ICS 321 Data Storage & Retrieval

Functional Dependencies

Prof. Lipyeow Lim
Information & Computer Science Department
University of Hawaii at Manoa
Example: Movies1

<table>
<thead>
<tr>
<th>title</th>
<th>year</th>
<th>length</th>
<th>genre</th>
<th>studioName</th>
<th>starName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Carrie Fisher</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Mark Hamill</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Harrison Ford</td>
</tr>
<tr>
<td>Gone With the Wind</td>
<td>1939</td>
<td>231</td>
<td>drama</td>
<td>MGM</td>
<td>Vivien Leigh</td>
</tr>
<tr>
<td>Wayne’s World</td>
<td>1992</td>
<td>95</td>
<td>comedy</td>
<td>Paramount</td>
<td>Dana Carvey</td>
</tr>
<tr>
<td>Wayne’s World</td>
<td>1992</td>
<td>95</td>
<td>comedy</td>
<td>Paramount</td>
<td>Mike Meyers</td>
</tr>
</tbody>
</table>

- What are the keys for this relation?
- What if you ignore the column starName?
- Can starName be a key?
Functional Dependency

• A **functional dependency** $X \rightarrow Y$ holds over relation $R$ if, for every allowable instance $r$ of $R$:
  – for all tuples $t1, t2$ in $r$,
    
    $\pi_X(t1) = \pi_X(t2)$ implies $\pi_Y(t1) = \pi_Y(t2)$
  – i.e., given two tuples in $r$, if the $X$ values agree, then the $Y$ values must also agree. ($X$ and $Y$ are sets of attributes.)

• An FD is a statement about *all* allowable instances.
  – Must be identified based on semantics of application.
  – Given some allowable instance $r1$ of $R$, we can check if it violates some FD $f$, but we cannot tell if $f$ holds over $R$!

• $K$ is a candidate key for $R$ means that $K \rightarrow R$
  – However, $K \rightarrow R$ does not require $K$ to be *minimal*!
Keys & Superkeys

- A set of one or more attributes \( \{A_1, A_2, \ldots, A_n\} \) is a key for a relation \( R \) if:
  - 1. Those attributes functionally determine all other attributes of the relation.
  - 2. No proper subset of \( \{A_1, A_2, \ldots, A_n\} \) functionally determines all other attributes of \( R \).
- A key must be minimal.
- When a key consists of a single attribute \( A \), we often say that \( A \) (rather than \( \{A\} \)) is a key.
- **Superkey**: a set of attributes that contain a key.
### FD Example: Movies

<table>
<thead>
<tr>
<th>title</th>
<th>year</th>
<th>length</th>
<th>genre</th>
<th>studioName</th>
<th>starName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Carrie Fisher</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Mark Hamill</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>124</td>
<td>SciFi</td>
<td>Fox</td>
<td>Harrison Ford</td>
</tr>
<tr>
<td>Gone With the Wind</td>
<td>1939</td>
<td>231</td>
<td>drama</td>
<td>MGM</td>
<td>Vivien Leigh</td>
</tr>
<tr>
<td>Wayne’s World</td>
<td>1992</td>
<td>95</td>
<td>comedy</td>
<td>Paramount</td>
<td>Dana Carvey</td>
</tr>
<tr>
<td>Wayne’s World</td>
<td>1992</td>
<td>95</td>
<td>comedy</td>
<td>Paramount</td>
<td>Mike Meyers</td>
</tr>
</tbody>
</table>

- What are the FDs for this relation?
- What are the keys for this relation?
- Can starName be a key?
Reasoning about FDs

• Given some FDs, we can usually infer additional FDs:
  – \( ssn \rightarrow deptID, \) \( deptID \rightarrow building \) implies \( ssn \rightarrow building \)

• \( T \) implies \( S \), or \( S \) follows from \( T \)
  – Every relation instance that satisfies all the FDs in \( T \) also satisfies all the FDs in \( S \)

• \( S \) is equivalent to \( T \)
  – The set of relation instances satisfying \( S \) is exactly the same as the set satisfying \( T \)
  – Alternatively, \( S \) implies \( T \) AND \( T \) implies \( S \)
Armstrong’s Axioms

Let X, Y, Z are sets of attributes:

• **Reflexivity**
  – If X is a subset of Y, then Y -> X

• **Augmentation**
  – If X -> Y, then XZ -> YZ for any Z

• **Transitivity**
  – If X -> Y and Y -> Z, then X -> Z

These are *sound* and *complete* inference rules for FDs!
Example: Armstrong’s Axioms

Hourly_Emps

<table>
<thead>
<tr>
<th>SSN</th>
<th>Name</th>
<th>Lot</th>
<th>Rating</th>
<th>Hourly_Wages</th>
<th>Hours_worked</th>
</tr>
</thead>
<tbody>
<tr>
<td>123-22-2366</td>
<td>Attishoo</td>
<td>48</td>
<td>8</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>231-31-5368</td>
<td>Smiley</td>
<td>22</td>
<td>8</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>131-24-3650</td>
<td>Smethurst</td>
<td>35</td>
<td>5</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>434-26-3751</td>
<td>Guldu</td>
<td>35</td>
<td>5</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>612-67-4134</td>
<td>Madayan</td>
<td>35</td>
<td>8</td>
<td>10</td>
<td>40</td>
</tr>
</tbody>
</table>

- **Reflexivity**: If X is a subset of Y, then Y -> X
  - SNLR is a subset of SNLRWH, SNLRWH -> SNLR
- **Augmentation**: If X -> Y, then XZ -> YZ for any Z
  - S -> N, then SLR -> NLR
- **Transitivity**: If X -> Y and Y -> Z, then X -> Z
  - S -> R, R -> W, then S -> W
Two More Rules

<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>DOB</th>
<th>Address</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Smith</td>
<td>Sep 9 1979</td>
<td>Honolulu,HI</td>
<td>808-343-0809</td>
</tr>
</tbody>
</table>

- **Union / Combining**
  - If \( X \rightarrow Y \) and \( X \rightarrow Z \), then \( X \rightarrow YZ \)
  - Eg. FLD \( \rightarrow \) A and FLD \( \rightarrow \) T, then FLD \( \rightarrow \) AT

- **Decomposition / Splitting**
  - If \( X \rightarrow YZ \), then \( X \rightarrow Y \) and \( X \rightarrow Z \)
  - Eg. FLD \( \rightarrow \) AT , then FLD \( \rightarrow \) A and FLD \( \rightarrow \) T

- **Trivial FDs**
  - Right side is a subset of Left side
  - Eg. F \( \rightarrow \) F, FLD \( \rightarrow \) FD

- **Does “XY \( \rightarrow \) Z imply X \( \rightarrow \)Z and Y \( \rightarrow \)Z” ?**
Closure

• **Implication**: An FD $f$ is *implied by* a set of FDs $F$ if $f$ holds whenever all FDs in $F$ hold.
  
  – $f=A \rightarrow C$ is *implied by* $F=\{ A\rightarrow B, B \rightarrow C \}$ (using Armstrong’s transitivity)

• **Closure $F^+$**: the set of all FDs implied by $F$
  
  – **Algorithm**:
    
    • start with $F^+=F$
    
    • keep adding new implied FDs to $F^+$ by applying the 5 rules (Armstrong’s Axioms + union + decomposition)
    
    • Stop when $F^+$ does not change anymore.
Example: Closure

- Given FLD is the primary key and C → Z
- Find the closure:
  - Start with \{ FLD → FLDSCZT, C → Z \}
  - Applying reflexivity, \{ FLD → F, FLD → L, FLD → D, FLD → FL, FLD → LD, FLD → DF, FLDSCZT → FLD, ... \}
  - Applying augmentation, \{ FLDS → FS, FLDS → LS, ... \}
  - Applying transitivity ...
  - Applying union ...
  - Applying decomposition ...
  - Repeat until \( F^+ \) does not change
Attribute Closure

• Computing the closure of a set of FDs can be expensive. (Size of closure is exponential in # attrs!)

• Typically, we just want to check if a given FD $X \rightarrow Y$ is in the closure of a set of FDs $F$. An efficient check:
  – Compute *attribute closure* of $X$ (denoted $X^+$) wrt $F$:
    • Set of all attributes $A$ such that $X \rightarrow A$ is in $F^+$
    • There is a linear time algorithm to compute this.
  – Check if $Y$ is in $X^+$

• Does $F = \{ A \rightarrow B, \ B \rightarrow C, \ C \ D \rightarrow E \}$ imply $A \rightarrow E$?
  – i.e., is $A \rightarrow E$ in the closure $F^+$? Equivalently, is $E$ in $A^+$?