

VO2012

Virtual Observations 2012

WERKCOLLEGE 2: SQL & DATA MODELING

Document identifier: **VO2012-W2-01**

Date: **September 10, 2012**

Activity:

Document status:

Document link:

Abstract: This werkcollege shows how to use ER and UML for databases and SQL for statistics on your data

1. TASKS

Tasks:

- draw ER diagrams for TWOMASS, USNOA2, CROSSID
- draw UML diagrams for TWOMASS, USNOA2, CROSSID
- draw SADT diagram for cross-identification process
- draw UML diagram for cross-identification start mysql server
- create database and programs according to the planned data processing
- create simple aggregate functions with the use of SQL

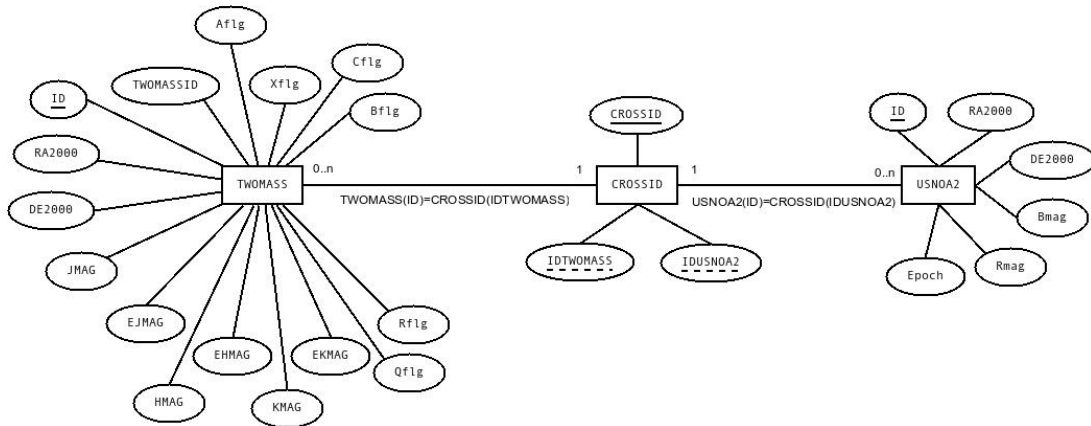


Figure 1: Full scheme for the cross-identification.

2. CROSS-IDENTIFICATION

Create a scheme according to ER diagram and make cross-identification according to the scheme.

3. REGRESSION IN MYSQL

Use the simple formula for linear regression with LSM to find dependence of magnitudes of cross-identified catalog. Assuming

$$y = ax + b \quad (1)$$

Find

$$a = (n \sum xy - \sum x \sum y) / (n \sum x^2 - \sum x \sum x) \quad (2)$$

and b using SQL functions SUM, COUNT and AVG.

Please, note that you can define variables in mysql as

```
SELECT @AVR_COORD := AVG(RA2000) FROM TWOMASS;
```

and reuse them in new statement. You have to be careful with initial setting of values of these variables.

4. USE OF AN XML

Define an XML scheme for your database using example below:

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns="sample.xsd" xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:sample="sample/0.1"
  targetNamespace="sample/0.1" elementFormDefault="qualified" version="0.1">

  <xs:complexType name="TWOMASS">
    <xs:annotation>
      <xs:documentation>Example of the schema for 2MASS simplified format</xs:documentation>
      <xs:appinfo/>
    </xs:annotation>
    <xs:sequence>
      <xs:element name="ID" type="xs:long"/>
      <xs:element name="RA2000" type="xs:float" />
      <xs:element name="DEC2000" type="xs:float" />
      <xs:element name="Jmag" type="xs:float" />
      <xs:element name="Hmag" type="xs:float" />
      <xs:element name="Kmag" type="xs:float" />
      <xs:element name="eJmag" type="xs:float" />
      <xs:element name="eHmag" type="xs:float" />
      <xs:element name="eKmag" type="xs:float" />
      <xs:element name="Qflg" type="xs:string" />
      <xs:element name="Rflg" type="xs:string" />
      <xs:element name="Bflg" type="xs:string" />
      <xs:element name="Cflg" type="xs:string" />
      <xs:element name="Xflg" type="xs:string" />
      <xs:element name="Aflg" type="xs:int" />
    </xs:sequence>
  </xs:complexType>

  <xs:element name="TWOMASSRow" type="sample:TWOMASS" />
</xs:schema>
```

Note the dummy record according to the scheme.

```
<?xml version="1.0" encoding="UTF-8"?>
<TWOMASSRow xmlns="sample/0.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="sample/0.1 file:/home/belikov/VO_lectures/2012/W2/sample.xsd">
  <ID>-1073741773</ID>
  <RA2000>0</RA2000>
  <DEC2000>0</DEC2000>
  <Jmag>0</Jmag>
  <Hmag>0</Hmag>
  <Kmag>0</Kmag>
  <eJmag>0</eJmag>
  <eHmag>0</eHmag>
  <eKmag>0</eKmag>
  <Qflg>Qflg0</Qflg>
  <Rflg>Rflg0</Rflg>
  <Bflg>Bflg0</Bflg>
  <Cflg>Cflg0</Cflg>
  <Xflg>Xflg0</Xflg>
  <Aflg>0</Aflg>
</TWOMASSRow>
```

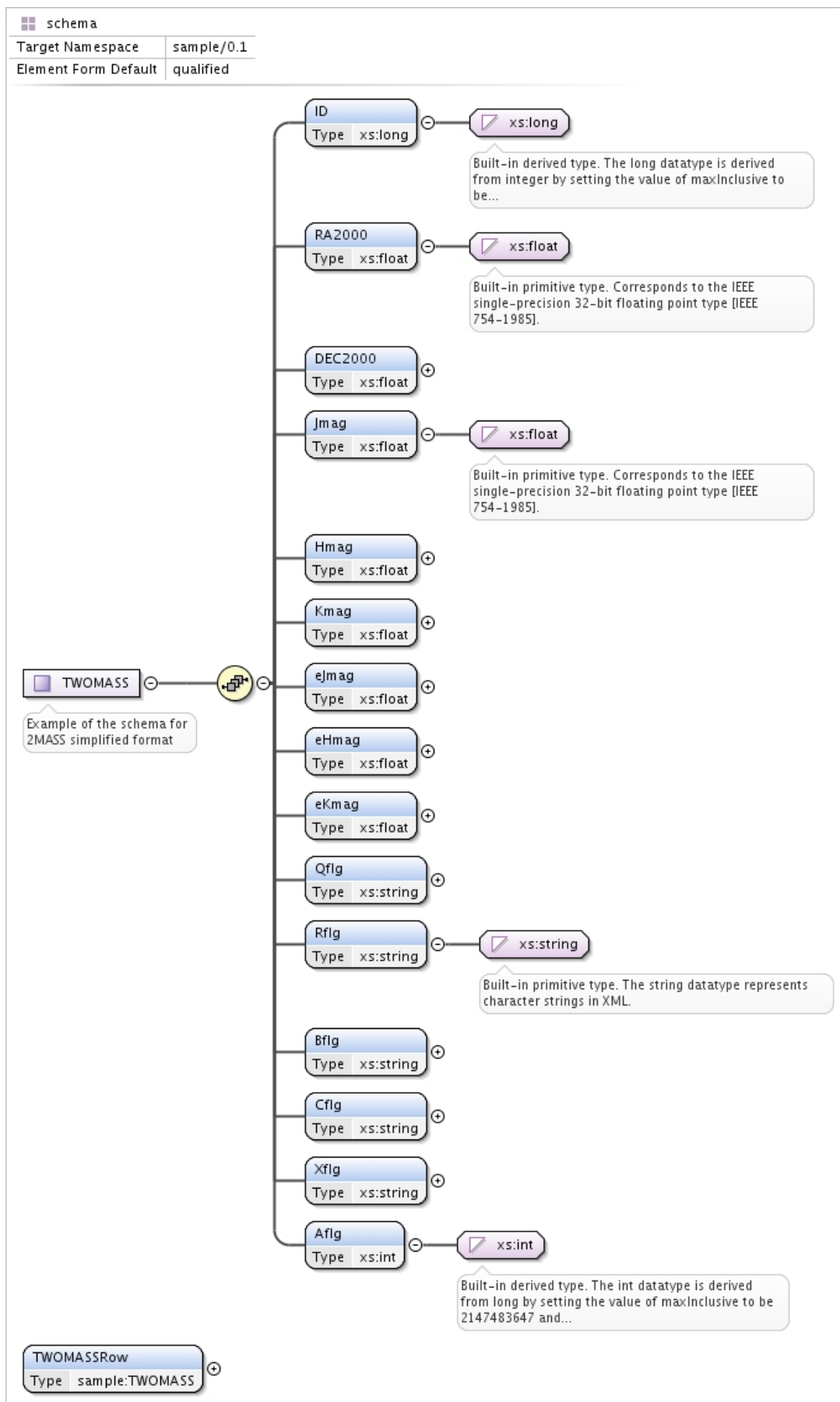


Figure 2: Sample XSD for 2MASS.