DATABASES I

Multi-valued Dependencies and 4NF

Marek Sergot
Department of Computing
Imperial College

November 2001

CTT

Course	Teacher	Text
physics	Prof. Green	Basic Mechanics
physics	Prof. Green	Principles of Optics
physics	Prof. Brown	Basic Mechanics
physics	Prof. Brown	Principles of Optics
maths	Prof. Green	Basic Mechanics
maths	Prof. Green	Vector Analysis
maths	Prof. Green	Algebra

Primary key : (Course, Teacher, Text)

CTT is in 3NF and BCNF.
There are no non-trivial FDs. Only trivial ones, such as Course ∅ Course.

However CTT contains certain patterns of data that **resemble** FDs.

For example, for every course, there is a particular set of texts. And for every course there is a particular set of teachers of that course.

- texts are dependent on courses (but not functionally dependent);
- teachers are dependent on courses (but not functionally dependent);

The dependencies in CTT are **multi-valued dependencies** (MVDs) written:

COURSE $\rightarrow\rightarrow$ TEXT
COURSE $\rightarrow\rightarrow$ TEACHER

A $\rightarrow\rightarrow$ B is read:

attribute B of relation R is **multi-dependent** on attribute A of relation R, or equivalently, attribute A of relation R **multi-determines** attribute B.

To illustrate the idea further, here is how CTT would look in non-1NF form:

CTT-non-1NF

Course	Teacher	Text
physics	{Prof. Green, Prof. Brown}	{Basic Mechanics, Principles of Optics}
maths	{Prof. Green}	{Basic Mechanics, Vector Analysis, Algebra}

The relation CTT contains much redundancy, leading to update anomalies.
For example, to add that maths is also taught by Prof. White, we must make sure we add **three** tuples, one for each of the three maths texts.

Better design
Replace CTT by the two projections

CT (<u>COURSE</u>, <u>TEACHER</u>)
CTEXT (<u>COURSE</u>, <u>TEXT</u>).

Both are in BCNF.

Multi-valued dependencies (MVDs)

Definition: (Multi-valued dependency)

Let R be a relation scheme and let A and B be two sets of attributes in R (i.e. $A \subseteq R$ and $B \subseteq R$).

Note: in these notes, when R is a relation we also write R for the set of attributes of R and let context determine the intended meaning. (Standard practice in many texts.)

Then the **MVD A $\rightarrow\rightarrow$ B** holds on R if for all pairs of tuples t_1 and t_2 in R such that

$t_1 [A] = t_2 [A]$,

there exist tuples t_3 and t_4 (not necessarily distinct) in R such that:

$t_1 [A] = t_2 [A] = t_3 [A] = t_4 [A]$
$t_3 [B] = t_1 [B]$
$t_3 [R-B] = t_2 [R-B]$
$t_4 [B] = t_2 [B]$
$t_4 [R-B] = t_1 [R-B]$

Pictorially

For all t_1 and t_2 (not necessarily distinct) such that

	A	B	R-A-B
t_1	$\underline{a}$	$\underline{b}_1$	$\underline{c}_1$
t_2	$\underline{a}$	$\underline{b}_2$	$\underline{c}_2$

($\underline{a}$ stands for a 'vector' of values.)

there exist t_3 and t_4 (not necessarily distinct) such that:

	A	B	R-A-B
t_3	$\underline{a}$	$\underline{b}_1$	$\underline{c}_2$
t_4	$\underline{a}$	$\underline{b}_2$	$\underline{c}_1$

Remarks

The following all follow immediately from the definition of MVD. (Check!)

1. **Functional dependencies are a special case of MVDs.**

Multivalued dependencies (MVDs) are a generalisation of functional dependencies (FDs), i.e. every FD is an MVD, but the converse is not true.

Suppose an FD A$\longrightarrow$ B holds on a relation R. Then the MVD A $\rightarrow\!\!\rightarrow$ B also holds on R such that the set of values of B that are dependent on each value of A always consists of a single value.

2. For every relation R and (set of) attributes A, A $\rightarrow\!\!\rightarrow$ A holds. (it is a **trivial** MVD).

3. For every relation R and (set of) attributes A, A $\rightarrow\!\!\rightarrow$ R holds (where R stands for all attributes of R). This is also a a **trivial** MVD.

Remarks (contd)

4. **MVDs always come in pairs.**
   If MVD A $\rightarrow\!\!\rightarrow$ B holds on R then the MVD A $\rightarrow\!\!\rightarrow$ R–A–B also holds on R.

> **Definition:**
>
> An MVD A $\rightarrow\!\!\rightarrow$ B on R is **trivial** if B$\subseteq$A or A $\cup$ B = R (i.e. A $\cup$ B is all attributes of R).

There are axioms concerning MVDs similar to those we have seen for functional dependencies. We won't bother with them in this course.

Fourth Normal Form (4NF)

> **Definition:**
>
> A relation R is in **fourth normal form (4NF)** iff whenever there is a non-trivial multivalued dependency A ∅∅ B in R, then all attributes of R are functionally dependent on A.
>
> **Equivalently**, R is in **4NF** if and only if every non-trivial multivalued dependency A $\rightarrow\!\!\rightarrow$ B in R is in fact a functional dependency A $\longrightarrow$ B and A contains a candidate key.
>
> **Equivalently**, R is in **4NF** if and only if it is in BCNF and all non-trivial multivalued dependencies in R are in fact functional dependencies.

Any relation which is not in 4NF can be nonloss decomposed into a set of 4NF relations.

How? Decompose to get rid of all non-tivial MVDs that are not FDs. (See example above.) Then decompose to BCNF.