

Tabla de Contenido

1. Introducción	1
1.1. Contexto y oportunidad a abordar	1
1.1.1. Centro Tecnológico Ucampus	2
1.1.2. Trabajo a realizar	2
1.2. Objetivos del trabajo	4
2. Marco Teórico	5
2.1. Procesamiento del lenguaje natural	5
2.2. Topic Modeling	6
2.2.1. Asignación Latente de Dirichlet (ALD)	6
2.2.2. Word embedding	6
2.2.3. Doc2Vec	7
2.2.4. BERT	8
2.2.5. Top2Vec	8
2.2.6. Clustering	8
2.2.6.1. K-Means Clustering	8
2.2.6.2. HDBSCAN Clustering	9
2.2.7. Técnicas de reducción de dimensionalidad	9
2.2.7.1. t-SNE: T-distributed Stochastic Neighbor Embedding	9
2.2.7.2. UMAP: Uniform Manifold Approximation and Projection	10
2.3. Sistemas de recomendación	10
2.3.1. User-Based Colaborative Filtering	10
2.3.1.0.1 Coeficiente de correlación de Pearson	10
2.4. Herramientas y conceptos para la validación.	11
2.4.1. Flask	11
2.4.2. Flask-Restful	12
2.4.3. Intercambio de recursos de origen cruzado (CORS)	12
2.4.4. Secure Sockets Layer (SSL)	12
2.4.5. Localtunnel	12
2.4.6. Escala de Likert	12
3. Situación Actual	13
3.1. Plataforma Ucampus	13
3.2. Experiencia de estudiantes	16
3.3. Estado del Arte	20
4. Análisis y diseño	24
4.1. Sistema de categorización	24

4.1.1.	Identificando alternativas	25
4.1.1.1.	Asignación Latente de Dirichlet	25
4.1.1.2.	Word2Vec	25
4.1.1.3.	Doc2Vec	26
4.1.1.4.	Top2Vec	26
4.1.2.	Evaluación de los métodos	26
4.1.3.	Abordando el problema con Top2Vec	28
4.1.4.	Diseño del pipeline	28
4.2.	Sistema de recomendación	31
4.2.1.	Análisis de resultados	33
4.2.2.	Diseño del pipeline	34
4.3.	Resumen	36
5.	Implementación	38
5.1.	Sistema de categorización	38
5.1.1.	Selección entre entrenar un modelo de embedding o usar uno preentrenado	40
5.1.2.	Generando modelo de Top2Vec	41
5.2.	Sistema de recomendación	42
5.2.1.	Preparando el dataset	42
5.2.2.	Estableciendo pipeline	43
5.3.	Integración de sistemas	46
5.3.1.	Montando servidor y ambiente de desarrollo	46
5.3.2.	Preparando API REST (Lectura de modelos)	47
5.3.3.	Montando interfaz básica para mostrar datos	48
5.3.4.	Flujo de credenciales (CORS)	48
5.4.	Resumen	49
6.	Validación	50
6.1.	Herramientas de validación	51
6.1.1.	Sistema de categorización	53
6.1.2.	Sistema de recomendación	54
6.2.	Validación y resultados	57
6.2.1.	Sistema de Categorización	58
6.2.2.	Sistema de Recomendación	60
6.3.	Análisis de resultados	62
6.3.1.	Sistema de categorización	62
6.3.2.	Sistema de recomendación	63
6.4.	Resumen	63
7.	Conclusiones	64
7.1.	Trabajo futuro	66
7.1.1.	Mejoras al modelo del Sistema de Categorización	66
7.1.2.	Mejoras al modelo del Sistema de Recomendación	66
	Bibliografía	68
	Anexos	70

Anexo A. Prueba de Word Embedding, usando Word2vec	70
A.1. Construcción de conjunto de datos de entrada	70
A.1.1. Preparando Word2vec	72
A.1.1.1. preprocesamiento	72
A.1.1.2. Procesamiento de palabras	73
A.1.1.3. Word Embeddings promediado	75
A.1.1.4. Resultados preliminares	75