CSC 370

Activity Worksheet: Functional Dependencies & Keys

Sean Chester

Fall 2022

Notes

This worksheet provides extra practice questions for determining functional dependencies, computing attribute set closures, and identifying (super)keys.

Questions

1. FD's in a Sales Database: You are provided with the (non-normalised) relation below. Identify at least four non-trivial functional dependencies. Justify each FD with a one sentence explanation. (A response is correct if the justification is both plausible and based on the formal definition of an FD.) One of the four FD's is provided as a sample.

```
Sales(
    product_id,
    product_name,
    purchase_timestamp,
    customer_id,
    customer_name,
    shipping_address,
    billing_address,
    subtotal,
    taxes,
    delivery_fee,
    total,
    has_loyalty_number,
    quantity_purchased
)
```

subtotal taxes delivery_free → total
The total is uniquely determined by summing the subtotal, taxes, and delivery fee.

product_id ⟶ product.name

— a product has a single id
so if we know th id we know name, but not th other way.

Custumer_id → costumer.name , has_loyalty_number

— same as above and has_loyalty. number follow th same logic

shipping_address subtotal delivery.fee → taxes

— possibly if has_loyalty. number doesn't effect price
taxes determined by address and based on th total price
before taxes.

2. FD's in a Library Catalogue: Which of the following functional dependencies are suitable for a database that tracks the items at a library and which of those are on loan? The first question is answered as a sample solution.

   (a) date_checked_out $\longrightarrow$ due_date
   No. It is likely that different items (e.g., a book and a ukulele) have different loan periods.

   (b) ISBN $\longrightarrow$ title author publication_year

   Yes. ISBN is unique to each book so can determin title, author and publication_year if ISBN is updated at each edition, or it is initial publication_year.

   (c) customer_id $\longrightarrow$ hold_on_account

   Yes. A single hold on the whole account of a single customer.

   (d) author publication_year $\longrightarrow$ title

   No. author can publish multiple titles in a year.

   (e) author title $\longrightarrow$ publication_year

   No. Author can publish a book named "Bear" twice in different years.

   (f) due_date $\longrightarrow$ is_overdue

   X No, since we would need item and current date.

   due_date should be consistent based despite current_date
   is_overdue (TRUE FALSE)

3. Closures: In the following questions, you are given a set F of functional dependencies and an attribute set X. Determine the closure of X, X^+.

   (a) Find the closure of $\{A, C\}$ given the following FD's:
   A $\rightarrow$ B
   C $\rightarrow$ D
   D $\rightarrow$ A

   $\{A, C\}^+ \supseteq \{A, C\}$ **// trivially**
   $\{A, C\}^+ \supseteq \{A, B, C\}$ **// because A $\rightarrow$ B**
   $\{A, C\}^+ \supseteq \{A, B, C, D\}$ **// because C $\rightarrow$ D**

   (b) Find the closure of $\{A\}$, given the following FD's:
   A $\rightarrow$ B
   C $\rightarrow$ B

   $\{A\}^+ = \{\}$ trivial
   $= \{A\}$ trivial
   $= \{A, B\}$ Rule

   (c) Find the closure of $\{D\}$, given the following FD's:
   A $\rightarrow$ BD
   C $\rightarrow$ BD
   D $\rightarrow$ A

   $\{D\}^+ = \{\}$ trivial
   $= \{D\}$ trivial
   $= \{D, A\}$ Rule
   $= \{D, A, B\}$ Rule

   (d) Find the closure of $\{C\}$, given the following FD's:
   A $\rightarrow$ BD
   AC $\rightarrow$ D
   D $\rightarrow$ A

   $\{C\}^+ = \{\}$ trivial
   $= \{C\}$ trivial

— Not doing all Keys =

4. Keys: In the questions below, you are given a relation R, its schema, and a set F of function dependencies. Identify all the keys of R.

(a) Find all keys of $R_1(A,B,C,D)$ given the following functional dependencies:

A $\rightarrow$ B
B $\rightarrow$ C
C $\rightarrow$ D

A = {A,B,C,D}   B = {B,C,D}   C = {C,D}   D = {D}
AB "   BC "   CD "
ABC "   BCD "
ABCD "

Superkey and Key

(b) Find all keys of $R_2(A,B,C,D)$ given the following set of functional dependencies:

A $\rightarrow$ B
B $\rightarrow$ AC
C $\rightarrow$ D

A = {A, B, C, D}   B = {A, B, C, D}

C = {C, D}   D = {D}

Superkey: A, B

Key: A, B

(c) Find all keys of $R_3(A,B,C)$ given the following set of functional dependencies:

A $\rightarrow$ B

A = A, B   AC = A, B, C   superkey, key
B = B   ABC = ABC   superkey

A, B = A, B

C = C

(d) Find all keys of $R_4(A,B,C,D,E)$ given the following set of functional dependencies:

A $\rightarrow$ C
B $\rightarrow$ BC
CD $\rightarrow$ E

A = A, C   ABCD = trivial   B = B, C   C = C   D = D
AB = A, B, C   ABCE = trivial   E = E
AC = A, C
AD = A, C, D, E
AE = A, C, E
ABC = A, B, C   NO   no
superkey and key   ABD = A, B, C, D, E ✓   AB   BD = B, C, D, E
ABE = A, B, C, E   AD
NO

Solutions

Question 1

Quite a few solutions are possible, depending on the assumptions of uniqueness. For example:

product_id → product_name

Each product only goes by one name and the product is uniquely identified by the product_id.

customer_id purchase_timestamp → shipping_address billing_address has_loyalty_number

While a customer may have multiple addresses, only one shipping and one billing address would be associated with a single transaction.

subtotal purchase_timestamp shipping_address has_loyalty_number → delivery_fee taxes total

This store calculates the delivery fee based on the subtotal for the order, where it is shipped to, whether the customer has a loyalty program discount, and the date of the order; moreover, because the subtotal, delivery_fee, and taxes determine the total (as in the first FD), we can also state that any two tuples with the same subtotal, purchase_timestamp, shipping_address, and has_loyalty_number will also have the same taxes and total.

Question 2

Part A

date_checked_out → due_date

No. It is likely that different items (e.g., a book and a ukulele) have different loan periods.

Part B

ISBN → title author publication_year

Yes, the ISBN uniquely identifies all of this information, down to the edition and format.

Part C

customer_id → hold_on_account

Probably yes. An example of "no" would be if this were historical transactions indicating a hold at the time.

Part D

author publication_year → title

No. Apparently Barbara Cartland wrote 23 novels per year.

Part E

author title → publication_year

No. It is possible to publish a second edition of a textbook, for example.

Part F

due_date → is_overdue

Yes. While is_overdue may change depending on the current date, it should be consistent for everything with the same due_date.

Question 3

Part A

$\{A, C\}^+ \supseteq \{A, C\}$ **// trivially**
$\{A, C\}^+ \supseteq \{A, B, C\}$ **// because A $\rightarrow$ B**
$\{A, C\}^+ \supseteq \{A, B, C, D\}$ **// because C $\rightarrow$ D**

Part B

$\{A\}^+ \supseteq \{A\}$ **// trivially**
$\{A\}^+ = \{A, B\}$ **// because A $\rightarrow$ B**

We cannot apply C $\rightarrow$ B because the antecedent, $\{C\}$, is not a subset of $\{A, B\}$. The fact that the consequent $\{B\}$ is a subset of $\{A, B\}$ doesn't mean that we can use the rule; it does, however, mean that the rule would not expand the closure, even if we could apply it.

Part C

$\{D\}^+ \supseteq \{D\}$ **// trivially**
$\{D\}^+ \supseteq \{A, D\}$ **// because D $\rightarrow$ A**
$\{D\}^+ \supseteq \{A, B, D\}$ **// because A $\rightarrow$ BD and** $\{A, D\} \cup \{B, D\} = \{A, B, D\}$
$\{D\}^+ = \{A, B, D\}$ **// because we cannot use the rule C $\rightarrow$ BD since** $\{C\} \not\subseteq \{A, B, D\}$

Part D

$\{C\}^+ \supseteq \{C\}$ **// trivially**
$\{C\}^+ = \{C\}$ **// because we cannot apply AC $\rightarrow$ D as we need both A and C, i.e.,** $\{A, C\} \not\subseteq \{C\}$

Question 4

Part A

We obtain the following closures:

$\{A\}^+ \supseteq \{A\}$ **// trivially**
$\{A\}^+ \supseteq \{A, B\}$ **// because A $\rightarrow$ B**
$\{A\}^+ \supseteq \{A, B, C\}$ **// because B $\rightarrow$ C**
$\{A\}^+ \supseteq \{A, B, C, D\}$ **// because C $\rightarrow$ D**
Therefore, $\{A\}$ **is a superkey**

$\{B\}^+ \supseteq \{B\}$ **// trivially**
$\{B\}^+ \supseteq \{B, C\}$ **// because B $\rightarrow$ C**
$\{B\}^+ \supseteq \{B, C, D\}$ **// because C $\rightarrow$ D**
At this point, we are stuck and cannot determine A. Therefore, $\{B\}$ **is not a superkey.**

$\{C\}^+ \supseteq \{B\}$ **// trivially**
$\{C\}^+ \supseteq \{C, D\}$ **// because C $\rightarrow$ D**
At this point, we are stuck and cannot determine A nor B. Therefore, $\{C\}$ **is not a superkey.**

$\{B, C\}^+ \supseteq \{B, C\}$ **// trivially**
$\{B, C\}^+ \supseteq \{B, C, D\}$ **// because C $\rightarrow$ D**
At this point, we are stuck and cannot determine A. Therefore, $\{B, C\}$ **is not a superkey.**

Any other subset of attributes is either a superset of $\{A\}$ **or does not enable us to use any new FD's compared to the attribute sets that we have already tried. Therefore, we can conclude that** $\{A\}$ **is the only key for** R_1**.**

Part B

Note that always $\{\}^+ = \{\}$; so, the empty set can only be a superkey for the empty relation $R()$. R_2 is not empty; so, any singleton set is minimal.

$\{A\}$ is a key given A $\rightarrow$ B, B $\rightarrow$ AC, and C $\rightarrow$ D and it is a singleton set (i.e., minimal).

$\{B\}^+ = \{B, A, C, D\}$, given B $\rightarrow$ AC and C $\rightarrow$ D. It is therefore a key because it is also a singleton set (i.e., minimal).

$\{C\}^+ = \{C, D\}$, given C $\rightarrow$ D. Therefore it is not a superkey.

- $\{C, A\}^+ = \{A, B, C, D\}$, because A $\rightarrow$ B. Therefore $\{C, A\}$ is a superkey. It is not a key because it is a proper superset of $\{A\}$, which is a key.
- $\{C, B\}^+ = \{A, B, C, D\}$, because B $\rightarrow$ AC. Therefore $\{C, B\}$ is a superkey. It is not a key because it is a proper superset of $\{B\}$, which is a key.

$\{D\}^+ = \{D\}$, because there are no FD's with $\{D\}$ as a superset of their antecedent. Therefore it is not a superkey.

- $\{D, A\}^+ = \{A, B, C, D\}$, because A $\rightarrow$ B and B $\rightarrow$ AC. Therefore $\{D, A\}$ is a superkey. It is not a key because it is a proper superset of $\{A\}$, which is a key.
- $\{D, B\}^+ = \{A, B, C, D\}$, because B $\rightarrow$ AC. Therefore $\{D, B\}$ is a superkey. It is not a key because it is a proper superset of $\{B\}$, which is a key.
- $\{D, C\}^+ = \{C, D\}$, because no FD's with an antecedent that is a subset of $\{C, D\}$ can expand the closure. Therefore $\{D, C\}$ is not a superkey.
- $\{D, C, A\}^+ = \{A, B, C, D\}$, because A $\rightarrow$ B. Therefore $\{D, C, A\}$ is a superkey, but not a key because it is a proper superset of $\{A\}$.
- $\{D, C, B\}^+ = \{A, B, C, D\}$, because B $\rightarrow$ AC. Therefore $\{D, C, B\}$ is a superkey, but not a key because it is a proper superset of $\{B\}$.

Part C

$\{A\}^+ = \{A, B\}$, given A $\rightarrow$ B. Therefore, it is not a superkey.

$\{A, C\}^+ = \{A, B, C\}$, given A $\rightarrow$ B. Therefore, it is a superkey. Moreover, it is minimal because neither $\{A\}$, $\{C\}$, nor $\{\}$ is a superkey

$\{A, B\}^+ = \{A, B\}$, since there are no FD's with $\{B\}$ as antecedent. Therefore, it is not a superkey

$\{A, B, C\}^+ = \{A, B, C\}$, trivially. It is a superkey but also a proper superset of the key $\{A, C\}$.

We can confirm similarly that the smallest sets containing B or C that are superkeys are ones that we have already found above. Therefore, $\{A, C\}$ is the only key for R_3.

Part D

Using a similar process to the above questions, we can confirm the following keys: $\{A, B, D\}$.