

Pauta control 2

Bases de datos

Otoño de 2008

Asignación de puntajes

El propósito de la evaluación es determinar el progreso del aprendizaje por parte del alumno. Este aprendizaje se ve en dos niveles: un aprendizaje *exterior*, en donde los contenidos son conocidos como información (memoria), y un aprendizaje *interior*, que concierne al nivel de asimilación de los contenidos.

El estar informado de los contenidos (conocimiento exterior) es un requisito mínimo. El estar en desconocimiento de la materia acredita puntaje mínimo.

Si el alumno está informado de los contenidos, el puntaje refleja el dominio que el alumno demuestra de los contenidos (conocimiento interior).

El significado de las notas usado en la evaluación, en coherencia con la escala de la FCFM, es:

| Nota | Significado |
|------|---|
| 1.0 | Precario. <i>Completamente inaceptable. Malo.</i> |
| 3.0 | Insuficiente. <i>No alcanza a ser aceptable. Malo.</i> |
| 4.0 | Suficiente. <i>Apenas aceptable. Pero fuera del rango de lo malo.</i> |
| 5.0 | Bueno. <i>Demuestra autonomía intelectual. Buena tasa de acierto.</i> |
| 5.7 | Sobresaliente. <i>O destacado. Responde bastante bien para el tiempo dado.</i> |
| 7.0 | Perfecto o experto. |

El puntaje de cada pregunta se definirá y asignará en coherencia con la escala anterior. (En la FCFM, 4.0 es suficiente, 5.0 es bueno y 5.7 es destacado. 1.0 y 7.0 son implícitos.)

Pregunta 1

Se trata de determinar si el alumno:

- Sabe SQL (requisito mínimo).
- Desarrolló una habilidad avanzada en el uso de SQL (producto de un buen estudio).

El puntaje de cada pregunta va de 0 a 2, con un dígito decimal. La nota de la pregunta es la suma de los puntajes más el punto base.

La escala de puntajes por pregunta se presenta en el siguiente cuadro, sin perjuicio de la existencia de puntajes intermedios:

| Puntaje | Características |
|---------|---|
| 0.0 | Entendimiento mínimo de SQL <i>No sabe agregar/no sabe agrupar</i> |
| 0.5 | Sabe SQL, pero no lo aplica bien <i>Sólo correcto en la forma, la cual es coherente con lo pedido.</i> |
| 1.0 | Responde parcialmente <i>Posible error de SQL, pero hace más del 50% de lo pedido</i> |
| 1.5 | Bastante bien, pero malas prácticas de SQL <i>Debió anidar algo, la respuesta podía ser mucho más simple</i> |
| 2.0 | Perfecto. |

Conviene indicar que cada pregunta no se revisa en independencia de las otras. De esta manera se pueden ver errores de interpretación *intencionados* (un error de interpretación del enunciado es consistente pregunta a pregunta). Si el error de interpretación no parece intencionado, pero tampoco es claro mal interpretar el enunciado en tal magnitud, se considerará el impacto en la asignación de puntajes. El impacto en el puntaje será mayor a mayor es la disonancia de interpretación. Esta pregunta, en especial, fue explicada por el profesor durante el control.

Pregunta 2

En este caso se requiere tener conocimiento y aplicar:

- Dependencias funcionales.
- Identificación de llaves.
- Equivalencia de conjuntos de dependencias funcionales.
- Evaluación de formas normales (3FN y FNBC).
- Normalización (3FN ó FNBC).

Cada pregunta tiene objetivos distintos, así que se verá la asignación de puntajes por cada pregunta.

Para 2.1, se requiere conocer la noción de equivalencia de clausuras, utilizar un método para probarla (computar las clausuras o deducir las dependencias funcionales de un conjunto desde el otro) y concluir satisfactoriamente, siempre usando correctamente el álgebra de dependencias funcionales. Lo mínimo es conocer el álgebra de dependencias funcionales y la noción de dependencias, requisitos para hacer todo el desarrollo de esta pregunta.

| Puntaje | Características |
|---------|---|
| 0.0 | No tiene el conocimiento para proceder. <i>No sabe álgebra de d.f. o no comprende la noción de equivalencia.</i> |
| 0.5 | Destreza insuficiente. <i>Usa +/- bien las d.f., trata de responder, pero el avance es escaso.</i> |
| 1.0 | Responde parcialmente. <i>Hace más del 50% de lo pedido.</i> |
| 1.5 | Faltó poco para concluir. <i>Poco faltó para desarrollar la clausura o deducir dependencias.</i> |
| 2.0 | Demuestra. |

Para 2.2, se requieren evaluar formas normales. Para ello es necesario obtener la llave candidata (que es sólo una, y se debe ver que sólo una es), entender lo que es una superllave, conocer la definición de las formas normales y evaluarlas bien. Un esquema de solución es:

- Determinar que AF es llave.
- Ver que AF es la única llave minimal (se puede argumentar que todo atributo no determinado en una d.f. debe ser parte de toda llave).
- Probar que no está en 3FN.
- Y sale que no está en FNBC ($3FN \Rightarrow FNBC$).

Lo más conveniente es usar N en vez de M . Si el alumno argumenta que la pregunta es ambigua por no especificar si el conjunto de dependencias no se especificó, la pregunta está mala. Si los conjuntos son equivalentes, no hay ambigüedad alguna.

| Puntaje | Características |
|---------|---|
| 0.0 | No tiene el conocimiento para proceder. <i>No sabe evaluar formas normales ni buscar llaves (usando d.f.).</i> |
| 0.5 | Destreza insuficiente. <i>Apenas parece aplicar alguna FN, pero busca llaves.</i> |
| 1.0 | Aplica medianamente 3FN y FNBC. <i>Aplica sólo una bien, o ambas medianamente bien.</i> |
| 1.5 | Faltó poco para concluir. <i>Pequeña falta de rigor en un FN, y aplicó perfectamente la otra FN.</i> |
| 2.0 | Aplica correctamente 3FN y FNBC. |

Para 2.3, es importante saber normalizar y entender los criterios de calidad de un esquema relacional. El argumento se debe basar en los criterios de calidad, que se explicaron ampliamente en cátedra y en clase auxiliar: la preservación de las dependencias y de la información. La última es la importante, y se deben acoger a esa. La preservación de dependencias sólo cubre parte de la calidad. La respuesta perfecta es lograr FNBC y preservar la información, viendo que también cumple 3FN. De hecho, usando N , la normalización en 3FN también cumple con FNBC. Ojo que no se pedía un gran argumento; sólo indicar el criterio de calidad y mostrar cómo lo cumplía la normalización. (Normalización en 3FN cumple *de fábrica*.)

| Puntaje | Características |
|---------|--|
| 0.0 | No tiene el conocimiento para proceder. <i>No sabe normalizar ni sabe los criterios de calidad.</i> |
| 0.5 | Destreza insuficiente. <i>Apenas parece normalizar / normaliza bien pero argumenta mal.</i> |
| 1.0 | Mala argumentación. <i>Normaliza correctamente pero argumenta pobremente.</i> |
| 1.5 | No maximiza. <i>Argumenta coherentemente, y logra 3FN. Pero faltó estudiar FNBC.</i> |
| 2.0 | Perfecto. |

Pregunta 3

Esta pregunta requiere saber:

- FNBC, en definición y evaluación.
- Dependencias funcionales.

Cada pregunta tiene un esquema propio de puntaje. A diferencia de los casos anteriores, su puntaje va de 0 a 3.

En la 3.1, hay que analizar cuatro casos y lograr concluir. Cada caso correctamente evaluado tiene un puntaje de 0.7, y la conclusión (que sale de que se estudiaron todos los casos posibles) agrega 0.2. Si un caso se estudia insatisfactoriamente, su nota bajará a 0.2 ó 0.1 (cada caso era muy sencillo, por lo que un error estropea todo). Los casos, todos en FNBC, son:

- Caso 1: sólo hay dependencias triviales, llave AB .
- Caso 2: $F = \{A \rightarrow B\}$, llave A .
- Caso 3: $F = \{B \rightarrow A\}$, pudiéndose argumentar simetría con el caso anterior. Llave B .
- Caso 4: $F = \{A \rightarrow B, B \rightarrow A\}$, A y B llaves candidato.

En la 3.2, hay que encontrar un contraejemplo. Se da 3 puntos para el contraejemplo correcto, 1 punto para algo incorrecto pero similar a un contraejemplo (esto es un regalo) y 0 si no se cumple lo anterior.

Ejemplo de solución

Pregunta 1

Recorrido más largo: localidad de salida, de destino, velocidad promedio

```
select s.localidadsalida, p.localidad, (p.distancia/p.llegada) as velocidad
from salida s, parada p
where s.numrecorrido=p.numrecorrido
and p.horas >= ALL(select horas from parada);
```

Por recorrido: trayectorias (inicio, fin), número localidades visitadas y tiempo que toma

```
select s.localidadsalida, p.localidad, count(distinct q.localidad) as nvisitadas, p.horallegada
from salida s, parada p, parada q
where s.numrecorrido=p.numrecorrido and p.numrecorrido=q.numrecorrido
group by s.numrecorrido, s.localidadsalida, p.localidad, p.horallegada
having max(p.horallegada)>=max(q.horallegada);
```

Número de recorridos directos (s/paradas intermedias)

```
select count(p.numrecorrido) from parada p
where p.numrecorrido in (select q.numrecorrido from parada q
group by q.numrecorrido having count(q.localidad)=1);
```

Pregunta 2

Mostrar que M^+ y N^+ son equivalentes

M produce N:

$A \rightarrow BD$: En M están $A \rightarrow B$ y $B \rightarrow C \Rightarrow A \rightarrow C \Rightarrow A \rightarrow AC$. Como en M $AC \rightarrow D$, $A \rightarrow AC \rightarrow D$. Como $A \rightarrow B$ y $A \rightarrow D$, $A \rightarrow BD$.

$B \rightarrow C$: $B \rightarrow C$ está en M.

$D \rightarrow E$: $D \rightarrow E$ está en M.

N produce M:

$A \rightarrow B$: En N está $A \rightarrow BD \Rightarrow A \rightarrow B$.

$B \rightarrow C$: $B \rightarrow C$ está en N.

$AC \rightarrow D$: En N está $A \rightarrow BD \Rightarrow AC \rightarrow BCD \Rightarrow AC \rightarrow D$.

$D \rightarrow E$: $D \rightarrow E$ está en N.

$A \rightarrow E$: En N están $A \rightarrow BD$ y $D \rightarrow E \Rightarrow BD \rightarrow BE$. Luego $A \rightarrow BD \rightarrow BE \Rightarrow A \rightarrow E$.

Se concluye su equivalencia porque sus dependencias funcionales se pueden generar a partir del otro.

Forma normal de (R,M)

Llave: los únicos atributos no determinados son A y F, por lo son parte de toda llave (de lo contrario ninguna llave lo determinaría). Su clausura es: $AF \rightarrow ABDF \rightarrow ABCDF \rightarrow ABCDEF$. Entonces AF es llave. Cualquier otra llave será superllave pues AF es parte de toda llave y es candidato.

Forma normal (test de 3FN): Usando N por sencillez, se contrasta con 3FN.

- $B \rightarrow C$ no es del tipo superllave \rightarrow algo y C no es atributo primo (sería llave, y la única es A). R no está en 3FN.

Entonces R no está en 3FN. Y tampoco puede estar en FNBC.

Normalización óptima

Se usará el algoritmo de normalización en 3FN, que garantiza preservación de la información. La normalización queda: $R1(\underline{A}, B, D)$, $R2(\underline{B}, C)$ y $R3(\underline{D}, E)$. Y cada una está en FNBC: en $R1$, $A \rightarrow BD$ es de la forma superllave \rightarrow algo, en $R2$, $B \rightarrow D$ es superllave \rightarrow algo, y en $R3$, $D \rightarrow E$ es superllave \rightarrow algo. Como falta la llave, se agrega: $R4(\underline{A}, E)$. Naturalmente, $R4$ está en FNBC. Luego, se ha logrado una descomposición en FNBC.

Normalización de buena calidad: se ha logrado una descomposición en FNBC que preserva la información (y las dependencias), con lo que se garantiza que la información sea recuperable y no se altere. Esto ocurrió por usar el algoritmo de normalización en 3FN que garantiza esta propiedad.

Pregunta 3

Toda relación binaria está en FNBC

Hay como máximo dos dependencias funcionales no triviales: $A \rightarrow B$ y $B \rightarrow A$. Esto genera cuatro subconjuntos F de dependencias funcionales no triviales ($2^2=4$):

- Sólo triviales (conjunto no trivial vacío): como toda dependencia funcional es trivial, no se viola FNBC.
- $\{A \rightarrow B\}$: Luego A es llave. Así que $A \rightarrow B$ es de la forma superllave \rightarrow algo, y no se viola FNBC.
- $\{B \rightarrow A\}$: Simétrico al anterior. Es cambiar $R(A, B)$ por $R(B, A)$. Luego está en FNBC.
- $\{A \rightarrow B, B \rightarrow A\}$: A y B son llaves alternas ($A \rightarrow AB=R$, $B \rightarrow AB=R$). De nuevo, ambas dependencias son de la forma superllave \rightarrow algo y no se viola FNBC.

Como en ninguno de los casos posibles de (R, F) se viola FNBC, siempre estará en FNBC.

No toda relación ternaria está en FNBC

Por ejemplo, sea $R(A, B, C)$ y $F=\{A \rightarrow B\}$. La llave es AC , los dos atributos no determinados. Luego, $A \rightarrow B$ no es del tipo superllave \rightarrow algo (y es no trivial), por lo que viola FNBC. Con esto se concluye.

Este ejemplo de solución representa una de las tantas maneras de resolver el control. En particular, 3.1 es solución de referencia casi en su totalidad. En demás, existen varias maneras de contestar, por lo que el valor referencial de este ejemplo es menor.