



Tecnológico de Costa Rica  
Escuela de Computación  
Recuperación de Información Textual  
Grupo 20

**Primer Proyecto Programado:**  
**Arañador Web**

Profesor:  
Aurelio Sanabria Rodríguez

Estudiantes:  
Rodolfo Cruz Vega - 2013235955  
Juan Pablo Martínez Brenes - 2016140404  
Luis Ortiz Rua - 2016117738  
Yerling Pérez Álvarez - 2018002082

**I Semestre 2022**

## **Tabla de Contenido**

<b>Motivación</b>	<b>3</b>
<b>Justificación de la implementación</b>	<b>4</b>
Descripción y justificación de las políticas elegidas para el modelo.	4
Uso del calendarizador, multihilo y partes del preprocesamiento.	4
<b>Análisis de resultados</b>	<b>5</b>
<b>Memes</b>	<b>6</b>
Yerling Pérez Álvarez	6
Rodolfo Cruz Vega	7
Luis Ortiz Rua	8
Juan Pablo Martínez Brenes	9
<b>Biografía</b>	<b>10</b>

## **Manejo de presupuesto, finanzas, aplicación de regla fiscal y FEES en las actas de las sesiones del Consejo Nacional de Rectores (CONARE).**

### **Motivación**

El Consejo Nacional de Rectores (CONARE) fue creado mediante “Convenio de Coordinación de la Educación Superior Universitaria Estatal en Costa Rica”, suscrito por las Instituciones de Educación Superior Universitaria Estatal el 4 de diciembre de 1974 y reformado por éstas el 20 de abril de 1982. En él se regulan aspectos de coordinación para el ejercicio conjunto de la autonomía universitaria en diversos ámbitos (CONARE, 2022).

Tiene entre sus funciones distribuir las rentas globales asignadas a la Educación Superior Universitaria Estatal y aprobar el presupuesto anual de la Oficina de Planificación de la Educación Superior.

En los últimos periodos de elecciones nacionales uno de los temas centrales ha sido el manejo de las finanzas públicas y el déficit fiscal así como las acciones que se han tomado para solventar estas situaciones. El CONARE al tener entre sus funciones actividades presupuestarias y ser un agente político que se involucra en varios temas nacionales como las finanzas del estado, el presupuesto de la educación (Zúñiga Rivero, 2022), negociación del FEES (Soto Sibaja, 2021), ley de empleo público (Jiménez Mata, 2019), entre otros; hace de interés el conocer que se discute en sus sesiones sobre presupuestos, aplicación de regla fiscal (Gómez Aguilar, 2021), y negociación del FEES.

Este tema es de alta importancia ya que el manejo de las finanzas en relación con la educación y en específico con la educación superior, por lo que se decidió basar el arañador web alrededor de este.

## **Justificación de la implementación**

### **Descripción y justificación de las políticas elegidas para el modelo.**

Debido a la cantidad histórica de actas haciendo uso de hilos es posible descargar y pre procesar los documentos de forma más rápida, posterior a la primera captura de las actas históricas sólo es necesario ejecutar la descarga una vez cada 30 días para obtener las actas nuevas del mes en curso.

Para evitar que los archivos previamente descargados se descarguen de nuevo cada 30 días, se utiliza un archivo txt donde se guardan los links que ya se visitaron y ya fueron descargados, el contenido de este archivo es leído y almacenado en una lista para un manejo más fácil. Al realizar una nueva ejecución dentro de 30 días se descargan y procesan sólo aquellos documentos que no están en el archivo de visitados.

### **Uso del calendarizador, multihilo y partes del preprocesamiento.**

El CONARE sesiona entre 30 y 40 veces por año, por que se hace innecesario una recurrencia alta de descargas de las actas. Por lo tanto, después de recolectar la información de los documentos ya existentes se considera que solo es necesario obtener los documentos nuevos una vez al mes para que de esta manera se puedan obtener los nuevos documentos del mes.

Se utilizó hilos para la descarga de los documentos, ya que al momento existen un historial de varios años de actas, por tanto descargarlas de forma secuencial es un proceso lento que tardaría demasiado tiempo y al hacer uso de hilos el proceso se vuelve mucho más rápido y efectivo.

Para el preprocesamiento se eliminaron las mayúsculas, las tildes, emojis, los números, signos de puntuación, saltos de línea, urls (ya que estos no aportan a la búsqueda) y se aplicó stopwords y lematización. Finalmente, se almacenan en un nuevo documento en formato txt con el texto procesado, como una lista, para posteriormente en las demás fases poder manejarlo de manera más simple.

## **Análisis de resultados**

Una vez implementado el arañador web, y puesto en funcionamiento, se obtuvieron correctamente todas las actas disponibles en la biblioteca de recursos disponible del CONARE hasta el periodo actual. Corresponde a un registro de 432 actas que se han acumulado desde el año 1974 hasta la fecha (omitiendo algunos años), el número de actas procesadas será actualizado cada mes resultado de la ejecución del calendarizador según se expuso anteriormente.

Se logró que éstas fueran pre procesadas correctamente con los métodos que se mencionaron anteriormente y que fueran guardadas como la nueva colección el formato txt. Adicionalmente, analizando los textos preprocesados, se encontró que 198 de los textos poseen la potencial palabra de búsqueda 'FEES' al menos una vez, durando un promedio de 0.16 segundos realizando la búsqueda, utilizando el comando (1) adjunto a continuación, mientras que la palabra 'Fiscal' tiene 145 ocurrencias, con una duración promedio de 0.24 segundos.

(1) `grep -lr fees ./preprocessed_files`

El preprocesamiento de texto permite que las búsquedas más sean ágiles y acertadas, evitando que, en este caso, el comando grep tenga que hacer comparaciones extra como lo son las comparaciones para que las búsquedas sean insensibles a las mayúsculas. Añadiendo el flag -i a la consulta, comando (2), el tiempo de ejecución de la consulta sube alrededor de 0.04 segundos por consulta producto de las comparaciones extra mencionadas.

(2) `grep -lr -i fees ./preprocessed_files`

El análisis anterior resalta la importancia del preprocesamiento de texto para optimizar futuras búsquedas, un sólo método puede optimizar una búsqueda hasta un 20% y considerando que el número de registros crece exponencialmente, este porcentaje es cada vez más relevante. El conjunto de métodos de preprocesamiento aplicados es tan importante como la extracción de registros en la disciplina de la recuperación de información textual.

## Memes

Yerling Pérez Álvarez



Yo al inicio del semestre bien motivado  
|| Yo ahorita



## Rodolfo Cruz Vega



Luis Ortiz Rua

Puedes recomendarme un buen  
Libro de programación  
Para principiantes?





Juan Pablo Martínez Brenes



## Biografía

[1] NLTK. (2022, Mar 25). NLTK::Natural Language Toolkit. [Online]. Available:

<https://www.nltk.org/>

[2] Python. Threading. [Online]. Available: <https://docs.python.org/3/library/threading.html>

[3] Python. Requests: HTTP for Humans. [Online]. Available:

<https://docs.python-requests.org/en/latest/>

[4] Beautiful Soup. Beautiful Soup Documentation. [Online]. Available:

<https://beautiful-soup-4.readthedocs.io/en/latest/>

[5] Soto Sibaja, J. (2021, Dec 07). Nuevo presidente de CONARE aspira a que negociación del FEES sea cada cinco años. [Online]. Available:

<https://www.monumental.co.cr/2021/12/07/nuevo-presidente-de-conare-aspira-a-que-negociacion-del-fees-sea-cada-cinco-anos/>

[6] Zúñiga Rivero , A. (2022, Mar 02). Figueres: “Me gusta la propuesta de negociar un FEES a cinco años plazo”. [Online]. Available:

<https://semanariouniversidad.com/universitarias/figueres-me-gusta-la-propuesta-de-negociar-un-fees-a-cinco-anos-plazo/>

[7] CONARE. Actas por año. [Online]. Available:

<https://biblioteca.conare.ac.cr/index.php/recursos-electronicos/actas-conare/actas-por-ano.html>

[8] Hoy en el TEC. (2019, Sept 26). Reducción de presupuesto destinado al FEES podría ser inconstitucional. [Online]. Available:

<https://www.tec.ac.cr/hoyeneltec/2019/09/26/reduccion-presupuesto-destinado-fees-podria-ser-inconstitucional>

[9] CrHoy.com. (2021, Sept 27). Fuerte recorte en presupuesto MEP contrasta con aumento de ¢35.000 millones para U públicas. [Online]. Available:

<https://www.crhoy.com/nacionales/fuerte-recorte-en-presupuesto-mep-contrasta-con-aumento-de-35-000-millones-para-u-publicas/>

[10] Seminario Universidad. (2022, Jan 117). CeNAT-Conare otorgará becas de hasta ₡3 millones para trabajos de investigación de estudiantes universitarios. Available: <https://semanariouniversidad.com/universitarias/cenat-conare-otorgara-becas-de-hasta-%E2%82%A13-millones-para-trabajos-de-investigacion-de-estudiantes-universitarios/#:~:text=%E2%80%9CEs%20una%20beca%20muy%20rica,equipo%20de%20profesionales%20del%20CeNAT.>

[11] Seminario Universidad. (2022, Feb 25). Conare rechaza afirmación de Chaves de que quitaron dinero al MEP para darlo a Universidades. [Online]. Available: <https://semanariouniversidad.com/universitarias/conare-rechaza-afirmacion-de-chaves-de-que-quitaron-dinero-al-mep-para-darlo-a-universidades/>

[12] Seminario universidad. (2021). Consejo Institucional solicita excluir al TEC de la Regla Fiscal. [Online]. Available: <https://www.tec.ac.cr/hoyeneltec/2021/12/16/consejo-institucional-solicita-excluir-tec-regla-fiscal#:~:text=Si%20no%20excluye%2C%20al%20menos,tecnolog%C3%ADa%20impulsada%20por%20el%20TEC.>