



UNIMINUTO
Corporación Universitaria Minuto de Dios
Educación de calidad al alcance de todos
Vigilada MinEducación

Proyecto

Proyecto. Utilizar las librerías de Python para tokenizar

Descripción: El estudiante elaborará un proyecto sobre los principios del procesamiento de lenguaje natural, por medio del uso de las librerías de Python y Anaconda realizará la actividad de tokenizar las frases previamente brindadas para el desarrollo práctico, con el objetivo de determinar los principios del procesamiento natural del lenguaje basados en IA. Esta actividad le permitirá desarrollar los resultados de aprendizaje para usar las librerías de NLP en el lenguaje Python.

Ver: Tenga en cuenta las siguientes preguntas orientadoras para desarrollar la actividad:



¿Cuáles son las librerías más utilizadas para el procesamiento de lenguaje natural?

¿Cómo es el proceso de tokenizar un texto y cómo influyen las Stop Words?

Juzgar: Consulte los siguientes recursos y con base en una lectura minuciosa, realice lo que se indica a continuación.

- Campesato, O. (2021). Working with Text: POS. En: *Natural language processing fundamentals for developers* [versión electrónica]. (pp. 69-74). Mercury Learning and information.
- Campesato, O. (2021). Chapter 4. Algorithms and Toolkits (I). En: *Natural language processing fundamentals for developers* [versión electrónica]. (pp. 123-160). Mercury Learning and information.
- Ganegedara, T. (2018). Chapter 2: Understanding TensorFlow. En: *Natural language processing with TensorFlow: Teach language to machines using python's deep learning library* [versión electrónica]. (pp. 27-65). Packt Publishing.
- Moreira, D., et al. (2021). Análisis del estado actual de procesamiento de lenguaje natural. *Revista ibérica de sistemas e tecnologias De informação, E.42*. 126 a 136.

- Srinivasa, B. (2018). Chapter 1. What is Text Analysis. En: *Natural language processing and computational linguistics: A practical guide to text analysis with python, gensim, spacy, and kera* [versión electrónica]. (pp. 9-20). Packt Publishing.

Tome apuntes o subraye las ideas principales de los documentos consultados, con el fin de sintetizar la información de forma precisa. Además, realice una investigación adicional en fuentes de información verídicas y confiables que le permita complementar la información.

Actuar: A partir de la consulta de los recursos básicos, complementarios y la infografía de la semana: *Librerías de Python PNL*, elabore un proyecto, teniendo en cuenta los siguientes aspectos:

1. Realice la actividad en la herramienta Google Colaboratory.

- A partir del *framework* llamado NLTK.ipynb, que se encuentra adjunto, complete los pasos de código que sean necesarios, de igual manera, si encuentra un código que se acondicione más a sus necesidades lo puede emplear.
- Comience con la instalación de NLTK.
- Posteriormente, elabore un método de lectura en una página html con el [resumen del libro "Cien años de soledad"](#).
- Para desarrollar este ejercicio, cree *tokens* para las palabras, quite los signos de puntuación; complete la función para verificar las secuencias de las palabras, elimine las *Stop Words* y visualice las palabras nuevamente; por último, cree los sinónimos de las palabras encontradas utilizando WordNet, utilice una palabra que encuentre en el texto para los sinónimos.
- Finalmente, incluya apoyos visuales que apoyen el contenido y ordénelos acorde a la información presentada.

2. Considere utilizar al menos cinco fuentes confiables para el desarrollo de su proyecto.

3. El trabajo deberá contener:

- **Portada.** Incluya el título del trabajo, su nombre completo, el nombre del curso y de la semana, el nombre del docente y fecha de entrega.
- **Índice.** Anexe un índice que permita visualizar los títulos y subtítulos del trabajo. Este deberá tener una secuencia ordenada y lógica de los temas.
- **Introducción.** Describa brevemente la problemática a abordar, el objetivo del proyecto, etc.
- **Planeación.** Presente una descripción detallada del plan de trabajo para la solución del problema, el producto a desarrollar o la producción del proyecto. Deberá incluir y especificar todas las actividades y procesos por realizar en el proyecto.
- **Desarrollo.** Presente la exposición de las problemáticas de manera ordenada, clara y lógica, incluya ideas, soluciones y resultados de principios del procesamiento de lenguaje natural.
- **Resultados.** Cierre su escrito presentando la resolución del problema, el producto desarrollado o la producción que realizó mediante el proyecto.
- **Conclusiones.** Redacte brevemente los beneficios que podría tener su solución al problema, producto o producción.
- **Referencias.** Incluya un apartado en el que enliste todas las fuentes que consultó para realizar la actividad.

4. Cuide su redacción y ortografía.

5. Recuerde que sus referencias y citas deben apearse a las normas APA.

Devolución creativa-metacognición: Las librerías como NLTK y Spacy, permiten desarrollar tokenización, lematización y etiquetado de las palabras. Esto hace que las tareas de procesamiento natural del lenguaje sean más sencillas, pero a la vez más claras para la máquina. Dichas tareas son la base del procesamiento de texto y, por tanto, la parte más importante, ya que si no se comprende a fondo el proceso gramatical es imposible llegar a resultados de buena calidad.

Tiempos y recursos educativos para el desarrollo de la actividad

Tipo de actividad	Individual
Desarrollo de la actividad	Desarrollo de la actividad: 8 horas Consulta de materiales: 4 horas
Recursos	Recurso(s) básico(s): • Campesato, O. (2021). Working with Text: POS. En: <i>Natural language processing fundamentals for developers</i> [versión electrónica]. (pp. 69-74). Mercury Learning and information. • Campesato, O. (2021). Chapter 4. Algorithms and Toolkits (I). En: <i>Natural language processing fundamentals for developers</i> [versión electrónica]. (pp. 123-160). Mercury Learning and information. • Ganegedara, T. (2018). Chapter 2: Understanding TensorFlow. En: <i>Natural language processing with TensorFlow: Teach language to machines using python's deep learning library</i> [versión electrónica]. (pp. 27-65). Packt Publishing. • Moreira, D., et al. (2021). Análisis del estado actual de procesamiento de lenguaje natural. <i>Revista ibérica de sistemas e tecnologias De informação, E.42</i>. 126 a 136. • Srinivasa, B. (2018). Chapter 1. What is Text Analysis. En: <i>Natural language processing and computational linguistics: A practical guide to text analysis with python, gensim, spacy, and kera</i> [versión electrónica]. (pp. 9-20). Packt Publishing.

Recurso(s) complementario(s):

- Arumugam, R., y Shanmugamani, R. (2018). Chapter 2. Text Classification and POS Tagging Using NLTK, En: *Hands-on natural language processing with python: A practical guide to applying deep learning architectures to your NLP applications*. (pp. 28-51). Packt Publishing.
- Echeverri, M., y Manjarres, R. (2020). Asistente Virtual Académico Utilizando Tecnologías Cognitivas De Procesamiento De Lenguaje Natural. *Revista Politécnica*, **16**(31), 85–96.
- Guerrero, F., y Marco, M. (2014). Parte II. Gramáticas de contexto libre, Parsing, unificación de rasgos y semántica y análisis semántico. Parte III. Discurso, extracción de información y resúmenes. En: *Procesamiento del Lenguaje Natural. Monografías de la revista Neoinstrumenta*. pp. 30-87.

Rol del tutor: A través de los canales de comunicación del curso, acompañará y guiará la elaboración del proyecto, orientando el proceso de intervención en la tarea de complementar el código, análisis de las funciones utilizadas y verificación de los recursos creados para el proyecto. Asimismo, aclarará inquietudes, ampliará conceptos, brindará retroalimentación y, finalmente, evaluará la actividad de acuerdo con los criterios establecidos en la rúbrica.

Forma de entrega:

- Guarde el documento en el formato correspondiente .docx., si su elección fue una herramienta digital, obtengan el enlace proporcionado por la herramienta que utilizó y ubíquelo en una hoja Word con sus datos.
- Nombre el archivo de la siguiente manera:
primerapellido_primernombre_tipodeactividad. Por ejemplo: romero_luis_proyecto.
- Dé clic en **Agregar entrega** para subir el trabajo a la plataforma.
- Arrastre y suelte el documento en **Archivos enviados**.
- Dé clic en el botón **Guardar cambios** para finalizar la entrega. De lo contrario, el archivo no se guardará y no se habrá entregado.

Consulte la [rúbrica](#) para conocer a detalle los criterios de evaluación.

Rúbrica. Proyecto

Criterios		No evaluable (0)	Insuficiente (1.9)	Regular (2.9)	Bueno (3.9)	Excelente (5)	Puntos
Apariencia, organización, ortografía y gramática		Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	- El número de páginas se aleja demasiado del criterio solicitado. - No utiliza títulos para organizar visualmente el material. - No contiene un índice o una tabla de contenidos. - Se observan ocho o más errores gramaticales y de ortografía. NO SE CONSIDERAN LOS ERRORES REPETIDOS.	- El número de páginas se aleja significativamente al criterio solicitado. - En algunas ocasiones utiliza títulos para organizar visualmente el material. - Presenta un índice o tabla de contenidos sin una secuencia ordenada y lógica de los temas. - Se observan de cinco a siete errores gramaticales y de ortografía. NO SE CONSIDERAN LOS ERRORES REPETIDOS.	- El número de páginas se acerca al criterio solicitado. - Utiliza títulos para organizar visualmente el material. - Presenta un índice o tabla de contenidos con una secuencia ordenada y lógica de los temas. - Se observan de uno a cuatro errores gramaticales y de ortografía. NO SE CONSIDERAN LOS ERRORES REPETIDOS.	- El número de páginas concuerda con el criterio solicitado. - Usa títulos y subtítulos para organizar visualmente el material. - Presenta un índice o tabla de contenidos con una secuencia ordenada y lógica de los temas. - No se observan errores gramaticales ni faltas de ortografía.	
	Puntaje	0	0.27	0.40	0.50	0.71	
Introducción		Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	- No describe el problema a resolver, el tipo de producto o la producción que realizará con el proyecto. - No indica las características de la población involucrada en el problema a resolver o el mercado al que irá dirigido el producto o la producción del proyecto. - No describe las razones que justifican la realización del proyecto, resaltando su importancia, implicaciones y /o beneficios. - No formula el objetivo o propósito de su proyecto, o no se comprende.	- Describe de forma incompleta el problema a resolver, el tipo de producto o la producción que realizará con el proyecto. - Indica algunas de las características de la población involucrada en el problema a resolver o el mercado al que irá dirigido el producto o la producción del proyecto. - Describe una o algunas de las razones que justifican la realización del proyecto, resaltando su importancia, implicaciones y /o beneficios. - Formula el objetivo o propósito de su proyecto, pero casi no se comprende.	- Describe de forma casi completa el problema a resolver, el tipo de producto o la producción que realizará con el proyecto. - Indica la mayoría de las características de la población involucrada en el problema a resolver o el mercado al que irá dirigido el producto o la producción del proyecto. - Describe la mayor parte de las razones que justifican la realización del proyecto, resaltando su importancia, implicaciones y /o beneficios. - Formula de forma comprensible el objetivo o propósito de su proyecto.	- Describe de forma amplia el problema a resolver, el tipo de producto o la producción que realizará con el proyecto. - Indica la totalidad de las características de la población involucrada en el problema a resolver o el mercado al que irá dirigido el producto o la producción del proyecto. - Describe la totalidad de las razones que justifican la realización del proyecto, resaltando su importancia, implicaciones y /o beneficios. - Formula claramente y de manera precisa el objetivo o propósito de su proyecto.	
	Puntaje	0	0.27	0.50	0.60	0.71	

Fondo	Planeación	Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	- No presenta una descripción del plan de trabajo para solucionar el problema, desarrollar el producto o la producción del proyecto; identifica las actividades, los tiempos de inicio y término y delimita responsables de cada actividad. - No se definen y calendarizan las actividades y procesos por realizar en el proyecto.	- Presenta una descripción muy general del plan de trabajo para la solución del problema, el producto a desarrollar o la producción del proyecto; identifica las actividades, los tiempos de inicio y término y delimita responsables de cada actividad. - Se definen y calendarizan algunas de las actividades y procesos por realizar en el proyecto.	- Presenta una descripción casi total del plan de trabajo para la solución del problema, el producto a desarrollar o la producción del proyecto; identifica las actividades, los tiempos de inicio y término y delimita responsables de cada actividad. - Se definen y calendarizan con precisión la mayor parte de las actividades y procesos por realizar en el proyecto.	- Presenta una descripción detallada del plan de trabajo para la solución del problema, el producto a desarrollar o la producción del proyecto; identifica las actividades, los tiempos de inicio y término y delimita responsables de cada actividad. - Se definen y calendarizan con precisión todas las actividades y procesos por realizar en el proyecto.	
	Puntaje	0	0.27	0.40	0.60	0.71	
	Congruencia del proyecto con el objetivo	Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	El desarrollo de las ideas en la redacción del proyecto no tiene una secuencia lógica y no se observa congruencia con el objetivo o propósito.	El desarrollo de algunas de las ideas en la redacción del proyecto tiene una excelente secuencia lógica y se observa cierta congruencia con el objetivo o propósito.	El desarrollo de la mayor parte de las ideas en la redacción del proyecto tiene una excelente secuencia lógica y se observa una buena congruencia con el objetivo o propósito.	El desarrollo de la totalidad de las ideas en la redacción del proyecto tiene una excelente secuencia lógica y se observa plena congruencia con el objetivo o propósito.	
		0	0.27	0.40	0.60	0.71	
	Diagramas, tablas, gráficas e ilustraciones*	Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	- No tiene diagramas o ilustraciones. Si presenta diagramas e ilustraciones no son: - Ordenados - Correctos - Favorecen el entendimiento del tema	- Algunos de los diagramas e ilustraciones son: - Ordenados - Correctos - Favorecen el entendimiento del tema	- La mayoría de los diagramas e ilustraciones son: - Ordenados - Correctos - Favorecen el entendimiento del tema	- Todos los diagramas e ilustraciones son: - Ordenados - Correctos - Favorecen el entendimiento del tema	
		0	0.27	0.40	0.60	0.71	
	Bibliografía	Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	No contiene fuentes o no son confiables.	Contiene al menos de una a tres fuentes confiables. Están citadas con menos de cuatro errores en estilo APA.	Contiene al menos cuatro fuentes confiables. Están citadas con menos de dos errores, en estilo APA.	Contiene al menos cinco fuentes confiables. Están citadas correctamente en estilo APA.	
		0	0.27	0.40	0.50	0.71	

Presentación del proyecto	Se detecta plagio total, se envió otra actividad o la actividad no cumple con los criterios mínimos solicitados para ser evaluada, la actividad está incompleta, no entregó.	- No muestra la forma en que resolvió el problema, el producto desarrollado o la producción que realizó mediante el proyecto. - No describe los beneficios que podría tener su solución al problema, producto o producción. - Su solución, producto o producción no es original, incluso podría ser una copia de otro. - No cumple con el objetivo o los propósitos para los que fue concebida la solución, el producto o la producción. - No solucionó el problema, su producto o producción no es relevante. - No soluciona el problema, el producto o la producción no son relevantes.	- Muestra de manera incompleta la forma en que resolvió el problema, el producto desarrollado o la producción que realizó mediante el proyecto. - Describe algunos de los beneficios que podría tener su solución al problema, producto o producción. - Su solución, producto o producción no es del todo original, le falta creatividad. - Cumple de forma parcial con el objetivo o los propósitos para los que fue concebida la solución, el producto o la producción. - La aportación a la solución del problema, la relevancia del producto o de la producción es aceptable.	- Muestra de manera completa la forma en que resolvió el problema, el producto desarrollado o la producción que realizó mediante el proyecto. - Describe la mayoría de los beneficios que podría tener su solución al problema, producto o producción. - Es original y creativo en su solución, producto o producción. - Cumple casi de forma total con el objetivo o propósitos para los que fue concebida la solución, el producto o la producción. - La aportación a la solución del problema, la relevancia del producto o de la producción es buena.	- Muestra de manera completa la forma en que resolvió el problema, el producto desarrollado o la producción que realizó mediante el proyecto. - Describe todos los beneficios que podría tener su solución al problema, producto o producción. - Es muy original y creativo en su solución, producto o producción. - Cumple con el objetivo o propósitos para los que fue concebida la solución, el producto o la producción. - Es notoriamente clara la aportación a la solución del problema, la relevancia del producto o de la producción.	
	0	0.28	0.40	0.50	0.74	
					Calificación total	