los reconocedores de *Sphinx 3*. Su enfoque está más orientado a los programadores que a los usuarios de *Sphinx*.

SPHINX III Signal Processing Front End Specification Describe brevemente el front-end del procesamiento de señales de voz del sistema de reconocimiento de voz *Sphinx 3*. Se centra en las etapas sin profundizar más de lo necesario. Considero que este es otro de los pocos documentos bien organizados y concisos.

Sphinx Decoders Sirve para conocer las distintas versiones del decodificador decodificadores que ha realizado *Sphinx Group*.

Sphinx-3 s3.X Decoder ($X = 6$) Página que describe la versión actual del decodificador de *Sphinx 3*. Desglosa sus parámetros, sus recursos y describe la forma en que opera el decodificador.

The CMU-Cambridge Statistical Language Modeling Toolkit v2 (CMU-SLM) Página que contiene al conjunto de herramientas para el modelado del lenguaje. Contiene todos los aspectos de estas herramientas, desde su instalación, la terminología, los formatos de archivos, el desglose de la aplicación de cada herramienta con sus respectivos parámetros, la forma de uso y las estrategias de descuento que se pueden usar en los modelos del lenguaje. Considero que éste es uno de los pocos documentos que se encuentra bien organizado, es conciso y trata los puntos esenciales que se requieren a la hora de diseñar el modelo del lenguaje.

The Hieroglyphs Documento que trata de agrupar todo el conocimiento acerca de *Sphinx 3*, *SphinxTrain* y *CMU-SLM* y sus recursos relacionados para construir aplicaciones de voz. Es un documento bien explicado y detalla información pertinente acerca del uso de *Sphinx 3*, *SphinxTrain* y *CMU-SLM*, sin embargo, se encuentra incompleto.

Sphinx-II User Guide Manual del usuario de *Sphinx 2*. Entre otras cosas contiene la explicación del funcionamiento de la API de *Sphinx 2*. Es útil sólo como referencia, ya que en determinados documentos de *Sphinx 3* se hace referencia a *Sphinx 2*, pero no a éste manual. En esta documentación también se encuentra una sección acerca de cómo se debe elaborar una gramática de estados finitos para usarse en el reconocedor. Para trabajar con pequeños vocabularios, mi intención era la de desarrollar una gramática de éste tipo y así mejorar la precisión del reconocimiento, sin embargo, al no haber encontrado referencia alguna en el momento en que elaboré las gramáticas para este trabajo decidí utilizar gramáticas basadas en 3-gramas, construidas con las herramientas SLM.

Sourceforge Sphinx Página donde se pueden descargar todas las versiones de *Sphinx* y de los recursos asociados con él.

Sphinx Presentation Existen una serie de diapositivas que se incluyen con la documentación de *Sphinx 3*, las cuáles hacen referencia a las mejoras entre distintas versiones. Me ayudaron a elegir cuál versión del decodificador de *Sphinx* usar. Así, elegí la versión *s3.6* porque tiene:

- Mejor arquitectura de búsqueda.
- Mayor soporte en la adaptación de locutores.
- Mayor portabilidad y mejor creación del código ejecutable cuando se compila en otra plataforma.
- Soporte y seguridad para gramáticas de estados finitos.

El código fuente de las aplicaciones de *SphinxTrain*, *Sphinx Decoder* y *CMU-SLM* En estos archivos se encuentran, además de los derechos de autor y las revisiones que se le han hecho a cada software, comentarios acerca de lo qué hace cada programa. Algunos contienen comentarios específicos acerca de los elementos esenciales de cómo se programó y otros no, por lo que fue necesario leer las líneas de código para entender dichos programas, sin embargo, la notación y uso de las variables son bastante descriptivas lo que facilitó su comprensión. Aún así, el código en ocasiones no sólo es extenso sino que la cantidad de archivos que tuve que revisar también requirieron de muchas horas de análisis y si se carece de los fundamentos necesarios tanto de Procesamiento de voz continua como de programación en lenguaje C resulta bastante difícil.

Sphinx Knowledge Base Tool Página de *CMU Sphinx Group* que contiene una herramienta para obtener el modelo del lenguaje para el sistema *Sphinx* directamente. El inconveniente es que dicha herramienta trabaja con el conjunto de fonemas base del idioma inglés lo cuál no fue útil para los propósitos de esta tesis, sin embargo, me permitió resolver algunas dudas acerca de qué estrategia de descuento usar.