A short summary in English of "Database for inf3100"

The following is a short summary in English of the document originally written by Igor V. Rafienko and used in conjunction with the mandatory exercise in inf3100.

Background

The test database to be used in the inf3100 mandatory exercises is a fragment of the Internet Movie Database (IMDB) [1], which is a large database containing information on approximately 400 000 movies, persons related to the movies, descriptions of the movies, different dates, ranging of the films etc. All the movies chosen to be included in the database for inf3100 are in some way related to France. The database runs on the Oracle 9.2i on the server delphinium.ifi.uio.no.
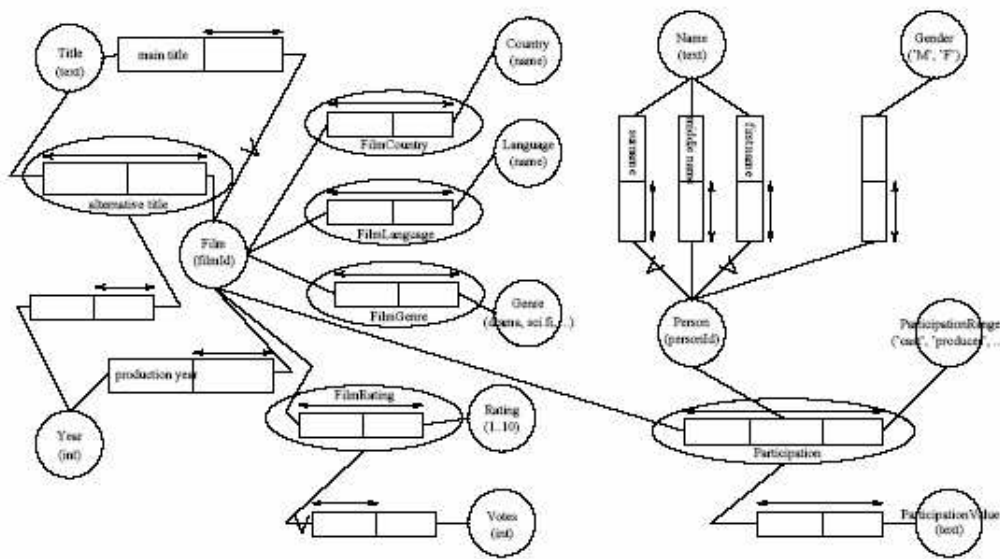

Figure 1: NIAM diagram for the relation database

Design

The database is centered around three entity types – movies, persons, and participation which links the first two together. The NIAM diagram can be seen in figure 1.

Each concept in a NIAM model translates into its own table during the realization stage. Not all concepts always warrants a table, and in this particular case, Country, Language, Rating, Votes, Year, Title, Name and Gender have been suppressed. The resulting tables are as follows: AlternativeFilmTitle, Film, FilmCountry, FilmGenre, FilmLanguage, FilmRating, Genre, Participation, ParticipationRange, ParticipationValue and Person.

The table Rating reflects how the movie has been rated in IMDB. This is based on people rating the films (1 to 10) over the Internet. The table Participation represents the participation of persons in movies. The participation can be of different types (all types are listed in ParticipationRange). For instance, the following SQL query describes the participation of Luc Besson in the film "The Fifth Element":

```
SQL> select * from Participation
  2 where personId = 26513 and filmId = 76914 ;
       PID   PERSONID PARTNAME                          FILMID
```

```
          ---------- ---------- ------------------------------ ----------
     2278671      26513 director                             76914
     2749250      26513 writing credits                      76914
SQL>
```

pid is an artificially created attribute, to facilitate joins between Participation and
ParticipationValue (the attribute is superfluous since <personId,partName,filmId> constitutes
a candidate key. But it is easier to join on one attribute rather than three, and less information
is duplicated in this fashion.)

ParticipationValue contains the values describing participation, e.g. roles played by actors in a
movie. The following example shows the role played by Milla Jovovich in the movie "The
Fifth Element":

```
SQL> select * from ParticipationValue
  2 where pid = 1846849 ;

       PID            VALUE
---------- ---------------
   1846849          Leeloo
SQL>
```

Examples
The following example shows all the roles played by Milla Jovovich with the corresponding
movies.

```
SQL> select pv.value, f.mainTitle
  2 from ParticipationValue pv, Participation pa,
  3      Person p, Film f
  4 where pv.pid = pa.pid and pa.filmId = f.filmId and
  5       pa.personId = p.personId and
  6       p.surName = 'Jovovich' and
  7       p.firstName = 'Milla' ;

       VALUE                                    MAINTITLE
--------------- ---------------------------------------
Mildred Harris                                   Chaplin
Lucia                                         Claim, The
Leeloo                                 Fifth Element, The
Joan of Arc    Messenger: The Story of Joan of Arc, The
SQL>
```

Source Code
The source code for the database is available from the inf3100 course pages.

Resources
It may be useful to familiarize yourself with the Oracle SQL dialect. A lot of documentation
can be found at /store/share/doc/ifi/oracle. The most interesting part is most likely the
"SQL Reference" manual. Ask your group teachers if you have questions related to SQL.

References
[1] The Internet Movie Database, http://www.imdb.com/