

"Ontology" and "Taxonomy": Short Definitions.

Dr. Tom Johnston
MindfulData.com
Written April, 2008.



Definition: "Ontology" (as it applies to data).

A set of structures and rules that express what that data means in a form that is usable by software applications that manage and provide query access to that data.

note: relational databases, and the SQL query language, also do this. But the term "ontology" refers to software that goes beyond what relational databases can do in this regard. Software that does this is usually called an "inference engine". So another definition of "ontology" would be "a set of structures and rules that make inference engines possible".

Example.

(a) If employee X earns more than employee Y, and Y earns more than employee Z, then it follows that X earns more than Z. But:

(b) if employee X regularly meets with employee Y, and Y regularly meets with employee Z, it does not follow that X regularly meets with Z.

This is obvious to us. But no relational database can capture this difference, and computer code can do so only by hardcoding a special rule.

The difference between "earns more than" and "regularly meets with" that accounts for the "follows that" vs. "does not follow that" difference is that the former relationship is transitive while the latter is not.

Inference engines do not need to hardcode specific rules about specific relationships. Based on metadata which characterizes each relationship as transitive or not, an inference would tell you that "X earns more than Z" even if that fact were not stored in a database, just as long as the two other facts in (a) were stored in that database. But based on the same metadata, that inference engine would not tell you that "X meets regularly with Z" even though the database it has access to does contain the two other facts in (b).

Definition: "Taxonomy" (as it applies to data).

A hierarchy made up of parent/child relationships, in which each child represents a type or a "kind of" the parent.

note: there is a lot of sloppy usage of this term going around, but this is the correct definition. So, for example, a hierarchy in which Personal Computer is a parent, and Screen, Keyboard, Motherboard and Power Supply are children, is not a taxonomy. Reason: screens, keyboards, motherboards and power supplies are not types, or kinds, of personal computers. They are parts of a personal computer.

Example.

A hierarchy in which Databases is a parent, and Hierarchical, Network, Object Oriented and Relational are children, is a taxonomy. Reason: a hierarchical database is a type, or kind, of database, and so too for network, object oriented and relational databases.