



POLITÉCNICA

Año Académico 2008/2009

**UNIVERSIDAD POLITECNICA DE MADRID  
E.T.S. DE INGENIEROS DE TELECOMUNICACION**

**ACTA DE EXAMEN**

**Asignatura: PROYECTO FIN DE CARRERA**

**TÍTULO DEL PROYECTO:** DISEÑO, IMPLEMENTACIÓN Y EVALUACIÓN DE METODOLOGÍAS DE ADAPTACIÓN DE MODELOS DE LENGUAJE PARA SISTEMAS MULTIDOMINIO

<b>APELLIDOS Y NOMBRE:</b>  FERRO LLANOS, VICENTE	<b>CALIFICACIÓN</b> 10p MATRICULA
---	---

**Tutor: D. LUIS FERNANDO D`HARO**

**Ponente: D. RICARDO DE CORDOBA HERRALDE**

Madrid, a 7 de octubre de 2009

**EL VOCAL PRIMERO**

**EL PRESIDENTE**

**EL VOCAL SECRETARIO**

**MIEMBROS DEL TRIBUNAL**

**Presidente: D. RICARDO DE CORDOBA HERRALDE**

**Vocal: D. JAVIER MACIAS GUARASA**

**Secretario: D. JUAN MANUEL MONTERO MARTINEZ**

*Suivent: JOSE MANUEL PARDO MONTE*

## PROYECTO FIN DE CARRERA

**Título:** Diseño, implementación y evaluación de metodologías de adaptación de modelos de lenguaje para sistemas multidominio

**Autor:** D. Vicente Ferro Llanos

**Tutor:** D. Luis Fernando D'Haro Enríquez

**Ponente:** D. Ricardo de Córdoba Herralde

**Departamento:** Departamento de Ingeniería Electrónica (DIE).

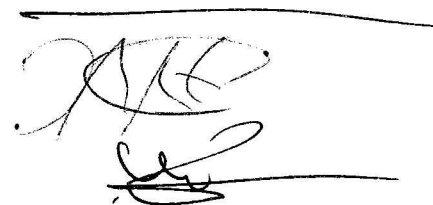
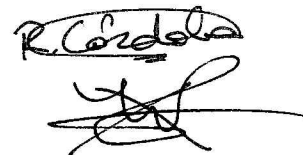
### TRIBUNAL CALIFICADOR

**Presidente:** D. Ricardo de Córdoba Herralde

**Vocal:** D. Javier Macías Guarasa

**Secretario:** D. Juan Manuel Montero Martínez

**Suplente:** D. José Manuel Pardo Muñoz



*Madrid, 7 de Octubre de 2009*

*Calificación obtenida*

*10 Matrícula de Honor*

## Resumen

El presente proyecto de fin de carrera se ha centrado en el estudio de técnicas de generación de modelos de lenguaje (basados tanto en n-gramas de palabras como de categorías gramaticales) y la búsqueda de metodologías de adaptación de modelos para sistemas multidominio en tareas específicas.

En la actualidad, uno de los métodos más importantes para dotar de conocimiento lingüístico a los sistemas de reconocimiento de habla es la utilización de modelos de lenguaje. El objetivo perseguido con la creación de estos modelos es determinar las restricciones existentes en las combinaciones de palabras que permiten crear las distintas frases de un lenguaje. Esta tarea se simplifica cuando se desarrolla una aplicación específica, puesto que el dominio semántico de ésta es mucho menor y sus restricciones lingüísticas mayores. La caracterización de dichas restricciones con modelos de lenguaje permite especificar la complejidad de la tarea de reconocimiento, medida finalmente en tasas de acierto y error de palabras reconocidas por el sistema tras realizar una primera aproximación en términos de perplejidad del modelo lingüístico.

El trabajo desarrollado ha consistido en la definición, generación y postprocesado de una base de datos que permitiese llevar a cabo los experimentos relacionados con los modelos de lenguaje, la generación de modelos de lenguaje basados en n-gramas de palabras o en n-gramas de categorías gramaticales POS (gracias a las herramientas *SRILM* y *Freeling* respectivamente), la adaptación de los modelos de lenguaje obtenidos y su posterior evaluación mediante experimentos de recuperación de la información y, finalmente, la puesta en marcha de un sistema de reconocimiento usando HTK que ha permitido la evaluación de modelos de lenguaje en términos de tasas de acierto en reconocimiento de voz.

## Palabras clave

Modelo de lenguaje estadístico (SLM), adaptación de modelos de lenguaje, clases automáticas, Part-Of-Speech (POS), perplejidad, tasa de reconocimiento, Reconocimiento Automático del Habla (RAH).