

Curso

Lingüística Aplicada II **(Sistemas de aplicación lingüística)**

DURACIÓN:

80 horas

AUTOR DEL PROGRAMA:

Dr. Alexander Gelbukh

OBJETIVO GENERAL

Describir que tipos de sistemas de aplicación lingüística existen y como se usan. Analizar a nivel detalle los sistemas de desambiguación de sentidos de palabras y de resolución de homonimia de partes de oración.

BIBLIOGRAFÍA

1. I. Bolshakov, A. Gelbukh. Computational linguistics. CIC-IPN, 2004, 187 p.
2. J. Allen. Natural Language Understanding. The Benjamin/Cummings Publishing Company, 1995, 654 p.
3. R. Grishman. Computational Linguistics. Cambridge University Press, 1986.
4. Oxford Handbook on computational linguistics. Oxford University Press, 2003, 808 p.
5. Manning C. and H. Schütze. Foundations of statistical natural language processing. MIT Press, 1999, 680 p.
6. Jurafsky, D. and J. Martin. Speech and language processing. Prentice Hall, 2000, 934 p.

CONTENIDO DEL CURSO

1. Inventario de sistemas de lingüística aplicada

- 1.1. Sistemas de formateo y de corrección de errores
- 1.2. Sistemas de referencia y educativos
- 1.3. Sistemas de extracción de información y minería de datos
- 1.4. Sistemas de traducción por computadora
- 1.5. Interfaces de lenguaje natural
- 1.6. Sistemas de reconocimiento
- 1.7. Sistemas de comprensión de textos y extracción de conocimientos

2. Clasificación de los sistemas aplicados

- 2.1. Clasificación de acuerdo a la función

- 2.1.1. Correctores
- 2.1.2. Sistemas de extracción de información y minería de datos
- 2.1.3. Sistemas de traducción
- 2.1.4. Interfaces de lenguaje natural
- 2.1.5. Otros sistemas lingüísticos y áreas vecinas
- 2.2. Clasificación de acuerdo a los métodos utilizados
 - 2.2.1. Métodos formales
 - 2.2.2. Métodos inteligentes
- 2.3. Sistemas inteligentes y no inteligentes
- 3. Formateo y separación de palabras para final de renglón**
 - 3.1. Separación por sílabas
 - 3.2. Separación por sílabas basada en la fonética
 - 3.3. Separación inteligente
 - 3.3.1. Separación basada en la morfología
 - 3.3.2. Diccionarios con separaciones
 - 3.3.3. Problemas de ambigüedad
- 4. Verificación de errores y corrección**
 - 4.1. Tareas de verificación de errores
 - 4.1.1. Clasificación de acuerdo a la función
 - 4.1.2. Clasificación de acuerdo a los métodos utilizados
 - 4.2. Verificación de ortografía y corrección
 - 4.2.1. Tareas de verificación de ortografía
 - 4.2.2. Detección de errores
 - 4.2.2.1. Método Trigram
 - 4.2.2.2. Método de las funciones Hash
 - 4.2.2.3. Diccionario de palabras
 - 4.2.2.4. Diccionarios morfológicos
 - 4.2.3. Corrección de errores
 - 4.2.3.1. Tipos de errores
 - 4.2.3.1.1. Errores de una letra
 - 4.2.3.1.2. Otro tipo de errores
 - 4.2.3.2. Modelos de los errores que se cometen
 - 4.2.3.2.1. Modelo basado en el teclado
 - 4.2.3.2.2. Modelo SOUNDEX
 - 4.2.3.2.3. Modelo gramático
 - 4.2.3.3. Métodos estadísticos basados en Trigram
 - 4.2.3.4. El uso de un diccionario de palabras
 - 4.2.3.5. El uso de un diccionario morfológico
 - 4.2.4. Verificador inteligente de ortografía
 - 4.2.4.1. Ambigüedad y el papel del contexto
 - 4.2.4.2. Contexto inmediato
 - 4.2.4.2.1. El uso de artículos
 - 4.2.4.2.2. El uso de preposiciones
 - 4.2.4.3. Contexto sintáctico
 - 4.2.4.4. Contexto semántico
 - 4.3. Métodos de verificación gramatical
 - 4.3.1. Tareas de verificación gramatical
 - 4.3.2. Contexto inmediato

- 4.3.3. El uso de los signos de puntuación
- 4.3.4. Verificación gramatical inteligente
 - 4.3.4.1. Análisis sintáctico
 - 4.3.4.2. El papel del contexto semántico
- 4.3.5. Corrección de los errores gramaticales
- 4.4. Métodos de verificación de estilo
 - 4.4.1. Tareas de verificación de estilo
 - 4.4.2. Tipos de estilo de documentos
 - 4.4.3. Métodos formales
 - 4.4.3.1. Índices de legibilidad
 - 4.4.3.1.1. Grado de facilidad de lectura según Flesch
 - 4.4.3.1.2. Grado de facilidad según Flesch-Kincaid
 - 4.4.3.1.3. Índice de claridad según Gunning (Gunning Fog's Index)
 - 4.4.4. Diccionario de combinaciones de palabras
 - 4.4.5. Diccionarios de construcciones y expresiones erróneas
 - 4.4.6. Corrección de errores de estilo

5. Sistemas de referencia y educativos

- 5.1. Diccionarios tradicionales.
 - 5.1.1. Diccionarios y thesauri explicativos
 - 5.1.1.1. Ejemplo: DRAE
 - 5.1.2. Diccionarios bilingües
 - 5.1.2.1. Ejemplo: MultiLex
- 5.2. Tipos modernos de diccionarios
 - 5.2.1. Diccionarios de combinaciones de palabras y de patrones de gobierno
 - 5.2.2. Diccionarios universales
 - 5.2.2.1. Importancia de las fuentes de información múltiple. Inferencia de información léxica
 - 5.2.2.2. Ejemplo: CrossLexica™

6. Sistemas de extracción de información y minería de datos

- 6.1. Tareas de extracción de información y minería de datos
 - 6.1.1. Extracción de documentos
 - 6.1.2. Navegación de datos
 - 6.1.3. Generalización de la información
 - 6.1.3.1. Resúmenes
 - 6.1.3.2. Tendencias, promedio y valores extremos
- 6.2. Sistemas de extracción de documentos
 - 6.2.1. Para calificar la calidad de la búsqueda
 - 6.2.1.1. Respuesta
 - 6.2.1.2. Exactitud
 - 6.2.2. Sistemas no inteligentes de extracción de información
 - 6.2.2.1. Métodos formales de búsqueda
 - 6.2.2.1.1. Búsqueda literal
 - 6.2.2.1.2. Expresiones comunes
 - 6.2.2.2. Sistemas basados en las características
 - 6.2.2.2.1. Tipos de descriptores
 - 6.2.2.2.2. Thesauri de descriptores

- 6.2.3. Sistemas inteligentes de extracción de información
 - 6.2.3.1. Normalización morfológica
 - 6.2.3.2. Normalización basada en la formación de palabras
 - 6.2.3.3. Sistemas basados en un Thesaurus y búsqueda confusa
 - 6.2.3.4. Sistemas para la comprensión de textos
- 6.2.4. Motores para la búsqueda en internet
- 6.3. Sistemas de navegación de datos
 - 6.3.1. Métodos para la navegación de datos
 - 6.3.1.1. Sistemas de hipertextos
 - 6.3.1.2. Otros sistemas basados en la red
 - 6.3.1.2.1. Ejemplo: TextAnalist
 - 6.3.2. Agrupamiento de documentos según bases lingüísticas
- 6.4. Sistemas de generalización de información
 - 6.4.1. Para determinar el (los) tema(s) principal(es) del documento
 - 6.4.1.1. Ejemplo: CLASITEX
 - 6.4.2. Sistemas para la comprensión de textos
 - 6.4.2.1. Ejemplo: Libretto
 - 6.4.3. Sistemas para la compilación de resúmenes
 - 6.4.4. Sistemas para la extracción de conocimiento

7. Sistemas de traducción

- 7.1. Generación de sistemas de traducción
- 7.2. Sistemas de traducción no inteligentes
 - 7.2.1. Problemas de la traducción palabra por palabra
 - 7.2.2. Resolución de la ambigüedad
 - 7.2.2.1. El uso del contexto inmediato
 - 7.2.2.2. El uso de temas locales y globales
 - 7.2.2.2.1. Ejemplo: CLASITEX
- 7.3. Traducción inteligente
 - 7.3.1. La traducción como una transformación de dos lados, Significado \Leftrightarrow Texto
 - 7.3.2. Traducción a nivel sintáctico
 - 7.3.3. Traducción a nivel semántico
 - 7.3.3.1. Nivel de estructuras sintácticas normalizadas
 - 7.3.3.1.1. El uso de las funciones léxicas
 - 7.3.3.1.2. Ejemplo: ETAP-2
- 7.4. Métodos de traducción automática.
 - 7.4.1. Traducción con transformación.
 - 7.4.2. Traducción directa.
 - 7.4.3. Interlingua.
 - 7.4.4. Traducción estadística.
 - 7.4.5. Traducción basada en ejemplos.

8. Interfaces del lenguaje natural

- 8.1. Un sistema de interface del lenguaje natural como un sistema de Significado \Leftrightarrow Texto
- 8.2. Características del discurso en diálogo
 - 8.2.1. Procesamiento de tiempo real

- 8.2.2. Consideraciones psicológicas. Para hacer el diálogo más natural y menos aburrido
- 8.3. Sistemas gobernados por la pragmática
 - 8.3.1. Sistemas de traducción directa
 - 8.3.2. De interfaces del lenguaje natural a sencillas bases de datos
 - 8.3.3. Sistemas de diálogos basados en estadísticas en un campo limitado
- 8.4. Sistemas inteligentes de diálogo
 - 8.4.1. Modelo del usuario
 - 8.4.2. Estrategias para el procesamiento del diálogo
- 9. Sistemas de generación de textos**
 - 9.1. Sistemas no inteligentes de generación de textos
 - 9.2. Generación de textos como la transformación de Significado \Rightarrow Texto
 - 9.3. Problemas en la generación de textos
 - 9.4. El uso de los pronombres
 - 9.5. Moldear las implicaciones y expectativas del lector
- 10. Sistemas de reconocimiento**
 - 10.1. Reconocimiento de un caracter óptico
 - 10.2. Reconocimiento del habla
 - 10.2.1. Sistemas de control de la voz
 - 10.2.2. Sistemas de dictado
 - 10.2.2.1. Homófonos
 - 10.2.3. Comprensión del habla
 - 10.2.3.1. HMM.
 - 10.2.3.2. Modelos de gramática. N-grams.
- 11. Métodos de desambiguación de sentidos de palabras.**
 - 11.1. Senseval.
 - 11.2. Métodos estadísticos.
 - 11.3. Algoritmo de Lesk.
 - 11.4. Métodos con optimización para el algoritmo de Lesk.
- 12. Sistemas de investigación científica**
 - 12.1. Bases de datos léxicas
 - 12.1.1. Bases de datos de ejemplos de textos completos
 - 12.2. Codificadores de partes de oración
 - 12.2.1. Codificar como se hace el *parsing*
 - 12.2.2. Esquemas de marcado diferentes.
 - 12.2.3. Brill tagger.
 - 12.2.4. Porter tagger.
 - 12.2.5. TNT tagger.
 - 12.2.6. Métodos heurísticos.
 - 12.2.7. Evaluación de los tagger.
 - 12.3. *Parsers*
- 13. Appendix. Clases prácticas y ejercicios**
 - 13.1. Lenguajes de programación para la descripción de las gramáticas formales (opcional)
 - 13.1.1. Lenguaje Lex
 - 13.1.1.1. Construcciones básicas
 - 13.1.1.2. Construcciones y técnicas avanzadas

- 13.1.1.3. Ejercicios
- 13.1.2. Lenguaje YACC
 - 13.1.2.1. Construcciones básicas
 - 13.1.2.2. Construcciones avanzadas
 - 13.1.2.3. Ejercicios
- 13.1.3. El uso de Lex y YACC juntos
 - 13.1.3.1. Ejercicios
- 13.1.4. Limitaciones de los lenguajes de gramática formal
- 13.2. Lenguajes Macro (opcional)
 - 13.2.1. Lenguaje AWK
 - 13.2.1.1. Construcciones básicas
 - 13.2.1.2. Construcciones avanzadas
 - 13.2.1.3. Ejercicios
 - 13.2.2. Lenguaje Perl
 - 13.2.2.1. Construcciones básicas
 - 13.2.2.2. Construcciones avanzadas
 - 13.2.2.3. Ejercicios
- 13.3. Sistemas de lingüística aplicada
 - 13.3.1. Implementación del Software de un sencillo corrector de ortografía
 - 13.3.2. Implementación del Software de un sencillo sistema de diálogo