## Métricas

En esta sección se realiza un análisis cuantitativo que permite establecer métricas de los resultados obtenidos a través de la herramienta. Para realizar dicho análisis, se utilizaron 2 tipos de métricas: el lapso de tiempo para efectuar la identificación, y aquellas especificas extraídas de la rama de *Recuperación de Información*. Estas últimas se pueden entender desde dos puntos de vistas: como un escenario de recuperación de información o como una tarea de clasificación estadística. Estos son análogos, y ambos serán explicados para cada métrica. Las métricas utilizadas para evaluar el desempeño de los algoritmos son *Precision*, *Recall* y *F-Measure* [15, 16, 18]. *Precision* puede ser interpretada como una métrica de exactitud o fidelidad, mientras que *Recall* es una métrica de completitud. *F-Measure* es una combinación de ambas.

Desde el punto de vista de la recuperación de información: una *Precision* perfecta se puntuaría con 1.0, y significa que cada resultado recuperado por una búsqueda fue relevante (pero no dice nada acerca de si todos los documentos relevantes fueron recuperados). En contraste, un *Recall* perfecto se puntuaría con 1.0, y significa que todos los documentos relevantes fueron recuperados por la búsqueda (pero no dice nada acerca de cuantos documentos irrelevantes también fueron recuperados). Para realizar el análisis, se utilizan los siguientes conceptos: *Relevant* *Documents*, el cual representa la cantidad de documentos relevantes que hay sobre una búsqueda dada, y *Documents* *Retrieved*, que representa aquellos documentos que fueron recuperados en una búsqueda.

Análogamente, desde el punto de vista de la clasificación estadística: se definen los siguientes conceptos: *True Positives*, que representa el número de ítems etiquetados como pertenecientes a la clase que efectivamente pertenecen a esa clase; *True Negatives*, que representa el número de ítems no etiquetados como pertenecientes a la clase que efectivamente no pertenecen a esa clase; *False Positives*, que representa el número de ítems etiquetados como pertenecientes a la clase que efectivamente no pertenecen a esa clase; y *False Negatives*, que representa el número de ítems no etiquetados como pertenecientes a la clase que efectivamente pertenecen a esa clase. En este contexto, *Precision* de una clase es el número de *True Positives* (por ejemplo, el numero de ítems correctamente etiquetados como pertenecientes a la clase) dividido por el número total de elementos etiquetados como pertenecientes a la clase (por ejemplo, la suma de *True Positives* y *False Positives*, que son los ítems incorrectamente etiquetados como pertenecientes a la clase). *Recall* en este contexto es definido como el número de *True Positives* dividido el número total de elementos que actualmente pertenecen a la clase (por ejemplo, la suma de *True Positives* y *False Negatives*, que son aquellos ítems que no fueron etiquetados como pertenecientes a esa clase pero debieron haberlo estado).

Frecuentemente, hay una relación inversa entre *Precision* y *Recall*, donde es posible incrementar una de ellas con el costo de reducir la otra, y viceversa. Usualmente, y debido a esta relación, estas dos métricas no son discutidas aisladamente. En cambio, ambos valores para una métrica son comparados a un nivel fijado de la otra métrica (por ejemplo, *Precision* a un nivel de 0.75 de *Recall*), o son combinadas en una medida única, como *F-Measure*, que representa la media armónica ponderada de *Precision* y *Recall*.

### Falsos y verdaderos, positivos y negativos

Para nuestro análisis, especificaremos las siguientes definiciones, basándonos en las establecidas previamente:

|  |  |
| --- | --- |
| *QA Verdaderos Positivos* (QVP) | Son aquellos atributos de calidad **identificados** por la herramienta, que **son realmente** atributos de calidad del sistema. |
| *QA Verdaderos Negativos*(QVN) | Son aquellos atributos de calidad **no identificados** por la herramienta, que **no son realmente** atributos de calidad del sistema. |
| *QA Falsos Positivos* (QFP) | Son aquellos atributos de calidad **identificados** por la herramienta, que **no son realmente** atributos de calidad del sistema. |
| *QA Falsos Negativos* (QFN) | Son aquellos atributos de calidad **no identificados** por la herramienta, que **son realmente** atributos de calidad del sistema. |
| *QA Verdaderos* (AV) | Son atributos de calidad que **son realmente** atributos de calidad del sistema. |

Se podría decir que los atributos de calidad en los casos de estudio anteriormente se corresponden con la definición de la siguiente manera: los QVP serian “los atributos de calidad correctos” y los QFP serian los “atributos de calidad incorrectos y mal etiquetados”. En la Tabla V‑43 se exhiben los datos utilizados para la realización del análisis estadístico de los datos. La cantidad total de los QV se estableció realizando inspecciones manuales de las especificaciones. Estos valores se determinaron para cada caso de estudio. Se debe tener en cuenta que aquellos atributos de calidad identificados redundantemente no se agregan ni en los QVP ni en los QFP. Como se dijo antes, los atributos de calidad mal etiquetados son considerados como QFP. Los QVN no son considerados porque no tiene sentido en este tipo de evaluación.

|  |  |
| --- | --- |
| **Caso de Estudio** | **HWS** |
| QVP | 4 |
| QVN | 2 |
| QFP | 3 |
| QV | 6 |

Tabla V‑43 - HWS - Resultados (QVP, QVN, QFP, QV)

### 

### Recall, Precision y F-Measure

Utilizando la nomenclatura utilizada anteriormente, las formulas para calcular estas métricas se reescriben de esta manera:

Se debe resaltar que según nuestra opinión, en el caso particular de la identificación de atributos de calidad, creemos que el *Recall* es la métrica más importante de las tres. Los costos generados por la omisión de identificar un QA pueden llegar a ser muy costosos. Por eso, se debe tratar de detectar la mayor cantidad de QA posibles y en consecuencia, siempre que sea viable, conseguir el mayor *Recall* que se pueda.

La formula de *F-Measure* se mantiene como se definió anteriormente. En las Tabla V‑45 se muestra los resultados de las métricas *Precision*, *Recall* y *F-Measure* (medidos en porcentajes) en el caso de estudio:

|  |  |
| --- | --- |
| **Caso de Estudio** | **HWS** |
| Precision |  |
| Recall |  |
| F-Measure |  |

Tabla V‑45 – HWS - Resultados (Precision, Recall , F-Measure)

### Tiempo de ejecución

De manera de tener una noción de la velocidad de procesamiento de la identificación por parte de cada algoritmo, en las Tabla V‑47 y Tabla V‑48 se muestran los lapsos de *Tiempo* necesarios para efectuar el análisis de identificación de aspectos (medidos en segundos) de cada caso de estudio con las tres técnicas evaluadas:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Caso de Estudio** | **SIMS** | | | **CHMS** | | | **ARS** | | |
| ***Técnica*** | ***AET*** | ***AEOT*** | ***AEST*** | ***AET*** | ***AEOT*** | ***AEST*** | ***AET*** | ***AEOT*** | ***AEST*** |
| Tiempo | 1 | 1 | 11 | 1 | 2 | 13 | 1 | 4 | 1026 |

Tabla V‑47 - SIMS, CHMS y ARS - Resultados (Tiempo)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Caso de Estudio** | **CRS** | | | **CPSP** | | |
| ***Técnica*** | ***AET*** | ***AEOT*** | ***AEST*** | ***AET*** | ***AEOT*** | ***AEST*** |
| Tiempo | 1 | 3 | 87 | 1 | 3 | 38 |

Tabla V‑48 - CRS y CSPS - Resultados (Tiempo)