

Tecnologías básicas

MÁSTER UNIVERSITARIO EN FONÉTICA Y FONOLOGÍA

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



DATOS GENERALES

Breve descripción

Conocimiento profundo de las tecnologías del habla básicas, esto es, las referidas al reconocimiento automático, a la síntesis y a los sistemas de diálogo.

Programa de la asignatura:

1. Reconocimiento del habla
2. Síntesis del habla (común con el itinerario D)
3. Sistemas de diálogo

Título asignatura

Tecnologías básicas

Código asignatura

101658

Curso académico

2016-17

Planes donde se imparte

[MÁSTER UNIVERSITARIO EN FONÉTICA Y FONOLOGÍA](#)

Créditos ECTS

3

Carácter de la asignatura

OPTATIVA

Duración

Anual

Idioma

Castellano

CONTENIDOS

Contenidos

Esta asignatura forma parte del Itinerario A - "Tecnologías del habla" (8 ECTS), que es uno de los cuatro itinerarios optativos del Máster entre los que el alumno puede optar, de acuerdo con sus preferencias y el asesoramiento que reciba de los profesores.

RECONOCIMIENTO DEL HABLA

Objetivos

Se pretende que el alumno se familiarice con conceptos básicos de reconocimiento de formas y, más en concreto, con las técnicas específicas que se usan en reconocimiento del habla, así como su utilización en aplicaciones, empleando interfaces de usuario estándar.

Programa

1. La señal de voz y su modelo digital
2. Extracción de las características espectrales
3. Modelos de Markov ocultos: entrenamiento y algoritmo de Viterbi
4. Modelado acústico-fonético con dependencia contextual
5. Modelado del lenguaje
6. Dificultades del reconocimiento en condiciones reales y técnicas que mejoran las prestaciones: gran vocabulario, robustez, adaptación, etc.
7. Paquetes de programas de reconocimiento
8. Aplicaciones: dictado, call centers, transcripción, etc.

SÍNTESIS DEL HABLA (común con el Itinerario D)

Objetivos

- Comprender lo que es un sistema de conversión texto-voz (CTV), y los módulos que lo constituyen.
- Reconocer la relativa complejidad de los problemas interdisciplinarios que es necesario abordar, y las soluciones que se dan.
- Conocer la utilidad de los CTV, familiarizarse con su uso adecuado, y saber lo que

pueden y lo que no pueden hacer.

Programa

1. Definición de conversión texto-voz (CTV)
2. Relación de la CTV con otras tecnologías
3. Antecedentes históricos
4. Algunas aplicaciones
5. Planteamiento general y arquitectura de un sistema CTV
6. Descripción de los módulos componentes
 - Normalizador
 - Preproceso
 - Categorizador
 - Estructurador / pausador
 - Conversor grafema-alófono
 - Generador de parámetros prosódicos
 - Sintetizador de voz
7. Clasificación de los sistemas CTV
8. Panorámica de sistemas actuales
9. Evaluación de la calidad
10. Construcción de nuevas voces ("locutores")
11. Tendencias de futuro

SISTEMAS DE DIÁLOGO

Objetivos

Se pretende que los alumnos adquieran los conocimientos fundamentales de las tecnologías empleadas hoy en día para desarrollar este tipo de sistemas. Asimismo, se pretende que conozcan diversas herramientas *software* disponibles en el mercado hoy en día para facilitar la implementación de tales sistemas.

Programa

Temario de Teoría

1. Introducción

Con este tema se pretende presentar a los alumnos una versión panorámica del curso que incluirá un repaso histórico de las diversas técnicas de interacción con sistemas informáticos, hasta llegar a los sistemas de diálogo actuales. La introducción incluirá una breve descripción de los temas a tratar a lo largo del curso y contará con varios vídeos ilustrativos relacionados con este tipo de sistemas.

2. Interfaz de entrada de sistemas de diálogo

Los sistemas de diálogo más sofisticados de hoy en día permiten utilizar diversos dispositivos para proporcionar datos al ordenador, como por ejemplo teclados, ratones, teléfonos, micrófonos, cámaras de video, guantes de datos o pantallas sensibles al tacto. Ello representa una gran ventaja, puesto que los datos se pueden introducir usando varios dispositivos simultáneamente o de forma concurrente. Además, los usuarios pueden utilizar aquellos dispositivos que les resulten más adecuados según sus preferencias o necesidades. En este tema se realizará una introducción a las tecnologías que permiten implementar este tipo de interacción persona-ordenador, entre las que destacan: procesamiento de lenguaje natural, lectura de labios, localización y seguimiento de la cara del usuario, seguimiento de su mirada, reconocimiento de gestos, y reconocimiento de escritura a mano.

3. Análisis y gestión del diálogo

Dado que actualmente las tecnologías empleadas en la interfaz de entrada no proporcionan un 100% de exactitud, los sistemas de diálogo deben incorporar mecanismos que les permitan solventar los posibles errores que se puedan producir durante la interacción con el usuario. Con este tema se pretende que los alumnos conozcan las diversas aproximaciones al modelado del diálogo, así como las estrategias de interacción y confirmación que se suelen emplear para implementar sistemas de diálogo capaces de funcionar satisfactoriamente en condiciones reales.

4. Interfaz de salida de sistemas de diálogo

Los sistemas de diálogo actuales pueden utilizar diversos canales de comunicación para interactuar con los usuarios, como por ejemplo, gráficos de diversos tipos, lenguaje natural en forma de texto o voz humana generada artificialmente. De esta forma se logra una mayor comprensión de la información proporcionada por el ordenador, y a la vez, se logra que éste sea más amigable y adaptativo a las necesidades y/o preferencias de los usuarios. En este tema se realizará una introducción a las tecnologías empleadas para lograr este tipo de interacción tan sofisticada, destacando entre éstas la relacionada con los agentes animados (animated agents o talking heads).

5. Estándares

Durante los últimos años se han llevado a cabo notables esfuerzos destinados a crear estándares que faciliten la implementación de sistemas de diálogo. Las empresas de desarrollo y los proveedores de servicios han apostado por potenciar estándares que permiten solventar, en cierta medida, las dificultades existentes a la hora de implementar este tipo de sistemas. Con este tema se pretende que los alumnos conozcan y aprendan a utilizar el estándar VoiceXML para la implementación de sistemas de diálogo.

6. Herramientas de implementación

Recientemente se han llevado a cabo notables esfuerzos destinados a crear herramientas que faciliten la implementación de sistemas de diálogo. Con este tema se presente que los alumnos conozcan la existencia de tales herramientas (CSLU Toolkit, Voxeo Designer y Xface), se familiaricen con su uso y puedan decidir entre usar una u otra a la hora de implementar un sistema de diálogo para una aplicación concreta.

Temario de Prácticas

Práctica nº 1: Implementación de un sistema de diálogo mediante redes de transiciones entre estados

Las redes de transiciones entre estados constituyen una técnica de modelado del diálogo que permite implementar el flujo de la interacción con un sistema de diálogo de una manera muy estructurada, teniendo en cuenta la información que se debe intercambiar en cada momento. Mediante esta práctica, los alumnos podrán aprender a utilizar una herramienta que permite implementar sistemas de diálogo mediante redes de transiciones entre estados.

Práctica nº 2: Implementación de sistemas mediante VoiceXML

VoiceXML (Voice Extensible Markup Language), también llamado VXML, es un lenguaje basado en marcas similar a HTML, que permite gestionar fácilmente la lógica de los sistemas de diálogo mediante navegadores Web que soportan voz. El objetivo principal de VoiceXML es favorecer la implementación de estos sistemas aprovechando las ventajas ofrecidas por los sistemas de transmisión de información basados en Internet. Mediante esta práctica, los alumnos podrán aprender a utilizar una herramienta que permite implementar sistemas de diálogo mediante este lenguaje.

COMPETENCIAS

Generales

CG1.- Que los estudiantes adquieran conocimientos sistemáticos, avanzados y actualizados de los conceptos clave y de los problemas principales de la Fonética y de la Fonología generales, así como de los de sus aplicaciones. Una formación sólida que les permita dirigir su carrera tanto hacia los sectores profesionales específicamente vinculados con los distintos itinerarios del Máster, como hacia los ámbitos académicos o de investigación básica con ellas relacionados.

CG2.- Que los estudiantes sean capaces de evaluar críticamente las aportaciones de la investigación actual -basándose en un sólido entendimiento de las materias estudiadas- y puedan, asimismo, presentar adecuada y coherentemente los resultados de la investigación, adquiriendo con todo ello la capacidad para integrarse en una línea de investigación concreta para llevar a cabo la tesis doctoral.

Transversales

CT1.- Aprender a formular hipótesis de investigación de manera adecuada.

CT2.- Aprender a obtener los datos para una investigación controlando todas las variables implicadas en ella.

CT3.- Aprender a juzgar el grado de significatividad estadística de los resultados de investigación.

CT4.- Dominar los programas de análisis informático útiles para fonética y/o fonología teóricas y aplicadas.

CT5.- Aprender a discutir e interpretar los resultados de una investigación.

CT6.- Aprender a formular conclusiones de modo adecuado.

Específicas

Las competencias específicas que adquirirán los estudiantes que opten por cursar las asignaturas que componen el Itinerario A - "Tecnologías del habla" son:

CEA1.- Conocer y saber evaluar los sistemas actuales de reconocimiento de voz

CEA2.- Saber juzgar la calidad de los asistentes virtuales

CEA3.- Conocer y saber analizar los sistemas de habla existentes para personas con necesidades especiales

CEA4.- Dominar el uso del *software*

PLAN DE APRENDIZAJE

Actividades formativas

AF2.- Visitas a centros externos o a centros privados.

AF4.- Seminarios y tutorías extracurriculares impartidos por los propios profesores del Máster.

AF5.- Horas de estudio en biblioteca y en el hogar.

AF10.- Docencia en el aula.

Metodologías docentes

MD1.- Cursos teóricos y talleres.

MD2.- Encuentros con representantes de empresas relacionadas de un modo u otro con el fenómeno del habla.

MD3.- Horas de experimentación en laboratorio.

MD5.- Etiquetado de registros sonoros mediante PRAAT.

MD6.- Exposiciones de los propios alumnos sobre las prácticas realizadas.

MD11.- Horas de prácticas de estadística y de análisis computacional.

MD12.- Horas de prácticas de evaluación de recursos en red.

MD14.- Horas de prácticas con software especializado.

Resultados de aprendizaje

Al concluir las asignaturas que componen el Itinerario A - "Tecnologías del habla", los estudiantes deben haber aprendido lo siguiente:

- Entender en qué medida el conocimiento fonético ayuda (y fundamenta) a las tecnologías del habla.
- Comprender y dominar las tecnologías básicas, como el reconocimiento y síntesis del habla y los sistemas de diálogo.
- Conocer los recursos de que dispone para trabajar en tecnologías del habla: las aplicaciones de la estadística a estos efectos, el procesado y el análisis de la señal de voz, etc.
- Saber evaluar la calidad de un producto tecnológico.
- Saber de qué aplicaciones son susceptibles las tecnologías del habla: traducción automática, identificación del locutor, creación de programas específicos de ayuda a personas discapacitadas, etc.
- Saber realizar una investigación en este campo, de suficiente envergadura como para constituir su Trabajo de fin de Máster.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

- La asistencia a las clases es obligatoria, y esto es válido para cualquier asignatura. Si las ausencias no justificadas de un alumno alcanzan el 15% de las horas de clase programadas, no podrá aprobar la materia.
- La calificación de esta asignatura dependerá de la nota obtenida en la prueba de control (examen tipo test de 40 preguntas) que se realizará una vez finalizadas todas las asignaturas que componen el itinerario A - "Tecnologías del habla". En este examen hay preguntas relativas a todas las asignaturas que lo componen, en número proporcional a las horas que a cada una de ellas se le han dedicado.
- Alguno de los profesores puede pedir de los alumnos que realicen algunas prácticas o algún trabajo pequeño relativo a los contenidos concretos que él explica dentro de la asignatura. Estas tareas no reciben una calificación, pero sí influyen en la valoración global del estudiante, que cada profesor transmite al coordinador de la asignatura.

Calendario de exámenes

Asignatura no ofertada en 2016/17

PROFESORADO

Profesor responsable

Llisterri Boix, Joaquim

*Profesor Titular de Lingüística General
Universidad Autónoma de Barcelona*

Profesorado

Pardo Muñoz, José Manuel

*Catedrático de Universidad
Universidad Politécnica de Madrid*

Rodríguez Crespo, Miguel Ángel

*Ingeniero Superior de Telecomunicación
Telefónica Investigación y Desarrollo*

López-Cózar Delgado, Ramón

*Profesor Titular de Lenguajes y Sistemas Informáticos
Universidad de Granada*

de Córdoba Herralde, Ricardo

*Profesor Titular de Tecnología Electrónica y Tecnologías del Habla
Universidad Politécnica de Madrid*

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Reconocimiento del habla

- L. R. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.
 X. Huang et al., *Spoken Language Processing*, Prentice-Hall, 2001.
 J. Coleman, *Introducing speech and language processing*, Cambridge University Press (Cambridge introductions to language and linguistics), 2005.
 J.B. Mariño, C. Nadeu, "La representación de la voz para el reconocimiento del habla", en Martí, M. A. - Llisterri, J. (Eds.), *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona - Fundació Duques de Soria, pp. 187-224.

Síntesis del habla (común con el Itinerario D)

- Conversor Texto-Voz multilingüe para español, catalán, gallego y euskera, M.Á. Rodríguez, J.G. Escalada y D. Torre, Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), revista nº 23, 1998.
An Introduction to Text-to-Speech Synthesis, Thierry Dutoit, Kluwer Academic Publishers 1997.
Text-to-Speech: The MITalk System, J. Allen, M. Hunnicut y D. Klatt, Cambridge University Press 1987.
Multilingual Text-to-Speech Synthesis: The Bell Labs Approach ed. Richard Sproat.
Talking Machines, ed. G. Bailly and C. Benoit, Hort Holland 1992.
Progress in Speech Synthesis ed. J. van Santen, R. Sproat, J. Olive y J. Hirschberg. Springer Verlag 1996.
Text-to-Speech Synthesis. New Paradigms and Advances. ed. S. Narayanan y A. Alwan, Prentice Hall 2004.

Sistemas de diálogo

- Kraiss, K. F. (Ed.). 2006. *Advanced Man-Machine Interaction: Fundamentals and Implementation*. Springer
 Kuppevelt, J., Dybkjaer, L., Bernsen, N. O. 2005. *Advances in Natural Multimodal Dialogue Systems*. Springer
 López-Cózar, R., Araki, M. 2005. *Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment*. Wiley
 Minker, W., Bühler, D., Dybkjaer, L. 2005. *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*. Springer
 Prendinger, H., Ishizuka, M. (Eds.). 2004. *Life-Like Characters. Tools, Affective Functions, and Applications*. Springer
 Wahlster, W. (Ed.) 2006. *SmartKom: Foundations of Multimodal Dialogue Systems*. Springer