REPORTE TAREA 2 y 3
ALGORITMOS Y COMPLEJIDAD

«Explorando la Distancia entre Cadenas, una Operación a la Vez»

Alejandro Vergara

13 de noviembre de 2024

16:31

Resumen

Un resumen es un breve compendio que sintetiza todas las secciones clave de un trabajo de investigación: la introducción, los objetivos, la infraestructura y métodos, los resultados y la conclusión. Su objetivo es ofrecer una visión general del estudio, destacando la novedad o relevancia del mismo, y en algunos casos, plantear preguntas para futuras investigaciones. El resumen debe cubrir todos los aspectos importantes del estudio para que el lector pueda decidir rápidamente si el artículo es de su interés.

En términos simples, el resumen es como el menú de un restaurante que ofrece una descripción general de todos los platos disponibles. Al leerlo, el lector puede hacerse una idea de lo que el trabajo de investigación tiene para ofrecer [1].

La extensión del resumen, para esta entrega, debe ser tal que la totalidad del índice siga apareciendo en la primera página. Recuerde que NO puede modificar el tamaño de letra, interlineado, márgenes, etc.

Índice

1. Introducción

La extensión máxima para esta sección es de 2 páginas.

La introducción de este tipo de informes o reportes, tiene como objetivo principal **contextualizar el problema que se va a analizar**, proporcionando al lector la información necesaria para entender la relevancia del mismo.

Es fundamental que en esta sección se presenten los antecedentes del problema, destacando investigaciones previas o principios teóricos que sirvan como base para los análisis posteriores. Además, deben explicarse los objetivos del informe, que pueden incluir la evaluación de un algoritmo, la comparación de métodos o la validación de resultados experimentales.

Aunque la estructura y el enfoque siguen principios de trabajos académicos, se debe recordar que estos informes no son publicaciones científicas formales, sino trabajos de pregrado. Por lo tanto, se busca un enfoque claro y directo, que permita al lector comprender la naturaleza del problema y los objetivos del análisis, sin entrar en detalles excesivos.

Introduction Checklist de *How to Write a Good Scientific Paper* [7], adaptada a nuestro contexto:

- Indique el **campo del trabajo** (Análisis y Diseño de algoritmos en Ciencias de la Computación), por qué este campo es importante y qué se ha hecho ya en este área, con las **citas** adecuadas de la literatura académica o fuentes relevantes.
- Identifique una **brecha** en el conocimiento, un desafío práctico, o plantee una **pregunta** relacionada con la eficiencia, complejidad o aplicabilidad de un algoritmo particular.
- Resuma el propósito del informe e introduzca el análisis o experimento, dejando claro qué se está investigando o comparando, e indique **qué es novedoso** o por qué es significativo en el contexto de un curso de pregrado.
- Evite; repetir el resumen; proporcionar información innecesaria o fuera del alcance de la materia (limítese al análisis de algoritmos o conceptos de complejidad); exagerar la importancia del trabajo (recuerde que se trata de un informe de pregrado); afirmar novedad sin una comparación adecuada con lo enseñado en clase o la bibliografía recomendada.

Recuerde que este es su trabajo, y sólo usted puede expresar con precisión lo que ha aprendido y quiere transmitir. Si lo hace bien, su introducción será más significativa y valiosa que cualquier texto automatizado. ¡Confíe en sus habilidades, y verá que puede hacer un mejor trabajo que cualquier herramienta que automatiza la generación de texto!

2. Diseño y Análisis de Algoritmos

Para los dos paradigmas de programación utilizados, se utilizo de estrategia la idea de calcular todos los costos posibles recursivamente, para finalmente elegir el minimo, es decir, se tendra una llamada recursiva cuando la funcion sea sustituir, insertar, eliminar o transponer, donde cada una obtendra su respectivo coste, eligiendo asi el minimo entre estos.
Las diferencia con el problema de minima distancia entre cadenas es que ahora se podra ejecutar con distintos costos, y tambien se le agrega la funcion de transponer, donde a esta solo se le considero intercambiar con el caracter que tiene adelante solo si los dos caracteres resultantes quedan en la posicion que deberian segun la cadena 2,
por ejemplo:

palabra: ab

objetivo: ba

en este casi si se podra ejecutar una transposición,
sin embargo si se tiene algo del tipo:

palabra: acb

objetivo: ba

no se podra resolver con transposiciones, ya que primero necesitamos eliminar 'c' para que luego sea factible la transposición, sin embargo esto no estara soportado por los algoritmos
Se tomaron otros supuestos, los cuales son, si los dos caracteres a comparar son iguales, no se comparara con ninguna funcion (sustituir, insertar, eliminar o transponer), ya que ya son iguales se considero que el costo sera 0.

La entrada consiste en dos cadenas de caracteres, la primera sera la que buscaremos transformar a la segunda, estas dos las etiquetaremos, la primera sera simplemente **palabra**, y la segunda sera **objetivo**, estas dos serán guardadas como variables globales para las dos implementaciones, también los costos de cada funcion estaran designados de forma matricial/vectorial que especificara cada costo especifico dependiendo de cual/es caracteres esten implicados.

La extensión máxima para esta sección es de 5 páginas.

Diseñar un algoritmo por cada técnica de diseño de algoritmos mencionada en la sección de objetivos. Cada algoritmo debe resolver el problema de distancia mínima de edición extendida, dadas dos cadenas `S1` y `S2`, utilizando las operaciones y costos especificados.

- Describir la solución diseñada.
- Incluir pseudocódigo (ver ejemplo algoritmo 1)
- Proporciones un ejemplo paso a paso de la ejecución de sus algoritmos que ilustren cómo sus algoritmos manejan diferentes escenarios, particularmente donde las transposiciones o los costos

variables afectan el resultado. Haga referencias a los programas expresados en psudocódigo (además puede hacer diagramas).

- Analizar la Complejidad temporal y espacial de los algoritmos diseñados en términos de las longitudes de las cadenas de entrada $S1$ y $S2$
- Discute cómo la inclusión de transposiciones y costos variables impacta la complejidad.

Los pseudocódigos los he diseñado utilizando el paquete *Algorithm2e documentation* [3] para la presentación de algoritmos. Se recomienda consultar *Algorithm2e on CTAN* [4] y *Writing Algorithms in LaTeX* [8].

Todo lo correspondiente a esta sección es, digamos, en "lapiz y papel", en el sentido de que no necesita de implementaciones ni resultados experimentales.

Recuerde que lo importante es diseñar algoritmos que cumplan con los paradigmas especificados.

Si se utiliza algún código, idea, o contenido extraído de otra fuente, este **debe** ser citado en el lugar exacto donde se utilice, en lugar de mencionarlo al final del informe.

2.1. Fuerza Bruta

> *"Indeed, brute force is a perfectly good technique in many cases; the real question is, can we use brute force in such a way that we avoid the worst-case behavior?"*
>
> — Knuth, 1998 [6]

Algoritmo 1: Distancia Minima de Edición - Fuerza Bruta

```
1  String palabra, objetivo
2  Procedure DistanciaEdicionFB(i, j)
3      if palabra vacia (i <0) and objetivo vacio (j <0) then
4          return 0
5      if palabra vacia (i <0) then
6          return (costo insertar objetivo_j) + DistanciaEdicionFB(i, j-1)
7      if objetivo vacio (j <0) then
8          return (costo eliminar palabra_i) + DistanciaEdicionFB(i-1, j)
9      if palabra[i] == objetivo[j] then
10         return DistanciaEdicionFB(i-1, j-1)
11     eliminar ← DistanciaEdicionFB(i-1, j) + (costo eliminar palabra_i)
12     insertar ← DistanciaEdicionFB(i, j-1) + (costo insertar objetivo_j)
13     sustituir ← DistanciaEdicionFB(i-1, j-1) + (costo sustituir palabra_i por objetivo_j)
14     transponer ← ∞
15     if i y j son >0 and palabra_i == objetivo_{j-1} and palabra_{i-1} == objetivo_j then
16         transponer ← DistanciaEdicionFB(i-2, j-2) + (costo transponer palabra_i por palabra_{i-1})
17     return (minimo entre eliminar, insertar, sustituir, transponer)
```

2.2. Programación Dinámica

Dynamic programming is not about filling in tables. It's about smart recursion!

Erickson, 2019 [2]

1 Describa la solución recursiva.

2 Escriba la relación de recurrencia, incluyendo condiciones y casos base.

3 Identifique subproblemas.

4 Defina estructura de datos a utilizar y especifique el orden de calculo que realiza su programa que utiliza programación dinámica.

Para este enfoque se utilizo una

Algoritmo 2: Distancia Minima de Edición - Programación Dinamica

```
1  String palabra, objetivo
2  Matriz tabla
3  Procedure DistanciaEdicionPD(i, j)
4      if tabla_ij != -1 then
5          return tabla_ij
6      if palabra vacia (i == 0) then
7          tabla_ij ← (costo insertar objetivo_j) + DistanciaEdicionPD(i, j-1)
8          return tabla_ij
9      if objetivo vacio (j == 0) then
10         tabla_ij ← (costo eliminar palabra_i) + DistanciaEdicionPD(i-1, j)
11         return tabla_ij
12     if palabra[i] == objetivo[j] then
13         return DistanciaEdicionPD(i-1, j-1)
14     eliminar ← DistanciaEdicionPD(i-1, j) + (costo eliminar palabra_i)
15     insertar ← DistanciaEdicionPD(i, j-1) + (costo insertar objetivo_j)
16     sustituir ← DistanciaEdicionPD(i-1, j-1) + (costo sustituir palabra_i por objetivo_j)
17     transponer ← ∞
18     if i y j son >0 and palabra_i == objetivo_{j-1} and palabra_{i-1} == objetivo_j then
19         transponer ← DistanciaEdicionPD(i-2, j-2) + (costo transponer palabra_i por palabra_{i-1})
20     tabla_ij ← (minimo entre eliminar, insertar, sustituir, transponer)
21     return tabla_ij
22 Procedure set()
23     ...
24     palabra ← " – " + leertxt
25     objetivo ← " – " + leertxt
26     setear tabla con valores -1
27     tabla[0][0] ← 0
28     ...
```

3. Implementaciones

La extensión máxima para esta sección es de 1 página.

Aquí deben explicar la estructura de sus programas haciendo referencias a los archivos y funciones de su entrega. No adjunte código en esta sección.

Si se utiliza algún código, idea, o contenido extraído de otra fuente, este **debe** ser citado en el lugar exacto donde se utilice, en lugar de mencionarlo al final del informe.

4. Experimentos

La extensión máxima para esta sección es de 6 página.

> *"Non-reproducible single occurrences are of no significance to science."*
>
> —Popper,2005 [9]

En la sección de Experimentos, es fundamental detallar la infraestructura utilizada para asegurar la reproducibilidad de los resultados, un principio clave en cualquier experimento científico. Esto implica especificar tanto el hardware (por ejemplo, procesador Intel Core i7-9700K, 3.6 GHz, 16 GB RAM DDR4, almacenamiento SSD NVMe) como el entorno software (sistema operativo Ubuntu 20.04 LTS, compilador g++ 9.3.0, y cualquier librería relevante). Además, se debe incluir una descripción clara de las condiciones de entrada, los parámetros utilizados y los resultados obtenidos, tales como tiempos de ejecución y consumo de memoria, que permitan a otros replicar los experimentos en entornos similares. *La replicabilidad es un aspecto crítico para validar los resultados en la investigación científica computacional* [5].

4.1. Dataset (casos de prueba)

La extensión máxima para esta sección es de 2 páginas.

Es importante generar varias muestras con características similares para una misma entrada, por ejemplo, variando tamaño del input dentro de lo que les permita la infraestructura utilizada en ests informe, con el fin de capturar una mayor diversidad de casos y obtener un análisis más completo del rendimiento de los algoritmos.

Aunque la implementación de los algoritmos debe ser realizada en C++, se recomienda aprovechar otros lenguajes como Python para automatizar la generación de casos de prueba, ya que es más amigable para crear gráficos y realizar análisis de los resultados. Python, con sus bibliotecas como `matplotlib` o `pandas`, facilita la visualización de los datos obtenidos de las ejecuciones de los distintos algoritmo bajo diferentes escenarios.

Debido a la naturaleza de las pruebas en un entorno computacional, los tiempos de ejecución pueden variar significativamente dependiendo de factores externos, como la carga del sistema en el momento de la ejecución. Por lo tanto, para obtener una medida más representativa, siempre es recomendable ejecutar múltiples pruebas con las mismas características de entrada y calcular el promedio de los resultados.

4.2. Resultados

La extensión máxima para esta sección es de 4 páginas.

En esta sección, los resultados obtenidos, como las gráficas o tablas, deben estar respaldados por los datos generados durante la ejecución de sus programas. Es fundamental que, junto con el informe, se adjunten los archivos que contienen dichos datos para permitir su verificación. Además, se debe permitir y especficiar como obtener esos archivos desde una ejecución en otro computador (otra infraestructura para hacer lso experimentos).

No es necesario automatizar la generación de las gráficas, pero sí es imprescindible que se pueda confirmar que las visualizaciones presentadas son producto de los datos generados por sus algoritmos, aunque la trazabilidad de los datos hasta las visualizaciones es esencial para garantizar que su validez: describa cómo se generaron los datos, cómo se procesaron y cómo se visualizaron de manera que pueda ser replicado por quien lea su informe.

Agregue gráficas que muestren los resultados de sus experimentos. La cantidad de páginas es limitada, por lo tanto escoja las gráficas más representativas y que muestren de manera clara los resultados obtenidos. Esta elección es parte de lo que se evaluara en la sección de presentación de resultados. Referencie las figuras en el texto, describa lo que se observa en ellas y por qué son relevantes.

En la fig. 1 se muestra un scatterplot hecho con TikZ con el tamaño ideal cuando se incluyen dos figuras. Queda a criterio de usted el decidir qué figuras incluir.

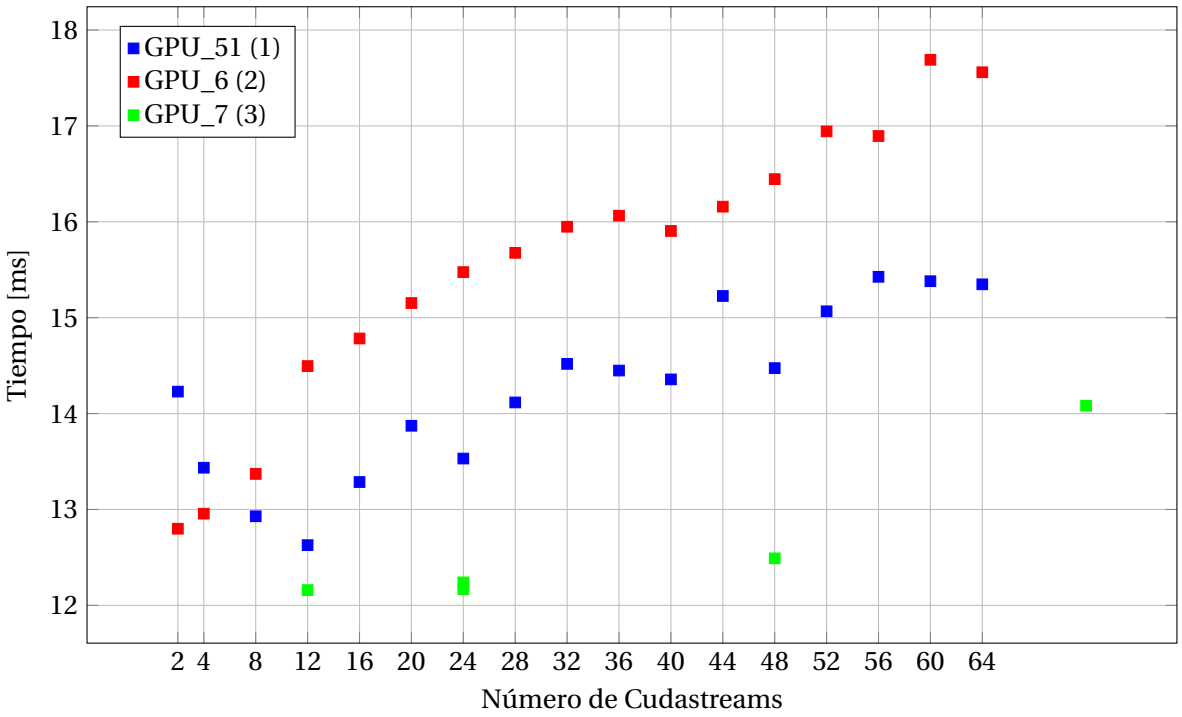

Figura 1: Ejemplo de scatterplot hecho con tikz. Tamaño ideal 1.

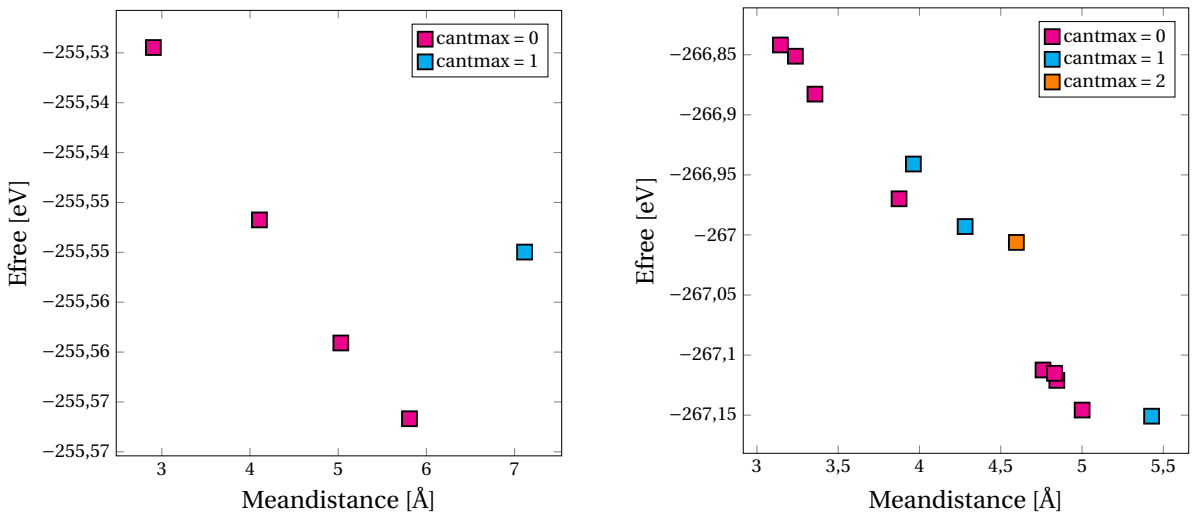

Figura 2: Ejemplo de scatterplot hecho con tikz. Tamaño ideal 2.

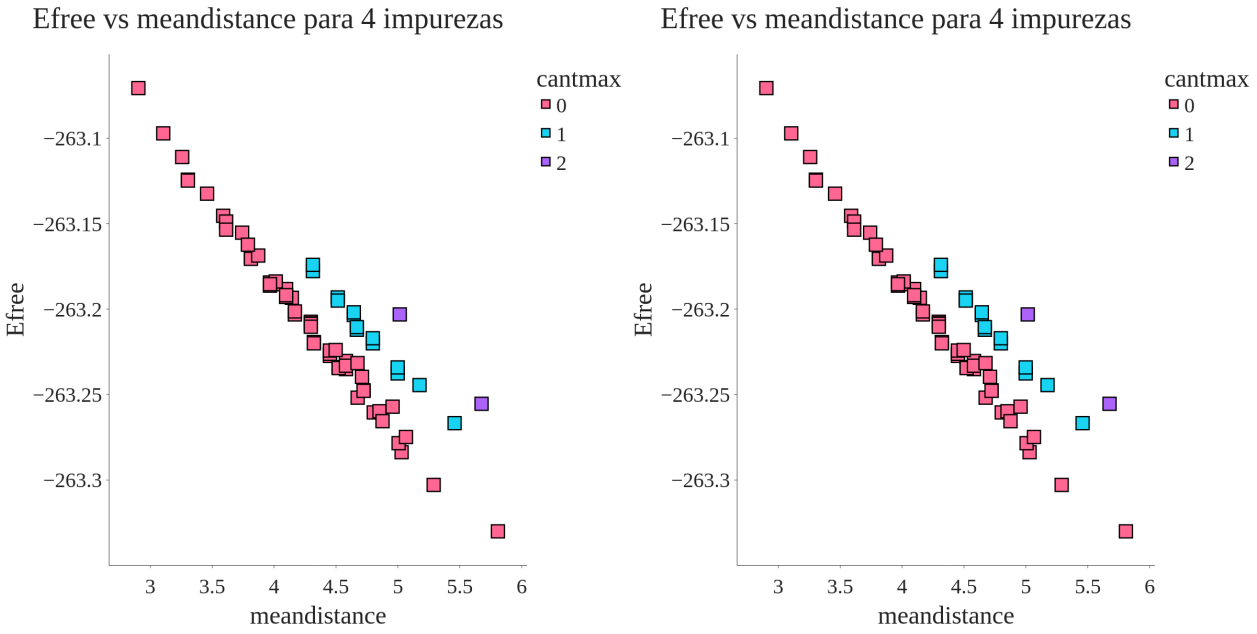

Figura 3: Ejemplo de scatterplot hecho con matplotlib.

Recuerde que es imprescindible que se pueda replicar la generación de las gráficas, por lo que usted debe incluir cómo generó esos datos y cómo podría generarlos la persona que revisa su entrega y ejecuta sus programas. Por ejemplo, si genera un scatterpolot con Tikz, usted debe explicar cómo obtener la tupla de valores que se usaron para generar la gráfica.

5. Conclusiones

La extensión máxima para esta sección es de 1 página.

La conclusión de su informe debe enfocarse en el resultado más importante de su trabajo. No se trata de repetir los puntos ya mencionados en el cuerpo del informe, sino de interpretar sus hallazgos desde un nivel más abstracto. En lugar de describir nuevamente lo que hizo, muestre cómo sus resultados responden a la necesidad planteada en la introducción.

- No vuelva a describir lo que ya explicó en el desarrollo del informe. En cambio, interprete sus resultados a un nivel superior, mostrando su relevancia y significado.
- Aunque no debe repetir la introducción, la conclusión debe mostrar hasta qué punto logró abordar el problema o necesidad planteada en el inicio. Reflexione sobre el éxito de su análisis o experimento en relación con los objetivos propuestos.
- No es necesario restablecer todo lo que hizo (ya lo ha explicado en las secciones anteriores). En su lugar, centre la conclusión en lo que significan sus resultados y cómo contribuyen al entendimiento del problema o tema abordado.
- No deben centrarse en sí mismos o en lo que hicieron durante el trabajo (por ejemplo, evitando frases como "primero hicimos esto, luego esto otro...").
- Lo más importante es que no se incluyan conclusiones que no se deriven directamente de los resultados obtenidos. Cada afirmación en la conclusión debe estar respaldada por el análisis o los datos presentados. Se debe evitar extraer conclusiones generales o excesivamente amplias que no puedan justificarse con los experimentos realizados.

6. Condiciones de entrega

- La tarea se realizará individualmente (esto es grupos de una persona), sin excepciones.
- La entrega debe realizarse vía `http://aula.usm.cl` en un tarball en el área designada al efecto, en el formato `tarea-2 y 3-rol.tar.gz` (`rol` con dígito verificador y sin guión).

  Dicho tarball debe contener las fuentes en LaTeX 2ε (al menos `tarea-2 y 3.tex`) de la parte escrita de su entrega, además de un archivo `tarea-2 y 3.pdf`, correspondiente a la compilación de esas fuentes.
- Si se utiliza algún código, idea, o contenido extraído de otra fuente, este **debe** ser citado en el lugar exacto donde se utilice, en lugar de mencionarlo al final del informe.
- Asegúrese que todas sus entregas tengan sus datos completos: número de la tarea, ramo, semestre, nombre y rol. Puede incluirlas como comentarios en sus fuentes LaTeX (en TeX comentarios son desde % hasta el final de la línea) o en posibles programas. Anótese como autor de los textos.
- Si usa material adicional al discutido en clases, detállelo. Agregue información suficiente para ubicar ese material (en caso de no tratarse de discusiones con compañeros de curso u otras personas).
- No modifique `preamble.tex`, `tarea_main.tex`, `condiciones.tex`, estructura de directorios, nombres de archivos, configuración del documento, etc. Sólo agregue texto, imágenes, tablas, código, etc. En el códigos funte de su informe, no agregue paquetes, ni archivos .tex (a excepción de que agregue archivos en `/tikz`, donde puede agregar archivos .tex con las fuentes de gráficos en `TikZ`).
- La fecha límite de entrega es el día 10 de noviembre de 2024.

NO SE ACEPTARÁN TAREAS FUERA DE PLAZO.

- Nos reservamos el derecho de llamar a interrogación sobre algunas de las tareas entregadas. En tal caso, la nota de la tarea será la obtenida en la interrogación.

NO PRESENTARSE A UN LLAMADO A INTERROGACIÓN SIN JUSTIFICACIÓN PREVIA SIGNIFICA AUTOMÁTICAMENTE NOTA 0.

A. Apéndice 1

Aquí puede agregar tablas, figuras u otro material que no se incluyó en el cuerpo principal del documento, ya que no constituyen elementos centrales de la tarea. Si desea agregar material adicional que apoye o complemente el análisis realizado, puede hacerlo en esta sección.

> Esta sección es solo para material adicional. El contenido aquí no será evaluado directamente, pero puede ser útil si incluye material que será referenciado en el cuerpo del documento. Por lo tanto, asegúrese de que cualquier elemento incluido esté correctamente referenciado y justificado en el informe principal.

Referencias

[1] Elsevier. *Differentiating between an introduction and abstract in a research paper*. Accessed: 2024-10-02. 2024. URL: https://scientific-publishing.webshop.elsevier.com/manuscript-preparation/differentiating-between-and-introduction-research-paper/.

[2] Jeff Erickson. *Algorithms*. Jun. de 2019. ISBN: 978-1-792-64483-2.

[3] Christophe Fiorio. *Algorithm2e documentation*. http://ctan.math.illinois.edu/macros/latex/contrib/algorithm2e/doc/algorithm2e.pdf. 2023.

[4] Christophe Fiorio. *Algorithm2e on CTAN*. https://ctan.org/pkg/algorithm2e. 2023.

[5] Jorge Fonseca y Kazem Taghva. «The State of Reproducible Research in Computer Science». En: ene. de 2020, págs. 519-524. ISBN: 978-3-030-43019-1. DOI: 10.1007/978-3-030-43020-7_68.

[6] Donald E. Knuth. *The art of computer programming, volume 3: (2nd ed.) sorting and searching*. USA: Addison Wesley Longman Publishing Co., Inc., 1998. ISBN: 0201896850.

[7] Chris Mack y Lola Muxamedova. *How to Write a Good Scientific Paper*. Nov. de 2019.

[8] Overleaf. *Writing Algorithms in LaTeX*. https://www.overleaf.com/learn/latex/Algorithms. 2023.

[9] K. Popper. *The Logic of Scientific Discovery*. Routledge Classics. Taylor & Francis, 2005. ISBN: 9781134470020. URL: https://books.google.cl/books?id=LWSBAgAAQBAJ.