

Universidad de Costa Rica

Escuela de Ciencias de la Computación e Informática

Examen 2

Respuestas Teóricas

---

### **jaccard\_omp**

**- Explique cada una de las directivas de OpenMP que utilizó y cómo funcionan dentro del programa**

```
#pragma omp parallel for
```

Lo anterior especifica que es un ciclo for y que debe dividir sus iteraciones en hilos para que corran en paralelo.

```
default(none) private(i) shared(matrix_row,test_vector)
```

El uso de default(none) obliga a especificar las variables que serán utilizadas como memoria compartida, en este caso serán matrix\_row que tiene que ser compartida para que cada hilo en el for pueda comparar si ya llegó a su última iteración o no. Además debe usar esa variable para hacer las operaciones lógicas en los ifs del ciclo. El i es privado porque cada hilo trabaja con su propia iteración i y no deben tener el mismo varos hilos.

```
reduction(+: words_union, words_intersection) \  
    schedule(static,1)
```

Se utiliza un reduction para que se divida entre los hilos y el schedule es estático y de forma que le toque una iteración a cada hilo. Se usa una reducción en

words\_union y words\_intersection ya que son los datos que están editando en el ciclo con una suma.

**- Explique de forma general cómo se está dando la repartición de tareas entre hilos en su solución**

Al usar el schedule de forma estática, la división de tareas es de forma que se asigna antes de la ejecución del programa para que a cada hilo le toque un chunk de iteraciones de tamaño 1.

---

**jaccard\_mpi**

**- ¿Cuál o cuáles funciones de MPI consideró necesarias para la solución de este problema y explique cómo funciona(n) en su código?**

```
MPI_Allreduce(&max_jaccard_similarity_local, &max_jaccard_similarity_global, 1, MPI_DOUBLE, MPI_MAX, MPI_COMM_WORLD);
```

Se utilizó únicamente la función ALLREDUCE dentro del código paralelo, de forma que cada proceso calcula una similaridad de jaccard local, y la guarda en una variable la cuál envía para la reducción y el resultado se guarda en max\_jaccard\_similarity\_global, el cual se imprime al final del programa.

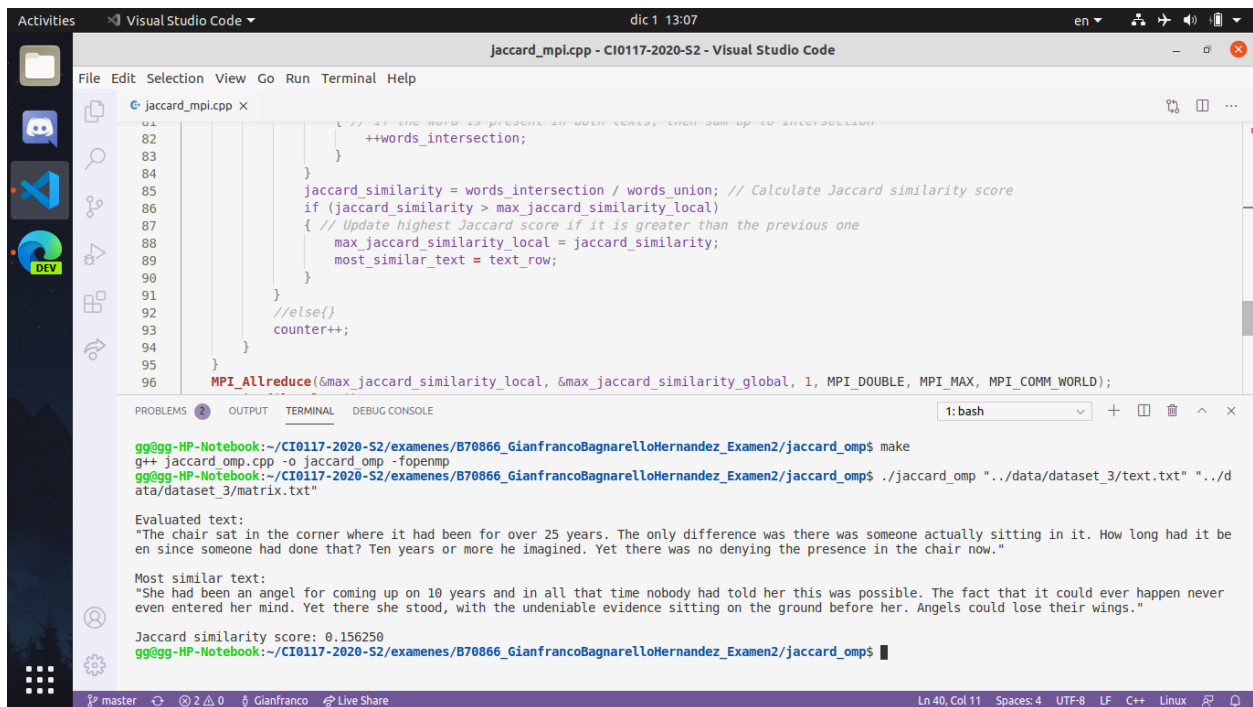
**- Explique cómo realizó la repartición de tareas entre los procesos.**

Las tareas se reparten de forma que se tiene un contador que lleva el control de los procesos. Los procesos se rotan de forma circular, usando su identificador "myid" . Si tengo N procesos, a cada proceso le toca una línea, y voy incrementando el contador. Cuando dicho contador llega a N, lo devuelvo a 0, deja que los procesos corran de manera concurrente del 0 a N. Cuando llega a N, repite el mismo proceso, mientras hayan líneas.

- Documente el resultado obtenido luego de correr el programa con cada uno de los conjuntos de datos propuestos y explique si tuvo que realizar alguna modificación a su programa. (3 pts)

## jaccard\_omp

### Conjunto de datos 3 (propuesto, 50 párrafos sin empate)



The screenshot shows the Visual Studio Code editor with the file `jaccard_mpi.cpp` open. The code is a C++ program that calculates the Jaccard similarity score between a given text and a dataset. It uses MPI for parallel processing. The terminal output shows the execution of the program, including the compilation of `jaccard_omp.cpp` and the execution of `jaccard_omp` with the dataset `dataset_3/text.txt`. The output displays the evaluated text, the most similar text, and the Jaccard similarity score: 0.156250.

```
82         ++words_intersection;
83     }
84 }
85 jaccard_similarity = words_intersection / words_union; // Calculate Jaccard similarity score
86 if (jaccard_similarity > max_jaccard_similarity_local)
87 { // Update highest Jaccard score if it is greater than the previous one
88     max_jaccard_similarity_local = jaccard_similarity;
89     most_similar_text = text_row;
90 }
91 }
92 //else{}
93 counter++;
94 }
95 }
96 MPI_Allreduce(&max_jaccard_similarity_local, &max_jaccard_similarity_global, 1, MPI_DOUBLE, MPI_MAX, MPI_COMM_WORLD);
```

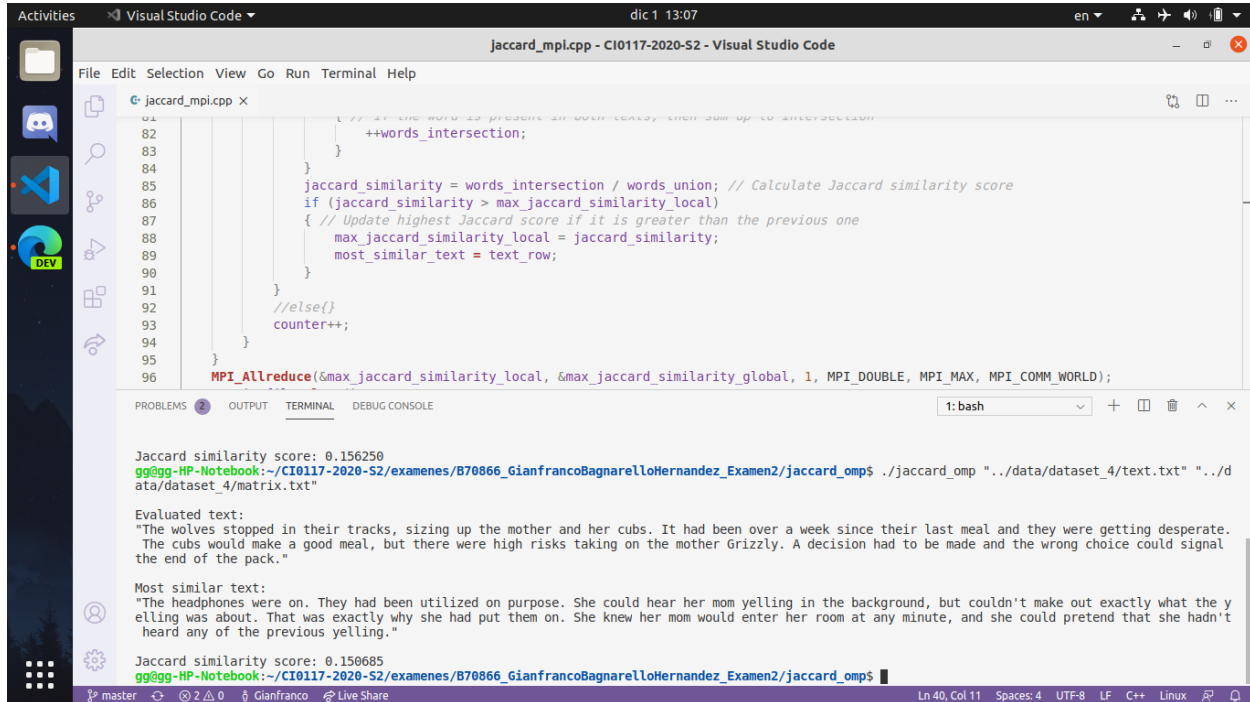
```
gg@gg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866_GianfrancoBagnarelloHernandez_Examen2/jaccard_omp$ make
g++ jaccard_omp.cpp -o jaccard_omp -fopenmp
gg@gg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866_GianfrancoBagnarelloHernandez_Examen2/jaccard_omp$ ./jaccard_omp ../data/dataset_3/text.txt ../data/dataset_3/matrix.txt

Evaluated text:
"The chair sat in the corner where it had been for over 25 years. The only difference was there was someone actually sitting in it. How long had it been since someone had done that? Ten years or more he imagined. Yet there was no denying the presence in the chair now."

Most similar text:
"She had been an angel for coming up on 10 years and in all that time nobody had told her this was possible. The fact that it could ever happen never even entered her mind. Yet there she stood, with the undeniable evidence sitting on the ground before her. Angels could lose their wings."

Jaccard similarity score: 0.156250
gg@gg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866_GianfrancoBagnarelloHernandez_Examen2/jaccard_omp$
```

## Conjunto de datos 4 (propuesto, 50 párrafos con empate)



```
82         ++words_intersection;
83     }
84 }
85 jaccard_similarity = words_intersection / words_union; // Calculate Jaccard similarity score
86 if (jaccard_similarity > max_jaccard_similarity_local)
87 { // Update highest Jaccard score if it is greater than the previous one
88     max_jaccard_similarity_local = jaccard_similarity;
89     most_similar_text = text_row;
90 }
91 }
92 //else{}
93 counter++;
94 }
95 }
96 MPI_Allreduce(&max_jaccard_similarity_local, &max_jaccard_similarity_global, 1, MPI_DOUBLE, MPI_MAX, MPI_COMM_WORLD);
```

Jaccard similarity score: 0.156250  
gggg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866\_GianfrancoBagnarelloHernandez\_Examen2/jaccard\_omp\$ ./jaccard\_omp "../data/dataset\_4/text.txt" "../data/dataset\_4/matrix.txt"

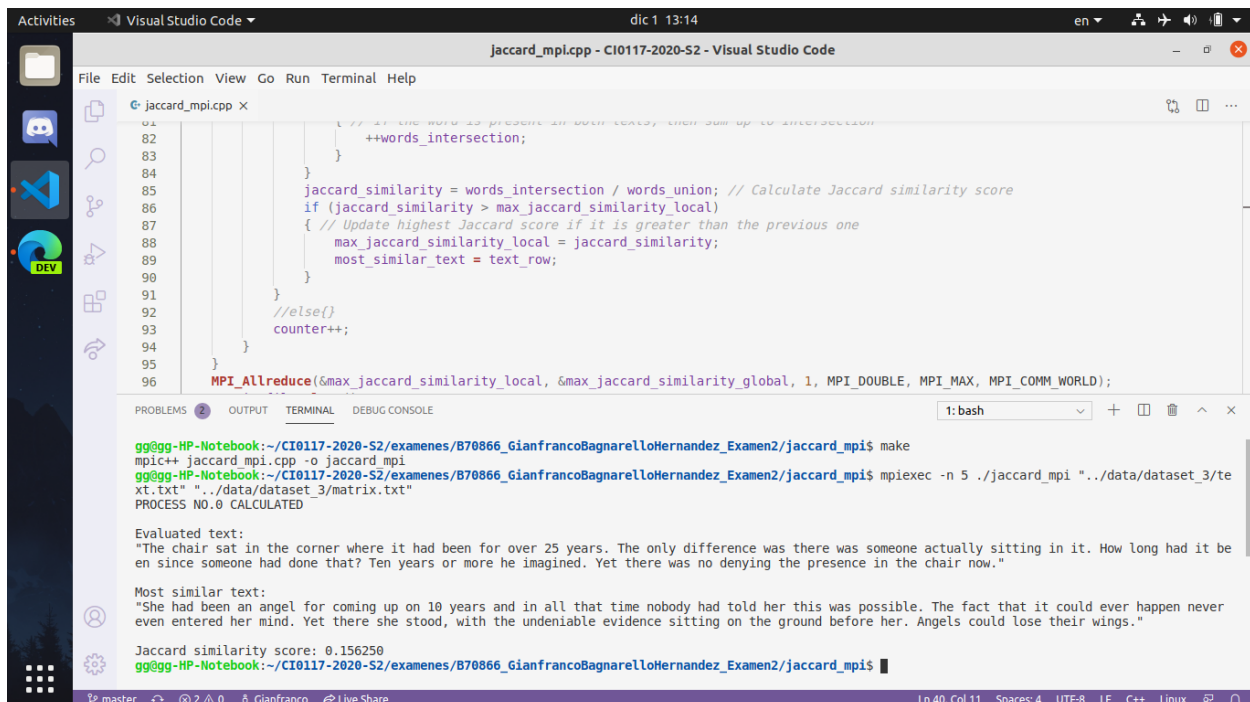
Evaluated text:  
"The wolves stopped in their tracks, sizing up the mother and her cubs. It had been over a week since their last meal and they were getting desperate. The cubs would make a good meal, but there were high risks taking on the mother Grizzly. A decision had to be made and the wrong choice could signal the end of the pack."

Most similar text:  
"The headphones were on. They had been utilized on purpose. She could hear her mom yelling in the background, but couldn't make out exactly what the yelling was about. That was exactly why she had put them on. She knew her mom would enter her room at any minute, and she could pretend that she hadn't heard any of the previous yelling."

Jaccard similarity score: 0.150685  
gggg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866\_GianfrancoBagnarelloHernandez\_Examen2/jaccard\_omp\$

## jaccard\_mpi

## Conjunto de datos 3 (propuesto, 50 párrafos sin empate)



```
82         ++words_intersection;
83     }
84 }
85 jaccard_similarity = words_intersection / words_union; // Calculate Jaccard similarity score
86 if (jaccard_similarity > max_jaccard_similarity_local)
87 { // Update highest Jaccard score if it is greater than the previous one
88     max_jaccard_similarity_local = jaccard_similarity;
89     most_similar_text = text_row;
90 }
91 }
92 //else{}
93 counter++;
94 }
95 }
96 MPI_Allreduce(&max_jaccard_similarity_local, &max_jaccard_similarity_global, 1, MPI_DOUBLE, MPI_MAX, MPI_COMM_WORLD);
```

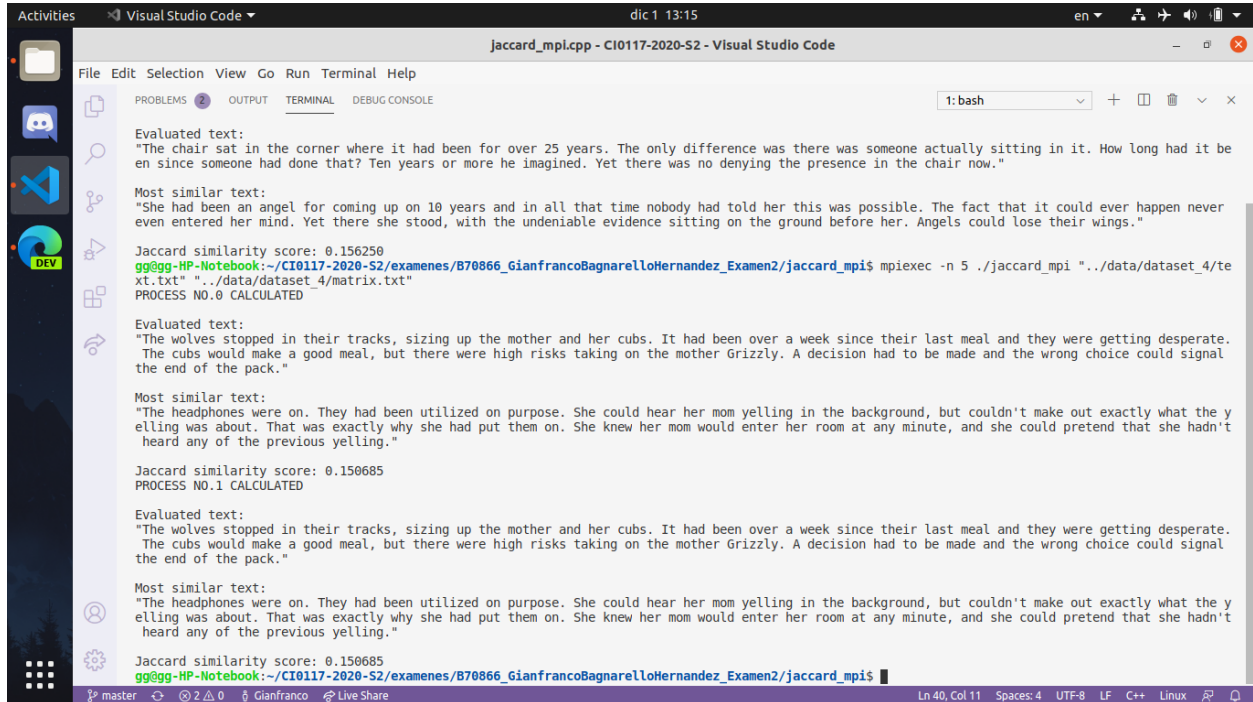
gggg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866\_GianfrancoBagnarelloHernandez\_Examen2/jaccard\_mpi\$ make  
mpic++ jaccard\_mpi.cpp -o jaccard\_mpi  
gggg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866\_GianfrancoBagnarelloHernandez\_Examen2/jaccard\_mpi\$ mpiexec -n 5 ./jaccard\_mpi "../data/dataset\_3/text.txt" "../data/dataset\_3/matrix.txt"  
PROCESS NO.0 CALCULATED

Evaluated text:  
"The chair sat in the corner where it had been for over 25 years. The only difference was there was someone actually sitting in it. How long had it been since someone had done that? Ten years or more he imagined. Yet there was no denying the presence in the chair now."

Most similar text:  
"She had been an angel for coming up on 10 years and in all that time nobody had told her this was possible. The fact that it could ever happen never even entered her mind. Yet there she stood, with the undeniable evidence sitting on the ground before her. Angels could lose their wings."

Jaccard similarity score: 0.156250  
gggg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866\_GianfrancoBagnarelloHernandez\_Examen2/jaccard\_mpi\$

## Conjunto de datos 4 (propuesto, 50 párrafos con empate)



The screenshot shows the Visual Studio Code interface with a terminal window open. The terminal displays the output of a Jaccard similarity calculation. It shows three paragraphs of text, each followed by its evaluated text, most similar text, and Jaccard similarity score. The scores are 0.156250, 0.150685, and 0.150685, indicating a tie between the second and third paragraphs.

```
Jaccard similarity score: 0.156250
gg@gg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866_GianfrancoBagnarelloHernandez_Examen2/jaccard_mpi$ mpxec -n 5 ./jaccard_mpi ../data/dataset_4/text.txt ../data/dataset_4/matrix.txt
PROCESS NO.0 CALCULATED

Evaluated text:
"The chair sat in the corner where it had been for over 25 years. The only difference was there was someone actually sitting in it. How long had it been since someone had done that? Ten years or more he imagined. Yet there was no denying the presence in the chair now."

Most similar text:
"She had been an angel for coming up on 10 years and in all that time nobody had told her this was possible. The fact that it could ever happen never even entered her mind. Yet there she stood, with the undeniable evidence sitting on the ground before her. Angels could lose their wings."

Jaccard similarity score: 0.150685
PROCESS NO.1 CALCULATED

Evaluated text:
"The wolves stopped in their tracks, sizing up the mother and her cubs. It had been over a week since their last meal and they were getting desperate. The cubs would make a good meal, but there were high risks taking on the mother Grizzly. A decision had to be made and the wrong choice could signal the end of the pack."

Most similar text:
"The headphones were on. They had been utilized on purpose. She could hear her mom yelling in the background, but couldn't make out exactly what the yelling was about. That was exactly why she had put them on. She knew her mom would enter her room at any minute, and she could pretend that she hadn't heard any of the previous yelling."

Jaccard similarity score: 0.150685
gg@gg-HP-Notebook:~/CI0117-2020-S2/examenes/B70866_GianfrancoBagnarelloHernandez_Examen2/jaccard_mpi$
```

**Nota:** para generar los párrafos de forma aleatoria se utilizó la siguiente herramienta:

<https://randomwordgenerator.com/paragraph.php>