

## Database Systems Lab 1

### *Data vs. Information*

Imagine there were a database of every instance where some form of media has referenced the Star Wars series. The raw data would be a list of descriptions, documenting very visual, dialog, or stylistic reference to Star Wars. To arrange this data into coherent information, the database needs to supply some context for these scenarios. For example, each description should be accompanied by the movie/television series/game/book/etc. that the reference comes from, as well as the timestamp or page it could be found at.

An example of the kind of no-context data we could find in this database, we might find an entry that reads, "Luke, I am your father. I always wanted to say that." While this kind of raw data might make for a vigorous fan quiz, it has no inherent organization or context, so it is of little use for users of this database. However, if we linked this quote with the source material ("The Fairly OddParents"), the media type (television cartoon), and episode title ("The Fairly OddParents!"), the quote is a much more valuable piece of information.

### *Data Models*

The hierarchical data model consists of layers that get more specific as they are traversed. The data is stored in relation to categories "above" them, mimicking the canonical structure of a tree. One of the shortcomings of this model is that it is potentially wasteful of space. For instance, if two branches of the hierarchy need to contain the same item as a child, that item must be stored twice. The network data model is a slight improvement to the hierarchical model, because it allowed more dynamic interconnectivity between elements. It therefore eliminates the duplicate problem that the hierarchical model has.

However, both of these models pale in comparison to the relational model. Relational is based on tables, which allows for more intuitive methods of organizing data. It is also more versatile, because by using intermediate tables that connect the tables that contain the raw data, one can traverse the database in many ways, instead of the strictness of hierarchical or network model.

In light of the above evaluations, XML would be as good a data model as the hierarchy model, i.e. not a very good one. XML is very structured language that is fairly similar to a hierarchy. It would not even have the benefits of a network model, because an XML tag can only be the child of one other XML tag, with no "diamond" relations or tag sharing allowed.