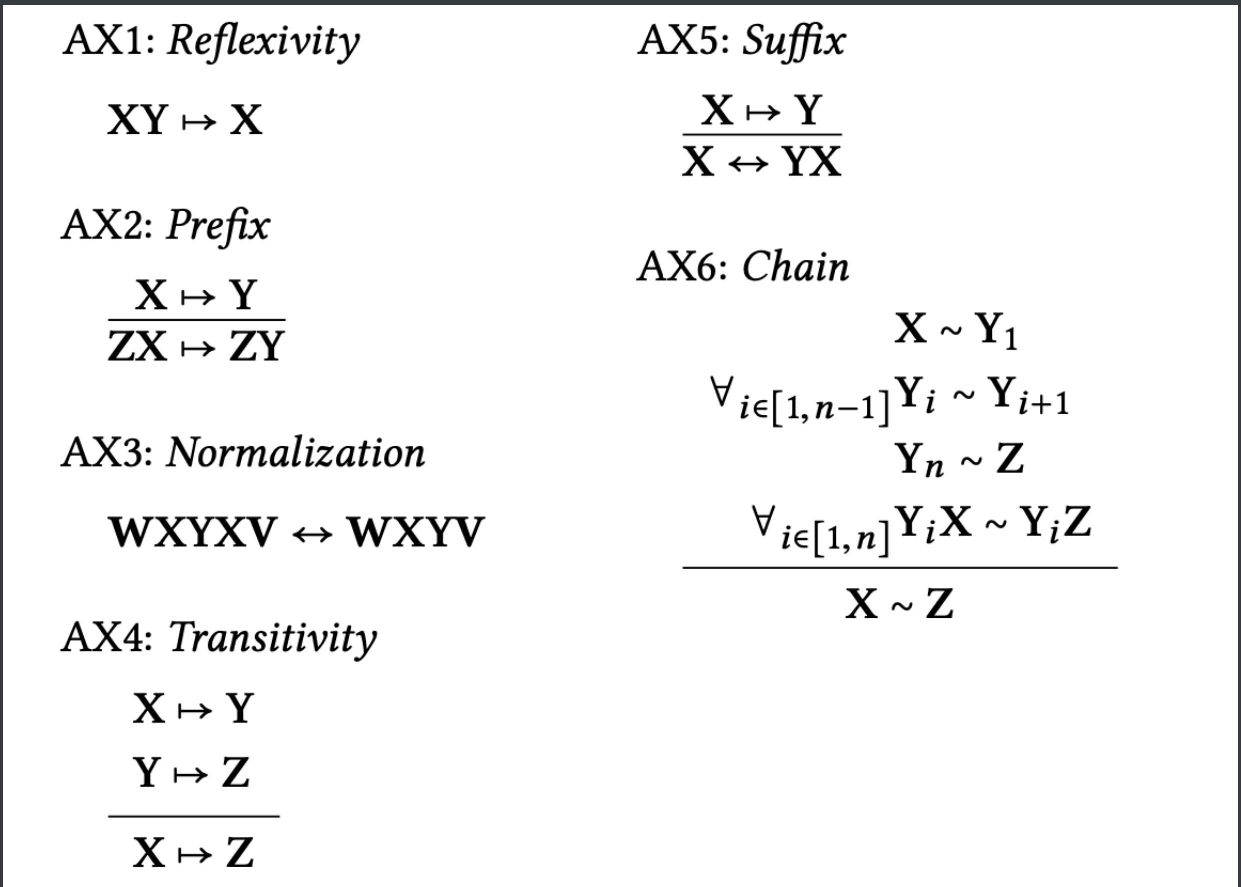
Discover Order Dependencies through Order Compatibility笔记

重要公理



AX1: *Reflexivity*

$$\mathbf{XY} \mapsto \mathbf{X}$$

AX2: *Prefix*

$$\frac{\mathbf{X} \mapsto \mathbf{Y}}{\mathbf{ZX} \mapsto \mathbf{ZY}}$$

AX3: *Normalization*

$$\mathbf{WXYXV} \leftrightarrow \mathbf{WXYV}$$

AX4: *Transitivity*

$$\frac{\begin{array}{c}\mathbf{X} \mapsto \mathbf{Y}\\ \mathbf{Y} \mapsto \mathbf{Z}\end{array}}{\mathbf{X} \mapsto \mathbf{Z}}$$

AX5: *Suffix*

$$\frac{\mathbf{X} \mapsto \mathbf{Y}}{\mathbf{X} \leftrightarrow \mathbf{YX}}$$

AX6: *Chain*

$$\frac{\begin{array}{c}\mathbf{X} \sim \mathbf{Y}_1\\ \forall_{i \in [1,n-1]} \mathbf{Y}_i \sim \mathbf{Y}_{i+1}\\ \mathbf{Y}_n \sim \mathbf{Z}\\ \forall_{i \in [1,n]} \mathbf{Y}_i\mathbf{X} \sim \mathbf{Y}_i\mathbf{Z}\end{array}}{\mathbf{X} \sim \mathbf{Z}}$$

~~公理1～公理4会证明~~

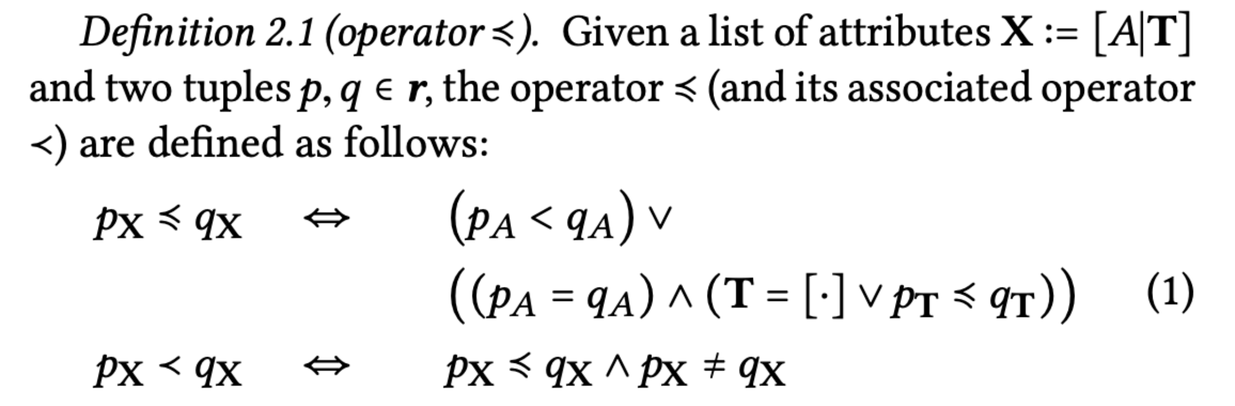
公理5～公理6还不会

重要定义

p_A： 元组p在属性A上的投影

投影：从原有关系生产一个新的关系，包含原来关系的部分列

定义2.1 操作符号

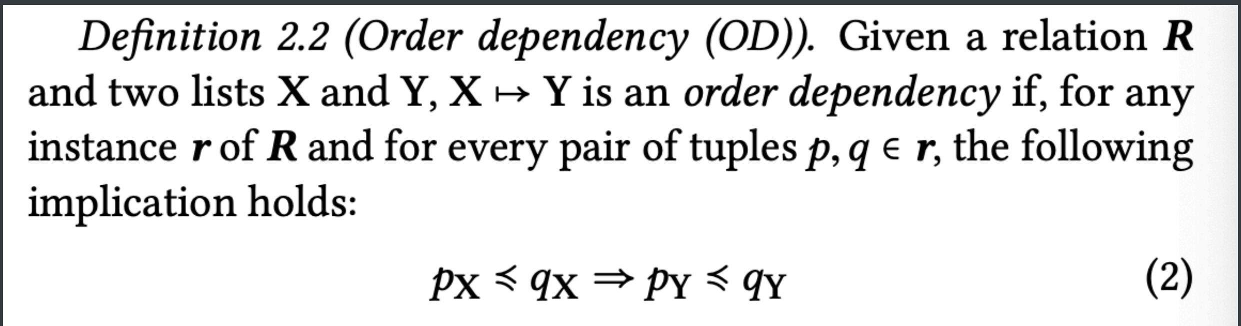


Definition 2.1 (operator ≼). Given a list of attributes $\mathbf{X} := [A|\mathbf{T}]$ and two tuples $p, q \in \boldsymbol{r}$, the operator ≼ (and its associated operator ≺) are defined as follows:

$$\begin{aligned} p_{\mathbf{X}} \preccurlyeq q_{\mathbf{X}} \quad &\Leftrightarrow \quad \big(p_A < q_A\big) \vee \\ &\quad\;\; \big(\big(p_A = q_A\big) \wedge \big(\mathbf{T} = [\cdot] \vee p_{\mathbf{T}} \preccurlyeq q_{\mathbf{T}}\big)\big) \\ p_{\mathbf{X}} \prec q_{\mathbf{X}} \quad &\Leftrightarrow \quad p_{\mathbf{X}} \preccurlyeq q_{\mathbf{X}} \wedge p_{\mathbf{X}} \neq q_{\mathbf{X}} \end{aligned} \tag{1}$$

该操作符与将数据按字典序排序效果一样，即将X按字典序进行排序

定义2.2 顺序依赖(OD)



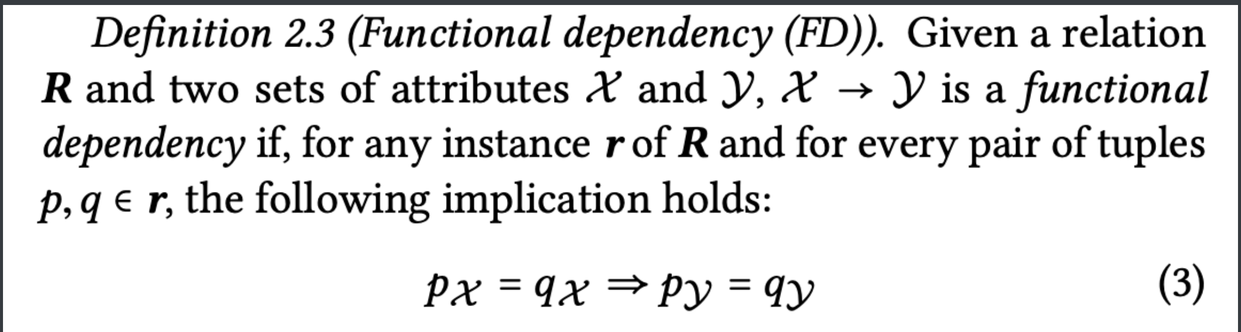
Definition 2.2 (Order dependency (OD)). Given a relation $\boldsymbol{R}$ and two lists $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X} \mapsto \mathbf{Y}$ is an *order dependency* if, for any instance $\boldsymbol{r}$ of $\boldsymbol{R}$ and for every pair of tuples $p, q \in \boldsymbol{r}$, the following implication holds:

$$p_{\mathbf{X}} \preccurlyeq q_{\mathbf{X}} \Rightarrow p_{\mathbf{Y}} \preccurlyeq q_{\mathbf{Y}} \tag{2}$$

即对于关系模式R中的每一个r，在对X属性列进行字典序排序的同时也会将Y属性列进行字典序排序。记为：

$$X \mapsto Y$$

定义2.3 函数依赖(FD)



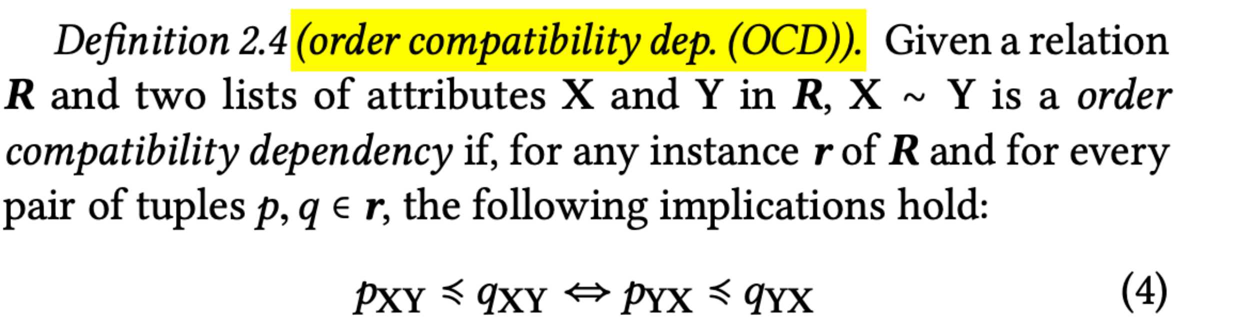
Definition 2.3 (Functional dependency (FD)). Given a relation $\boldsymbol{R}$ and two sets of attributes $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{X} \rightarrow \mathcal{Y}$ is a *functional dependency* if, for any instance $\boldsymbol{r}$ of $\boldsymbol{R}$ and for every pair of tuples $p, q \in \boldsymbol{r}$, the following implication holds:

$$p_{\mathcal{X}} = q_{\mathcal{X}} \Rightarrow p_{\mathcal{Y}} = q_{\mathcal{Y}} \tag{3}$$

即每一个X属性列的值相等时，Y属性列的值也相等。

表现为函数关系：一个X对应一个Y。

定义2.4 函数一致性依赖(OCD)



Definition 2.4 (order compatibility dep. (OCD)). Given a relation $\boldsymbol{R}$ and two lists of attributes $\mathbf{X}$ and $\mathbf{Y}$ in $\boldsymbol{R}$, $\mathbf{X} \sim \mathbf{Y}$ is a *order compatibility dependency* if, for any instance $\boldsymbol{r}$ of $\boldsymbol{R}$ and for every pair of tuples $p, q \in \boldsymbol{r}$, the following implications hold:

$$p_{\mathbf{XY}} \preccurlyeq q_{\mathbf{XY}} \Leftrightarrow p_{\mathbf{YX}} \preccurlyeq q_{\mathbf{YX}} \tag{4}$$

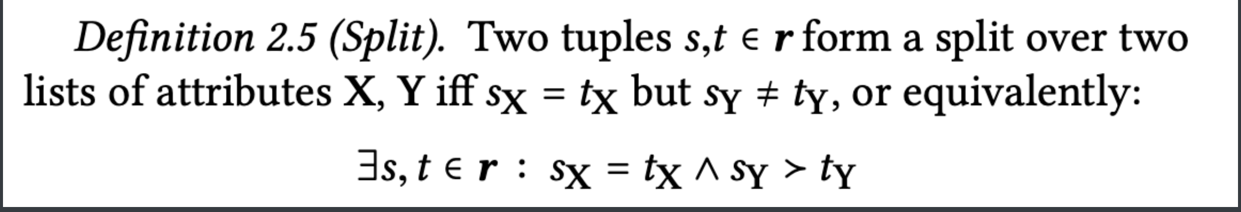
即对XY(YX)属于列进行字典序排序的同时也会对YX(XY)属性列进行字典序排序。

表达式为：$XY \mapsto YX \wedge YX \mapsto XY$，也记作$X \sim Y$

查看两列之间的OCD的另一种方法是，当将它们视为一对时，它们的值都是单调非递减的。

OCD编码为这样的事实：在一个关系中，两个属性列表显示相同的单调性。如果我们以非递减顺序对列表的任何一个进行排序，则它们最后都变为两个顺序排序，因此它们的值都是单调非递减的。

定义2.5 Split

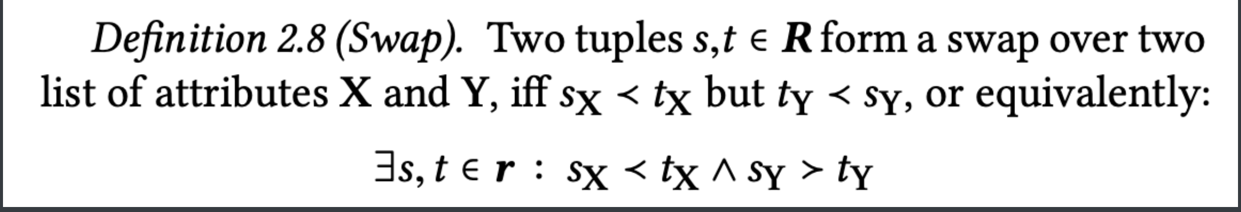


Definition 2.5 (Split). Two tuples $s,t \in \boldsymbol{r}$ form a split over two lists of attributes $\mathbf{X}$, $\mathbf{Y}$ iff $s_{\mathbf{X}} = t_{\mathbf{X}}$ but $s_{\mathbf{Y}} \neq t_{\mathbf{Y}}$, or equivalently:

$$\exists s, t \in \boldsymbol{r} \ : \ s_{\mathbf{X}} = t_{\mathbf{X}} \wedge s_{\mathbf{Y}} > t_{\mathbf{Y}}$$

即对于(X,Y)这个属性列，两个元组"前等后不等"。

定义2.8 Swap

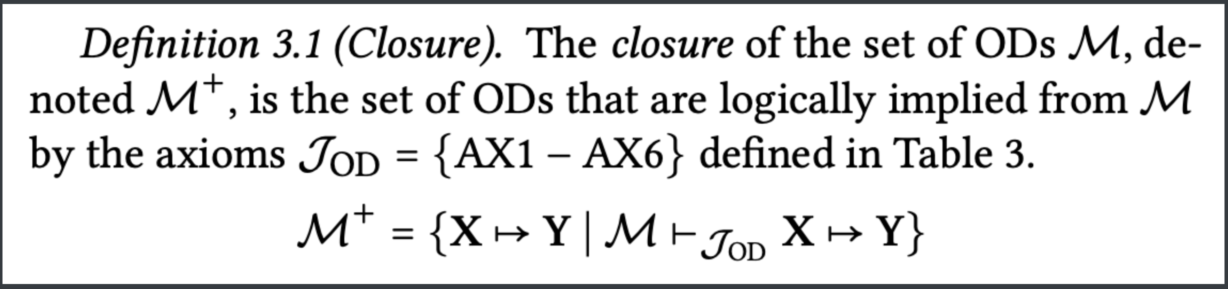


Definition 2.8 (Swap). Two tuples $s,t \in \boldsymbol{R}$ form a swap over two list of attributes $\mathbf{X}$ and $\mathbf{Y}$, iff $s_{\mathbf{X}} \prec t_{\mathbf{X}}$ but $t_{\mathbf{Y}} \prec s_{\mathbf{Y}}$, or equivalently:

$$\exists s, t \in \boldsymbol{r} \ : \ s_{\mathbf{X}} \prec t_{\mathbf{X}} \wedge s_{\mathbf{Y}} \succ t_{\mathbf{Y}}$$

即对于(X,Y)这个属性列，两个元组"前小后大"。

定义3.1 闭包(Clouse)

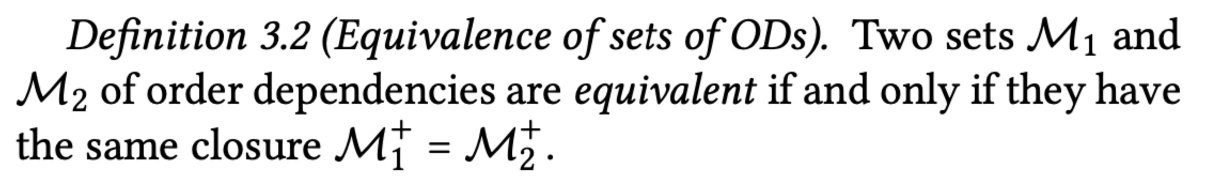


Definition 3.1 (Closure). The *closure* of the set of ODs $\mathcal{M}$, denoted $\mathcal{M}^+$, is the set of ODs that are logically implied from $\mathcal{M}$ by the axioms $\mathcal{J}_{\text{OD}} = \{\text{AX1} - \text{AX6}\}$ defined in Table 3.

$$\mathcal{M}^+ = \{\mathbf{X} \mapsto \mathbf{Y} \mid \mathcal{M} \vdash_{\mathcal{J}_{\text{OD}}} \mathbf{X} \mapsto \mathbf{Y}\}$$

一个ODs集合的闭包是该集合通过上述公理可以推导出来的所有ODs的集合。

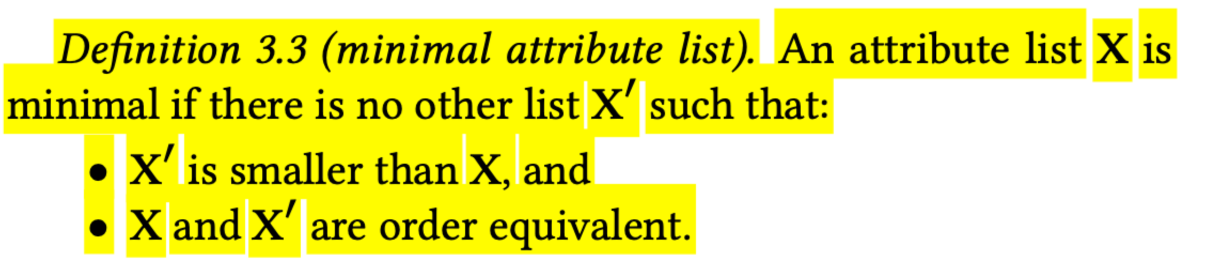
定义3.2 等价ODs集合



Definition 3.2 (Equivalence of sets of ODs). Two sets $\mathcal{M}_1$ and $\mathcal{M}_2$ of order dependencies are *equivalent* if and only if they have the same closure $\mathcal{M}_1^+ = \mathcal{M}_2^+$.

当且仅当两个ODs集合的闭包相等时，这两个ODs集合等价

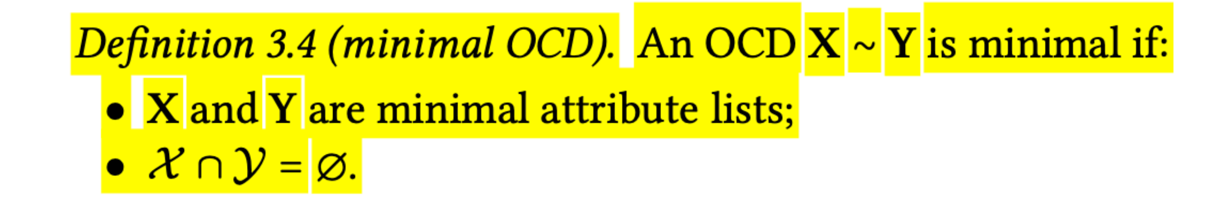
定义3.3 最小属性列



Definition 3.3 (minimal attribute list). An attribute list $\mathbf{X}$ is minimal if there is no other list $\mathbf{X}'$ such that:

- $\mathbf{X}'$ is smaller than $\mathbf{X}$, and
- $\mathbf{X}$ and $\mathbf{X}'$ are order equivalent.

定义3.4 最小OCD

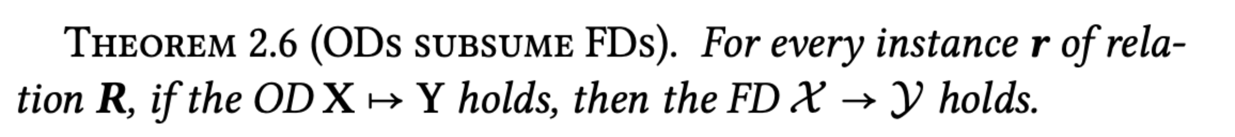


Definition 3.4 (minimal OCD). An OCD $\mathbf{X} \sim \mathbf{Y}$ is minimal if:

- $\mathbf{X}$ and $\mathbf{Y}$ are minimal attribute lists;
- $\mathcal{X} \cap \mathcal{Y} = \varnothing$.

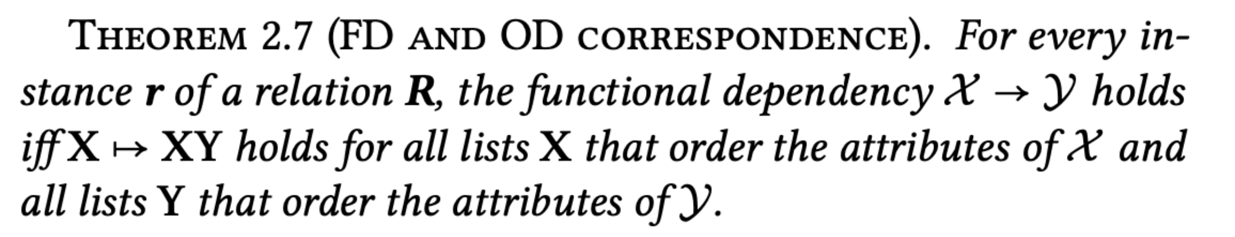
重要定理

定理2.6 OD包含FD



Theorem 2.6 (ODs subsume FDs). *For every instance* $\mathbf{r}$ *of relation* $\mathbf{R}$, *if the OD* $\mathbf{X} \mapsto \mathbf{Y}$ *holds, then the FD* $\mathcal{X} \rightarrow \mathcal{Y}$ *holds.*

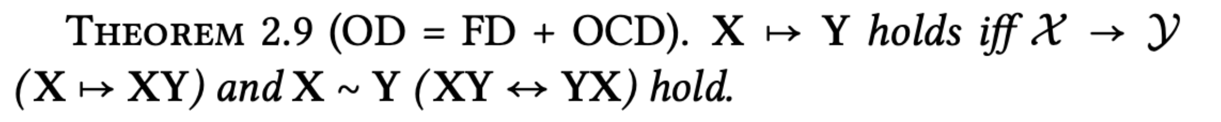
定理2.7 当且仅当OD$X \mapsto XY$，FD$X \rightarrow Y$成立



Theorem 2.7 (FD and OD correspondence). *For every instance* $\mathbf{r}$ *of a relation* $\mathbf{R}$, *the functional dependency* $\mathcal{X} \rightarrow \mathcal{Y}$ *holds iff* $\mathbf{X} \mapsto \mathbf{XY}$ *holds for all lists* $\mathbf{X}$ *that order the attributes of* $\mathcal{X}$ *and all lists* $\mathbf{Y}$ *that order the attributes of* $\mathcal{Y}$.

尝试证明

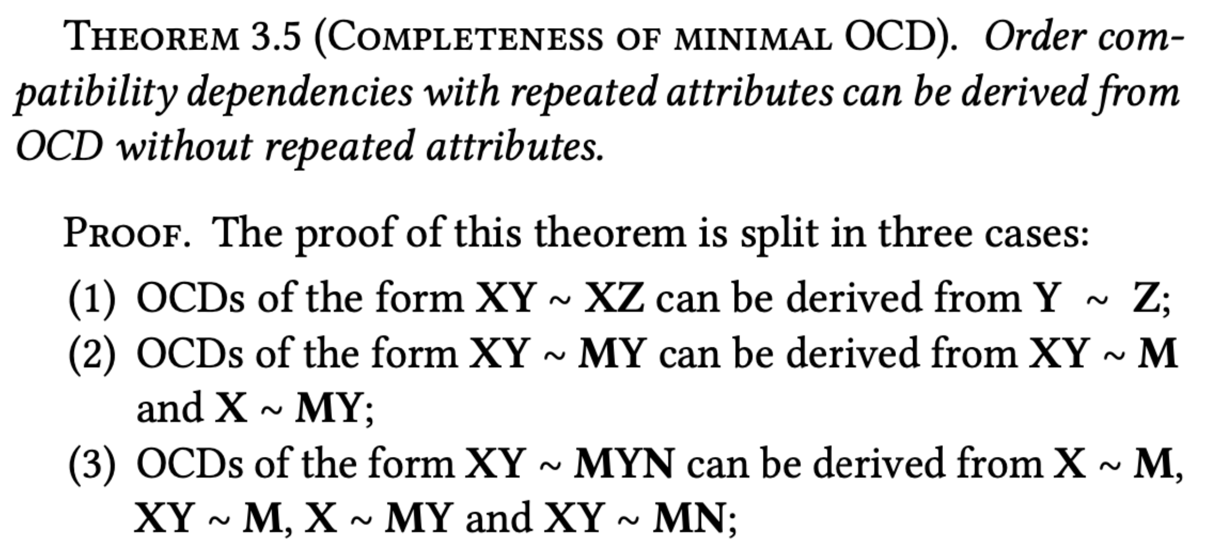
定理2.9 OD=FD + OCD



THEOREM 2.9 (OD = FD + OCD). $\mathbf{X} \mapsto \mathbf{Y}$ *holds iff* $\mathcal{X} \to \mathcal{Y}$ $(\mathbf{X} \mapsto \mathbf{XY})$ *and* $\mathbf{X} \sim \mathbf{Y}$ $(\mathbf{XY} \leftrightarrow \mathbf{YX})$ *hold.*

- FD $X \to Y$有效，表明没有split
- OCD $X \sim Y$有效，表明没有swap

定理3.5 有重复属性的OCD可由没有重复元素的OCD推导得出



THEOREM 3.5 (COMPLETENESS OF MINIMAL OCD). *Order compatibility dependencies with repeated attributes can be derived from OCD without repeated attributes.*

PROOF. The proof of this theorem is split in three cases:

(1) OCDs of the form $\mathbf{XY} \sim \mathbf{XZ}$ can be derived from $\mathbf{Y} \sim \mathbf{Z}$;

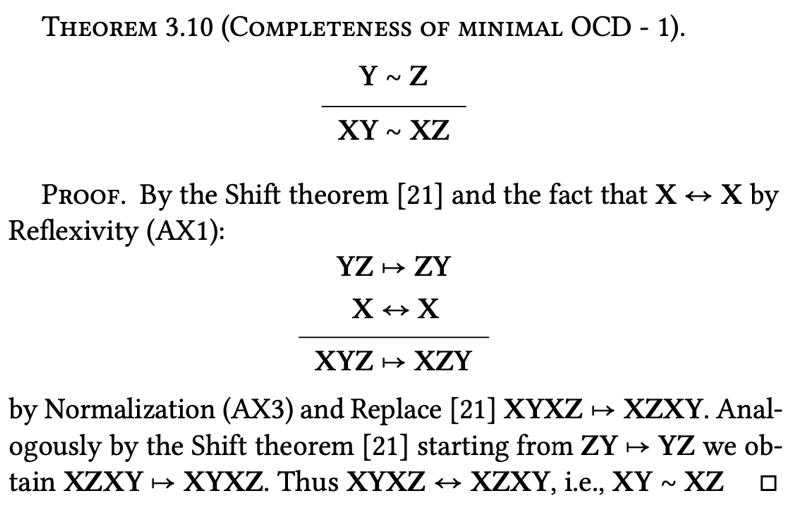
(2) OCDs of the form $\mathbf{XY} \sim \mathbf{MY}$ can be derived from $\mathbf{XY} \sim \mathbf{M}$ and $\mathbf{X} \sim \mathbf{MY}$;

(3) OCDs of the form $\mathbf{XY} \sim \mathbf{MYN}$ can be derived from $\mathbf{X} \sim \mathbf{M}$, $\mathbf{XY} \sim \mathbf{M}$, $\mathbf{X} \sim \mathbf{MY}$ and $\mathbf{XY} \sim \mathbf{MN}$;

证明（2，3不会证）：

分为三种情况：

1. 证明开头具有重复属性的属性列表是多余的：



THEOREM 3.10 (COMPLETENESS OF MINIMAL OCD - 1).

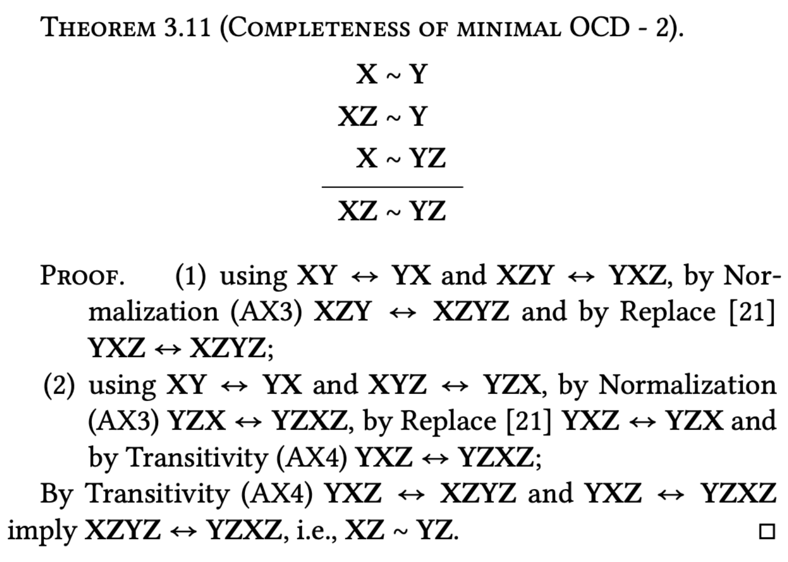
$$\frac{\mathbf{Y} \sim \mathbf{Z}}{\mathbf{XY} \sim \mathbf{XZ}}$$

PROOF. By the Shift theorem [21] and the fact that $\mathbf{X} \leftrightarrow \mathbf{X}$ by Reflexivity (AX1):

$$\frac{\begin{array}{c}\mathbf{YZ} \mapsto \mathbf{ZY} \\ \mathbf{X} \leftrightarrow \mathbf{X}\end{array}}{\mathbf{XYZ} \mapsto \mathbf{XZY}}$$

by Normalization (AX3) and Replace [21] $\mathbf{XYXZ} \mapsto \mathbf{XZXY}$. Analogously by the Shift theorem [21] starting from $\mathbf{ZY} \mapsto \mathbf{YZ}$ we obtain $\mathbf{XZXY} \mapsto \mathbf{XYXZ}$. Thus $\mathbf{XYXZ} \leftrightarrow \mathbf{XZXY}$, i.e., $\mathbf{XY} \sim \mathbf{XZ}$ □

2. 证明结尾具有重复属性的属性列表是多余的：



Theorem 3.11 (Completeness of minimal OCD - 2).

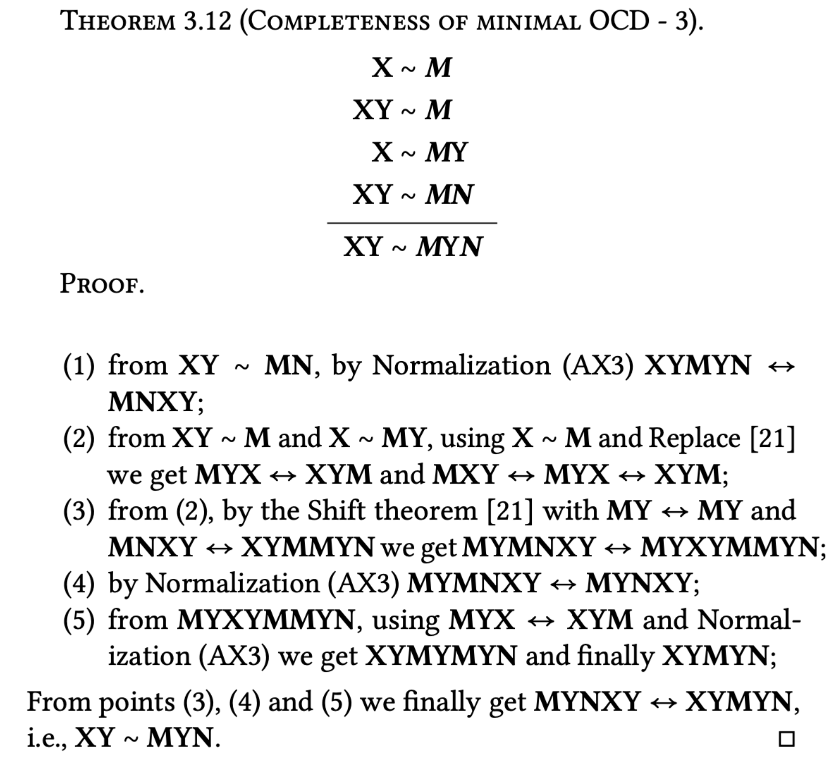
$$\frac{\begin{array}{c} \mathbf{X} \sim \mathbf{Y} \\ \mathbf{XZ} \sim \mathbf{Y} \\ \mathbf{X} \sim \mathbf{YZ} \end{array}}{\mathbf{XZ} \sim \mathbf{YZ}}$$

Proof. (1) using **XY** ↔ **YX** and **XZY** ↔ **YXZ**, by Normalization (AX3) **XZY** ↔ **XZYZ** and by Replace [21] **YXZ** ↔ **XZYZ**;

(2) using **XY** ↔ **YX** and **XYZ** ↔ **YZX**, by Normalization (AX3) **YZX** ↔ **YZXZ**, by Replace [21] **YXZ** ↔ **YZX** and by Transitivity (AX4) **YXZ** ↔ **YZXZ**;

By Transitivity (AX4) **YXZ** ↔ **XZYZ** and **YXZ** ↔ **YZXZ** imply **XZYZ** ↔ **YZXZ**, i.e., **XZ** ~ **YZ**. □

3. 证明中间具有重复属性的属性列表是多余的：



Theorem 3.12 (Completeness of minimal OCD - 3).

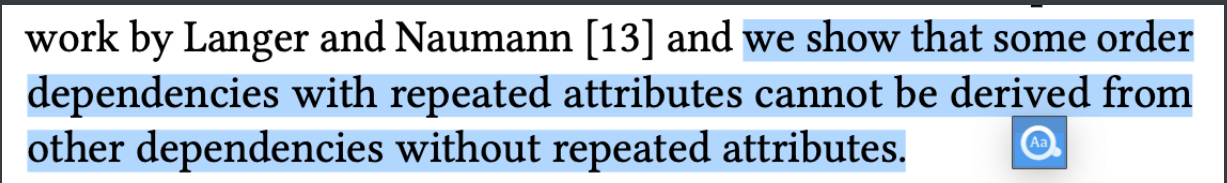
$$\frac{\begin{array}{c} \mathbf{X} \sim \boldsymbol{M} \\ \mathbf{XY} \sim \boldsymbol{M} \\ \mathbf{X} \sim \boldsymbol{MY} \\ \mathbf{XY} \sim \boldsymbol{MN} \end{array}}{\mathbf{XY} \sim \boldsymbol{MYN}}$$

Proof.

(1) from **XY** ~ **MN**, by Normalization (AX3) **XYMYN** ↔ **MNXY**;
(2) from **XY** ~ **M** and **X** ~ **MY**, using **X** ~ **M** and Replace [21] we get **MYX** ↔ **XYM** and **MXY** ↔ **MYX** ↔ **XYM**;
(3) from (2), by the Shift theorem [21] with **MY** ↔ **MY** and **MNXY** ↔ **XYMMYN** we get **MYMNXY** ↔ **MYXYMMYN**;
(4) by Normalization (AX3) **MYMNXY** ↔ **MYNXY**;
(5) from **MYXYMMYN**, using **MYX** ↔ **XYM** and Normalization (AX3) we get **XYMYMYN** and finally **XYMYN**;

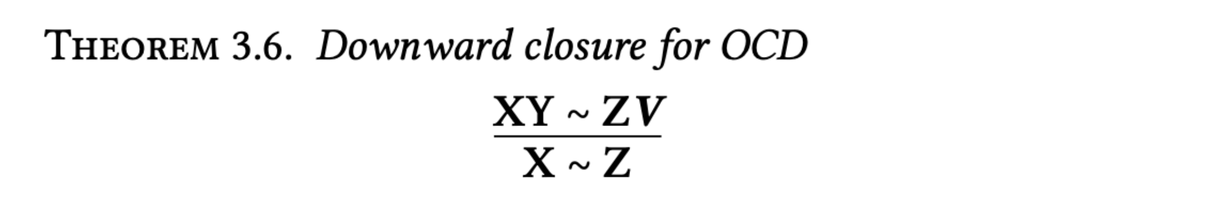
From points (3), (4) and (5) we finally get **MYNXY** ↔ **XYMYN**, i.e., **XY** ~ **MYN**. □

下文中又说：有些具有重复属性的ODs不能由没有重复属性的ODs推导得出



work by Langer and Naumann [13] and we show that some order dependencies with repeated attributes cannot be derived from other dependencies without repeated attributes.

定理3.6 OCD向下闭包

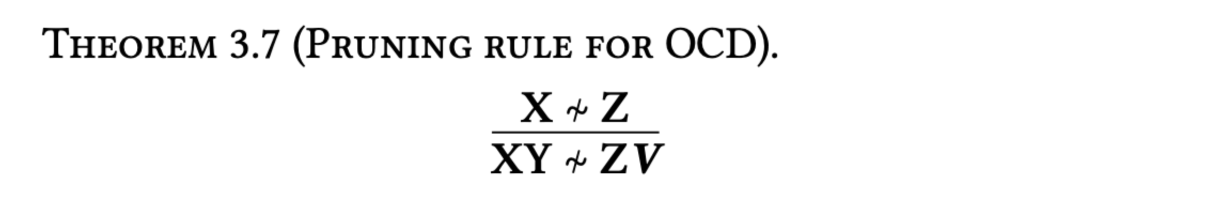


THEOREM 3.6. *Downward closure for OCD*

$$\frac{\mathbf{XY} \sim \mathbf{Z}\boldsymbol{V}}{\mathbf{X} \sim \mathbf{Z}}$$

不会证

定理3.7 OCD的剪枝规则

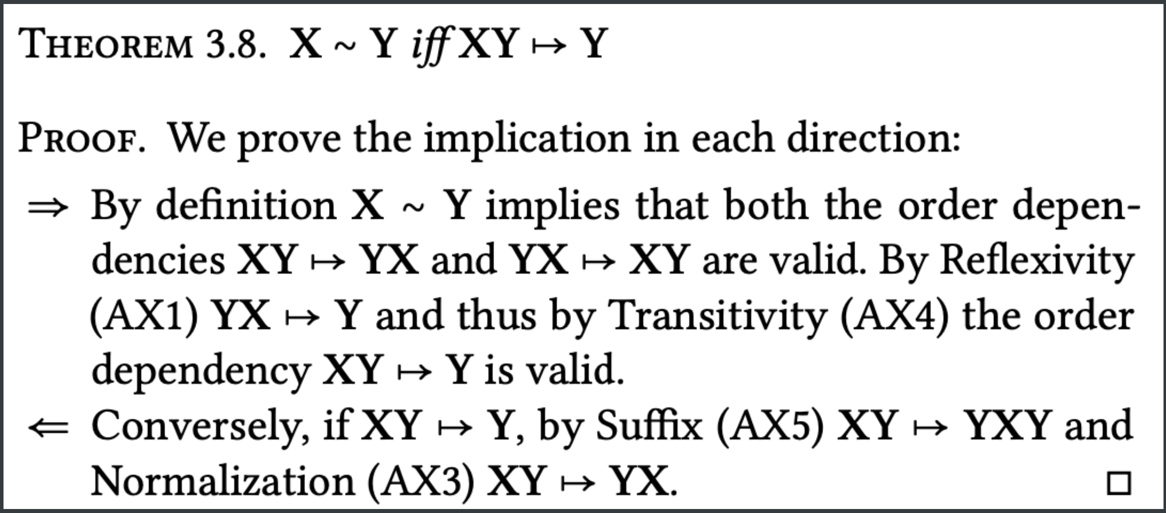


THEOREM 3.7 (PRUNING RULE FOR OCD).

$$\frac{\mathbf{X} \nsim \mathbf{Z}}{\mathbf{XY} \nsim \mathbf{Z}\boldsymbol{V}}$$

3.6的逆命题

定理3.8 当且仅当$XY \mapsto Y$有效，$X{\sim}Y$有效



THEOREM 3.8. $\mathbf{X} \sim \mathbf{Y}$ *iff* $\mathbf{XY} \mapsto \mathbf{Y}$

PROOF. We prove the implication in each direction:

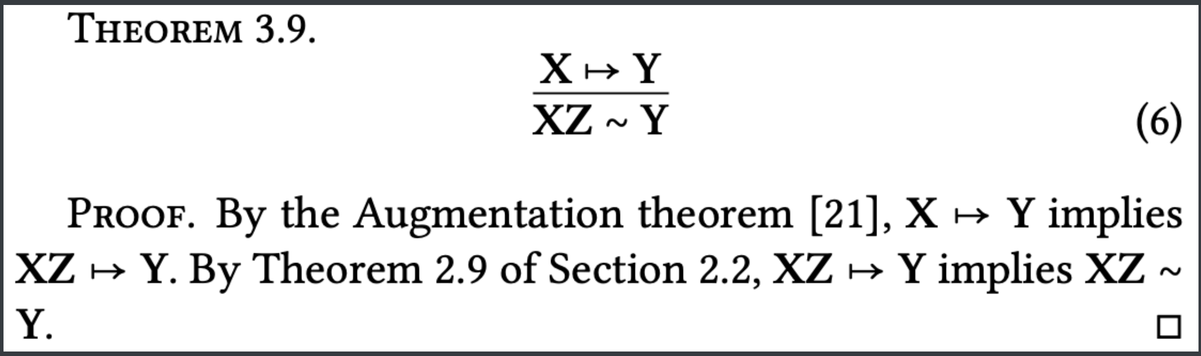
⇒ By definition $\mathbf{X} \sim \mathbf{Y}$ implies that both the order dependencies $\mathbf{XY} \mapsto \mathbf{YX}$ and $\mathbf{YX} \mapsto \mathbf{XY}$ are valid. By Reflexivity (AX1) $\mathbf{YX} \mapsto \mathbf{Y}$ and thus by Transitivity (AX4) the order dependency $\mathbf{XY} \mapsto \mathbf{Y}$ is valid.

⇐ Conversely, if $\mathbf{XY} \mapsto \mathbf{Y}$, by Suffix (AX5) $\mathbf{XY} \mapsto \mathbf{YXY}$ and Normalization (AX3) $\mathbf{XY} \mapsto \mathbf{YX}$. □

反向证明应该为：$XY \mapsto Y$**，由(AX5)**$XY \leftrightarrow YXY$**，再由(AX3)得**$XY \leftrightarrow YX$**，即**$X{\sim}Y$

这意味着ODs：$XY \mapsto Y$**，与OCDs：**$X{\sim}Y$**等价**

定理3.9



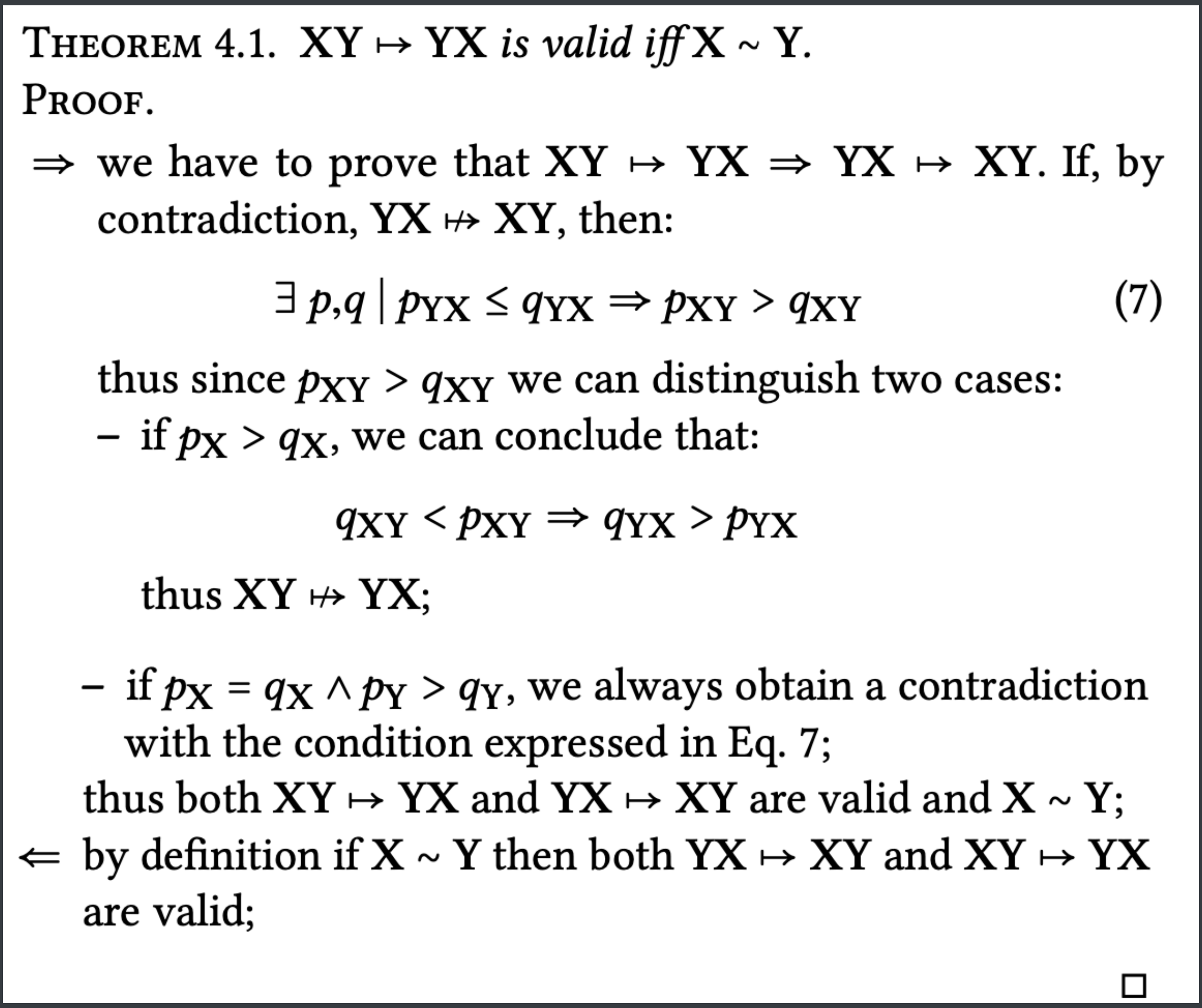
Theorem 3.9.

$$\frac{\mathbf{X} \mapsto \mathbf{Y}}{\mathbf{XZ} \sim \mathbf{Y}} \tag{6}$$

Proof. By the Augmentation theorem [21], $\mathbf{X} \mapsto \mathbf{Y}$ implies $\mathbf{XZ} \mapsto \mathbf{Y}$. By Theorem 2.9 of Section 2.2, $\mathbf{XZ} \mapsto \mathbf{Y}$ implies $\mathbf{XZ} \sim \mathbf{Y}$. □

不懂

定理4.1 当且仅当X~Y，$XY \mapsto YX$成立



Theorem 4.1. $\mathbf{XY} \mapsto \mathbf{YX}$ *is valid iff* $\mathbf{X} \sim \mathbf{Y}$.

Proof.

⇒ we have to prove that $\mathbf{XY} \mapsto \mathbf{YX} \Rightarrow \mathbf{YX} \mapsto \mathbf{XY}$. If, by contradiction, $\mathbf{YX} \not\mapsto \mathbf{XY}$, then:

$$\exists\, p,q \mid p_{\mathbf{YX}} \leq q_{\mathbf{YX}} \Rightarrow p_{\mathbf{XY}} > q_{\mathbf{XY}} \tag{7}$$

thus since $p_{\mathbf{XY}} > q_{\mathbf{XY}}$ we can distinguish two cases:

- if $p_{\mathbf{X}} > q_{\mathbf{X}}$, we can conclude that:

  $$q_{\mathbf{XY}} < p_{\mathbf{XY}} \Rightarrow q_{\mathbf{YX}} > p_{\mathbf{YX}}$$

  thus $\mathbf{XY} \not\mapsto \mathbf{YX}$;
- if $p_{\mathbf{X}} = q_{\mathbf{X}} \wedge p_{\mathbf{Y}} > q_{\mathbf{Y}}$, we always obtain a contradiction with the condition expressed in Eq. 7;

thus both $\mathbf{XY} \mapsto \mathbf{YX}$ and $\mathbf{YX} \mapsto \mathbf{XY}$ are valid and $\mathbf{X} \sim \mathbf{Y}$;

⇐ by definition if $\mathbf{X} \sim \mathbf{Y}$ then both $\mathbf{YX} \mapsto \mathbf{XY}$ and $\mathbf{XY} \mapsto \mathbf{YX}$ are valid;

□

证明见草稿（以后上传）

THE OCDDISCOVER ALGORITHM

减少搜索空间的策略：

- 移除常量列
  - 因为常量列可以被任意列X排序，能产生大量ODs
- 移除order-equivalent列
  - 若$A \leftrightarrow B$，则当属性A出现在的ODs中，我们都可以用属性B去代替属性A，会产生大量ODs
  - 给所有的order-equivalent属性建立一个等价类，并选取一个代表，移除其他所有的列，然后储存这些删除掉的列的信息。

搜索树

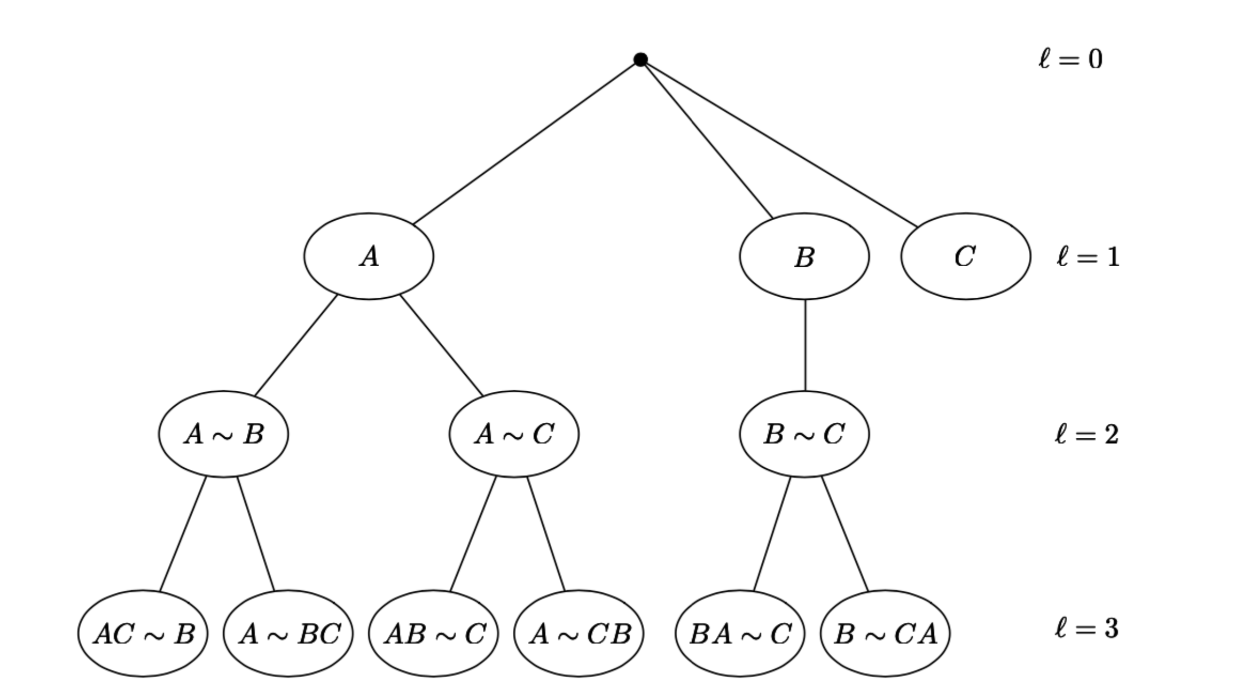


Figure 1: Permutation tree for a table with $n = 3$ attributes.

使用广度优先策略，使较短的最小依赖项在较长的最小依赖项之前被发现。

每个OCD候选集会被检查是否有效，会出现下面2种情况：

1. 不有效，则不从该候选集节点产生任何别的候选集了。
2. 有效，则按照以下方式生成新候选集
   - 在一定条件下（下面解释），生成$XA{\sim}Y$
   - 在一定条件下（下面解释），生成$X{\sim}YA$

剪枝原则

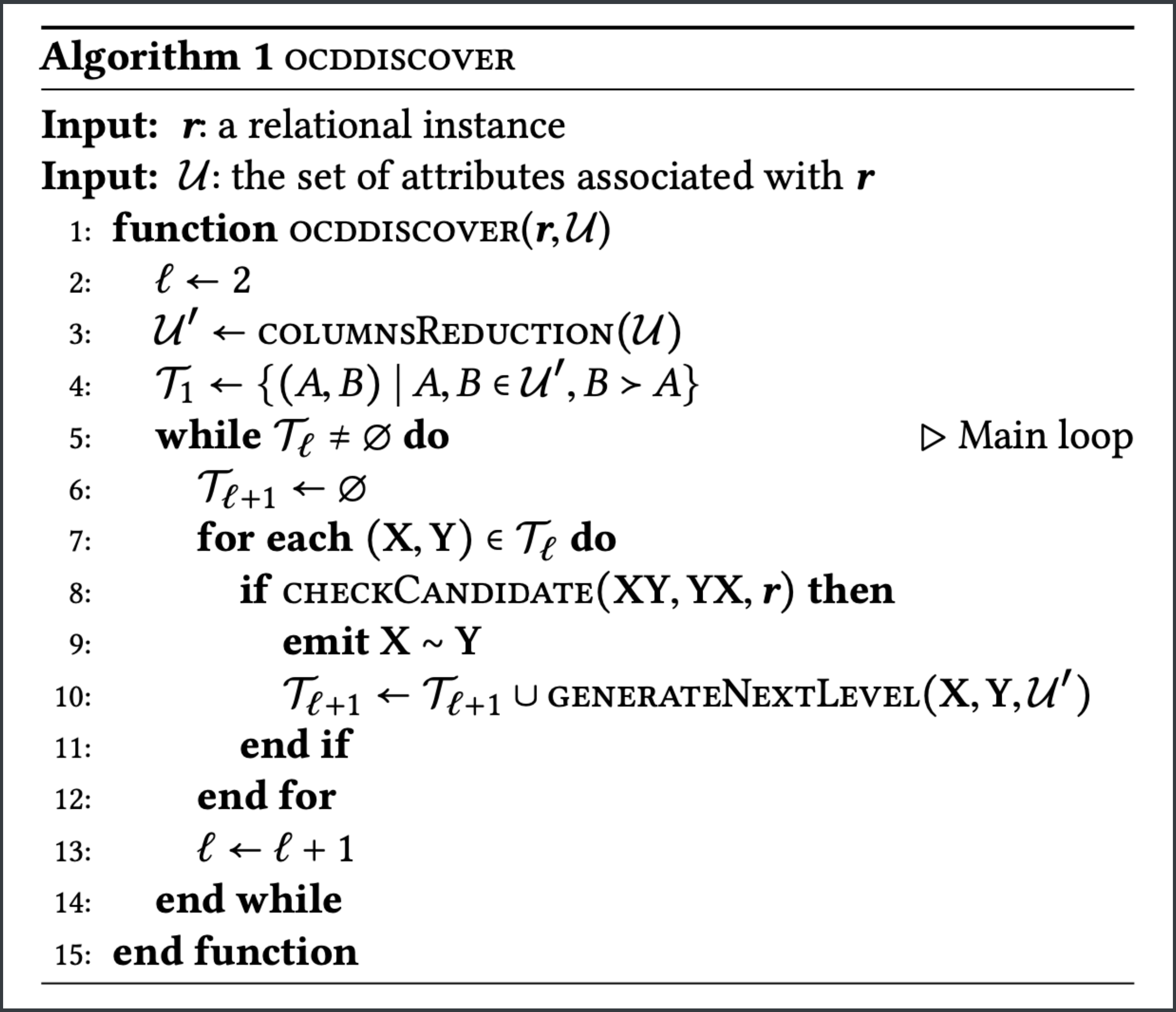
如果发现了新OCD $X \sim Y$,我们进一步检查OD$X \mapsto Y$和$Y \mapsto X$的有效性。

- 如果$X \mapsto Y$，则不产生形如$XZ \sim Y$的候选集，即$X \sim Y$的左儿子被剪掉
- 如果$Y \mapsto X$，则不产生形如$X \sim YZ$的候选集，即$X \sim Y$的右儿子被剪掉
- 如果上述两个都有效，则删除该候选集节点的所有字树

索引

> 计算一个OD候选集是否有效，先将候选集左侧属性进行排序，建立一个只包含各元组在该排序中的位置，然后依次检查右侧属性是否违反排序规则。

OCDDISCOVER

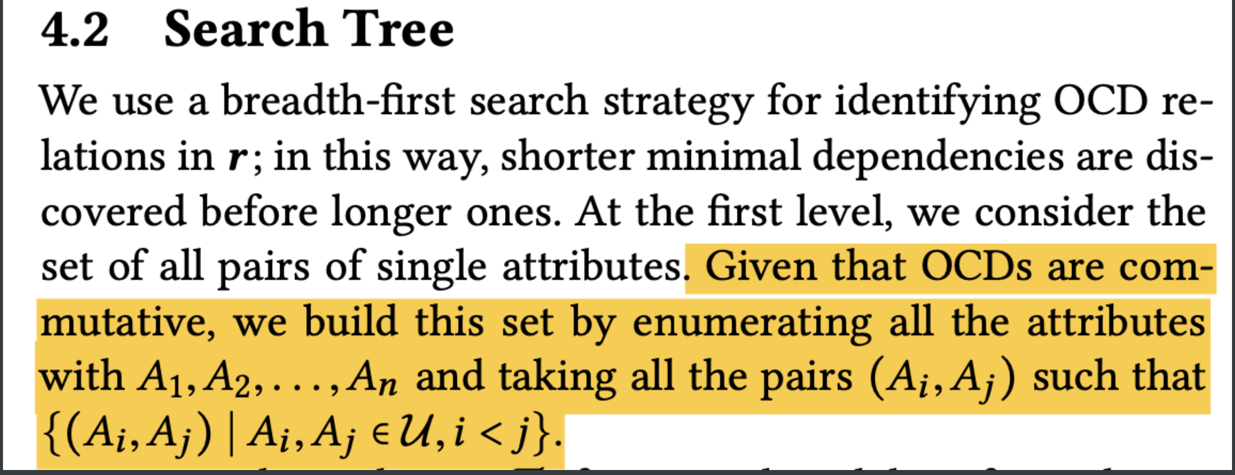


Algorithm 1 OCDDISCOVER

```
Input: r: a relational instance
Input: $\mathcal{U}$: the set of attributes associated with r
 1: function OCDDISCOVER(r, $\mathcal{U}$)
 2:     $\ell \leftarrow 2$
 3:     $\mathcal{U}' \leftarrow$ COLUMNSREDUCTION($\mathcal{U}$)
 4:     $\mathcal{T}_1 \leftarrow \{(A, B) \mid A, B \in \mathcal{U}', B > A\}$
 5:     while $\mathcal{T}_\ell \neq \varnothing$ do                          ▷ Main loop
 6:         $\mathcal{T}_{\ell+1} \leftarrow \varnothing$
 7:         for each $(X, Y) \in \mathcal{T}_\ell$ do
 8:             if CHECKCANDIDATE(XY, YX, r) then
 9:                 emit $X \sim Y$
10:                 $\mathcal{T}_{\ell+1} \leftarrow \mathcal{T}_{\ell+1} \cup$ GENERATENEXTLEVEL($X, Y, \mathcal{U}'$)
11:             end if
12:         end for
13:         $\ell \leftarrow \ell + 1$
14:     end while
15: end function
```

Line 3：删除u中的常量属性，删除等价属性，并选取一个作为代表保留

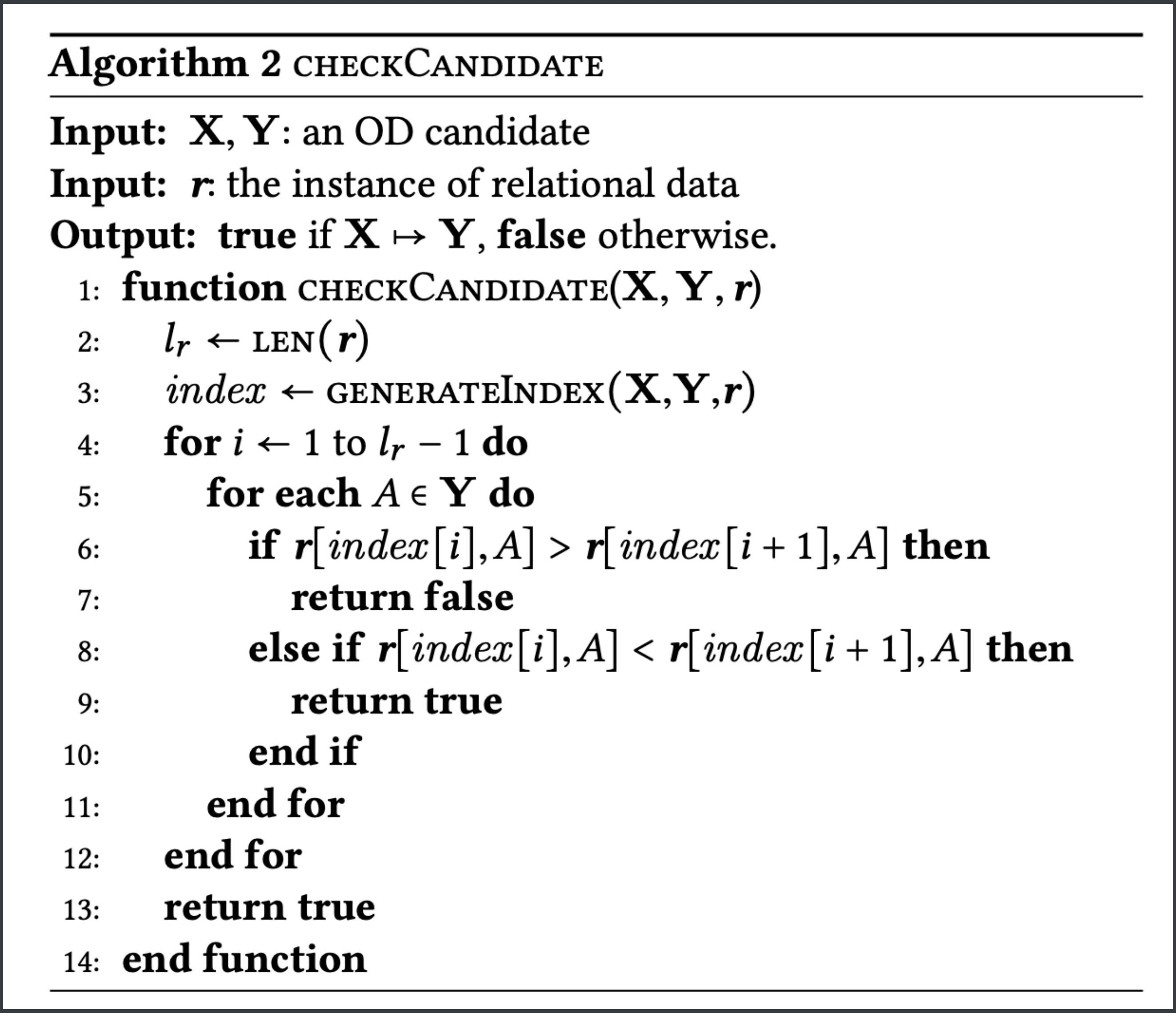
Line 4：B>A不太懂是什么意思。（4.2节黄色标记，B>A大概就是在关系中B属性排在A属性后）



4.2 Search Tree

We use a breadth-first search strategy for identifying OCD relations in $\boldsymbol{r}$; in this way, shorter minimal dependencies are discovered before longer ones. At the first level, we consider the set of all pairs of single attributes. Given that OCDs are commutative, we build this set by enumerating all the attributes with $A_1, A_2, \ldots, A_n$ and taking all the pairs (A_i, A_j) such that $\{(A_i, A_j) \mid A_i, A_j \in \mathcal{U}, i < j\}$.

总体思想：层次遍历搜索树，检查每个OCD候选集节点的有效性（无效则不从该节点产生新子节点候选集；有效则根据剪枝规则继续判断是否产生子节点候选集，产生某种候选集）；再将该节点产生的新的子节点候选集加入下一层。直到某层为空，无候选集则算法结束。



Algorithm 2 CHECKCANDIDATE

Input: $\mathbf{X}, \mathbf{Y}$: an OD candidate

Input: $\boldsymbol{r}$: the instance of relational data

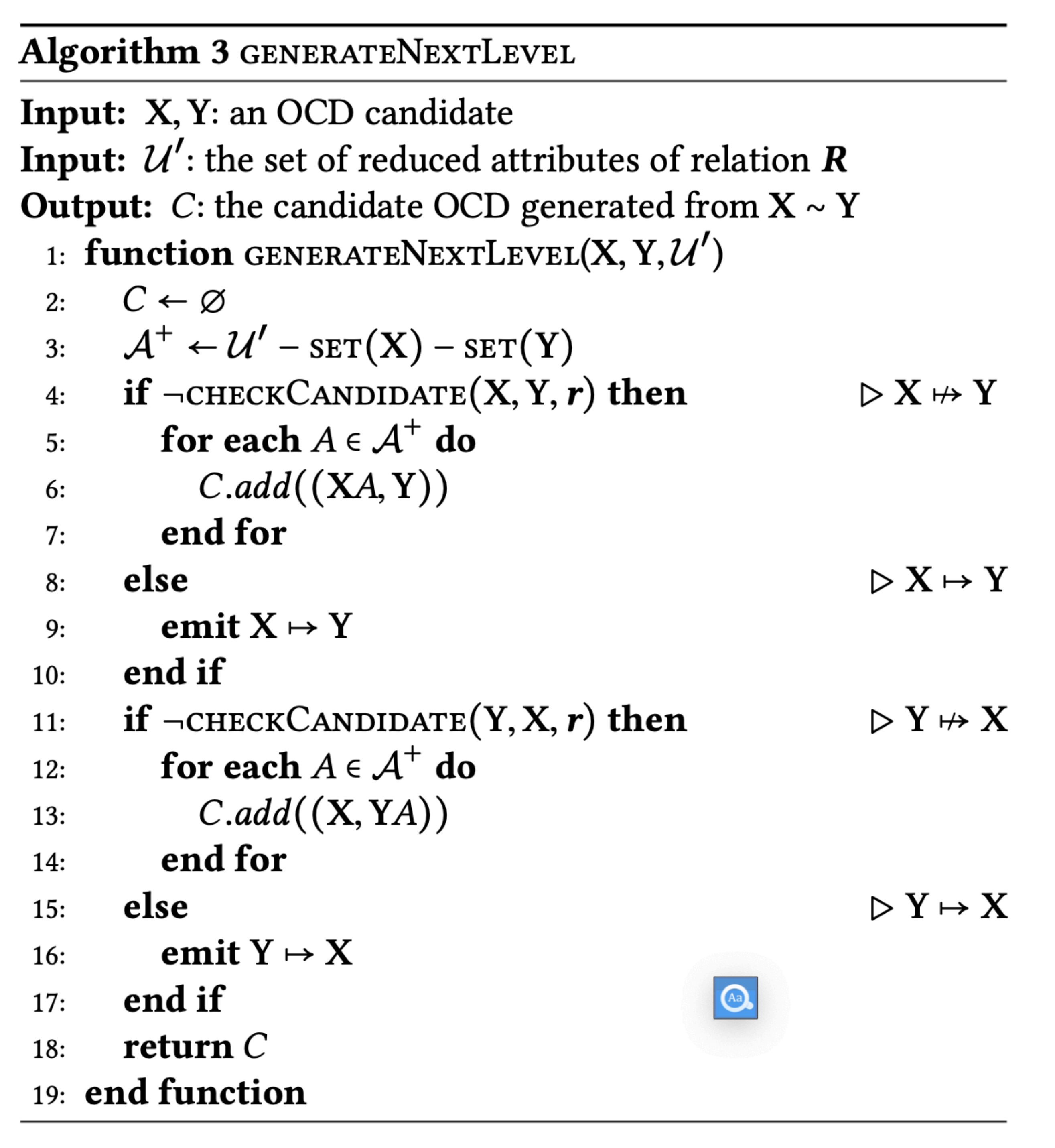
Output: **true** if $\mathbf{X} \mapsto \mathbf{Y}$, **false** otherwise.

```
1:  function CHECKCANDIDATE($\mathbf{X}, \mathbf{Y}, \boldsymbol{r}$)
2:      $l_r \leftarrow$ LEN($\boldsymbol{r}$)
3:      $index \leftarrow$ GENERATEINDEX($\mathbf{X}, \mathbf{Y}, \boldsymbol{r}$)
4:      for $i \leftarrow 1$ to $l_r - 1$ do
5:          for each $A \in \mathbf{Y}$ do
6:              if $\boldsymbol{r}[index[i], A] > \boldsymbol{r}[index[i+1], A]$ then
7:                  return false
8:              else if $\boldsymbol{r}[index[i], A] < \boldsymbol{r}[index[i+1], A]$ then
9:                  return true
10:             end if
11:         end for
12:     end for
13:     return true
14: end function
```

Line 2：LEN(r)代表r表中的元组数？或者理解为二维数组的大小？

Line3：重要函数，产生位置索引，还不太清楚怎么编写，需要继续研究

Line5～12：已经对X进行字典序排序，检查Y是否也是顺序一致，若存在违规(Swap)，则X～Y无效，终止循环返回false。



Algorithm 3 GENERATENEXTLEVEL

Input: X, Y: an OCD candidate

Input: $\mathcal{U}'$: the set of reduced attributes of relation $\boldsymbol{R}$

Output: C: the candidate OCD generated from **X** ~ **Y**

```
 1: function GENERATENEXTLEVEL(X, Y, $\mathcal{U}'$)
 2:     $C \leftarrow \varnothing$
 3:     $\mathcal{A}^+ \leftarrow \mathcal{U}' - \text{SET}(X) - \text{SET}(Y)$
 4:     if ¬CHECKCANDIDATE(X, Y, r) then                ▷ $X \not\mapsto Y$
 5:         for each $A \in \mathcal{A}^+$ do
 6:             $C.add((XA, Y))$
 7:         end for
 8:     else                                           ▷ $X \mapsto Y$
 9:         emit $X \mapsto Y$
10:     end if
11:     if ¬CHECKCANDIDATE(Y, X, r) then                ▷ $Y \not\mapsto X$
12:         for each $A \in \mathcal{A}^+$ do
13:             $C.add((X, YA))$
14:         end for
15:     else                                           ▷ $Y \mapsto X$
16:         emit $Y \mapsto X$
17:     end if
18:     return $C$
19: end function
```

Line 3：A^+新候选集补充的属性A（在$u^{'}$中，不在X、Y中的集合）的集合

Line4～17：根据剪枝规则产生下一层的候选集节点

总结

思想是通过发现OCD顺道来发现OD

发现的OCD是完整的，且最小的。

使用了广度优先来遍历搜索树，并且证明了有开头，中间，结尾重复属性的OCD是多余的。所以只用在下一层生成形如XZ～Y，X~YZ的OCDs。再利用剪枝策略，验证$X \to Y$，$Y \to X$并根据结果进行剪枝。从而得到该关系实例中所有的最小OCD，以及OD。

发现的OD是完整的么，是最小的么

形如$X \to Y$的ODs（X与Y无交集），在广度优先遍历的时候产生。(文章中是如何保证最小的？OCD无效的节点就不会再产生子节点了，这样会漏掉一些ODs么？)

LHS与RHS有交集的ODs，由OCDs X～Y产生（$XY \leftrightarrow YX$）。OCDs是最小的且是完整的，所以该形式的ODs应该也是完整的。