

Database and Information Systems

Homework 5

October 30, 2003; Due November 6 at 1:30 PM

Suppose you are starting a new website called **Nile.biz**, a clone of Amazon.com. You have three suppliers, two of whom supply music (and whose schemas were actually excerpted from CIS 550 Homework 4 submissions), and the third of which supplies CDs, movies, and books. You would like to have a unified, integrated view of this data so you can support your website. The three sources have the following schemas:

1. **CD**(key, title, publisherName, genre, artistXref: refs Artist)
CDsong(key: refs CD, song)
shippingRate)
Participates(artistKey: refs Artist, publisherName)
Artist(artistKey, name, street, city, state, zip, country)
ArtistFormerName(artistKey: refs Artist, name)
2. **CD**(cdkey, artistKey, title, genre, year)
Track(cdkey: refs CD, song)
CDsubtitle(cdkey: refs CD, subtitle)
CDlabel(cdkey: refs CD, label)
Artist(artistKey, group, website, street_address, box, city, state, country, zip)
KnownAs(artistKey, name)
Plays(artistKey: refs Member, instrumentType)
Employee)
refs Employee)
3. **Item**(iKey, classification, title, category, year)
SubItem(iKey, title)
Publishes(iKey: refs Item, pKey: refs Publisher)
Publisher(pKey, name, street, city, state, zip)
DevelopedBy(iKey: refs Item, aKey: refs AuthorOrContributor, role)
AuthorOrContributor(aKey, name, nickname, website, street, city, state, zip, country)
Stock(iKey, count, location)

Notes: **Item** classifications are “book,” “CD,” or “movie.” Roles include “artist,” “star,” “director,” and “author.” **SubItems** are either chapters or songs, depending on what the referenced item is. **Publishers** in this source are *only* based on the USA.

Propose a mediated relational schema for these three data sets. This mediated schema should be able to capture most of the relevant information in the schemas, and it should be flexible enough to incorporate movies, CDs, and books (as well as most of their associated information).

Note that it may be challenging to support *every* possible attribute, so you may wish to drop certain attributes. Assume that CD keys, book ISBN numbers, and movie keys will not collide (hence you can safely union them together). You may assume that all key values are global across all databases, i.e., if CD keys in two sources are the same, then the CDs are the same. You may also assume that publisher names are globally unique (i.e., they're candidate keys for Publisher).

Given the schema, propose a set of view mappings between your data sources and your mediated schema. You are free to use either local-as-view or global-as-view mappings, in the Datalog language.

Briefly summarize (in English) what you think are the main shortcomings of your mediated schema and mappings. What did you have the hardest time representing? What did you have to omit? Where do the mappings not describe the relationship well?