

Plan de Evaluación de Sistemas ALBAYZIN-08

Conversión Texto a Voz

(ALBAYZIN-08 CTV)

Objetivo

El objetivo principal de esta evaluación es promover el intercambio de ideas, estimular la creatividad, favorecer la colaboración entre los grupos de investigación que trabajan en la conversión de texto a voz y comparar distintas técnicas utilizadas. Con este fin, se propone una competición entre sistemas de síntesis, similar al Blizzard Challenge para Inglés y Chino Mandarín. La construcción de la voz se realizaría sobre una base de datos común y para lengua castellana.

Condiciones generales

Los grupos participantes dispondrán de hasta siete semanas para desarrollar sus sistemas, a partir de materiales de desarrollo que les serán suministrados. También podrán utilizar cualesquiera otros materiales y subsistemas (modelos acústicos para la segmentación, modelos prosódicos externos, lexicones, etc.), que deberán hacer constar en la descripción del sistema, incluyendo las referencias pertinentes. A continuación se les enviará un conjunto de frases de test y dispondrán de una semana para sintetizarlas y enviar los archivos de audio resultantes. No se permitirá ninguna manipulación manual de las frases de test (manipulación de los textos de entrada, utilización de distintas configuraciones o subconjuntos de la BD en cada frase de test salvo que el sistema lo haga automáticamente, etc...).

Se espera que cada grupo participante proporcione al menos 5 expertos en tecnologías del habla y nativos de lengua castellana, para que colaboren en la evaluación de los sistemas de síntesis.

Corpus de Evaluación

No se harán públicos los datos relativos a las frases de test hasta la fecha de inicio de la evaluación para que la adaptación del sistema de síntesis al entorno de test no sea posible. En cualquier caso, el corpus constará de varias centenas de frases de las cuales sólo una parte serán utilizadas en la evaluación subjetiva para determinar tanto la inteligibilidad como la naturalidad de las voces.

Materiales de desarrollo

El corpus de desarrollo constará de aproximadamente una hora de grabación por parte de una locutora profesional (Marta). Junto a las señales de voz se suministrará la

segmentación automática de las mismas en formato sampa. Los grupos participantes tendrán libertad para utilizar directamente la segmentación proporcionada o generar una segmentación nueva por sus propios medios, debiendo indicar en la descripción del sistema, la alternativa tomada.

Calendario provisional

- 5 de mayo de 2008
 - Publicación del plan de evaluación detallado en la web de las JTH2008
 - Apertura del plazo de inscripción
 - Habilitación de un formulario de inscripción en línea accesible desde la web de las JTH2008.
- 15 de junio de 2008
 - Comienzo del envío de los materiales de desarrollo a los grupos participantes
- 15 de julio de 2008
 - Fin de plazo de inscripción
- 5 de septiembre de 2008
 - Envío de las frases de test a los grupos participantes
- 12 de septiembre de 2008 (hasta las 24:00, GMT+2)
 - Fin del plazo de envío de resultados, según el método y formato establecidos en el plan de evaluación detallado
- 17 de septiembre de 2008
 - Inicio de la evaluación subjetiva
- 17 de octubre de 2008
 - Fin de la evaluación subjetiva e inicio del análisis estadístico de los resultados
- 31 de octubre de 2008
 - Notificación a cada grupo participante de los resultados de la evaluación de su sistema o sistemas
- 12 de noviembre de 2008

- Presentación de un resumen de los resultados de la Evaluación ALBAYZIN-08 CTV en dentro de las JTH2008

Inscripción

Los grupos que deseen participar en esta evaluación deberán hacer llegar sus datos al comité organizador a través de un formulario en línea accesible desde la web de las JTH2008 a partir del 5 de mayo de 2008. El plazo de inscripción termina el 15 de julio de 2008. Antes de inscribirse, se recomienda leer con atención el plan detallado de evaluación que aparecerá en la web de las JTH2008 a partir del 5 de mayo. Este plan podría sufrir modificaciones, debido a restricciones o necesidades no previstas inicialmente, a errores o a imprecisiones, que en todo caso serán comunicadas a todos los grupos participantes. Los interesados pueden dirigirse directamente al comité organizador para consultar aspectos que no estén claros sobre la evaluación, por e-mail a eva.navas@ehu.es y por teléfono al número 946017306.

Compromisos de los grupos participantes

1. Usar los datos suministrados exclusivamente con fines de investigación. En el futuro, podrán utilizar los datos para desarrollar o evaluar sus propios sistemas, pero no podrán distribuirlos ni venderlos. Si dieran lugar a publicaciones, deberán referenciar la base de datos como sigue:

Ignacio Hernández, Asunción Moreno. Diseño de un corpus para una base de síntesis de voz. Actas del XV Symposium Nacional de la U.R.S.I., Zaragoza (España), Septiembre 2000.

2. Enviar, junto a las frases de test, un documento (preferiblemente en formato PDF) donde se describan con suficiente detalle las características del sistema que se presenta a la evaluación. Para homogeneizar el aspecto de estas contribuciones, que serán distribuidas a todos los grupos participantes junto con el informe final del comité organizador de las Evaluaciones ALBAYZIN-08 CTV en las V Jornadas de Tecnología del Habla, se utilizarán como base las plantillas LaTeX o WORD de las JTH2008.
3. Permitir la distribución entre todos los grupos participantes del conjunto de señales sintéticas generadas para la realización de la evaluación para su uso en tareas de investigación.
4. Asistir a las V Jornadas de Tecnología del Habla (JTH2008), a celebrar el 12-14 de noviembre de 2008 en Bilbao, para la presentación de los detalles técnicos de su sistema ante los demás grupos participantes y el comité organizador.