Dimensional arrangement of knowledge structures

Sebastian Samaruga
Cognescent
Buenos Aires, Argentina
+54 9 11 6807 9874
cognescent@gmail.com

Abstract

This paper discusses dimensional modeling in respect to currently available semantic web / knowledge representation technologies and techniques.

1. Measures

Given known facts about the world, any scientific intelligence[1] capable of regarding about them even interpreting those facts, remembering them or communicating them to another scientific intelligence would need some kind of convention for talking (internally or externally) about that data, making it available as information for later processing it into knowledge.

The most elemental tool for given such intentions will be that of representing those facts given a representation convention for the dimension in which the quantities being measured took place. This could be regarded, in the most elemental sense, as a data type in which the data is available as information. Later it will be needed to process it to further become knowledge, for what is necessary this data type to be regarded in co-relation with others and with the actual facts or 'things' of the real world they resemble.

For the sake of simplicity, let's state here that any given measure of a quantity (call it Value) will be regarded as something (call it Property) of something (owner of the Property, an Object) given an observer (an Agent) who maps the measure (a Sample) and asserts the Sample facts in a given Context, being this last relative to the Agent and the Sample measure itself.

2. World's facts modeling

Given the approach took in semiotics[1] the model of the facts expressed using a sign/symbol mechanism for representing knowledge is similar as the one mentioned above. That is, the essence of a symbol, in semiotics, is that it is an Entity that stands for or represents another Entity for an Agent who interprets the symbol.

In our case, the agent would 'interpret' the sign (evaluate it according the Context) and 'register' this snapshot or sample he took of the 'meaning' of the sign, again, according it's Context (or previously available statements in it's knowledge).

Peirce says that a sign/symbol stands for an object and the interpretation process is the de-referencing of the concept in the memory of the agent interpreting the sign, bringing the sign to a 'live' representation of the concept in the brain of the agent.

Any dimensional modeling, even the most basic[2] should regard time as a dimension. In our attempt to make snapshots / samples of reality available as knowledge for Agents in Contexts, we could model our sampling of the world regarding three dimensions:

1. Time
2. Location
3. State

And given them we could ask to our set of statements, or knowledge base, for example, the value of (1) regarding values in pairs of (2) and (3) given a particular Context or, more specifically, a particular Sample.

A Sample will encode State given Location in Time. The whole value of the three items in the statements is regarded as the result of causal relationships that affected entities in the (evaluated) world that lead them to become the ones that are sampled. This could be regarded as a fourth dimension, in which we can relate something that happened before to the state of affairs in the actual contextually evaluable world, given that this information is available. Or else, we could infer why this is like it is right now given the appropriate statements in the KB that relate those facts.

Example:
a. The floor is wet (now)
b. If it rains then the floor is wet.
c. It rained (inferred)

This same technique, plus temporal logic / tense logic multiple truth values could allow to model 'behavior' as statements in the knowledge base: 'data', given that something that happens now makes truth a previous statement and allows new statements to be inferred. (Ala production system / RETE rule base)

An historical cause-effect relationship holds for the data known as facts, and they are related accordingly given the ordering of events in each dimension. A scenario, hypothesis or tendency flow could be modeled regarding knowledge of what could be the reason of something to happen.

3. Relations and propositions

Given way of modeling we can encode relations between objects, and map relations from one set to another in a table (later use for FCA analysis) and also encode logical propositions in the form of statements.

The first is implemented mapping all objects with the same property and aggregating range and domain of the property.

The propositions, or logic statements, are the way to encode some knowledge, using logical relations (in our case we'll use the five semantic primitives and the three logical operators mentioned by Peirce) in which the relation property is one of primitives or operators. Statements can be nested and their subjects and objects be statements in turn. Primitives are: Existence, co reference, relation, conjunction, negation. And operators are: Universal, implication and negation[1].

The truth values, given they are added and evaluated in a evaluation Context, for which they and their values become part, correspond to the rest of (related) statements they refer, or to an assignment made in insertion and, if it holds, will allow the context to infer new data. Also new statements could allow to the Context to 'react' producing new statements for a given 'behavior' being encoded in them.

4. 3-Tuples / Functional mappings

There are cases where measures about a fact, for example, a Travel, are related each other in a ternary form. For example, given a speed, some distance is traveled at a specific time. Another example could be, given a computer font or typeface, some glyph (representation of a character by a drawing on screen) maps to a character (example, in Unicode).

This ternary mappings are easily mapped to functions of the form f(x, y) -> z, where x, y and z can be any of the three mentioned items. And, furthermore, given mappings of this form in raw data, learners and classifiers could be trained
to map and infer the third (missing) value from the other two, knowing it's type.

This mapping technique can be useful in many similar situation with real (surrogate) keys like the given above. But it could also be applied if one carefully identifies the objects in the 3-Tuples with the adequate IDs, given a sequence, and use the IDs later for calculating the correspondences.

5. Representation

The measures and they values are easily stored as numbers for, for example, integer quantities, strings for names, etc. But if we want knowledge bases to inter operate, there must be a Representation layer where one can unequivocally represent 'something', being this representation absolute or relative.

For example, for something to be regarded as 'hot' (high temperature) we could state that it will be hot if temperature (another thing needing representation) is over some degree. Or if the temperature, compared to the temperature of the environment, is superior to +5 degrees.

Ways with such representations could be achieved range from implementing different grammars (for example, DSLs), using a numeric system convention, using semantic conventions in the pragmatics of the conversation and, of course, previously agreeing in meaning of terms and concepts represented in those ways.

6. Formal Concept Analysis (FCA)

The techniques shown in [3] are proven to be useful in the analysis of concept lattices that can be created from objects, their properties and the property's values. The technique of scaling is needed to convert from continuous values to discrete ones, using, in this case, ranges of values. It is also useful to group attributes by key, in key-value pairs, that are idempotent to attributes as used in conventional FCA (example, Sex: Male, Female) (TBD table).

7. Architecture / Meta model

For this approach to be useful, there is need for a common meta model for the data and information comming from different sources. At least we have to represent units of data (Entity), which correspond to a property (Role) in a surrounding entity (Mapping) for a given concept (Context). This roughly matches with a database's table cell, column, row and table. And to an RDF property, predicate, triple and subject.

There is also the need for meta data to be held about the aforementioned entities of the meta model, regarding its origin, context and meaning. This is like MDM (Master Data Management) would be to BI (Business Intelligence) and the governance and traceable aspects of knowledge modeling.

8. Functional user interface

At the user's end, inspired by MVC pattern creator, there is a 'Use case driven / Functional' user experience / interface approach. The user should be careless about the underlying representation of the information he's handling. And the whole user interaction process should be domain driven and not technology or implementation driven.

Aspects, metaphors and controls should be developed so the user's head is focused only in 'what' he is doing and not in 'how' he will do it. Multiple applications domain research should be done so the actual metaphor on this 'functional' paradigm will work.

9. References

1. "Ontology, Metadata, and Semiotics", Sowa, http://users.bestweb.net/~sowa/peirce/ontometa.htm
2. "A Brief Tutorial on Database Queries, Data Mining, and OLAP", Lutz Hamel
3. "A first course in formal concept analysis", Karl Wolff