

PLAN DE EVALUACIÓN PARA TRADUCCIÓN AUTOMÁTICA DE CASTELLANO A EUSKERA

1.- Introducción

El propósito de esta evaluación es el fomentar la investigación en sistemas de traducción automática entre castellano y euskera e impulsar la colaboración entre grupos teniendo en cuenta que existen menos recursos tanto bilingües como monolingües relacionados con el euskera.

Con dicho propósito se plantea un plan de evaluación que consiste en traducir artículos divulgativos de consumo de castellano a euskera. Para ello se plantean dos opciones. La primera consiste en desarrollar un traductor utilizando todos los grupos los mismos recursos para poder comparar diferentes aproximaciones bajo las mismas condiciones. La segunda opción permitirá presentar sistemas que utilicen otros recursos disponibles, como pueden ser corpus adicionales, diferentes procesadores de lenguaje o diccionarios.

Los participantes se comprometen a la presentación de los resultados de la evaluación en una sesión especial que tendrá lugar durante las V Jornadas en Tecnología del Habla. La participación se realiza a modo individual o equipo formado donde el representante del mismo deberá ser estudiante. Cada equipo podrá presentar uno o más sistemas.

2.- Medición de prestaciones

Las prestaciones se medirán usando varias técnicas de puntuación automática. Estas medidas comparan la salida de la traducción automática con las traducciones manuales de referencia. Se usarán las medidas habituales BLEU, NIST, WER y PER. BLEU¹ y NIST² son medidas basadas en la similitud de subsecuencias (N-gramas) de la traducción automática y la de referencia. La tasa de error por palabra WER mide las inserciones, supresiones y sustituciones entre la traducción automática y la de referencia. La tasa de error por palabra independiente de la posición PER calcula la distancia entre el conjunto de palabras que contiene la traducción automática y el de la traducción de referencia. Para la evaluación se utilizarán las herramientas de TC-STAR (www.tc-star.org).

3.- Condiciones de evaluación

Se evaluarán dos tipos de sistemas: los sistemas basados en los recursos proporcionados y los sistemas de recursos sin restricciones. Para los sistemas basados en recursos limitados se proporcionarán 58.000 frases para entrenamiento, aproximadamente 1.500 para ajustar el sistema y otras 1.500 para test. Se ofrecerá una única traducción por frase de origen.

El corpus se proporcionará en formato texto, tokenizado y alineado a nivel de frase y se entregará por separado la parte correspondiente a cada idioma. Alternativamente, se proporcionará el texto lematizado y etiquetado automáticamente. Cada línea corresponderá a una frase tanto en el fichero de origen como en el de destino. Todos los procesos aplicados son automáticos, por lo que el corpus contendrá errores. El test final consistirá en traducir varios artículos de divulgación del mismo dominio no incluidos en el material inicial y estará procesado de forma análoga al resto de material.

¹K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in Proc. of the 40th Annual Meeting of the ACL, Philadelphia, PA, July 2002, pp. 311–318.

²Doddington, G. "Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics". In Proc. ARPA Workshop on Human Language Technology, San Diego, California, March 2002.

4.- Procedimiento para la evaluación

El procedimiento con las fechas para la evaluación es el siguiente:

- El 1 de Mayo de 2008 se dispondrá de los planes de evaluación y se abre el periodo de inscripción.
- La fecha límite de inscripción será el 31 de Mayo de 2008.
- A partir del 16 de Junio de 2008 se podrá disponer del material de entrenamiento y desarrollo para las distintas evaluaciones. Es necesario estar inscrito en la evaluación para recibir el material.
- El 15 de Septiembre de 2008 se proporcionarán los datos para la evaluación.
- El 30 de Septiembre de 2008 a las 24:00 es la fecha límite para recibir los resultados.
- El 31 de Octubre de 2008 se publicarán los resultados de la evaluación entre los participantes.

5.- Envío de traducciones

Las traducciones se enviarán por correo electrónico a la organización. Deberán ser completas, conteniendo por tanto todo el conjunto de datos de evaluación. Deberán enviarse a: Nerea Ezeiza (nerea.ezeiza@gmail.com). Se debe enviar un fichero por artículo y por sistema que se desee evaluar.

Los resultados estarán disponibles una vez se hayan enviado dichos resultados a los participantes. Esto permitirá realizar análisis previos a la celebración de las V Jornadas de Tecnologías del Habla.

Cada participante deberá remitir una descripción del sistema enviado a la evaluación, que deberá incluir:

- Identidad elegida para el sistema (sysid)
- Condiciones de evaluación (datos de entrenamiento)
- En el caso de los sistemas sin restricciones, deberán indicarse las características de los recursos utilizados para el desarrollo del mismo:
 - En el caso del corpus debe indicarse:
 - si se ha utilizado el corpus proporcionado o no
 - si es un corpus alternativo, una breve descripción del contenido, tamaño aproximado, si es monolingüe o bilingüe y si está públicamente accesible
 - En el caso de los diccionarios, tamaño aproximado, si son monolingües o bilingües y si está públicamente accesible
 - En el caso de procesadores del lenguaje
 - tipo de procesador (analizador morfológico / POS / sintáctico superficial / semántico /...)
 - si está públicamente accesible
 - idiomas a los que se les ha aplicado
- Descripción de la aproximación algorítmica

Esta descripción se enviará con la misma plantilla utilizada para las comunicaciones regulares de las V Jornadas en Tecnología del Habla. Las descripciones recibidas se distribuirán como parte del material de análisis de la evaluación.