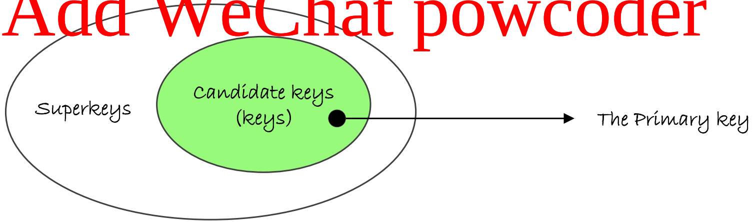
Functional Dependencies – Part 3

Finding Keys

A Bunch of Keys

- We will need keys for defining the normal forms later on.
  - *A subset of the attributes* of a relation schema R is a **superkey** if it uniquely determines all attributes of R.
  - A superkey K is called a **candidate key** if no proper subset of K is a superkey.
    - That is, if you take any of the attributes out of K, then there is not enough to uniquely identify tuples.
  - **Candidate keys** are also called **keys**, and the **primary key** is chosen from them.



Finding Keys

- Given a set Σ of FDs on a relation R, the question is:

How can we find all the (candidate) keys of R?

Implied Functional Dependencies

- To design a good database, we need to consider **all possible FDs**.

- If each student works on one project and each project has one supervisor, does each student have one project supervisor?

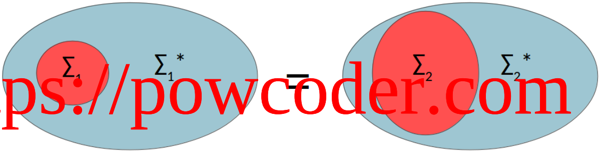
$$\begin{array}{c}\{\{\text{StudentID}\}\rightarrow\{\text{ProjectNo}\},\\ \{\text{ProjectNo}\}\rightarrow\{\text{Supervisor}\}\}\end{array} \models \{\text{StudentID}\}\rightarrow\{\text{Supervisor}\}$$

- We use the notation $\Sigma \models X \rightarrow Y$ to denote that $X \rightarrow Y$ is **implied** by the set Σ of FDs.

- We write Σ^* for all possible FDs **implied** by Σ.

Equivalence of Functional Dependencies

- Σ_1 and Σ_2 are **equivalent** if $\Sigma_1^* = \Sigma_2^*$.



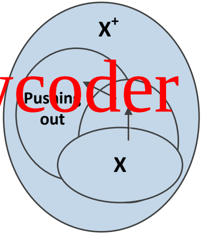
- **Example:** Let $\Sigma_1 = \{X \rightarrow Y, Y \rightarrow Z\}$ and $\Sigma_2 = \{X \rightarrow Y, Y \rightarrow Z, X \rightarrow Z\}$. We have $\Sigma_1 \neq \Sigma_2$ but $\Sigma_1^* = \Sigma_2^* = \{X \rightarrow Y, Y \rightarrow Z, X \rightarrow Z\}$. Hence, Σ_1 and Σ_2 are equivalent.

- **Questions:**
  1. Is it possible that $\Sigma_1^* = \Sigma_2^*$ but $\Sigma_1 \neq \Sigma_2$? **Yes**
  2. Is it possible that $\Sigma_1^* \neq \Sigma_2^*$ but $\Sigma_1 = \Sigma_2$? **No**

Implied Functional Dependencies

- Let Σ be a set of FDs. Check whether or not $\Sigma \models X \rightarrow W$ holds? We need to
  1. Compute **the set of all attributes** that are dependent on X, which is called the **closure** of X under Σ and is denoted by X^+.
  2. $\Sigma \models X \rightarrow W$ holds iff $W \subseteq X^+$.

- **Algorithm**[1]
  - $X^+ := X$;
  - repeat until no more change on X^+
    - for each $Y \rightarrow Z \in \Sigma$ with $Y \subseteq X^+$, add all the attributes in Z to X^+, i.e., replace X^+ by $X^+ \cup Z$.



[1] See Algorithm 15.1 on Page 538 in [Elmasri & Navathe, 7th edition] or Algorithm 1 on Page 555 in [Elmasri & Navathe, 6th edition]

Implied Functional Dependencies – Example

- Consider a relation schema $R = \{A, B, C, D, E, F\}$, a set of FDs $\Sigma = \{AC \rightarrow B, B \rightarrow CD, C \rightarrow E, AF \rightarrow B\}$ on R.
- Decide whether or not $\Sigma \models AC \rightarrow ED$ holds.

1. We first build the closure of AC:

$(AC)^+$	$\supseteq AC$	initialisation
	$\supseteq ACB$	using $AC \rightarrow B$
	$\supseteq ACBD$	using $B \rightarrow CD$
	$\supseteq ACBDE$	using $C \rightarrow E$
	$= ACBDE$	

2. Then we check that $ED \subseteq (AC)^+$. Hence $\Sigma \models AC \rightarrow ED$.

- **Can you quickly tell whether or not $\Sigma \models AC \rightarrow EF$ holds?**

Finding Keys

- **Fact:** A key K of R always defines a FD $K \rightarrow R$.
- **Algorithm**[2]:

  Input: a set Σ of FDs on R.

  Output: the set of all keys of R.

  - for every subset X of the relation R, compute its closure X^+
  - if $X^+ = R$, then X is a superkey.

  - if no proper subset Y of X with $Y^+ = R$, then X is a key.

- A **prime attribute** is an attribute occurring in a key, and a **non-prime attribute** is an attribute that is not a prime attribute.

[2] It extends Algorithm 15.2(a) in [Elmasri & Navathe, 7th edition, pp. 542], or Algorithm 2(a) or in Algorithm 2(a) in [Elmasri & Navathe, 6th edition pp. 558] to finding all keys of R

Exercise – Finding Keys

- Consider a relation schema $R = \{A, B, C, D\}$ and a set of functional dependencies $\Sigma = \{AB \rightarrow C, AC \rightarrow D\}$.
  1. List all the keys and superkeys of R.
  2. Find all the prime attributes of R.
- **Solution:**

  1. We compute the closures for all possible combinations of the attributes in R:
     - $(A)^+ = A$, $(B)^+ = B$, $(C)^+ = C$, $(D)^+ = D$;
     - $(AB)^+ = ABCD$, $(AC)^+ = ACD$, $(AD)^+ = AD$, $(BC)^+ = BC$, $(BD)^+ = BD$, $(CD)^+ = CD$
     - $(ABC)^+ = ABCD$, $(ABD)^+ = ABCD$, $(ACD)^+ = ACD$, $(BCD)^+ = BCD$

  2. Hence, we have
     - AB is the only key of R.
     - AB, ABC, ABD and $ABCD$ are the superkeys of R.
     - A and B are the prime attributes of R.

Exercise – Finding Keys

- Checking all possible combinations of the attributes is too tedious!

  Example: Still consider a relation schema $R = \{A, B, C, D\}$ and $\Sigma = \{AB \rightarrow C, AC \rightarrow D\}$. List all the keys of R.

- **Some tricks:**
  - If an attribute *never* appears in the dependent of any FD, this attribute must **be part of each key**.
  - If an attribute *never* appears in the determinant of any FD but appears in the dependent of any FD, this attribute must **not be part of each key**.
  - If a proper subset of X is a key, then X must **not be a key**.

Finding Keys - Example

- Consider ENROLMENT and the following FDs:
  - {StudentID} → {Name};
  - {StudentID, CourseNo, Semester} → {ConfirmedBy, Office};
  - {ConfirmedBy} → {Office}.

ENROLMENT					
Name	StudentID	CourseNo	Semester	ConfirmedBy	Office
Tom	123456	COMP2400	2010 S2	Jane	R301
Mike	123458	COMP2400	2008 S2	Linda	R203
Mike	123458	COMP2600	2008 S2	Linda	R203

- What are the keys, superkeys and prime attributes of ENROLMENT?
  - {StudentID, CourseNo, Semester} is the only key.
  - Every set that has {StudentID, CourseNo, Semester} as its subset is a superkey.
  - StudentID, CourseNo and Semester are the prime attributes.