Base de Datos II

Modelo Conceptual.

Profesor: ING. Julio Marcelo Gutierrez.

Alumnos:

* Emiliano Barg.
* Nicolás Santiago Gómez.

**Introducción:**

Para el proyecto que debemos llevar a cabo en esta materia decidimos utilizar fuentes de datos sobre descubrimientos de exoplanetas, con el objetivo de obtener información sobre el rendimiento de métodos y observatorio. Imaginamos que un supuesto inversor interesado en la astronomía, o tal vez una persona con poder de toma de decisión sobre esos observatorios, debe decidir sobre cómo va a alocar los recursos a su disposición. Este contexto nos llevó a plantearnos ciertas preguntas que esta supuesta persona podría llegar a hacerse, lo cual nos permitió realizar el análisis que explayamos en las siguientes páginas.

Desarrollo:

Comenzamos investigando sobre posibles fuentes de datos sobre los descubrimientos de exoplanetas, esta búsqueda nos llevo a dos páginas que utilizamos para obtener 3 archivos .csv con la información que queríamos encontrar:

* ExoplanetEU (<https://exoplanet.eu/home/>) : Es una página web dedicada a catalogar los descubrimientos de exoplanetas que son publicados en varias fuentes, el mayor filtro que realizan es sobre la masa medida del exoplaneta, debido a que si su masa es mayor a 60 veces la masa de Júpiter, ese exoplaneta debería ser una estrella, entonces con un margen de error de 1, filtran todos los descubrimientos de exoplanetas donde la masa del exoplaneta descubierto sea mayor a 61 veces la masa de Júpiter. Del catalogo disponible en esta página obtuvimos un .csv que vamos a referenciar como exoplanetEU.
* Exoplanet Archive (<https://exoplanetarchive.ipac.caltech.edu/index.html>) : Esta es una página web creada por la NASA, similar al caso anterior, ofrecen una gran variedad de información sobre los descubrimientos de exoplanetas, pero en este caso tienen dos fuentes principales, una tabla llamada Planetary Systems y otra llamada Planetary Systems Composite Data, en la primera cada descubrimiento de cada observatorio sobre cada exoplaneta es documentado por separado, permitiendo la comparación del mismo exoplaneta cuando es estudiado por varios observatorios, pero esto viene con la desventaja de que muchas veces hay vacíos de información, esto es resuelto con la tabla de Composite Data, donde la NASA trabaja las mediciones existentes para generar un valor promedio aceptable de acuerdo a lo medido. De esta fuente tomamos dos .csv, cada uno proviniendo de cada una de las tablas que mencionamos, haremos referencia a estos como exoplanetNASA y exoplanetCompositeNASA respectivamente.

Luego de familiarizarnos con la información extraída, planteamos las siguientes preguntas que podrían otorgar información valiosa para tomar una decisión:

1. ¿Cuál es el observatorio con más descubrimientos? ¿Y en este año?
2. ¿Cuál es el observatorio que ha llegado más lejos?
3. ¿Cuál es el observatorio con mayor precisión? ¿Y anualmente?
4. ¿Cuál es el método más preciso?
5. ¿Cuál es el método más utilizado?
6. ¿Cuál es el método que ha llegado más lejos?
7. ¿Cuál es el método que menos información nos ha dado sobre la estrella de ese sistema?
8. ¿Cómo han evolucionado estas características a través de los años para cada método?

Con estas preguntas, comenzamos a seguir la metodología HEFESTO para poder comenzar a elaborar el modelo conceptual. Entonces analizando estas preguntas logramos encontrar los siguientes indicadores y perspectivas:

**Indicadores:**

* Cantidad de descubrimientos.
* Precisión en medidas.
* Distancia a la Tierra.
* Existencia de información sobre la estrella.

**Perspectivas:**

* Observatorio.
* Método.
* Sistema.
* Planeta.
* Tiempo.

Con estas perspectivas e indicadores, realizamos una primera instancia para el modelo conceptual:

|  |
| --- |
| Observatorio. |
| Método. |
| Sistema. |
| Planeta. |
| Tiempo. |

|  |
| --- |
| Cantidad de descubrimientos. |
| Precisión de medida. |
| Distancia a la Tierra. |
| Existencia de información sobre la estrella. |

Continuando con el método HEFESTO, ahora comenzamos a construir una versión más elaborada para el modelo conceptual, para lograr esto revisitamos cada perspectiva e indicador y definimos los atributos que tendrán las perspectivas y cómo haremos para obtener los indicadores, además de especificar de qué fuente podemos obtener esa información.

**Perspectivas:**

* **Observatorio:**

**Id\_observatorio:** Será un número identificador autoincremental.

**Nombre:** Nombre del observatorio. Se puede extraer de la columna disc\_facility en exoplanetNASA.

* **Método:**

**Id\_metodo:** Será un número identificador autoincremental.

**Nombre**: Nombre del método utilizado. Se puede obtener desde la columna Detection\_type en exoplanetEU y de la columna discoverymethod en exoplanetNASA.

* **Sistema:**

**Id\_sistema:** Será un número identificador autoincremental.

**Nombre:** Nombre del Sistema. Se encuentra en la columna Star\_name en exoplanetEU y en la columna host\_name en exoplanetNASA.

* **Planeta:**

**Id\_planeta:** Será un número identificador autoincremental.

**Nombre:** Nombre asignado a ese planeta, puede llegar a tener varios nombres equivalentes, pero priorizaremos el que se encuentra en exoplanetCompositeNASA. Se puede extraer de la columna Name en exoplanetEU y de la columna pl\_name en exoplanetNASA y exoplanetCompositeNASA.

**Id\_sistema:** Identificador del sistema al que pertenece, la columna star\_name en exoplanetEU y host\_name en exoplanetNASA nos permitirían realizar esta conexión.

* **Tiempo:**

**Id\_tiempo:** Será un número identificador autoincremental.

**Año:** Año en el que sucedió el descubrimiento. Se encuentra en la columna discovered en exoplanetEU y en la columna disc\_year en exoplanetNASA.

**Indicadores:**

* **Cantidad de descubrimientos:** Contar la cantidad de planetas que se descubrieron ese año.
* **Precisión de medida:** Promediar el error porcentual de las mediciones, si no está disponible la información será un 100% de error, esto se hará solo con los parámetros del exoplaneta.
* **Distancia a la tierra:** Se puede obtener de la columna sy\_dist en exoplanetNASA y de la columna star\_distance en exoplanetEU.
* **Información sobre la estrella:** Será un valor booleano que indica si se conoce sobre la estrella del sistema. Para que sea TRUE se deberá saber:
  + Nombre de estrella: De la columna Star\_name en exoplanetEU y la columna host\_name en exoplanetNASA.
  + Radio de estrella: Presente en la columna st\_rad en exoplanetNASA y en la columna star\_radius en exoplanetEU.
  + Tipo de estrella: Se encuentra en la columna st\_spectype en exoplanetNASA y en la columna star\_sp\_type en exoplanetEU.

Con esta información volvimos a construir el modelo conceptual, obteniendo:

|  |
| --- |
| Observatorio:   * Nombre. |
| Método:   * Nombre. |
| Sistema:   * Nombre. |
| Planeta:   * Nombre. * Sistema. |
| Tiempo:   * Año. |

|  |
| --- |
| Cantidad de descubrimientos:   * COUNT(Descubrimientos). |
| Precisión de medida:   * AVG(Error\_porcentual’). |
| Distancia a la Tierra:   * sy\_dist y star\_distance. |
| Existencia de información sobre la estrella:   * Debe estar definido el nombre,   el radio y el tipo de la estrella. |