



ESCUELA POLITÉCNICA NACIONAL



FACULTAD DE SISTEMAS

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

RECUPERACIÓN DE LA INFORMACIÓN

IMPLEMENTACIÓN DE UN SISTEMA RAG (RETRIEVAL-AUGMENTED GENERATION)

PROYECTO BIMESTRAL II

Vickiann Jiménez

Gabriela Salazar

Jostin Vega



1.	Introducción	2
2.	Fases del Proyecto.....	2
2.1	Obtención del Corpus (20%)	2
2.2	Preprocesamiento del Corpus (25%)	6
2.3	Módulo de Recuperación (20%).....	9
2.4	Módulo de Generación e implementación técnica (30%).....	12
2.5	Evaluación del Sistema (5%).....	15
3.	Planificación Scrum	21
4.	Conclusiones y recomendaciones.....	23
5.	Anexos	25

Enlace al repositorio en GitHub: <https://github.com/JostinVega/proyectoRI>

1. Introducción

El objetivo de este proyecto es diseñar e implementar un sistema Retrieval Augmented Generation (RAG), que combina técnicas de recuperación de información (RI) con modelos de generación de texto para producir respuestas a partir de documentos relevantes. El sistema será capaz de recuperar documentos relevantes desde un corpus y generar respuestas coherentes y precisas basadas en estos documentos. El proyecto se desarrolló utilizando la metodología Scrum, adaptada para un ciclo de dos sprints de una semana cada uno, lo que permitió un enfoque ágil e iterativo. El proyecto se dividirá en varias fases, que se describen a continuación.

2. Fases del Proyecto

2.1 Obtención del Corpus (20%)

Objetivos:

- Recopilar los planes de trabajo de los 16 candidatos a la presidencia de Ecuador.
- Obtener entrevistas de los candidatos.
- Obtener biografías de los candidatos desde fuentes confiables como CNN y Primicias.

Tareas:

- Recolectar los planes de trabajo de los 16 candidatos.
- Obtener las entrevistas de los candidatos, ya sea en formato de texto o video/audio.
- Obtener biografías de los candidatos desde fuentes confiables.



- Organizar la información en un formato estructurado mediante DataFrames.
- Guardar los DataFrames en archivos CSV para facilitar su posterior procesamiento.

Paso 1: Recolección de los planes de trabajo

El primer paso consistió en la recolección de los planes de trabajo oficiales de los 16 candidatos a la presidencia de Ecuador. Este proceso se realizó mediante un script automatizado utilizando Selenium, que permitió navegar por el portal Voto Informado del Consejo Nacional Electoral (CNE) en <https://votoinformado.cne.gob.ec/procesoActual/partidos>.

Este proceso de recolección incluyó las siguientes etapas:

Navegación automatizada con Selenium: Se utilizó Selenium para navegar por la página del portal Voto Informado y acceder a la información de cada binomio electoral. El script identificó los partidos, números de lista, y los candidatos a presidente y vicepresidente de cada binomio electoral.

Descarga automatizada de planes de trabajo: Una vez obtenidos los enlaces a los planes de trabajo de los candidatos, el script automatizó la acción “clic” en cada uno de estos enlaces para descargar los archivos en formato PDF.

Procesamiento de los archivos PDF: Los archivos PDF descargados fueron procesados para extraer el texto. En los casos en que el texto no era directamente legible (por ejemplo, en documentos escaneados), se utilizó OCR (Reconocimiento Óptico de Caracteres) para convertir las imágenes de texto en texto editable.

Organización de la información en un DataFrame y guardado en CSV: Se creó un DataFrame con las siguientes columnas: PARTIDO, LISTA, PRESIDENTE, VICEPRESIDENTE, PLAN DE TRABAJO (texto extraído de los PDF) El DataFrame fue guardado en un archivo CSV para su uso posterior.

Este proceso, que combina web scraping, descarga automática de archivos y procesamiento de documentos (incluyendo OCR), permitió obtener todos los planes de trabajo de los candidatos.



Organización de la información en un DataFrame y guardado en CSV: Se creó un DataFrame con las siguientes columnas: PARTIDO, LISTA, PRESIDENTE, VICEPRESIDENTE, ENTREVISTA (donde se incluyeron las tres entrevistas

obtenidas por candidato). El DataFrame fue guardado en un archivo CSV para su posterior análisis.

[illegible]

Paso 3: Obtención de biografías

Extracción de biografías: Se recopiló la información disponible en las plataformas web de CNN y Primicias.

Procesamiento del texto: Se limpiaron los textos para eliminar información irrelevante.

Organización de la información en un DataFrame y guardado en CSV: Se creó un DataFrame con las siguientes columnas: PARTIDO, LISTA, PRESIDENTE, VICEPRESIDENTE, BIOGRAFÍA (texto extraído de las fuentes). El DataFrame fue guardado en un archivo CSV para su uso posterior.



	Partido	Lista	Presidente	Vicepresidente	Biografía
1	ENTRO DEMOCRÁTICO	1	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	te de la política pública.
2	TIDO UNIDAD POPULAR	2	JORGE ESCALA	PACHA TERAN	ndígenas y el sector rural.
3	TRIÓTICA 21 DE ENERO	3	ANDREA GONZALEZ	GALO MONCAYO	no de Andrea González.
4	AD DEMOCRACIA "PID"	4	VICTOR ARAUS	CRISTINA CARRERA	ncia y la institucionalidad.
5	ÓN CIUDADANA - RETO	5	LUISA GONZALEZ	DIEGO BORJA	a Revolución Ciudadana.
6	IDO SOCIAL CRISTIANO	6	RY KRONFLE KOZHAYA	LLYANA PASSAILAIGUE	cia junto a Henry Kronfle.
7	RATICA NACIONAL, ADN	7	DANIEL NOBOA AZIN	MARIA JOSE PINTO	nsparencia en el Estado.
8	PARTIDO AVANZA	8	LUIS FELIPE TILLERIA	ARLA PAULINA ROSERO	o de riesgos en seguridad.
9	UIERDA DEMOCRÁTICA	12	CARLOS RABASCALL	NDRA RIVAS MANTILLA	bernanza y transparencia.
10	ND OPORTUNIDADES	16	JUAN IVAN CUEVA	CRISTINA REYES	ca del Movimiento Amigo.
11	IALISTA ECUATORIANO	17	PEDRO GRANJA	VERONICA SILVA	ciudadana en el gobierno.
12	NACIONAL PACHAKUTIK	18	LEONIDAS IZA	KATIUSKA MOLINA	históricamente excluidos.
13	IENTO DEMOCRACIA SÍ	20	IVAN SAQUICELA	MARIA LUISA COELLO	a campaña de Saquicela.
14	ND OPORTUNIDADES	21	FRANCESCO TABACCHI	BLANCA SACANCELA	il del movimiento CREO.
15	DA MÁS ACCIÓN, SUMA	23	ENRIQUE GOMEZ	INES DIAZ	a los recursos del Estado.
16	VIMIENTO CONSTRUYE	25	HENRY CUCALON	CARLA LARREA	ercamiento a los votantes.

Scrum en esta fase:

Sprint 1: Durante el primer sprint, se recopiló y organizó el corpus textual con los planes de trabajo de los candidatos, las transcripciones de las entrevistas y las biografías en DataFrames estructurados, los cuales fueron almacenados en archivos CSV para facilitar su uso en futuras fases del proyecto.

2.2 Preprocesamiento del Corpus (25%)

Objetivos:

- Limpiar y normalizar los textos obtenidos en la fase de obtención del corpus.
- Tokenizar y eliminar elementos irrelevantes como signos de puntuación y palabras vacías.
- Organizar y estructurar los textos preprocesados

Tareas:

- Limpiar los textos de los planes de trabajo, entrevistas y biografías.
- Tokenizar los textos y eliminar palabras irrelevantes.
- Detectar y eliminar secciones no informativas.
- Guardar los textos preprocesados en archivos CSV.

Paso 1: Procesamiento del Plan de Trabajo

Lectura del archivo CSV: Se leyó el archivo dataframe_con_planes.csv.

Eliminación de caracteres especiales: Se eliminaron símbolos y caracteres no alfabéticos o irrelevantes.



Normalización de acentuación: Se convirtieron caracteres acentuados a su forma base utilizando unicodedata.

Eliminación de secciones numeradas: Se implementó una función para detectar y eliminar secciones no informativas mediante expresiones regulares.

Segmentación de texto: Se dividió el contenido en páginas y oraciones utilizando nltk.sent_tokenize.

Limpieza de oraciones: Se eliminaron stopwords en español para mejorar la calidad del corpus.

Generación de identificadores: Se creó un identificador único por oración considerando la lista, la página y la posición en el texto.

Almacenamiento de datos: Se generó un nuevo DataFrame con las oraciones procesadas y se guardó en un archivo CSV llamado planes_trabajo_procesados.csv.

	Lista	Partido	Presidente	Vicepresidente	id_oracion	pagina	oracion_original	oracion_limpia	oracion_sin_stopwords
1	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_1	1	rsidad humana y natural.	ersidad humana y natural	iversidad humana natural
2	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_2	1	uardarse en nuestro país.	uardarse en nuestro país	ros optado guardarse país
3	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_3	1	biodiversidad del mundo.	r biodiversidad del mundo	ayor biodiversidad mundo
4	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_4	1	nte de divisas importante.	nte de divisas importante	fuente divisas importante
5	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_5	1	ara financiar el desarrollo.	ara financiar el desarrollo	ursos financiar desarrollo
6	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_6	1	etido con la vida del país.	etido con la vida del país	o comprometido vida país
7	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_7	1	tejido social en el mundo.	tejido social en el mundo	mas tejido social mundo
8	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_8	1	mbien un territorio de paz.	mbien un territorio de paz	idor tambien territorio paz
9	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_9	1	hasta los recientes años.	hasta los recientes años	traumatica recientes años
10	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_10	1	e sus horas más oscuras.	te sus horas mas oscuras	mente horas mas oscuras
11	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_11	1	vive en estos momentos.	vive en estos momentos	mpararse vive momentos
12	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_12	1	uedan venir días mejores.	uedan venir días mejores	uedan venir días mejores
13	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_13	1	poder solo para repartirlo.	poder solo para repartirlo	gado poder solo repartirlo
14	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_14	1	sus modelos de negocios.	sus modelos de negocios	solidar modelos negocios
15	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_15	1	ad una política de Estado.	ad una política de estado	ivolidad banalidad politica
16	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_16	1	es esencialmente político.	es esencialmente politico	ais esencialmente politico
17	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_17	1	rativas y cálculos futuros.	ctativas y calculos futuros	rectativas calculos futuros
18	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_18	1	detrimento de los demás.	detrimento de los demas	lidades detrimento demas
19	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_1_19	1	incluso a quienes ganan.	incluso a quienes ganan	s perjudica incluso ganan
20	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_2_20	2	mpromiso real por el país.	mpromiso real por el país	ten compromiso real país
21	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_2_21	2	ral #Ecuador en el Centro!	oner al ecuadorenelcentro	poner ecuadorenelcentro
22	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_2_22	2	señalado anteriormente.	o senalado anteriormente	o senalado anteriormente
23	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_2_23	2	de la campaña electoral.	e de la campana electoral	cierre campana electoral
24	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1_2_24	2	ansversal y multisectorial.	ansversal y multisectorial	transversal multisectorial

Paso 2: Procesamiento de Entrevistas

Lectura del archivo CSV: Se leyó el archivo dataframe_con_entrevistas.csv.

Segmentación de entrevistas: Se identificaron y separaron los segmentos de entrevistas mediante expresiones regulares.

Limpieza de texto: Se eliminaron caracteres especiales y se normalizó la acentuación.

Tokenización: Se dividió el texto en oraciones con nltk.sent_tokenize y se eliminaron stopwords.

Procesamiento de descripciones y temas: Se limpiaron y almacenaron los segmentos de descripción y tema de cada entrevista.

Generación de identificadores: Se creó un identificador único por oración considerando la lista, el número de entrevista y la posición en el texto.



Almacenamiento de datos: Se organizó la información en un DataFrame y se guardó en un archivo CSV llamado entrevistas_procesadas.csv.

	Lista	Partido	Presidente	Vicepresidente	numero_entrevista	id_oracion	oracion_original	oracion_limpia	oracion_sin_stopwords	des
1	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_1	nta del Club de Nacional.	enta del club de nacional	presidenta club nacional	a las fam
2	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_2	toriano y el Movimiento 1.	toriano y el movimiento 1	ecuatoriano movimiento 1	a las fam
3	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_3	asiones, en 2004 y 2019.	casiones en 2004 y 2019	dos ocasiones 2004 2019	a las fam
4	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_4	, desde 2009 hasta 2018.	is desde 2009 hasta 2018	dos periodos 2009 2018	a las fam
5	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_5	stituyente de Montecristi.	nstituyente de montecristi	constituyente montecristi	a las fam
6	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_6	al figura como presidente.	al figura como presidente	pinion s figura presidente	a las fam
7	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_7	3 cuatro periodos fiscales.	s cuatro periodos fiscales	s cuatro periodos fiscales	a las fam
8	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_8	nos que el salario básico.	anos que el salario basico	ndo menos salario basico	a las fam
9	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_9	Buenos días, candidato.	buenos dias candidato	buenos dias candidato	a las fam
10	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_10	enido a Contacto Directo.	ivenido a contacto directo	envenido contacto directo	a las fam
11	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_11	Buenos días, Lenin.	buenos dias lenin	buenos dias lenin	a las fam
12	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_13	Muy bien, candidato.	muy bien candidato	bien candidato	a las fam
13	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_14	también con el correísmo.	tambien con el correismo	triotica tambien correismo	a las fam
14	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_17	er partido, efectivamente.	mer partido efectivamente	mer partido efectivamente	a las fam
15	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_18	emocrático en esa época.	lemocratico en esa epoca	centro democratico epoca	a las fam
16	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_19	indé Centro Democrático.	funde centro democratico	funde centro democratico	a las fam
17	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_20	ica camiseta ha sido esa.	ica camiseta ha sido esa	aqui unica camiseta sido	a las fam
18	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_21	zas, siempre en el centro.	nzas siempre en el centro	o alianzas siempre centro	a las fam
19	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_22	n, apoyamos a Allantopic.	on apoyamos a allantopic	cion apoyamos allantopic	a las fam
20	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_23	de nuestras propias filas.	sde nuestras propias filas	r candidatura propias filas	a las fam
21	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_24	sidencia de la República.	residencia de la republica	lato presidencia republica	a las fam
22	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_25	s candidato por descarte.	as candidato por descarte	mente candidato descarte	a las fam
23	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZZA	IA VALLECILLA SUAREZ	1	1_1_26	a en una tercera instancia.	a en una tercera instancia	ted llega tercera instancia	a las fam

Paso 3: Procesamiento de Biografías

Lectura del archivo CSV: Se leyó el archivo dataframe_con_biografias.csv.

Limpieza de texto: Se eliminaron caracteres especiales, tildes y se normalizó el formato de texto.

Segmentación de oraciones: Se dividió el texto en oraciones con nltk.sent_tokenize.

Eliminación de palabras irrelevantes: Se eliminaron stopwords en español para mejorar la calidad del corpus.

Generación de identificadores: Se creó un identificador único por oración considerando la lista y la posición en el texto.

Almacenamiento de datos: Se generó un DataFrame con las oraciones procesadas y se guardó en un archivo CSV llamado biografias_procesadas.csv.



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

	Lista	Partido	Presidente	Vicepresidente	id_oracion	oracion_original	oracion_limpia	oracion_sin_stopwords	
1	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_1	años en radio y televisión.	anos en radio y television	ai 40 anos radio television	
2	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_2	radial en los últimos años.	radial en los ultimos anos	grama radial ultimos anos	
3	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_3	no de Sixto Durán Ballén.	arno de sixto duran ballen	obierno sixto duran ballen	
4	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_4	yaquil, pero no tuvo éxito.	ayaquil pero no tuvo exito	er alcalde guayaquil exito	
5	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_5	uerza Compromiso Social.	fuerza compromiso social	fuerza compromiso social	
6	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_6	no candidato presidencial.	no candidato presidencial	ala candidato presidencial	
7	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_7	también es administrador.	tambien es administrador	o s tambien administrador	
8	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_8	.300 en ese mismo lapso.	.300 en ese mismo lapso	is usd 1.300 mismo lapso	
9	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_9	gia integral de seguridad.	agia integral de seguridad	ategia integral seguridad	
10	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_10	ograma nacional de salud.	ograma nacional de salud	programa nacional salud	
11	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_11	ra fortalecer la economía.	ara fortalecer la economia	ancia fortalecer economia	
12	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_12	administrativa y deportiva.	administrativa y deportiva	n administrativa deportiva	
13	1	ENTRO DEMOCRÁTICO	IMY JAIRALA VALLAZA	IA VALLECILLA SUAREZ	1_1_13	arte de la política pública.	parte de la politica publica	rtivo parte politica publica	
14	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_1	lo de Mérida (Venezuela).	lado de merida venezuela	ritorial merida venezuela	
15	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_2	al de Educadores (UNE).	cional de educadores une	nacional educadores une	
16	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_3	en Unidad Popular (UP).	rmo en unidad popular up	sformo unidad popular up	
17	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_4	ente en 2007 por el MPD.	yente en 2007 por el mpd	a constituyente 2007 mpd	
18	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_5	ea Nacional (2009-2013).	nblea nacional 2009 2013	nblea nacional 2009 2013	
19	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_6	nando mayores recursos.	gnando mayores recursos	gnando mayores recursos	
20	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_7	filas policiales y militares.	filas policiales y militares	on filas policiales militares	
21	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_8	inserción escolar masiva.	ainsercion escolar masiva	ainsercion escolar masiva	
22	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_9	gestión para el desarrollo.	gestion para el desarrollo	ecologia gestion desarrollo	
23	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_10	idad Central del Ecuador.	sidad central del ecuador	iversidad central ecuador	
24	2	PIDO UNIDAD POPULAR	JORGE ESCALA	PACHA TERAN	2_1_11	ndígenas y el sector rural.	indigenas y el sector rural	des indigenas sector rural	

Scrum en esta fase:

Sprint 1: Durante este sprint, se llevó a cabo el preprocesamiento del corpus, incluyendo la limpieza, tokenización y estructuración de los textos en archivos CSV. Esto permitió dejar los datos en un formato adecuado para su uso posterior.

2.3 Módulo de Recuperación (20%)

Objetivos:

- Diseñar un sistema que, dado un texto de consulta, recupere los documentos más relevantes del corpus.
- Implementar un índice eficiente para la búsqueda rápida de información.
- Utilizar modelos de aprendizaje automático como Sentence Transformers para generar embeddings y FAISS para la indexación y búsqueda eficiente de documentos relevantes.

Tareas:

- Cargar y organizar los datos procesados en los archivos CSV.
- Desarrollar un sistema de búsqueda eficiente combinando TF-IDF para recuperación rápida y embeddings de oraciones para mejorar la precisión semántica.
- Implementar FAISS para la indexación optimizada y la recuperación rápida de documentos relevantes mediante la generación de embeddings.



- Optimizar la recuperación de información aplicando pesos diferenciales a los distintos tipos de documentos (planes, entrevistas, biografías).
- Implementar una función de búsqueda que devuelva los resultados más relevantes con contexto adicional.

Paso 1: Carga y organización de datos

Lectura de archivos CSV: Se cargaron los archivos planes_trabajo_procesados.csv, biografías_procesadas.csv y entrevistas_procesadas.csv.

Filtrado de textos válidos: Se eliminaron textos muy cortos o irrelevantes.

Creación de metadatos: Se asignaron pesos y atributos clave a cada fragmento de texto según su origen.

Paso 2: Creación del sistema de búsqueda

Vectorización de textos: Se aplicó TF-IDF para pre-filtrar los textos más relevantes.

Generación de embeddings: Se utilizaron modelos de Sentence Transformers para generar representaciones vectoriales de los textos.

Indexación con FAISS: Se creó un índice eficiente para la búsqueda semántica de documentos relevantes.

Paso 3: Implementación de la función de búsqueda

Procesamiento de consultas: Se transformó el texto de la consulta en una representación vectorial.

Búsqueda en TF-IDF y FAISS: Se aplicó un modelo híbrido para obtener los documentos más relevantes.

Ajuste de relevancia: Se aplicaron pesos a los distintos tipos de documentos para mejorar la precisión de la recuperación.

Presentación de resultados: Se mostraron los documentos relevantes junto con el contexto en el que se encontraron.



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Sistema de búsqueda listo.

=====

Ingrese su búsqueda (o 'salir' para terminar): reducir impuestos
Número de resultados a mostrar (Enter para usar 5):

Buscando: reducir impuestos
Cargando sistema de búsqueda...

Resultados encontrados:

Ranking: 1
Tipo: entrevista
Lista: 3 - Partido: PARTIDO SOCIEDAD PATRIÓTICA 21 DE ENERO
Presidente: ANDREA GONZALEZ
ID: 3_2_76
Relevancia: 0.9786

Contexto:
Nuestra propuesta es recuperar el control del perfil costero. Si logramos esto, reduciremos el tráfico de drogas y el crimen organizado sin desgastarnos en la frontera norte y sur. [RELEVANTE] Sobre la economía y la reducción de impuestos
Ha propuesto reducir impuestos como el ISD. Pero el país tiene un déficit fiscal. Más empleos, no más impuestos.

Número de entrevista: 2
Tema: Candidatura presidencial, separación de Construye, alianza con Sociedad Patriótica, uso del petróleo para el desarrollo, minería legal e ilegal, impacto del oro ilegal, estrategia de seguridad territorial, narcotráfico y control marítimo, reducción de impuestos (ISD), fomento del empleo, reformas estructurales y nueva Constitución.
Descripción: En esta entrevista, Andrea González Nader, candidata presidencial por Sociedad Patriótica, explica su decisión de postularse y su separación de Construye, enfatizando que mantiene sus principios y un plan de gobierno propio. Defiende el uso del petróleo y la minería como herramientas para el desarrollo, pero advierte sobre la corrupción y los errores históricos en su manejo. Propone combatir la minería ilegal con una estrategia de seguridad territorial y recuperar el control del mar ecuatoriano para frenar el narcotráfico. En economía, aboga por reducir impuestos como el ISD para incentivar el empleo y fortalecer la recaudación fiscal a través del crecimiento del sector privado. Finalmente, subraya la necesidad de reformas estructurales y una nueva Constitución para incentivar la inversión y mejorar la gobernabilidad.

Ranking: 2
Tipo: entrevista
Lista: 25 - Partido: MOVIMIENTO CONSTRUYE
Presidente: HENRY CUCALON
ID: 25_2_85
Relevancia: 0.7780

Contexto:
Creo en los incentivos fiscales, no voy a aumentar impuestos. Es un compromiso. [RELEVANTE] Eliminaré varios impuestos, como el Impuesto a los Consumos Especiales y reduciré el Impuesto a la Salida de Divisas hasta llegar a cero, ya que este impuesto lo que hace es perjudicar la entrada de divisas al país. Con austeridad fiscal y disciplina, a partir del segundo o tercer año de gobierno, voy a reducir el IVA. También creo firmemente en la descentralización.

Número de entrevista: 2
Tema: Política exterior y migración, relaciones con Estados Unidos, seguridad, reforma judicial, lucha contra el crimen organizado, oportunidades laborales para jóvenes, reforma laboral, trabajo por horas, gestión de consensos políticos, eliminación de monopolios estatales, atracción de inversión extranjera, liberalización del sector energético, reforma al sector petrolero, apertura del catastro minero, concesión del sistema vial, infraestructura vial, reducción de impuestos, descentralización fiscal, modernización de la Policía Nacional, becas para educación técnica, seguridad jurídica, combate a la minería ilegal.
Descripción: En esta entrevista, Henry Cucalón expuso sus propuestas para abordar los problemas más críticos de Ecuador, desde la seguridad y el empleo hasta la reforma judicial y la atracción de inversión. Se mostró firme en la defensa de un Estado eficiente y pequeño, proponiendo medidas como la liberalización del sector energético y la apertura del catastro minero para fomentar la inversión y generar empleo. También se comprometió a eliminar monopolios estatales, modernizar la policía y la justicia, y dinamizar las reglas laborales para crear más empleos. Además, Cucalón habló de la necesidad de políticas migratorias coherentes y de trabajar en la integración de los ecuatorianos deportados desde EE. UU., al mismo tiempo que se comprometió a reducir impuestos y a mejorar la infraestructura vial y el sector energético.

Ranking: 3
Tipo: plan
Lista: 6 - Partido: PARTIDO SOCIAL CRISTIANO
Presidente: HENRY KRONFLE KOZHAYA
ID: 6_26_317
Relevancia: 0.6170

Contexto:



Ranking: 3
Tipo: plan
Lista: 6 - Partido: PARTIDO SOCIAL CRISTIANO
Presidente: HENRY KRONFLE KOZHAYA
ID: 6_26_317
Relevancia: 0.6170

Contexto:
El país

tiene que entender que solo se pueden cobrar tasas o contribuciones, contra la prestación de los servicios. Simplificación del sistema tributario

Reducir la complejidad y la carga administrativa tanto para los contribuyentes como para la administración tributaria. [RELEVANTE] Eliminación de impuestos como iSD, y simplificación de los procedimientos de presentación de declaraciones y pago de impuestos. De la misma manera buscar la eficacia y rapidez para asignar recursos a los GADs que puedan transformarse en obras y servicios. Ampliación de la base tributaria
Fortalecer la base de contribuyentes y reducir la evasión fiscal, por ejemplo, mejorando los mecanismos de fiscalización y fortaleciendo la cooperación entre las entidades tributarias y otras instituciones para el intercambio de información, beneficios por pronto pago, entre otras alternativas.

Ranking: 4
Tipo: entrevista
Lista: 6 - Partido: PARTIDO SOCIAL CRISTIANO
Presidente: HENRY KRONFLE KOZHAYA
ID: 6_2_67
Relevancia: 0.6031

Contexto:
Bajando impuestos, devolviendo el dinero a las familias ecuatorianas. No voy a arreglar al gobierno, voy a arreglar la familia ecuatoriana. [RELEVANTE] Bajando impuestos, subiendo salarios, eliminando trámites que afectan la competitividad de las empresas. Manteniendo el subsidio de la bombona de gas doméstico y reduciendo el IVA en la canasta básica, los insumos médicos y los productos agrícolas. Con esto, cuando las familias tienen más dinero, compran más, las empresas venden más, tributan más, y se genera más empleo

Scrum en esta fase:

Sprint 2: Durante este sprint, se implementó el sistema de recuperación de información, integrando técnicas de búsqueda híbrida con TF-IDF y embeddings. Se optimizó la indexación con FAISS y se mejoró la relevancia de los resultados mediante la asignación de pesos a los distintos tipos de documentos. Se realizaron pruebas iniciales para evaluar la eficacia del sistema en consultas diversas.

2.4 Módulo de Generación e implementación técnica (30%)

Objetivos

- Desarrollar un sistema de generación de respuestas naturales y coherentes basado en los documentos recuperados
- Implementar un sistema de clasificación de tipos de consultas para respuestas especializadas
- Crear prompts dinámicos adaptados al tipo de consulta
- Establecer un sistema de respaldo para casos de fallo en la generación principal

Tareas

- Implementar un sistema de identificación de tipos de consultas mediante patrones regex



- Desarrollar prompts específicos para cada tipo de consulta
- Integrar el modelo Mistral para la generación de respuestas
- Crear un sistema de fallback para garantizar respuestas en todo momento
- Implementar un sistema de extracción de fuentes para referencias

Paso 1: Identificación de Tipos de Consultas

Se implementó un sistema de clasificación de consultas mediante expresiones regulares (regex) que permite identificar y procesar seis tipos principales de consultas:

- **Biografías:** Maneja patrones como "quien es [candidato]" o "biografía de [candidato]", permitiendo extraer información biográfica específica.
- **Propuestas verbales:** Procesa consultas del tipo "qué candidatos proponen [tema]", facilitando la búsqueda temática de propuestas.
- **Entrevistas:** Identifica patrones como "temas de la entrevista de [candidato]", permitiendo recuperar información de entrevistas específicas.
- **Partido-candidato:** Maneja consultas sobre afiliación política como "¿de qué partido es [candidato]?".
- **Partido-nombre:** Procesa preguntas como "¿qué candidatos pertenecen al partido [nombre]?".
- **Propuestas de candidato:** Identifica patrones como "propuestas de [candidato]".

Paso 2: Generación de Prompts Dinámicos

El sistema genera prompts especializados adaptados a cada tipo de consulta, estructurados en cuatro componentes principales:

Contexto inicial:

- Define el objetivo de la respuesta
- Presenta la consulta normalizada

Documentos relevantes:

- Integra los documentos recuperados
- Estructura la información por tipo de documento
- Incluye metadatos relevantes (candidato, partido, tipo)

Instrucciones específicas:



- Define el formato de respuesta según el tipo
- Establece reglas de presentación
- Determina prioridades de información

Reglas generales:

- Mantiene consistencia en el tono
- Evita referencias a fuentes
- Garantiza fluidez narrativa

Paso 3: Integración con Modelo de Generación

La integración con el modelo Mistral se realizó mediante una API REST con las siguientes características:

- **Temperatura 0.7:** Balance entre creatividad y precisión
- **Top-p 0.9:** Diversidad controlada en la generación
- Contexto máximo de 4096 tokens
- Sistema de reintentos ante fallos

Paso 4: Sistema de Respaldo

Se implementó un sistema de respaldo multinivel que garantiza respuestas incluso en casos de fallo del generador principal:

Generadores especializados:

- Biografías: Extracción y estructuración de información personal y trayectoria
- Propuestas: Organización jerárquica por candidato y tema
- Entrevistas: Síntesis temática y puntos clave
- Información de partido: Datos de afiliación y contexto político

Sistema de fuentes:

- Extracción automática de referencias según tipo de documento
- Mapeo de medios para entrevistas (Ecuavisa, Teleamazonas, Vistazo)
- Identificación precisa de páginas en planes de trabajo

Integración mediante Endpoints:

El sistema expone dos endpoints principales en Flask:

1. /buscar:

- Gestiona la recuperación de documentos relevantes
- Implementa filtros y rankings de relevancia



- Retorna documentos estructurados con metadatos

2. /generar_respuesta:

- Coordina el proceso completo de generación
- Integra identificación, generación y respaldo
- Retorna respuestas estructuradas con fuentes

Scrum en esta fase:

Sprint 2: Durante este sprint, se diseñó e implementó el módulo de generación de respuestas, asegurando una integración fluida con el sistema de recuperación de información y optimizando la entrega de respuestas relevantes.

2.5 Evaluación del Sistema (5%)

Objetivos:

- Evaluar la precisión y relevancia del sistema de búsqueda y generación de respuestas.
- Comparar las respuestas generadas con respuestas de referencia establecidas previamente.
- Identificar posibles mejoras en la recuperación de información y la generación de respuestas.

Tareas:

- Definir un conjunto de consultas predefinidas que cubran distintos tipos de preguntas.
- Ejecutar el sistema con las consultas predefinidas y almacenar los resultados.
- Analizar la precisión de la recuperación de documentos y la relevancia de las respuestas generadas.
- Identificar errores o inconsistencias en la generación de respuestas y proponer mejoras.


El sistema permite responder consultas relacionadas con:

- Afiliación política:
 - El candidato [*nombre candidato*], ¿a qué partido pertenece?
- Entrevistas:
 - ¿Qué temas se tratan en la entrevista de [*nombre candidato*]?
- Información biográfica:
 - Biografía de [*nombre candidato*].
 - ¿Quién es [*nombre candidato*]?



- Propuestas de los candidatos:
 - ¿Qué candidatos proponen *[verbo]*?
 - ¿Quiénes proponen *[verbo]*?
 - Propuestas del candidato *[nombre candidato]*.

El candidato *Henry Kronfle*, ¿a qué partido pertenece?

**CandiTrack 2025**
AI Assistant

de EcuadorTV durante el año 2018-2019. En el ámbito empresarial, es accionista y administrador de varias compañías en el sector de telecomunicaciones.

11:39 PM

el candidato henry kronfle a que partido perteneces

11:40 PM

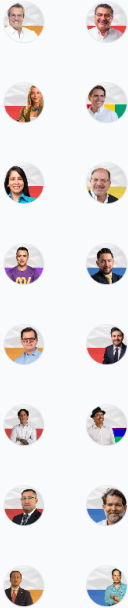
Henry Kronfle es un candidato político ecuatoriano que actualmente pertenece al Partido Social Cristiano (PSC). En su carrera empresarial, se desempeñó como presidente de la Asociación de Industriales Latinoamericanos y de la Cámara de Industrias de Guayaquil. Sin embargo, tras ingresar a la política en el año 2017, se ha desempeñado tres veces consecutivas como asambleísta nacional. Henry Kronfle es conocido por su lucha contra la corrupción y su compromiso con la reducción de la burocracia y el gasto público. Ha presentado propuestas para la restructuración de empresas públicas, aumentar la producción petrolera y minera, reducir los impuestos, aumentar el empleo y eliminar trámites burocráticos. Además, ha defendido la seguridad ciudadana, la justicia y la lucha contra la corrupción, así como la consulta popular para sancionar a jueces y fiscales corruptos. En su último periodo como asambleísta nacional, ocupó la presidencia de este organismo. Antes de entrar en política, fue cónsul honorario de México y dirigente gremial. Actualmente, encabeza el binomio socialcristiano que conforma con Dayana Pasleg para las elecciones 2025.

Fuentes:

- CNN Español y Primicias EC - Perfiles candidatos presidenciales 2025
- Entrevista en Ecuavisa con HENRY KRONFLE KOZHAYA
- Entrevista en Teleamazonas con HENRY KRONFLE KOZHAYA

11:43 PM

Message CandiTrack 2025...



¿Qué temas se tratan en la entrevista de *Carlos Rabascall*?



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN



CandiTrack 2025
AI Assistant

que temas se tratan en la entrevista de carlos rabascall

11:26 PM



En esta entrevista con Carlos Rabascall, se abordan varios temas relevantes en su carrera política y plataforma electoral. Como candidato a la Vicepresidencia por el Partido Izquierda Democrática (ID), Rabascall se enfoca en la modernización del Estado, la digitalización de procesos gubernamentales, y la gestión de riesgos en seguridad. En cuanto a su trayectoria, se menciona que ha sido conocido por su trabajo como presentador de noticias y entrevistador en medios estatales y privados, incluyendo Ecuavisa y EcuadorTV. En 2021, Rabascall fue compañero de fórmula de Andrés Arauz en la alianza correísta Unión por la Esperanza (UNES), pero perdieron las elecciones ante Guillermo Lasso. En la entrevista se habla sobre su motivación para ser candidato, ya que a su juicio, el país ecuatoriano sigue teniendo los mismos problemas estructurales después de 45 años de regreso a la democracia después de una dictadura. Rabascall enfatiza en las debilidades estructurales del Ecuador, como una economía débil, finanzas públicas débiles, incapacidad para generar trabajo formal, un sistema educativo obsoleto, salud pública catastrófica, hipercentralismo que no ha permitido la construcción de nuevos actores económicos y la captación política y el crimen organizado de la función judicial. Además se discuten temas como la seguridad y creación de una policía civil especializada, la reestructuración del sistema penitenciario, generación de empleo, reforma laboral, crisis del IESS y afiliación, aumento de la producción petrolera, atracción de inversión privada, reforma fiscal sin aumento de impuestos, eliminación del ISD en sectores estratégicos, reducción de aranceles y posible ajuste del IVA. En el discurso político, Rabascall aborda su postura frente al gobierno de Rafael Correa y las diferencias con el correísmo. También se habla sobre la necesidad de fortalecer el partido ID y reducir la polarización política en el país. En cuanto a los temas más concretos, Rabascall presenta propuestas para generar empleo temporal en zonas rurales, reactivar la economía, mejorar la educación y la salud pública, así como enfrentar la seguridad y el combate a la delincuencia. En resumen, los temas principales que se abordan en la entrevista de Carlos Rabascall incluyen la modernización del Estado, la seguridad y la creación de una policía civil especializada, la reestructuración del sistema penitenciario, la generación de empleo, la reforma laboral, la crisis del IESS y afiliación, el aumento de la producción petrolera, la atracción de inversión privada, la reforma fiscal sin aumento de impuestos, la eliminación del ISD en sectores estratégicos, la reducción de aranceles y posible ajuste del IVA, así como su

Message CandiTrack 2025...



Biografía de *Henry Cucalón*.



CandiTrack 2025
AI Assistant

biografía de henry cucalon

11:01 PM



Henry Cucalón, con 51 años, es un destacado activista por los derechos de los pueblos indígenas y una ferviente defensora de la equidad, el desarrollo rural y la representación de comunidades en la política. Cucalón ha disfrutado de una extensa trayectoria en el servicio público, habiendo desempeñado cargos importantes como Secretario General de la Procuraduría (1998-2002) y Secretario del Municipio de Guayaquil (2003-2012). Recientemente, se desempeñó como Ministro de Gobierno en la administración de Guillermo Lasso, tras su militancia por muchos años en el Partido Social Cristiano (PSC), del cual se desafiló en 2021 debido a diferencias internas. En la actualidad, Henry Cucalón es candidato a la Presidencia de Ecuador y su rol en la campaña política se centra en la inclusión social, los derechos indígenas y el acceso equitativo a los recursos del Estado. Su binomio lo completa Carla Larrea como candidata a Vicepresidenta. Entre sus propuestas están la seguridad integral, la lucha contra el crimen organizado, la reforma económica, el fomento de la inversión privada nacional y extranjera, el reactivación del sector privado, la reducción del gasto público sin afectar lo esencial, entre otras. También propone mecanismos adicionales de transparencia y gobierno abierto, así como rendición de cuentas permanente para evaluar el avance de sus propuestas. Además, cuenta mucho importancia a la dinamización de la economía, la reforma del sistema judicial, la seguridad, la salud y la educación, la austeridad fiscal y la recorte de la burocracia, entre otros temas políticos relevantes.

Fuentes:

- CNN Español y Primicias EC - Perfiles candidatos presidenciales 2025
- Entrevista en Ecuavisa con HENRY CUCALON
- Plan de trabajo de HENRY CUCALON, página 125
- Entrevista en Revista Vistazo con HENRY CUCALON

11:06 PM

Message CandiTrack 2025...




¿Quién es *Ivan Saquicela*?



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS


INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

**CandiTrack 2025**
AI Assistant

11:06 PM

quien es ivan saquicela

11:15 PM



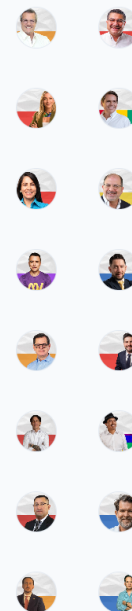
Ivan Saquicela es un político ecuatoriano de 49 años afiliado al Movimiento Democracia Sí. Obtuvo su doctorado en Jurisprudencia por la Universidad de Cuenca, y posee una maestría en Derecho Penal y Criminología por la Universidad Regional Autónoma de Los Andes (UNIANDES). Ha tenido una amplia trayectoria en la Función Judicial, desempeñándose como Fiscal en la provincia de Azuay y Conjuez de la Corte Nacional de Justicia. Además, fue concejal de Cuenca entre 2013 y 2017. En el presente, Saquicela se encuentra en una alianza electoral con María Luisa Coello como compañera de fórmula para los comicios de 2025. En esta campaña, buscan fortalecer el Estado Plurinacional y crear oportunidades económicas para sectores históricamente excluidos. Dentro de sus propuestas políticas actuales, Saquicela se enfoca en temas como la seguridad y lucha contra el crimen, combate a la corrupción, experiencia en el sector judicial, fortalecimiento de salud y educación, transparencia y eficiencia en la administración pública. Además, plantea un interés en mejorar la recaudación tributaria, cobrar a los morosos y mejorar la producción petrolera mediante inversión privada. En general, Saquicela y su compañera de fórmula luchan por convicciones, no dependientes de los sondeos, y buscan diagnosticar qué es lo que pasa en la política ecuatoriana para presentar propuestas efectivas.

Fuentes:


- CNN Español y Primicias EC - Perfiles candidatos presidenciales 2025
- Entrevista en Ecuavisa con IVAN SAQUICELA
- Entrevista en Telemazonas con IVAN SAQUICELA

11:18 PM

Message CandiTrack 2025...




¿Quiénes proponen *reducir impuestos*?

**CandiTrack 2025**
AI Assistant

10:48 PM

quienes proponen reducir impuestos

10:53 PM



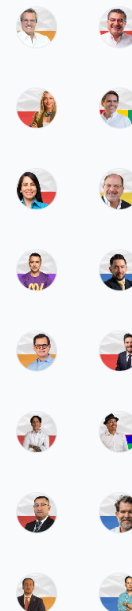
1. Propuesta de Henry Cudalon (Movimiento Construye) - Eliminará varios impuestos, incluyendo el Impuesto a los Consumos Especiales y reducirá el Impuesto a la Salida de Divisas hasta llegar a cero. - A partir del segundo o tercer año de gobierno, reducirá el IVA. - Intentará hacer una austeridad fiscal y disciplina para lograr estas reducciones de impuestos. - Creará en la descentralización fiscal. 2. Propuesta de Andrea Gonzalez (Partido Sociedad Patriótica 21 de Enero) - Reducirá el Impuesto al Servicio Domiciliario (ISD). - Realizará un análisis de impacto fiscal para determinar las implicaciones de la reducción de este impuesto. - Priorizará la agricultura y la sostenibilidad a largo plazo, aprovechando el gran potencial agrícola del país y construyendo las bases para una economía diversificada, más resiliente, sostenible en el tiempo y ecológica. 3. Propuesta de Henry Kronfle Kozhaya (Partido Social Cristiano) - Simplificación del sistema tributario: Reducirá la complejidad y la carga administrativa tanto para los contribuyentes como para la administración tributaria. - Eliminará impuestos como el ISD, y simplificará los procedimientos de presentación de declaraciones y pago de impuestos. - Fortalecerá la base de contribuyentes y reducirá la evasión fiscal. - Intentará hacer eficaz y rapidez para asignar recursos a los GADs que puedan transformarse en obras y servicios. - Ampliará la base tributaria fortaleciendo la cooperación entre las entidades tributarias y otras instituciones para el intercambio de información, beneficios por pronto pago, entre otras alternativas.

Fuentes:

- Plan de trabajo de ANDREA GONZALEZ, página 40
- Entrevista en Telemazonas con ANDREA GONZALEZ
- Entrevista en Telemazonas con HENRY CUDALON
- Plan de trabajo de HENRY KRONFLE KOZHAYA, página 26

10:53 PM

Message CandiTrack 2025...



¿Qué candidatos proponen *construir hospitales*?



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN



CandiTrack 2025
AI Assistant

que candidatos proponen construir hospitales

10:39 PM



1. Propuestas de Pedro Granja (Partido Socialista Ecuatoriano) - Construcción de hospitales en todo el país para mejorar la salud general de la población. - Reducir las disparidades en la atención médica y acceso a servicios médicos básicos, incluso en áreas urbanas y rurales remotas. 2. Propuestas de Víctor Araus (Movimiento Pueblo Igualdad Democracia "PID") - Garantizar acceso a un sistema de salud integral, de alta calidad y eficiente para todos los ciudadanos, independientemente de su ubicación o situación económica. - Modernización y equipamiento de hospitales y centros de salud con tecnología de punta. 3. Propuestas de Juan Ivan Cueva (Movimiento Amigo, Acción Movilizadora Independiente Generando Oportunidades) - Construcción de escuelas, hospitales, centros comunitarios y mejoras en vías de comunicación para aumentar la rentabilidad del sector sin explotar nuevas áreas. 4. Propuestas de Luis Felipe Tillería (Partido Avanza) - Reducción de tiempos de espera y mejora de la eficiencia en el sistema de salud del país mediante soluciones digitales como la telemedicina, historia clínica electrónica y una red digital unificada. - Construcción de 10 nuevos hospitales en zonas rurales y áreas vulnerables. 5. Propuestas de Carlos Rabascall (Partido Izquierda Democrática) - Modernización del sistema de salud mediante la digitalización y renovación de infraestructura. - Implementar telemedicina en zonas remotas y urbanas marginadas. - Construcción de un hospital de tercer nivel en Cuenca. - Desarrollo de un Plan Nacional de Salud Mental.

Fuentes:

- Plan de trabajo de JUAN IVAN CUEVA, página 38
- Plan de trabajo de VÍCTOR ARAUS, página 28
- Entrevista en Teleamazonas con PEDRO GRANJA
- Plan de trabajo de LUIS FELIPE TILLERÍA, página 55
- Plan de trabajo de CARLOS RABASCALL, página 77

10:44 PM

Message CandiTrack 2025...



Propuestas del candidato *Henry Cucalón*.



CandiTrack 2025
AI Assistant

propuestas del candidato henry cucalon

12:26 AM



En cuanto al candidato Henry Cucalón y su partido Movimiento Construye, se puede destacar que su propuesta política está enfocada en varios temas principales: 1. **Combate a la corrupción y política de integridad**: Proponen mecanismos adicionales de transparencia, gobierno abierto y participación ciudadana para evaluar el avance diario de sus propuestas. Esto incluye rendición de cuentas efectivas que garanticen los principios en la configuración de todos y cada uno de los procesos del gobierno en forma permanente. 2. **Estado eficiente y pequeño, honestidad**: El candidato Cucalón aboga por un estado eficiente y pequeño que respete la democracia, luchando contra el crimen organizado para construir un futuro mejor para los ecuatorianos. 3. **Experiencia, reforma económica**: Henry Cucalón cuenta con una extensa trayectoria en el servicio público en diferentes cargos, y su experiencia le permite propositar una reforma económica que incluye incentivos a la inversión privada nacional y extranjera. 4. **Sector privado como generador de empleo**: El candidato apoya el sector privado como fuente importante de empleo para los ecuatorianos. 5. **Crítica al populismo y demagogia**: Cucalón se opone al populismo y demagogia, buscando implementar políticas basadas en la razón y la evidencia. 6. **Seguridad**: El candidato enfatiza en la seguridad integral para combatir el crimen organizado en Ecuador. 7. **Salud y educación**: Cucalón propone mejorar la seguridad jurídica y ofrecer becas para la educación técnica. 8. **Austeridad fiscal, reducción de la burocracia**: El candidato apoya una austeridad fiscal y una reducción de la burocracia en el gobierno, buscando eliminar subsidios ineficaces y mejorar la eficiencia del estado. 9. **Desarrollo de la minería y el petróleo**: Proponen mejorar la conectividad vial, fomento del turismo, reforma del catastro minero, reducción del gasto público sin afectar lo esencial. 10. **Reforma judicial**: El candidato apoya una reforma del sistema judicial para mejorar la seguridad jurídica en Ecuador. 11. **Cooperación internacional, migración**: Cucalón aboga por una política exterior que permita relaciones estables y productivas con Estados Unidos y otros países, además de enfatizar en la lucha contra la minería ilegal. 12. **Fomento tecnológico y modernización de las Fuerzas Armadas**: El candidato también propone incentivar el desarrollo tecnológico y modernizar las Fuerzas Armadas para garantizar la seguridad del país.

Message CandiTrack 2025...





Se mide la eficiencia del sistema de recuperación mediante el uso de embeddings de similitud semántica. Los documentos recuperados se comparan con un ground truth que consiste en un conjunto de textos de referencia seleccionados para cada consulta, representan respuestas esperadas y están diseñados para reflejar las propuestas políticas en diversas áreas, como reducción de impuestos, seguridad ciudadana, educación pública, empleo juvenil y cambio climático.

El método de evaluación se basó en calcular la cosine similarity entre los embeddings de los documentos recuperados y los embeddings de los textos en el ground truth. Se consideró que si la similitud es mayor a 0.7, el documento es relevante. Tomando esto en cuenta, se calculan las métricas de: precisión, recall y F1-Score, que permiten medir cuántos de los documentos relevantes fueron recuperados y qué porcentaje de los recuperados son realmente relevantes.

```
*** Cargando sistema de búsqueda...

🔵 Evaluando query: reducción de impuestos
Cargando sistema de búsqueda...

🔴 **Documentos obtenidos y similitud:**
1. Sobre la economía y la reducción de impuestos
Ha propuesto reducir impuestos como el ISD... (✅ Relevante - Similitud: 0.85)
2. Economía: Reducción de impuestos especiales.... (✅ Relevante - Similitud: 0.84)
3. Disminución del Impuesto a Consumos Especiales y del Impuesto a la
Salida de Divisas
Acciones:
o Realizar un análisis de impacto fiscal para determina... (✅ Relevante - Similitud: 0.73)
4. No más incrementos de impuestos ni amnistías tributarias: No incrementar más
impuestos, pero sí controlar los créditos tributarios y devoluciones, sin... (✅ Relevante - Similitud: 0.82)
5. Eliminación de impuestos como ISD, y simplificación de los
procedimientos de presentación de declaraciones y pago de impuestos.... (✅ Relevante - Similitud: 0.75)

🔵 Evaluando query: seguridad ciudadana
Cargando sistema de búsqueda...

🔴 **Documentos obtenidos y similitud:**
1. SEGURIDAD CIUDADANA
La seguridad ciudadana se refiere a la protección de los ciudadanos
frente a las amenazas contra su vida, su propiedad, su liberta... (❌ No relevante - Similitud: 0.54)
...

📊 **Resultados Promedio**
🔴 Precision@5: 0.4800
🔴 Recall@5: 0.8000
🔴 F1-Score@5: 0.5333

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Scrum en esta fase:



Sprint 2: Se llevaron a cabo pruebas exhaustivas del sistema, evaluando su desempeño en términos de precisión, relevancia y coherencia de las respuestas generadas.

3. Planificación Scrum

3.1 Planificación inicial del proyecto

El proyecto fue desarrollado en un periodo de dos semanas utilizando la metodología Scrum. Se establecieron dos sprints, con tareas definidas para cada uno y una evaluación al final de cada iteración. El equipo acordó reuniones diarias para seguimiento y revisiones al final de cada sprint para evaluar el progreso y ajustar la estrategia si fuese necesario.

3.2 Roles asignados en el equipo

El equipo se organizó con los siguientes roles:

- **Scrum Master (Vickiann Jiménez):** Responsable de facilitar el proceso Scrum, gestionar reuniones y eliminar impedimentos.
- **Desarrollador de Módulo de Recuperación (Gabriela Salazar):** Encargada del diseño y optimización del sistema de recuperación de documentos.
- **Desarrollador de Módulo de Generación (Jostin Vega):** Responsable de la implementación del modelo de generación de respuestas.
- **Procesamiento de Datos (Todo el equipo):** Todos participaron en la recopilación, limpieza y organización del corpus textual.

3.3 Sprints definidos y retrospectivas

Sprint 1: (Semana 1)

Objetivo: Recopilar, organizar y preprocesar el corpus textual, asegurando que los datos estén limpios, estructurados y listos para su uso en la fase de recuperación y generación de respuestas.

Tareas:

- Recopilar los planes de trabajo de los candidatos desde el portal Voto Informado (CNE) utilizando Selenium.
- Obtener y transcribir las entrevistas desde YouTube mediante Whisper.
- Recopilar biografías de los candidatos desde fuentes confiables como CNN y Primicias.
- Limpiar, tokenizar y normalizar el texto de los planes de trabajo, entrevistas y biografías para reducir ruido y mejorar la coherencia del corpus.
- Segmentar el contenido relevante, asegurando que las secciones clave del texto sean fácilmente accesibles.



- Guardar los textos preprocesados en archivos CSV para su posterior uso.

Retrospectiva Sprint 1:

- **Éxitos:** Se logró recolectar, limpiar y estructurar el corpus con éxito, asegurando que estuviera listo para la fase de recuperación.
- **Desafíos:** Hubo dificultades en la transcripción de entrevistas debido a ruido en los audios, lo que requirió mayor tiempo de limpieza.
- **Acciones de mejora:** Se implementaron mejores filtros de ruido en la transcripción para mejorar la calidad de los datos.

Sprint 2: (Semana 2)

Objetivo: Implementar el sistema de recuperación de información y generación de respuestas, asegurando la optimización de ambos módulos y evaluando el desempeño del sistema RAG.

Tareas:

- Implementar el sistema de recuperación utilizando TF-IDF y embeddings de Sentence Transformers.
- Integrar FAISS para optimizar la indexación y búsqueda de documentos relevantes.
- Implementar un modelo de generación de respuestas basado en Mistral, configurando los prompts para cada tipo de consulta.
- Desarrollar un sistema de clasificación de consultas utilizando expresiones regulares para mejorar la interpretación de preguntas.
- Implementar endpoints en Flask para la recuperación de documentos y generación de respuestas.
- Evaluar el sistema mediante métricas de precisión como Precisión, Recall y F1Score.

Retrospectiva Sprint 2:

- **Éxitos:** Se logró integrar correctamente la generación de respuestas con la recuperación de documentos, mejorando la coherencia de las respuestas.
- **Desafíos:** Se identificaron algunas consultas ambiguas que generaban respuestas menos precisas.
- **Acciones de mejora:** Se implementó un sistema de clasificación de consultas para mejorar la interpretación de las preguntas.

3.4 Documentación de decisiones tomadas

Durante el desarrollo del proyecto, se tomaron las siguientes decisiones clave:



- Utilización de FAISS para mejorar la rapidez en la recuperación de documentos.
- Integración del modelo Mistral para generar respuestas más precisas y contextualizadas.
- Implementación de un sistema de respaldo que garantiza respuestas en todo momento.
- Ajuste de pesos diferenciales en los tipos de documentos (planes de trabajo, entrevistas, biografías) para mejorar la relevancia de los resultados.

3.5 Evaluación final del uso de Scrum

El uso de Scrum fue beneficioso para estructurar el desarrollo del proyecto en ciclos iterativos, permitiendo ajustes y mejoras de manera ágil.

Puntos positivos:

- Se logró mantener un flujo de trabajo organizado con tareas bien definidas.
- La división en sprints permitió enfocarse en fases específicas del desarrollo.
- La retrospectiva al final de cada sprint facilitó la identificación de problemas y mejoras.

Oportunidades de mejora:

- Se podría haber implementado un backlog de tareas más detallado para priorizar mejor los desarrollos.
- Incluir reuniones más frecuentes para revisar los avances de manera más estructurada.

En general, la metodología Scrum permitió alcanzar los objetivos del proyecto de manera eficiente, con entregables funcionales en cada iteración y una clara mejora en la calidad de los resultados generados.

4. Conclusiones y recomendaciones

Conclusiones:

Mediante un sistema de Recuperación Aumentada por Generación (RAG), se ha desarrollado una solución capaz de manejar consultas sobre candidatos y procesos electorales, utilizando técnicas de procesamiento de lenguaje natural que superan los métodos tradicionales de búsqueda de información.

Un componente central del sistema es su mecanismo de clasificación de consultas, implementado mediante expresiones regulares. Este sistema puede identificar y procesar seis tipos principales de consultas: biografías, propuestas verbales, entrevistas, afiliación partido-candidato, candidatos por partido y propuestas específicas. La flexibilidad de este



enfoque permite una recuperación y procesamiento de información y adaptable a diferentes tipos de preguntas electorales.

La arquitectura técnica integra múltiples componentes especializados, destacando el uso del modelo Mistral para generación de respuestas, configurado con parámetros específicos de temperatura y diversidad. El sistema incluye un respaldo multinivel que garantiza respuestas coherentes, endpoints en Flask para recuperación de información, y técnicas híbridas de búsqueda como TF-IDF y embeddings, optimizadas mediante FAISS.

La metodología Scrum ha sido fundamental para el desarrollo estructurado del proyecto, con dos sprints claramente definidos: el primero enfocado en recopilación, preprocesamiento del corpus textual y en la implementación del sistema de recuperación, y el segundo en la optimización del sistema de recuperación y generación de respuestas.

Las fuentes de información integradas incluyen planes de trabajo de candidatos, entrevistas transcritas y biografías de los candidatos, lo que permite una recuperación de información rica y contextualizada. El sistema ha superado desafíos como la integración de múltiples fuentes, transcripción de contenidos y generación de embeddings para representación vectorial.

Entre los logros más destacados se encuentran el desarrollo de un sistema de generación de respuestas naturales, la implementación de un sistema de clasificación de consultas, la creación de prompts dinámicos y el establecimiento de un sistema de respaldo para generación de respuestas.

Recomendaciones:

Es fundamental expandir el sistema de identificación de tipos de consultas más allá de las expresiones regulares actuales. Se recomienda desarrollar un modelo de machine learning que complemente la clasificación actual, permitiendo un procesamiento más dinámico y adaptable de consultas electorales. Esto implicaría entrenar el modelo con un conjunto ampliado de variaciones de preguntas, mejorando significativamente la capacidad de interpretación de consultas complejas y contextuales.

Un componente crítico para la mejora del sistema será desarrollar un módulo de retroalimentación robusto. Los usuarios podrán calificar la relevancia y precisión de las respuestas generadas, creando un ciclo de aprendizaje continuo. Esta implementación permitirá recopilar métricas de calidad en tiempo real, identificar patrones sistemáticos de error y realizar ajustes automáticos en los modelos de recuperación y generación.

Para aumentar la profundidad y contextualización de la información, se recomienda ampliar significativamente las fuentes de información. La integración de feeds de redes sociales de candidatos, comunicados de prensa oficiales, registros de debates y eventos

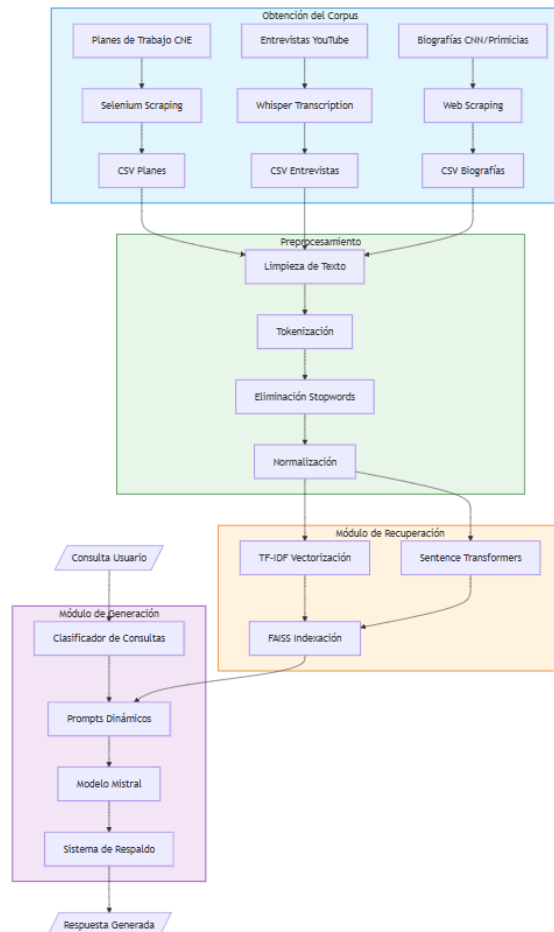


políticos, proporcionará una visión más completa y dinámica del panorama electoral. Un sistema de actualización automática garantizará la relevancia constante de la información. La interfaz debe incorporar herramientas de comparación entre candidatos que ofrezcan análisis multidimensionales. Se sugiere diseñar paneles interactivos que permitan comparar propuestas, trayectorias políticas y coherencia programática mediante gráficos, líneas de tiempo y visualizaciones de datos que faciliten la comprensión rápida y profunda de la información electoral.

En relación con las opciones de filtrado, se recomienda desarrollar un sistema avanzado y altamente personalizable. Los usuarios deberían poder explorar la información desde múltiples perspectivas, con opciones de segmentación por temas, fechas, candidatos o tipos de propuestas, manteniendo una navegación intuitiva y una estética moderna que incentive la exploración de la información.

5. Anexos

Diagrama de flujo del sistema RAG





El sistema RAG mostrado en el diagrama comienza con la obtención del corpus desde tres fuentes principales: planes de trabajo del CNE (usando Selenium Scraping), entrevistas de YouTube (procesadas con Whisper Transcription) y biografías de CNN/Primicias (mediante Web Scraping), generando archivos CSV para cada fuente. Estos datos pasan por un proceso de preprocesamiento que incluye limpieza de texto, tokenización, eliminación de stopwords y normalización.

Una vez procesados, el módulo de recuperación funciona a través de un proceso secuencial que comienza con el preprocesamiento y almacenamiento de textos y embeddings en un archivo ``sistema_busqueda.pkl``. Cuando recibe una consulta, primero aplica TF-IDF (implementado con ``sistema['vectorizer'].transform(query)``) para realizar un pre-filtrado inicial identificando palabras importantes, luego utiliza Sentence-Transformers con el modelo `'paraphrase-multilingual-mpnet-base-v2'` para convertir el texto en vectores densos de 768 dimensiones, capturando así el significado semántico.

Posteriormente, emplea FAISS (Facebook AI Similarity Search) mediante ``sistema['index'].search(query_embedding, k)`` para realizar una búsqueda eficiente de vectores similares, siguiendo el flujo: Query → TF-IDF pre-filtrado → Generar embedding → Búsqueda FAISS → Cálculo de relevancia → Ranking, lo que asegura que los documentos recuperados sean semánticamente relevantes para la consulta del usuario antes de pasarlos al módulo de generación, el cual procesa las consultas del usuario a través de un clasificador de consultas, prompts dinámicos, el modelo Mistral y un sistema de respaldo, produciendo finalmente una respuesta generada. Todo el flujo está diseñado para transformar eficientemente los datos crudos en respuestas coherentes y precisas para el usuario.