

Traducción Automática

Ivana Romina Altamirano

Directora: Paula Estrella

Grupo de Procesamiento de

Lenguaje Natural

Oficina 328



Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Presentación

- Egresada de FaMAF en Febrero 2010
 - Tesis con L. Alonso sobre marcos de sub-categorización
 - Resumen de la tesis presentado en artículo conferencia NAACL
 - North American Chapter of the Association for Computational Linguistics
- Traducción automática – Paula Estrella
- Inscripta en doctorado desde septiembre 2010
 - Beca PFDT SeCyT-UNC-FonCyT
 - Proyectos de Formación de Doctores en Áreas Tecnológicas Prioritarias

Traducción automática

- Aplicar técnicas de PLN para convertir un texto de idioma origen a idioma destino
 - Ej: Primera Jornada de Doctorandos de Computación
→ First Day of Computing PhD
 - Según Google Translate
- TA útil para: creación masiva de contenido multilingüe, comprensión de textos en idiomas desconocidos, comunicación oral/escrita, etc.

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Agregar conocimiento lingüístico

- Actualmente la mayoría de los sistemas solo tratan las palabras como strings
- Que es?
 - Información morfológica, sintáctica o semántica sobre el corpus utilizado para desarrollar un sistema de TA
 - Ej: marcos de subcategorización, pos tags, lemmas, ...
- Para que serviría?
 - Evitar malformaciones en traducción, aumentar fidelidad de las TAs, ...

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Simplificación textual

- Que es?
 - Trabajar el texto a traducir para convertirlo en uno de menor complejidad
- Ejemplo
 - sistema de detección de intrusos en el interior → alarma
- Para que serviría?
 - En principio si el texto es simple el traductor debería traducir mejor

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Combinación de modelos

- Existen distintos modelos de traducción: por reglas, ejemplos, estadística, híbridos, ...
- Que es la combinación de modelos?
 - Traducir texto con varios modelos y “elegir” la mejor combinación como salida de la traducción
- La idea es captar lo mejor de cada modelo y generar texto de mayor calidad

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Evaluación

- Evaluación automática

Para combinar partes de la traducción hechas por distintos motores

- Evaluación manual

Para evaluar la performance del sistema total

Plan de la charla

- Presentación
- Áreas de estudio propuestas
 - Incorporación de conocimiento lingüístico
 - Simplificación de texto
 - Combinación de modelos de traducción
 - Evaluación
- Resumen 2010

Cursos de posgrado

- Materia: “Traducción Automática y Evaluación”
Paula Estrella
- Materia: “Inteligencia Artificial”
Laura Alonso I Alemany y Luciana Benotti
- Materia: “Redes Neuronales”
Francisco Tamarit

Relación con mi tesis

- Creación y mejora de sistema de TA por ejemplos: permitió familiarizarme con modelos de TA estado del arte y técnicas de evaluación
- Resolución de correferencias: puede servir en alguna medida para simplificar el texto y va a ser integrada al preprocesamiento antes de la traducción

A decorative border surrounds the central text, composed of various stylized letters and symbols in shades of blue and black, arranged in a repeating pattern.

Gracias!