

**DOCUMENTACIÓN Y LINGÜÍSTICA APLICADA.
INTRODUCCIÓN AL PROCESAMIENTO DEL
LENGUAJE NATURAL EN TRADUCCIÓN**

Fuensanta M^a Guerrero Carmona
Azahara Veroz González
Manuel Marcos Aldón

UCOPress

Editorial Universidad de Córdoba

DOCUMENTACIÓN Y LINGÜÍSTICA APLICADA. INTRODUCCIÓN AL
PROCESAMIENTO DEL LENGUAJE NATURAL EN TRADUCCIÓN

Fuensanta M^a Guerrero Carmona, Azahara Veroz González, Manuel Marcos Aldón. -
Cordoba : UCOPress. Editorial Universidad de Córdoba, 2017

(Series Digitalia Nova Instrumenta; vol 5)

ISBN : 978-84-9927-289-4

Edit: UCOPress. Editorial Universidad de
Córdoba Campus Universitario de
Rabanales
Ctra. Nacional IV, km. 396 14071
– Córdoba, Spain

<http://www.uco.es/ucopress> ucopress@uco.es

ISBN: 978-84-9927-289-4

© UCOPress

© Los autores

All rights reserved. No part of this book may be reproduced, translated, stored in any retrieval system, nor transmitted in any form without written permission from the Publisher

Contenido

PARTE I. AUTÓMATAS FINITOS, PROCESAMIENTO DE UNIDADES MORFOLÓGICO-LÉXICAS Y ETIQUETADO SINTÁCTICO	5
1. Introducción.....	6
2. Expresiones Regulares y Autómatas Finitos.....	7
2.1 Expresiones regulares.....	7
2.2 Autómatas de estados finitos.....	7
2.3 Reconocimiento de cadenas	9
2.4 Autómatas de estados finitos no deterministas (NFSAs)	12
3. Procesamiento de unidades morfológico-léxicas	14
3.1 Estudio de la morfología	14
3.2 Lexicón de estado finito	19
3.3 Transductores de estado finito.....	20
3.4 Análisis morfológico y reglas ortográficas con transductores léxicos	21
3.5 Transductores de léxico libre: lematizador de Porter	22
3.6 Tokenización	22
4. N-Gramas.....	23
4.1 N-Gramas simples.....	23
4.2 Corpus de entrenamiento y de test	24
4.3 Smoothing (suavizado).....	25
4.4 Interpolación y Backoff.....	26
4.5 Análisis morfológico-léxico aplicado a motores de búsqueda textuales	26
5. Etiquetado.....	27
5.1 Conjunto de etiquetas para el inglés.....	27
5.2 Etiquetado <i>Part-of-Speech</i> o POS-tagging.....	28
5.3 POS-tagging basado en reglas.....	28
5.4 POS-tagging estocástico.....	29
5.5 Etiquetado basado en transformación (Transformation-Based tagging)	29
6. Conclusiones	30
PARTE II. GRAMÁTICAS DE CONTEXTO LIBRE, PARSING, UNIFICACIÓN DE RASGOS Y SEMÁNTICA Y ANÁLISIS SEMÁNTICO.....	33
7. Introducción.....	34
8. Gramáticas de contexto libre para el análisis del lenguaje natural	35
8.2 Reglas gramaticales.....	36

8.3	Treebanks	38
8.4	Equivalencia gramatical y forma normal	38
8.5	Estado finito y gramáticas de contexto libre	39
8.6	Gramáticas de dependencia	40
9.	Análisis sintáctico	40
9.1	El análisis como búsqueda	40
9.2	Ambigüedad	42
9.3	Buscar en la superficie de la ambigüedad	42
9.4	Métodos de análisis de programación dinámica.....	43
9.5	Análisis parcial.....	45
10.	Unificación de rasgos	46
10.1	Estructuras de rasgos.....	46
10.2	Unificación de estructuras de rasgos	48
10.3	Estructuras de rasgos en gramática	48
10.4	Implementación de la unificación	50
10.5	Análisis con restricciones de unificación	52
11.	Semántica y análisis semántico	52
11.1	Representación del significado.....	52
11.2	Semántica computacional.....	59
11.3	Semántica léxica.....	63
11.4	Semántica léxica computacional	66
12.	Conclusiones	66
	Glosario de términos	69
	PARTE III. DISCURSO, EXTRACCIÓN DE INFORMACIÓN Y RESÚMENES.....	72
13.	Introducción.....	73
14.	Discurso	74
14.1	Segmentación del discurso	74
14.2	Referencias	74
15.	Extracción de información	75
15.1	Reconocimiento de entidades nombradas	75
15.2	Detección y clasificación de relaciones entre entidades.....	76
15.3	Procesamiento temporal y de eventos	77
16.	Question Answering (Búsqueda de Respuestas)	78
17.	Sistemas de extracción de resúmenes	79

17.1	Componentes de un algoritmo de resumen	80
17.2	Evaluación de los sistemas de resúmenes	82
18.	Conclusiones	83
	Bibliografía	85