BASE DE DATOS

OPERADORES DEL ÁLGEBRA RELACIONAL

1. La idea es pensar a las **tablas** como **relaciones**, es decir, **conjuntos** de tuplas.
2. En este enfoque **no hay repeticiones** de tuplas (si no no serían conjuntos). De arranque nomás, tenemos todas las operaciones habituales de conjuntos (unión, intersección, diferencia, producto cartesiano, ...) pero se utilizan con cierto cuidado, como se explica a continuación.
3. Un **atributo** es un rótulo, por ejemplo nombre, apellido, edad, peso, domicilio, etc. Habitualmente utilizamos A, B, C, D, etc. para atributos.
4. Cada atributo tiene asociado un **dominio** de valores, por ejemplo, cadena de caracteres, números naturales, números enteros, números reales, etc. Se asume que todos los dominios incluyen un valor especial **null** que representa la indefinición. Usamos $\mathcal{D}$ para dominios.
5. Un **esquema de relación** es una secuencia de atributos sin repeticiones. Por ejemplo: dni,nombre,apellido,fechaDeNacimiento,género es un esquema de relación. Usaremos letras R, S, T, etc. para esquemas de relación.
6. Dadas dos esquemas de relaciones $R = A_1, A_2, \ldots, A_n$ y $S = B_1, B_2, \ldots, B_m$ disjuntos, es decir, tal que $A_i \neq B_j$ para todo i, j, se define $R, S = A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m$.
7. Dada una relación r y un esquema de relación R, escribimos $r(R)$ para decir que r **es una relación con esquema de relación** R. Para explicar esto, sea R el esquema de relación $A_1, A_2, \ldots, A_n$ donde cada atributo A_i tiene asociado el dominio de valores $\mathcal{D}_i$, $r(R)$ significa que $r \subseteq \mathcal{D}_1 \times \mathcal{D}_2 \times \ldots \times \mathcal{D}_n$. Dada $r(R)$, $t \in r$ y A uno de los atributos de R. Se escribe $t[A]$ al valor que la tupla t asocia al atributo A.
8. Una relación **se declara** con un esquema en particular. Luego las operaciones se definen de forma tal que el esquema del resultado queda determinado.
9. Dada $r(R)$ donde $R = A_1, A_2, \ldots, A_n$, se define por $r.R$ al esquema de relación cuyos atributos son $r.A_1$, $r.A_2$, ... y $r.A_n$, es decir, $r.R = r.A_1, r.A_2, \ldots, r.A_n$.
10. Las operaciones entre relaciones se realizan de manera de tener cuidado con los esquemas asociados a las relaciones.
11. La **unión**, **intersección** y **diferencia** de relaciones se podría hacer entre cualquier par de relaciones, la teoría de conjuntos permite hacer uniones, intersecciones y diferencias arbitrarias. Pero esto no suele ser necesario en base de datos: interesa aplicarlos entre relaciones de igual esquema, o de **esquema compatibles**. Dos esquemas $A_1, A_2, \ldots, A_n$ y $B_1, B_2, \ldots, B_m$ son compatibles si $n = m$ y para todo i el dominio asociado a A_i es el mismo que el dominio asociado a B_i. Es decir, no importa que los atributos tengan el mismo nombre, lo que importa es que sus dominio se compatibles.
12. Dadas $r(R)$ y $s(S)$, con R y S esquemas compatibles, las relaciones $r \cap s$, $r \cup s$ y $r - s$ son relaciones con esquema de relación compatible con R (y por ende, S), y **anónima**. Es decir, los nombres de las columnas son *frescos* y desconocidos para nosotros. Para poder realizar operaciones (como seleccionar los elementos de una tal columna que cumplan con cierto valor), se deberá renombrar (ver punto 15). Para evitar renombres innecesarios, en el caso que $R = S$, entonces la operación no será anónima y mantendrá el nombre de las columnas de R.
13. Dadas $r(R)$ y $s(S)$, $r \times s$ es el producto cartesiano de las relaciones r y s y su esquema es $r.R, s.S$, la concatenación de los esquemas de relación dados. Por simplicidad notacional,

suelen omitirse los prefijos "r." y "s." que no lleven a que dos atributos queden con el mismo rótulo. Por ejemplo, si el atributo A figura en R y en S, debemos dejar $r.A$ y $s.A$ en el esquema del producto cartesiano. En cambio si B figura en R y no en S podemos escribir B en vez de $r.B$. Similamente para atributos que sólo aparecen en S.

14. Esta notación no resuelve el problema de encontrarle un esquema de relación al producto cartesiano de la forma $r \times r$, para $r(R)$. En ese caso, conviene generar un nuevo nombre: sea $s = r$, ahora $r \times s$ es un producto con esquema $r.R, s.R$.
15. A veces uno quiere **renombrar** los rótulos de los atributos. Por ejemplo, sea $r(R)$, y S un esquema compatible con R, $\rho_S(r)$ es una relación igual a r, pero su esquema de relación es S en vez de R. A veces queremos cambiar el nombre de pocos atributos, puede convenir $\rho_{A \to B, C \to D}(r)$ para que el atributo A cambie su nombre por B, y el atributo C cambie su nombre por D.
16. El operador de **selección** se denota por σ. Dada una condición P construída utilizando los conectivos lógicos, a partir de los atributos de R y operaciones sobre los dominios de la base de datos (expresiones aritméticas, de cadenas de caracteres, de comparación de fechas, etc.), y dada $r(R)$, $\sigma_P(r)$ es una relación con esquema de relación R, y se define como el conjunto de las tuplas de r que satisfacen la condición P.
17. El operador de **proyección** se denota por Π. Dado un esquema de relación S que sea un subconjunto del esquema R, y dada $r(R)$, $\Pi_S(r)$ es una relación con esquema de relación S y se obtiene descartando de cada tupla de r las componentes que no corresponden a atributos de S, y reordenando los restantes componentes de la tupla según el orden en que se listen los atributos en el esquema S. Ojo que al hacer esto ingenuamente, pueden producirse tuplas repetidas en el resultado que deben eliminarse porque estamos trabajando con conjuntos.
18. La **proyección generalizada** permite utilizar expresiones como subíndices de Π. Por ejemplo, $r = \Pi_{eId, eSalario*13 \to eSalarioAnual}(empleados)$ para computar el salario anual de cada empleado multiplicando el salario mensual por 13 (12 meses más aguinaldo). Observar que en el caso de una expresión es necesario especificar el nombre del atributo nuevo (en este caso eSalarioAnual). El esquema de la relación r es $eId, eSalarioAnual$. Si la expresión utiliza un único atributo A, y se omite el nombre del atributo nuevo, se asume que el nombre del atributo del resultado sigue siendo A. Así, el ejemplo anterior puede escribirse $\Pi_{eId, eSalario*13}(empleados)$ cuyo esquema es $eId, eSalario$.
19. La operación de **agregación** se escribe $s = A_1, A_2, \ldots, A_n \mathcal{G}_{F_1(B_1), F_2(B_2), \ldots, F_m(B_m)}(r)$ donde $r(R)$, los A_i y las B_j son atributos de R, y las F_j son funciones de agregación: avg (promedio), min (mínimo), max (máximo), sum (suma) y count (contar). El esquema del resultado tiene $n + m$ elementos, con los dominios correspondientes a A_i y F_j, pero como con los operadores de conjuntos (unión, etc.) es anónima. La tupla $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m) \in s$ si y sólo si, considerando el conjunto $T_{x_1, \ldots, x_n}$ de tuplas de t tales que $t[A_i] = x_i$ para todo i, y_j se obtiene aplicando la función F_j a los atributos B_j de todas las tuplas del conjunto $T_{x_1, \ldots, x_n}$. En s solamente puede haber una tupla con el mismo comienzo $x_1, x_2, \ldots, x_n$, en otras palabras, $A_1, A_2, \ldots, A_n$ es superclave de s.

Las siguientes operaciones pueden definirse a partir de las anteriores.

1. Sean $r(R)$ y $s(S)$ donde hay atributos comunes entre R y S. Sean $A_1, A_2, \ldots, A_n$ todos los atributos comunes a R y S. La **reunión natural** entre r y s se define por la ecuación $r \bowtie s = \Pi_T(\sigma_{r.A_1 = s.A_1 \wedge r.A_2 = s.A_2 \wedge \ldots \wedge r.A_n = s.A_n}(r \times s))$ donde T es el esquema R, S' y S' es S sin los atributos $A_1, A_2, \ldots, A_n$. El esquema de $r \bowtie s$ es T.
2. Las filminas muestran cómo las **reuniones externas** pueden obtenerse a partir de la reunión natural y otros operadores.

3. Sean $r(R)$ y $s(S)$ con S un subconjunto de R. Sea T el esquema R sin los atributos de S. La **división** entre r y s se define por $r \div s = \Pi_T(r) - \Pi_T((\Pi_T(r) \times s) - r)$. Su esquema es T. Para ser precisos, la última r de la ecuación puede tener los atributos en otro orden (el de R), habría que modificar la ecuación: $r \div s = \Pi_T(r) - \Pi_T((\Pi_T(r) \times s) - \Pi_{T,S}(r))$.