Prioritized Preferences and Choice Constraints

Wilfred Ng

Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong
wilfred@cse.ust.hk

Abstract. It is increasingly recognised that user preferences should be addressed in many advanced database applications, such as adaptive searching in databases. However, the fundamental issue of how preferences impact the semantics and rankings in a relation is not resolved. In this paper, we model a user preference term involving one attribute as a hierarchy of its underlying data values and formalise the notion of Prioritized Preferences (PPs). We then consider multiple user preferences in ranking tuples in a relational table. We examine the impact of a given set of PPs on possible choices in ranking a database relation and develop a new notion of Choice Constraints (CCs) in a relation, r. Given two PPs, X and Y, a CC, $X \leq Y$, is satisfied in r, if the choice of rankings according to Y is no less than that of X. Our main results are related to these two notions of PPs and CCs and their interesting interactions with the well-known Functional Dependencies (FDs). First, we exhibit a sound and complete set of three inference rules for PPs and further prove that for each closed set of PPs, there exists a ranking that precisely satisfies these preferences. Second, we establish a sound and complete set of five inference rules for CCs. Finally, we show the soundness and completeness of two mixed systems of FD-PPs and FD-CCs. All these results are novel and fundamental to incorporating user preferences in database design and modelling, since PPs, CCs and FDs together capture rich semantics of preferences in databases.

1 Introduction

Preference is an important and natural constraint that captures human wishes when seeking information. However, the semantics of preferences were not adequately studied until the recent work in [7, 8, 2, 14]. In these papers, the fundamental nature of different preferences in the form of “I like A better than B” is modelled by a set of orderings defined over data. Still, the impact of preferences as a semantic constraint is not adequately addressed in many ways. For example, in database modelling, traditional constraints like Functional Dependencies (FDs) capture the semantics of the hard fact only, but preferences do not have such semantics as constraints that represent a priority of choices. However, as information becomes abundant over the web, there is a practical need for generating a ranking that satisfies some user preferences in the search result [7, 8]. In addition, although FDs are widely recognized as the most important integrity constraint in databases, the interactions of FDs with preferences, to our knowledge, have never been studied in literature.

In our modelling, we assume that a user preference is expressed in a sequence of attributes that associate with their respective *preference terms*. We call the attributes involved in preference terms *preference attributes*. The underlying idea is that a user preference is inherent to the ordering relationship between the data projected onto the preference attributes, and thus a *preference hierarchy* can be devised to capture the choices of *preference rankings*. Our approach is to transform a relation to a preference relation, r, which has only natural numbers according to the level of the preference hierarchy. Then a ranking of tuples, $\leq_r$, can be arbitrary defined on r whereas the consistency of $(r, \leq_r)$ is determined by the lexicographical order of the preference attributes. The following example illustrates the use of a preference relation.

Example 1. Suppose a second-hand car relation is defined by the preference attributes $PRICE\_RANGE$, $ENGINE\_POWER$ and $MILEAGE\_USED$, which assert the preferences specified by $YOUTH\_ CHOICE$ (the choice of young customers). The preference increases with first the price range and then the engine power and finally the car's mileage. We adopt the PREFERRING clause proposed in [7] to express the preference terms, which essentially impose an order over their corresponding data domains. The three terms together the respective preference hierarchies are assumed to be prioritized as follows:

First priority: LOWEST(price) $\Rightarrow$ \$$5001 - 6000 <$ \$$4001 - 5000 <$ \$$1001 - 2000$.
Second priority: HIGHEST(power) $\Rightarrow 1000cc < 2000cc < 3000cc$.
Third priority: mileage AROUND 30,000km $\Rightarrow 10000km < 20000km < 30000km$.

A preference relation, r, is generated by mapping the data values in the car relation to natural numbers according to the level of the preference hierarchies of the given preference terms, which is shown in the right-hand side of Figure 1. The overall preference ranking (which is unique in this simplified example but may be more than one in general) in the last column, **Rank**, is determined by the lexicographical order of PRICE, ENGINE and MILEAGE, which is consistent with the tuple ordering, $t_1 <_r \cdots <_r t_5$. Note that some attributes are abbreviated in the table due to width limits.

	$PRICE$	$ENGINE$	$MILEAGE$
t_1	1001-2000	$1500cc$	$20000km$
t_2	4001-5000	$3000cc$	$30000km$
t_3	4001-5000	$2000cc$	$20000km$
t_4	4001-5000	$1500cc$	$10000km$
t_5	5001-6000	$3000cc$	$10000km$

$\Longrightarrow$

	$PRICE$	$ENGINE$	$MILEAGE$	**Rank**
t_1	1	3	2	**1**
t_2	2	1	1	**2**
t_3	2	2	2	**3**
t_4	2	3	3	**4**
t_5	3	1	3	**5**

Fig. 1. Transforming the second-hand car relation into a preference relation according to the preference terms of $YOUTH\_CHOICE$

Middle-class adult customers may have different preferences. This gives rise to a different preference relation as shown in Figure 2, where the preference ranking (i.e. **Rank**) is not consistent with the tuple ranking (i.e. $<_r$). The preference terms of MIDDLE CLASS_CHOICE are assumed to be reprioritized as follows:

First priority: price AROUND \$4000-\$5000.
Second priority: HIGHEST(power).
Third priority: LOWEST(mileage).

Finally, pensioner customers may have another set of preference terms, which give rise to the different preference relation and ranking shown in Figure 3. The preference terms are assumed to be reprioritized as follows:

First priority: LOWEST(price).
Second priority: mileage BETWEEN 20,000km AND 30,000km.
Third priority: power AROUND 2000cc.

	$PRICE$	$ENGINE$	$MILEAGE$	**Rank**
t_1	3	3	2	**5**
t_2	1	1	3	**1**
t_3	1	2	2	**2**
t_4	1	3	1	**3**
t_5	2	1	1	**4**

Fig. 2. $MIDDLE\ CLASS\_CHOICE$

	$PRICE$	$ENGINE$	$MILEAGE$	**Rank**
t_1	1	2	1	**1**
t_2	2	3	1	**3**
t_3	2	1	1	**2**
t_4	2	2	2	**4**
t_5	3	3	2	**5**

Fig. 3. $PENSIONER\_CHOICE$

Any tuple ranking is *trivially satisfied* in a preference relation r when there are no imposed preference terms. When preference terms are stated by the users, we check if the tuple ranking in r are consistent with a (any) lexicographical order of the sequence of the preference attributes. This gives rise to the notion of *Prioritized Preferences* (PPs) (cf. see Definition 6 in [7] for the motivation for prioritized preferences), and in order to have PP satisfied in r, tuple rankings are restricted to the set of preference rankings. This also gives rise to another notion of Choice Constraints (CCs) being satisfied in a relation. Given two PPs, X and Y, a CC, $X \leq Y$, is satisfied in r, if the choice of preference rankings according to Y is no less than that of X. We focus on three interesting problems related to the semantics of preferences in relations:

1. When there is a tuple ranking that satisfies a set of PPs, what are the rules governing such preference satisfaction?
2. When there is more than one possible tuple ranking that satisfies different PPs, what are the rules of governing the ranking possibilities (CCs)?
3. What are the interactions between FDs, PPs and CCs?

Our main contribution is related to the above problems. We present a spectrum of interesting axiom systems in this paper. With respect to preference satisfaction, we exhibit a sound and complete set of three inference rules for PPs. It is further proved that for each closed set of PPs, there exists a ranking that satisfies these preferences and no others. With respect to the choice of tuple rankings for a given set of PPs, we establish a sound and complete set of five inference rules for CCs. Finally, we study the interactions between PPs and FDs, and between CCs and FDs and formally show the soundness and completeness of two mixed systems of FD-PPs and FD-CCs. All these results are novel and fundamental to incorporating user preferences in database design and modelling, since PPs, CCs and FDs together capture rich semantics of preferences in many database applications in reality.

The rest of the paper is organised as follows. In Section 2, we present some preliminary concepts and notation. In Section 3, we present a sound and complete system with respect to PP satisfaction. In Section 4, we introduce the concept of CCs and present a sound and complete system with respect to CC satisfaction. In Section 5, we discuss the interactions between FDs and PPs and those between FDs and CCs. We present two sound and complete systems of FD-PPs and FD-CCs. In Section 6, we review some related work. In Section 7, we give our concluding remarks.

2 Preliminaries

We assume throughout that X and Y are sequences of attributes and that $X \sim Y$ indicates the fact that X and Y have the same elements. XY denotes the *concatenation* of X and Y (appending Y to X). A *prefix* of X, denoted as $pre(X)$, is a sequence of the form $\langle A_1, \ldots, A_{m_1} \rangle$, where $X = \langle A_1, \ldots, A_m \rangle$ and $1 \leq m_1 \leq m$. A *shuffle* of X and Y, denoted as $shu(X, Y)$, is defined as a sequence of the form $\langle C_1, \ldots, C_{m+n} \rangle$, where there exists two *subsequences* of attributes $\langle C_{i_1}, \ldots, C_{i_m} \rangle = X$ and $\langle C_{j_1}, \ldots, C_{j_n} \rangle = Y$, and the order of the attributes in X and Y is preserved in $shu(X, Y)$.

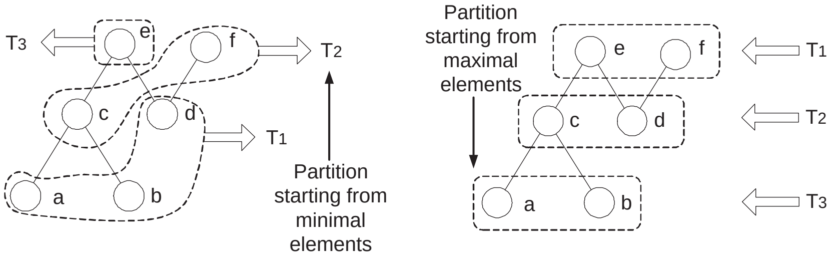
Lexicographical ordering is a fundamental property of prioritized preferences as illustrated in Example 1, where the preference in $YOUTH\_CHOICE$ can be modelled as a lexicographical ordering of the Cartesian product of the domains $PRICE \times ENGINE \times MILEAGE$ in the preference relation in Figure 1.

We assume the usual terminologies and notation used in the relational data model [1]. In particular, let $R = \{A_1, \ldots, A_n\}$ be the relation schema and $t[A_i]$ $(1 \leq i \leq n)$ denote the *projection* of t onto attribute A_i. A *relation* r defined over R is a finite set of tuples over R. We define $r[A_i] = \{t[A_i] \mid t \in r\}$.

Note that preference terms such as "BETWEEN AND", "HIGHEST", "LOWEST" and "IN" as defined in [7] are equivalent to defining a partial ordering over the tuples induced by the involved preference attributes. Thus, we are able to map the data values into natural numbers according to a preference hierarchy, resulting in a preference relation.

We now assume a relation having one preference attribute, $R = \{A\}$, to illustrate the idea. We first denote by $\mathcal{H}(r, A)$ a *partition* of r, which is a set of pairwise disjoint non-empty subsets of r such that $\bigcup_{T \in \mathcal{H}(r,A)} T = r$, and we call the element $T \in \mathcal{H}(r, A)$ a *preference level* of r induced by A. A *preference hierarchy* of r induced by A is a linearly ordered partition of r, corresponding to the preference term p imposed on A.

Example 2. Consider $r = \{a, b, c, d, e, f\}$ (6 tuples), where $a \leq_A^p c$, $b \leq_A^p c$, $c \leq_A^p e$, $d \leq_A^p e$ and $d \leq_A^p f$. We now show two possible internal hierarchies, $\mathcal{H}(r, A) = \{T_1, T_2, T_3\}$, given in Figure 4, in which each tuple is represented by a node.



(a) Bottom-up partition (b) Top-down partition

Fig. 4. Two possible preference hierarchies $\mathcal{H}(r, A)$

Using the bottom-up partition approach, we successively collect the sets of minimal tuples in the subsets of r and construct the preference hierarchy as illustrated in Figure 4(a). We remark that this method of constructing the preference hierarchy is essentially a matter of convention and another possibility is shown in Figure 4(b) as a comparison. The two conventions can also be used to represent the "like less" and "like more" preferences.

The idea of a preference hierarchies can be straightforwardly generalized to multiple preference attributes. Algorithm 1 shows how to generate a preference relation arising from the preference terms. Essentially, the algorithm collects the minimal tuples of a relation (or its subset) with respect to each preference order using a bottom-up partition.

Algorithm 1 ($PREFERENCE\_RELATION(r$, X))

Input: A relation r, a set of preference attributes X and a set of preference orderings $\leq^p_A$ for all $A \in X$.
Output: A preference relation of r.

begin
  for all $A \in X$, **do**
    $i = 0$;
    do until $r[A] = \emptyset$
      Increment i;
      Obtain T_i as the set of minimal tuples (wrt $\leq^p_A$) of $r[A]$;
      $r[A] := r[A] - T_i$ and $\mathcal{H}(r, A) := \{T_1 <_h \cdots <_h T_i\}$;
  for all $t \in r$, $A \in X$, **do**
    Map $t[A]$ to n in r whenever $t[A] \in T_n$ and $T_n \in \mathcal{H}(r, A)$;
  return r (with mapped natural numbers on $r[X]$);
end

Definition 1. **(Preference Relation)** Given a relation r over R, a prioritized preference, $X \subseteq R$ and a set of preference terms over X. A preference relation, $(r, \leq_r)$, is the relation (with mapped natural numbers on $r[X]$ returned by Algorithm 1) together with a tuple ranking, $\leq_r$. From now on, we simply call a preference relation a relation whenever no ambiguity arises.

The preference hierarchy generated in Step 7 by Algorithm 1 is unique and therefore Definition 1 is well-defined. The uniqueness of the result of Algorithm 1 is due to the fact that T_n is the unique set of all minimal tuples of $r[A]$ according to $\leq_A^p$. Intuitively, a level $T_n \in \mathcal{H}(r, A)$ captures the "equivalent choices" with respect to a preference term and the hierarchy observes the order arising from the preference term imposed on A. In the special case of linearly ordered preference terms such as HIGHEST(power) or LOWEST(price), T_n is the singleton containing the nth tuple sorted in numerical order.

In our running example, the preference hierarchies of $PRICE$, $MILEAGE$ and $ENGINE\_POWER$ corresponding to the $YOUTH\_CHOICE$ are $\{\{t_1\} <_h \{t_2, t_3, t_4\} <_h \{t_5\}\}$, $\{\{t_2, t_5\} <_h \{t_3\} <_h \{t_1, t_4\}\}$ and $\{\{t_2\} <_h \{t_1, t_3\} <_h \{t_4, t_5\}\}$, respectively.

3 Preferences and Choices

In this section, we present the notion of a *Prioritized Preference* (PP) and its satisfaction over a relation.

The semantics of a preference with multiple attributes, X, is defined according to lexicographical orderings, denoted as $\leq_X^{lex}$, on the Cartesian product of the mapped numerical values via the preference hierarchies of the attributes.

Definition 2. (Prioritized Preference and Choice) A *prioritized preference* (or simply a preference), X, is a sequence of attributes obtained from a relation schema, R. A preference, X, is satisfied in a relation, (r, $\leq_r$) over R, denoted as (r, $\leq_r$) $\models X$, if for all $t_1, t_2 \in r$, $t_1[X] <_X^{lex} t_2[X]$ implies that $t_1 <_r t_2$. We call any distinct $\leq_r$ such that (r, $\leq_r$) $\models X$ a *choice* of rankings wrt X (or simply a choice whenever (r, $\leq_r$) and X are understood), and denote the number of such distinct ranking choices as $\mid choice(r, X) \mid$. In particular, if the choice is unique, we call the satisfaction arising from the choice the unique satisfaction.

Notably, PPs allow the same attribute appearing several times in a preference X. This is necessary for studying the inference rules later on, since some rules may infer PPs having repeated attributes. However, by removing the repeated occurrence of a particular attribute after its first occurrence in a preference, we can obtain an "equivalent preference" in which each attribute appears at most once. This also implies there exists only a finite number of distinct PPs (up to equivalence) for a given relational schema.

The following proposition follows directly from Definition 2. It means that if a relation satisfies a unique choice, its tuples are simply ordered by $\leq_X^{lex}$. Remarkably, if $X = R$ the satisfaction must be unique, since $\leq_R^{lex}$ is a linear order on r. In addition, if we have all distinct (integer) values for all tuples under any attribute $A \in X$, the satisfaction is also unique. This follows that for any arity-1 relation, i.e. $\mid R \mid= 1$, the satisfaction, if any, must also be unique.

Proposition 1. Given $X = A_1 \cdots A_n$. If $\mid choice(r, X) \mid= 1$, then, for all $t_1, t_2 \in r$, $t_1 <_r t_2$, if and only if $\exists k$, $1 \leq k < n$, such that $t_1[A_1 \cdots A_k] = t_2[A_1 \cdots A_k]$ and $t_1[A_{k+1}] < t_2[A_{k+1}]$. $\square$

For example, it can be checked that the second-hand car relation has a unique satisfaction according to (unique) **Rank** in Figures 1 to 3. However, we may have another set of preference terms, which gives rise to two possible preference rankings, **Rank 1** and **Rank 2**, shown in the two right columns of Figure 5. The preference terms are assumed to be prioritized as follows:

First priority: LOWEST(price).
Second priority: mileage LESS THAN 30,000km.
Third priority: power BETWEEN 1500cc AND 2000cc.

	$PRICE$	$MILEAGE$	$ENGINE$	**Rank 1**	**Rank 2**
t_1	1	1	1	**1**	**1**
t_2	2	2	2	**4**	**4**
t_3	2	1	1	**2**	**3**
t_4	2	1	1	**3**	**2**
t_5	3	1	2	**5**	**5**

Fig. 5. Two choices or rankings satisfying the preference terms

In other words, the relation in the above example should rank as $\{t_1 <_r t_4 <_r t_2 <_r t_3 <_r t_5\}$ or $\{t_1 <_r t_4 <_r t_3 <_r t_2 <_r t_5\}$ in order to satisfy the imposed preference, i.e. we have $| \, choice(r, (price, mileage, power)) \, | = 2$.

We now illustrate some non-trivial aspects of preference satisfaction in the following example (assuming usual numerical order $0 < 1$).

Example 3. Let $r = \{t_1 <_r t_2\}$ and $X = ABCD$ as given in Figure 6. It is straight-

	A	B	C	D
t_1	1	0	1	0
t_2	1	1	0	1

Fig. 6. $r \models BC$ but $r \not\models CB$; $r \models ABC$ but $r \not\models AC$

forward to check that $r \models BC$ but not $r \models CB$ and that $r \models ABC$ but not $r \models AC$. However, we will prove later some interesting but non-trivial result such as that $r \models AB$ and $r \models DC$ imply $r \models ADB$ and $r \models ADBC$, as also illustrated in r.

The interesting interactions in the above example motivate our work of establishing a set of inference rules for deriving preferences. In the subsequent discussion, we assume preference satisfaction is restricted to a unique choice of ranking (i.e., $| \, Choice(r, X) \, | = 1$) and say that $(r, \leq_r) \models_u X$ if $(r, \leq_r) \models X$ and there exists no distinct $\leq'_r$ such that $(r, \leq'_r) \models X$. The study of a unique choice of ranking is important, since it affects the way to store and index a preference relation. It may also lead to more efficient evaluation of search queries, for example if the user asks follow-up questions based on existing preference and ranking then we need to evaluate only one relation.

We now begin to formalise the notion of PP satisfaction as follows.

Definition 3. **(PP Satisfaction and Implication)** Given a set of preferences, P, and a relation, $(r, \leq_r)$, we say that $(r, \leq_r)$ logically implies P, denoted as $(r, \leq_r) \models_u P$, if and only if $\forall X \in P$, $(r, \leq_r) \models_u X$. In addition, we say that P logically implies X, denoted as $P \models X$, if for any $(r, \leq_r)$, $(r, \leq_r) \models_u P$ implies that $(r, \leq_r) \models_u X$.

From now on, we may lighten the notation of $(r, \leq_r)$ and simply use r to mean a preference relation if $\leq_r$ can be understood.

An *axiom system* [1] for preferences over relations is a set of inference rules that can be used to derive new preferences from P. We denote by $P \vdash X$ the fact that either $X \in P$ or X can be inferred (or derived) from P by using one or more of the inference rules in Definition 4.

Definition 4. **(Inference Rules for Prioritized Preferences)** Let P be a set of preferences over *R*, $A \in R$. Let X, Y be non-empty sequences of attributes obtained from R. The inference rules for preferences are defined as follows:

(PP1) *Expansion*: If $P \vdash X$, then $P \vdash XA$.
(PP2) *Shuffle*: If $P \vdash X$ and $P \vdash Y$, then $P \vdash shu(X, pre(Y))$.
(PP3) *Compression*: If $P \vdash XAYAZ$, then $P \vdash XAYZ$.

Unlike most known database constraints, P consists of no reflexivity rule in Definition 4, since there is no trivial preference satisfaction in relations. We also remark that the axiom system comprising these rules is *minimal*, since the three rules given in Definition 4 are independent.

Lemma 1. The axiom system comprising inference rules PP1-PP3 is sound for the satisfaction of PPs in relations. $\Box$

We now show in next theorem that the axiom system comprising the inference rules in Definition 4 is *sound* and *complete* for preference satisfaction in preference relations. The underlying idea in this proof is first to assume that a preference, X, cannot be inferred from the axiom system and then to present a relation as a counter-example in which all the preferences of P' hold except for X (cf. see Theorem 3.21 in [1]). The result is significant since it indicates that the axiom system can be employed as a theorem-proving tool for preferences.

Theorem 1. The axiom system comprising rules PP1 to PP3 is sound and complete for preference satisfaction in relations.

Proof. We now establish the completeness by showing that if $P \not\vdash X$, then $P \not\models X$. Equivalently for the latter, it is sufficient to exhibit a relation as a counter-example, r^c, such that $r^c \models_u P$ but $r^c \not\models_u X$. Assuming that L is the largest prefix of X such that $P \vdash LQ$ for some $Q \subseteq R$. Let us call this the *L-assumption*.

There are two cases to consider.

In the first case, we assume that $L = X$. We consider the relation $r^c = \{t_1 <_r t_2\}$ shown in Figure 7. Obviously, we have $r^c \not\models_u X$, since $choice(r^c, X)$ is not unique. It remains to show that $r^c \models_u P$. Assume to the contrary that $r^c \not\models_u P$. So $\exists X' \in P$ such that $r^c \not\models_u X'$. By the construction of r^c, we have $X' \subseteq X$ (as a set inclusion). By the L-assumption and PP2, it follows that $P \vdash LX'$. So, we have $P \vdash L$ by PP3, which is a contradiction, since we derive X from P.

	X	$R - X$
t_1	$0 \cdots 0$	$0 \cdots 0$
t_2	$0 \cdots 0$	$1 \cdots 1$

Fig. 7. A counter-example relation r^c used in the case of $L = X$

	L	B	$R - BL$
t_1	$0 \cdots 0$	1	$0 \cdots 0$
t_2	$0 \cdots 0$	0	$1 \cdots 1$

Fig. 8. A counter-example relation r^c used in the case of $L \neq X$

In the second case, we assume that $L \neq X$. Let $X = LBQ'$ where $B \notin L$ and $BQ' \subseteq R$. Using a similar technique of the first case, we construct the relation r^c shown in Figure 8, in which $r^c \not\models_u X$.

We now show that $r^c \models_u P$. We assume to the contrary that $\exists p \in P$ such that $r^c \not\models_u p$, where $p = X'$. By the construction of r^c, we have the following two possible cases concerning X'.

(Case of $X' \subseteq L$). By PP1, we expand p by attaching the attribute B. It follows that $P \vdash X'B$. By the L-assumption and PP2, it follows that $P \vdash LX'BQ$. We thus have $P \vdash LBQ$. But LB is the prefix of X and strictly contains L. This leads to a contradiction, since we violate the L-assumption.

(Case of $X' \not\subseteq L$). Let $X' = VBW$ where $V \subseteq L$ and $W \subseteq R$. By the L-assumption and PP2, it follows that $P \vdash LX'$. So by PP3 we have $P \vdash LBW$. But LB is the prefix of X. This leads to the same contradiction, since we also violate the L-assumption. $\square$

4 Choice Constraints

In this section, we consider the case of more than one ranking of r that satisfy X and formalize the notion of a *Choice Constraint* (CC) and their satisfaction in relations. We formulate five inference rules that are proved to be sound and complete for CCs.

Definition 5. (Choice Constraint) Let X and Y be two sequences of non-empty attributes obtained from R. The Choice Constraint (CC), $Y \leq X$, is satisfied in r, written as $r \models Y \leq X$, if and only if, $\mid choice(r, Y) \mid \leq \mid choice(r, X) \mid$. Given a set of CCs, C, we say that r logically implies C, denoted as $r \models C$, if and only if $\forall (Y \leq X) \in C$, $r \models Y \leq X$. In addition, we say that C logically implies $Y \leq X$, denoted as $C \models Y \leq X$, if for any r, $r \models C$ implies $r \models Y \leq X$.

The study of CCs is related to maintaining the preference rankings in a database, since user preference terms may be removed or added. This is particular important for cache-conscious systems in a client-server architecture, in this case some possible rankings should be evaluated first in order to have quick response in the query evaluation. For example, referring to the $PENSIONER\_CHOICE$ ranking given in Figure 3, if the user is willing to drop the third priority of engine power, then we have two choices. However, dropping the second priority of mileage used does not offer more choices. It can be checked that the relation satisfies the CC, $PRICE, ENGINE \leq PRICE, MILEAGE$.

We are now ready to define a particular axiom system for CC satisfaction in relations.

Definition 6. (Inference Rules for Choice Constraints) Assume that X, Y, Z are non-empty sequences of attributes obtained from R.

(CC1) *Reflexivity*: $C \vdash X \leq X$.
(CC2) *Expansion*: If $C \vdash X \leq Y$ and X is a subsequence of W, then $C \vdash W \leq Y$.
(CC3) *Transitivity*: If $C \vdash X \leq Y$ and $C \vdash Y \leq Z$, then $C \vdash X \leq Z$.
(CC4) *Pseudo Augmentation*: If $C \vdash Y \leq XY$, then $C \vdash YZ \leq XYZ$.
(CC5) *Permutation*: If $X \sim X'$, $Y \sim Y'$ and $C \vdash X \leq Y$, then $C \vdash X' \leq Y'$.

Note that CCs do not have usual augmentation as FDs. The counter example in Figure 9 shows that the statement if $C \vdash B \leq A$, then $C \vdash BC \leq AC$ is false. It can also be checked that $\mid choice(r, A) \mid = \mid choice(r, B) \mid= 2$ but $\mid choice(r, AC) \mid= 1$ and $\mid choice(r, BC) \mid=\mid choice(r, CB) \mid= 2$.

	A	B	C
t_1	0	1	0
t_2	0	0	1
t_3	1	0	1

$$choice(r, A) = \{t_1 <_r t_2 <_r t_3; t_2 <_r t_1 <_r t_3\}$$
$$choice(r, B) = \{t_2 <_r t_3 <_r t_1; t_3 <_r t_2 <_r t_1\}$$
$$choice(r, AC) = \{t_1 <_r t_2 <_r t_3\}$$
$$choice(r, BC) = \{t_2 <_r t_3 <_r t_1; t_3 <_r t_2 <_r t_1\}$$
$$choice(r, CB) = \{t_1 <_r t_2 <_r t_3; t_1 <_r t_3 <_r t_2\}$$

Fig. 9. $r \models B \leq A$ but $r \not\models BC \leq AC$

Lemma 2. The following three inference rules can be derived from CC1 - CC5.

(CC6) *Projection I*: If Z is a subsequence of X, then $C \vdash X \leq Z$.
(CC7) *Projection II*: If $C \vdash X \leq Y$ and Z is a subsequence of Y, then $C \vdash X \leq Z$.
(CC8) *Pseudo Union*: If $C \vdash X \leq XY$ and $C \vdash X \leq XZ$, then $C \vdash X \leq XYZ$.

Lemma 3. The axiom system comprising inference rules CC1-CC5 is sound for the satisfaction of CCs in relations. □

We now establish the completeness of the rules given in Definition 6. First, we introduce two technical concepts of *CC closure* and *CC cover* for establishing the result. Given C, a CC closure, denoted as C^+, is given by $C^+ = \{X \leq Y \mid C \vdash X \leq Y\}$.

A CC cover of C, denoted as $cover(C)$, is the set of CCs that have maximal sets of attributes on the right side. Formally, $cover(C) = \{X \leq Y \mid X \leq Y \in C^+$ and $\forall(X \leq Z) \in C^+, Z \subseteq Y$ (as sets)$\}$.

Clearly, C and C^+ are equivalent. The following lemma shows that C and $cover(C)$ are equivalent with respect to CC inferencing.

Lemma 4. $C \vdash X \leq Y$ if and only if $cover(C) \vdash X \leq Y$.
Proof. The proof of the "if" part is trivial by the definition of C^+, since $C \vdash C^+$ and $cover(C) \subseteq C^+$. The "only if" part can be established as follow: let $(X \leq Y) \in C$. Then $\exists(X \leq Z) \in cover(C)$ such that $Y \subseteq Z$. If $Y \neq Z$, then we apply CC7 and thus it follows that $cover(C) \vdash X \leq Y$. $\square$

Theorem 2. The axiom system comprising inference rules CC1 to CC5 is sound and complete for the satisfaction of CCs in relations.
Proof. Let $X^+ = \{Y \mid X \leq Y \in cover(C)\}$ and $\mathcal{Y} = \bigcup_{Y \in X^+}(Y - X)$. We now define an *equivalence relation* $\mathcal{E}$ on $\mathcal{Y}$ as follows: for any pair of attributes $A_1, A_2 \in R$, $A_1 \approx_{\mathcal{E}} A_2$ if, for any $Y \in X^+$, $A_1 \in Y$ iff $A_2 \in Y$. Let C is an equivalence class (a set of attributes) induced by $\mathcal{E}$. The collection of all $E = (C - X)$, $\mathcal{P} = \{E_1, \ldots, E_n\}$, forms a partition of $\mathcal{Y}$. We now construct a counter example relation $r^c = \{t_0 <_r t_1 <_t \cdots <_t t_n\}$ as follows. Let $E_0 = X$. We generate an ith tuple for each E_i $(0 \leq i \leq n)$ as $t_i[A] = 0$ whenever $A \in E_i$, $t_i[A] = i$ whenever $A \in R - (X \cup \mathcal{Y})$, and 1 otherwise. The schema of r^c is valid, since all $E \in (\mathcal{P} \cup \{X\})$ do not overlap.

By Lemma 3, we know that CC1 to CC5 are sound for CCs. We prove the completeness by showing that if $C \not\vdash X \leq Y$, then $C \not\models X \leq Y$. Equivalently for the latter, it is sufficient to exhibit a relation r^c, such that $r^c \models C$ but $r^c \not\models X \leq Y$. Let r^c be the relation shown in Figure 10.

	X	E_1	E_2	$\cdots$	E_n	$R-(X \cup \mathcal{Y})$
t_0	$0 \cdots 0$	$1 \cdots 1$	$1 \cdots 1$	$\cdots$	$1 \cdots 1$	$0 \cdots 0$
t_1	$1 \cdots 1$	$0 \cdots 0$	$1 \cdots 1$	$\cdots$	$1 \cdots 1$	$1 \cdots 1$
t_2	$1 \cdots 1$	$1 \cdots 1$	$0 \cdots 0$	$\cdots$	$1 \cdots 1$	$2 \cdots 2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
t_{n+1}	$1 \cdots 1$	$1 \cdots 1$	$1 \cdots 1$	$\cdots$	$0 \cdots 0$	$n+1 \cdots n+1$

Fig. 10. A relation r^c showing that $C \not\models X \leq Y$

We first show that $r^c \models C$. Suppose to the contrary that $r^c \not\models C$ and thus there exists a CC, $V \leq W \in C$, such that $r^c \not\models V \leq W$. From the definition of X^+ and $\mathcal{P}$, V and W do not cross more than one E. It follows from the construction of r^c that $\exists A \in W$ such that $A \in R - (X \cup \mathcal{Y})$ and that $V \subseteq X$ or $V \subseteq E_i$ (as sets). In the first case, it follows by CC5 and CC6 that $C \vdash X \leq V$. By CC3, it follows that $C \vdash X \leq W$. Thus, it follows that $C \vdash X \leq A$ by CC6 again. This leads to a contradiction, since $A \in (X \cup \mathcal{Y})$. In the second case, it follows by the definition of E_i and by CC5 and CC6 that $C \vdash X \leq E_i$. By CC6, it follows that $C \vdash X \leq V$. By CC3, it follows that $C \vdash X \leq W$. Thus, $A \in (X \cup \mathcal{Y})$. This leads to the same contradiction again as the first case.

We conclude the proof by showing that $r^c \not\models X \leq Y$. Suppose to the contrary that $r^c \models X \leq Y$; by the construction of r^c, $Y \subseteq E_i$ (as sets). It follows by definition of E_i and by CC6 that $C \vdash X \leq E_i$. By CC3 and CC5, it follows that $C \vdash X \leq Y$. This leads to a contradiction, since we assume $C \not\vdash X \leq Y$. $\square$

5 Interaction Rules

In this section we investigate the interactions between FDs and PPs in Section 5.1 and those between FDs and CCs in Section 5.2.

We first state Armstrong's axiom, which is known to be sound and complete for FDs [1]. We also need to adapt the axiom to this context as follows.

Definition 7. (Armstrong's Axiom System) Let X, Y, Z be non-empty sequences of attributes obtained from R, $A \in R$ and F be a set of FDs.

(FD1) *Reflexivity:* If $Y \subseteq X$, then $F \vdash X \rightarrow Y$.
(FD2) *Augmentation:* If $F \vdash X \rightarrow Y$, then $F \vdash XA \rightarrow YA$.
(FD3) *Transitivity:* If $F \vdash X \rightarrow Y$ and $F \vdash Y \rightarrow Z$, then $F \vdash X \rightarrow Z$.
(FD4) *Permutation:* If $X \sim X'$, $Y \sim Y'$, and $F \vdash X \rightarrow Y$, then $F \vdash X' \rightarrow Y'$.

5.1 Interactions between FDs and PPs

We show that the axiom system that consists of PP rules in Definition 4, Armstrong's rules in Definition 7 and three new FD-PP interaction rules in Definition 8 is sound and complete for FDs and PPs.

Now, we present the"mixed rules" for the interactions between FDs and PEs.

Definition 8. (Inference Rules for Interactions between FDs and PPs) Let Γ be a mixed set FDs and PPs.

(FD-PPl) *Superkey:* If $\Gamma \vdash X$, then $\Gamma \vdash X \rightarrow R$.
(FD-PP2) *Absorption:* If $\Gamma \vdash X \rightarrow A$ and $\Gamma \vdash XAY$, then $\Gamma \vdash XY$.
(FD-PP3) *Generation:* If $\Gamma \vdash X \rightarrow A$ and $\Gamma \vdash XY$, then $\Gamma \vdash XAY$.

Similar to the concept of implication used in PPs and CCs, we say $r \models \Gamma$, if and only if $r \models \gamma$ for all $\gamma \in \Gamma$. Notably, the statement actually means $r \models f$ for any FD $f \in \Gamma$ and $r \models_u X$ and for any PP $X \in \Gamma$.

Lemma 5. The three interaction rules PF1 to PF3 are sound for the satisfaction of both FDs and PPs in relations. $\square$

We now show that the collection of the inference rules {PP1, PP2, PP3, FD1, FD2, FD3, FD4, FD-PP1, FD-PP2, FD-PP3} is a sound and complete set of rules for proving the implications of FDs and PPs taken together.

Theorem 3. The axiom system comprising inference rules PP1 to PP3, FD1 to FD4, and FD-PP1 to FD-PP3 is sound and complete for the satisfaction of both PPs and FDs in relations.

Proof. We only need to prove the completeness. Let $\Gamma = \Gamma_f \cup \Gamma_p$ where Γ_f is the set of all FDs and Γ_p is the set of all PPs. We now establish the completeness by showing that if $\Gamma \not\vdash \gamma$, then $\Gamma \not\models \gamma$, where γ is either f (an FD) or p (a PP). Equivalently for the latter, it is sufficient to exhibit a relation as a counter-example, r^c, such that $r^c \models \Gamma$ but $r^c \not\models \gamma$. We let $\Gamma_{p2f} = \{X \to R \mid X \in \Gamma_p\}$.

	X^+	$R - X^+$
t_1	$0 \cdots 0$	$0 \cdots 0$
t_2	$0 \cdots 0$	$1 \cdots 1$

Fig. 11. A counter example relation r^c used in the case $\gamma = X \to A$

	R
t_1	$1 \cdots 1$
t_2	$0 \cdots 0$

Fig. 12. A counter example relation r^c used in the case $\gamma = X$ when $\Gamma_p = \emptyset$

(*Case of* $\gamma = f$.) Let $\gamma = X \to A$. By FD-PP1, we have $\Gamma \vdash \Gamma_{p2f} \cup \Gamma_f$. Let $X^+ = \{B \mid \Gamma_{p2f} \cup \Gamma_f \vdash X \to B\}$. By the assumption of $\Gamma \not\vdash \gamma$, it follows that $A \notin X^+$. We consider the relation $r^c = \{t_1 <_r t_2\}$ shown in Figure 11. Clearly, $r^c \not\models_u X \to A$. We proceed to show $r^c \models \Gamma$. It is straightforward to check that $r^c \models \Gamma_f$. It remains for us to show that $r^c \models_u \Gamma_p$. Assume to the contrary that there exists $p \in \Gamma_p$ such that $r^c \not\models_u p$. Let $p = Z$. By construction of r^c, $Z \subseteq X^+$. It follows that $\Gamma_{p2f} \cup \Gamma_f \vdash X \to Z$. But $Z \to R \in \Gamma_{p2f}$. By FD3, it follows that $X \to R$. Thus, $X^+ = R$ and $A \in X^+$. This leads to a contradiction, since by assumption, $A \notin X^+$. This completes the proof of this case, since we have shown $r^c \models \Gamma_f \cup \Gamma_p$.

(*Case of* $\gamma = p$.) Let $\gamma = X$. There are two cases concerning Γ_p to consider.

First, if $\Gamma_p = \emptyset$, then the relation r^c in Figure 12 satisfies $r^c \models \Gamma_f$ but $r^c \not\models_u X$.

Second, if $\Gamma_p \neq \emptyset$, then we assume that X_0 is the largest prefix of X such that $\Gamma \vdash X_0 Q$ for some $Q \subseteq R$. Let us call this the X-assumption. We consider two further cases concerning X_0.

(Case 1:) When $X_0 = X$, we let $X^+ = \{B \mid \Gamma_f \vdash X \to B\}$ and $\Gamma \vdash Z$. We use again the relation shown in Figure 11. (But note that the definition of X^+ in this case is not the same.) It is clear that $r^c \not\models_u X$ but $r^c \models \Gamma_f$. It remains for us to show that $r^c \models_u \Gamma_p$. Assume to the contrary that there exists $p \in \Gamma_p$ such that $r^c \not\models_u p$. Let $p = Z$. By construction of r^c, $Z \subseteq X^+$. But $X_0 = X$ and thus, from the X-assumption, it follows that $\Gamma \vdash XQ$. By $\Gamma_f \vdash X \to X^+$ and FD-PP3, it follows that $\Gamma \vdash XX^+Q$. By PP2, it follows that $\Gamma \vdash XX^+ZQ$. By FD-PP1, $\Gamma \vdash Z \to R$. Thus, we have $\Gamma \vdash Z \to Q$. By FD-PP2, it follows that $\Gamma \vdash XX^+Z$ and by PP3 it follows that $\Gamma \vdash XX^+$. Thus, it follows that $\Gamma \vdash X$, since $\Gamma \vdash X \to X'$. This is a contradiction to the assumption of $\Gamma \not\vdash X$.

(Case 2:) When $X_0 \neq X$, we let $X = X_0 A Q$ where $A \notin X_0$. We let $X^+ = \{B \mid \Gamma_f \vdash X_0 \to B\}$. Note that $A \notin X^+$. Otherwise, it follows that $\Gamma_f \vdash X_0 \to A$ and by assumption $\Gamma \vdash X_0 Q$, it follows that $\Gamma \vdash X_0 A Q$ by FD-PP3. This leads to a violation of the X-assumption. We now consider the relation, r^c, shown in Figure 13. Clearly,

	X^+	A	$R - X^+A$
t_1	$0 \cdots 0$	1	$0 \cdots 0$
t_2	$0 \cdots 0$	0	$1 \cdots 1$

Fig. 13. A counter example relation r^c used in the case $\gamma = X$ when $\Gamma_p \neq \emptyset$ (Case 2)

$r^c \models \Gamma_f$ but $r^c \not\models_u X$. It remains for us to show that $r^c \models_u \Gamma_p$. Assume to the contrary that there exists $p \in \Gamma_p$ such that $r^c \not\models_u p$. Let $p = Z$. By construction of r^c, we have the following two possible cases of Z.

(Case of $Z \subseteq X^+$.) A contradiction can be established similar to the proof of Case 1 when $X = X_0$.

(Case of $Z \not\subseteq X^+$.) Let $Z = VAW$ where $V \subseteq X^+$ and $W \subseteq R$. Since $\Gamma \vdash X_0Q$ and $\Gamma \vdash VAW$, it follows by PP2 that $\Gamma \vdash X_0VAWQ$. Since $\Gamma_f \vdash X_0 \rightarrow X^+$, it follows by FD-PP3 that $\Gamma \vdash X_0X^+VAWQ$. Thus, by PP3 it follows that $\Gamma \vdash X_0X^+AWQ$. Finally, by FD-PP2 and $\Gamma_f \vdash X_0 \rightarrow X^+$, it follows that $\Gamma \vdash X_0AWQ$. This leads to a contradiction, since we violate the X-assumption. □

5.2 Interactions between FDs and CCs

We establish two new interaction rules for CCs and FDs. We show that the axiom system that consists of CC rules in Definition 6, Armstrong's rules in Definition 7 and the new FD-CC interaction rules in Definition 9 is sound and complete for FDs and CCs.

Definition 9. (Inference Rules for Interactions between FDs and CCs) Let Σ be a mixed set FDs and CCs.

(FD-CC1) *Reverse*: If $\Sigma \vdash X \rightarrow Y$ and $\Sigma \vdash Y \leq X$, then $\Sigma \vdash Y \rightarrow X$.
(FD-CC2) *Transformation*: If $\Sigma \vdash X \rightarrow Y$, then $\Sigma \vdash X \leq Y$.

Lemma 6. The inference rules FD-CC1 and FD-CC2 are sound for the satisfaction of both FDs and CCs in relations. □

We now prove the axiom system is sound and complete for unary CCs and unary FDs.

Theorem 4. The axiom system comprising inference rules CC1-CC5, FD1-FD4 and CC-FD1-CC-FD2 is sound and complete for the satisfaction of both CCs and FDs in relations.
Proof. We only need to prove the completeness. Let $\Sigma = \Sigma_f \cup \Sigma_c$ where Σ_f is the set of all FDs and Σ_c is the set of all CCs. We now establish the completeness by showing that if $\Sigma \not\vdash \sigma$, then $\Sigma \not\models \sigma$, where σ is either f (an FD) or c (an CC). Equivalently for the latter, it is sufficient to exhibit a relation as a counter-example r^c, such that $r^c \models \Sigma$ but $r^c \not\models \sigma$. We let $\Sigma_{f2c} = \{Y \leq X \mid X \rightarrow Y \in \Sigma_f\}$, which can be derived by the rule FD-CC1.

(*Case of* $\sigma = c$.) We now show that $\Sigma \models \sigma$ if and only if $\Sigma_c \cup \Sigma_{f2c} \models \sigma$. For the "if" part: by CC-FD1, it follows that $\Sigma \models \Sigma_c \cup \Sigma_{f2c}$. Thus, $\Sigma \models \sigma$. For the "only if" part: assume that $\Sigma_c \cup \Sigma_{f2c} \not\models \sigma$. We need to show that $\Sigma \not\models \sigma$.

Let $\sigma = X \leq Y$ and $X_f^+ = \{A \mid \Sigma_f \vdash X \rightarrow A\}$. We then modify the relation based on Figure 10 such that $\forall t \in r^c$, $t[A] = 1$ whenever $A \in X_f^+$. It follows by FD-CC2 that $X_f^+ \subseteq X \cup \mathcal{Y}$. Then, we can show that the following claim is true.

$(*)$ Claim: If $r^c \not\models X \rightarrow Y$ and $r^c \models X \leq Y$, then $r^c \not\models Y \rightarrow X$.

By using the claim $(*)$, we are able to check that $r^c \models \Sigma_f$. The proof of $r^c \models \Sigma_c$ but $r^c \not\models X \leq Y$ is similar to Theorem 2.

The result then follows by Theorem 2, since the set of inference rules for CCs in Definition 6 is complete.

(*Case of* $\sigma = f$.) Let FD be $X \rightarrow Y$. It can be shown that if $\Sigma \models X \rightarrow Y$, then $\Sigma_f \models X \rightarrow Y$, or else $\Sigma_f \models Y \rightarrow X$. Assume that $\Sigma_f \models X \rightarrow Y$. The result immediately follows by Armstrong's axiom. Otherwise, by the completeness of Armstrong's axiom it follows that $\Sigma \vdash Y \rightarrow X$. It also follows by FD-CC2 that $\Sigma \models X \leq Y$, since we assume that $\Sigma \models X \rightarrow Y$. Thus, it follows by the case of $(\sigma = c)$ in this proof that we have $\Sigma \vdash X \leq Y$. The result follows, since by FD-CC1 we have $\Sigma \vdash X \rightarrow Y$. $\Box$

6 Related Work

In literature, there is abundant work on data dependencies in relational databases [1] but they have not been used to capture user preferences. It is worth mentioning that in [5, 6] the axiom system for partial order dependencies is co-NP, which has limited the applicability of *order comparison dependencies* for decades. Here, with a given set of user preference terms, we override the partial order with a preference hierarchy and generate a preference relation, which simplifies much complex technicalities in establishing the axiom systems.

Preferences are receiving much attention in querying, since DBMSs need to provide better information services in advanced applications [7, 8]. In partiuclar, preference SQL [8] is equipped with a "preferring" clause that allows user to specify soft constraints reflecting multiple preference terms.

Our previous work [14] proposes Preference Functional Dependencies (PFDs) as an extension of FDs in relations, which captures the relationship between preferences and preference-dependent data. We emphasize that the constraints considered in this paper are entirely different from PFDs. We study the inference rules for preference constraints (PPs and CCs) in their own right. We neither incorporate preferences into FDs nor classify attributes as the assumptions in [14]. However, we thoroughly study the interactions between PPs, CCs and FDs.

7 Concluding Remarks

We model preference terms as partial orderings on a sequence of attributes and study the implication problem of preference satisfaction in a relation. We first formalize the concept of Prioritized Preferences (PPs), which is a sequence of preference attributes used for ranking a relation. We then establish a novel sound and complete inference system for PPs. The ranking choice is formalized as a set of possible rankings in a relation that satisfies a PP. We propose the concept of Choice Constraints (CCs) which capture the fact that the ranking choice resulting from one preference is less than or equal to another. We then establish a sound and complete inference system for CCs. Finally, we present interesting results on interactions between Functional Dependencies (FDs) and PPs, and between FDs and CCs. The main result of this paper is fundamental, which paves the way to transform the implication problem into a finite procedure for deriving PPs, CCs and FDs from a given set of such constraints. With the established axiom systems, efficient algorithms for checking various kinds of preference satisfaction are to be considered in our future work. It is also interesting to study how to infer and handle vague user preference [10, 11], since in real life the user may not be willing to detail and check all the preferences when querying.

Acknowledgements. This work is partially supported by RGC CERG under grant No. HKUST6185/03E and DAG04/05.EG10.

References

1. P. Atzeni and V. De Antonellis. *Relational Database Theory*. Benjamin/ Cummings Publishing Company, Inc., (1993).
2. J. Chomicki. Preference formulas in relational queries. *ACM Transaction Database System* **28**(4): pp. 427-466 (2003).
3. J. Chomicki. Semantic Optimization Techniques for Preference Queries. *Information Systems* **32**: pp. 670-684 (2006).
4. M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Co., New York, (1979).
5. S. Ginsburg and R. Hull. Order Dependency in the Relational Model. *Theoretical Computer Science* **26**(1-2), pp. 149-195, (1983).
6. S. Ginsburg and R. Hull. Sort Sets in the Relational Model. *Journal of the Association for Computing Machinery* **33**(3), pp. 465-488, (1986).
7. W. Kießling and G. Köstler. Preference SQL - Design, Implementation, Experiences. In: *Proc. of VLDB*, (2002).
8. W. Kießling and G. Köstler. Foundations of Preference in Database Systems. In: *Proc. of VLDB*, (2002).
9. H. Mannila and K-J Raiha. *The Design of Relational Databases.* Addison-Wesley, (1992).
10. A. Lu and W. Ng. *Mining hesitation information by vague association rules.* Proc. of ER 2007. This volume, (2007).
11. A. Lu and W. Ng. *Handling Inconsistency of Vague Relations with Functional Dependencies* Proc. of ER 2007. This volume, (2007).
12. W. Ng. Ordered Functional Dependencies in Relational Databases. *Information Systems* **24**(7), pp. 535-554, (1999).
13. W. Ng. An Extension of the Relational Data Model to Incorporate Ordered Domains. *ACM Transactions on Database Systems* **26**(3), (2001).
14. W. Ng. *Preference Functional Dependencies for Managing Choices*. In: Proc of ER 2006. LNCS Vol 4215, pp. 140-154,(2006).