

The Universal Data Cube Ontology

Curran Kelleher

Abstract

The Universal Data Cube Ontology is a conceptual model facilitating the creation, interchange and integration of collections of comparable data cubes and their metadata.

1 Classes

The Universal Data Cube Ontology is defined by the following classes:

A **DataCube** represents the multidimensional categorical space derivable from the set of all possible instances of a particular **owl:Class**, termed the *fact class* of the **DataCube**. The fact class is analogous to the *fact table* in a snowflake schema. A **DataCube** is defined by an **owl:Class** (its fact class), a set of **Dimensions** and a set of **Measures**.

A **Dimension** represents the set of all subsets of the range of a particular **owl:Property** of the fact class of its parent **DataCube**, termed the *dimension property*. The range subsets can have a hierarchical containment structure, termed a *dimensional hierarchy*, whose nodes can be classified by *hierarchy levels*. Each hierarchy level is represented by a **Level**, and each hierarchy node (range subset) is represented by a **Member**. This structure is a prototypical one,

A **Level** represents a level of the dimensional hierarchy of its parent **Dimension**. A **Level** contains many **Members**. A **Level** also is representative of a corresponding **owl:Class**.

A **Member** represents a subset of the range of the dimension property of its parent **Dimension**. A **Member** is also representative of a corresponding **owl:Individual**, whose type must be the **owl:Class** associated with the **Member's** parent **Level**.

A **Measure** is a mapping from a set of facts (instances of the fact class of its parent **DataCube**) to a numeric value derived from applying a particular **AggregationFunction** to a particular **owl:Property** of the fact class.

The classes described thus far represent structural components of *prototypical* data cubes, and do not represent any actual concrete data cube. Instances of these classes should be published in the Semantic Web and viewed as permanent and authoritative definitions of absolute objects found in the real world. These definitions must be reliable, as they will serve as the common vocabulary which data cube content providers will use to describe their dimensions and measures.

A **DataCubeInstance** represents a live and queryable data cube whose prototypical structure matches that of a particular **DataCube**. The structure of the concrete contents of the **DataCubeInstance** are defined by a set of **DimensionInstances** and a set of **MeasureInstances**. A **DataCubeInstance** also has a data property called **queryEndpointURL**, which contains the URL of the UDC Query Service exposing the actual data cube content.

A **DimensionInstance** represents the contents of a particular **DataCubeInstance** from a particular **Dimension**. A **DimensionInstance** has a set of **LevelInstances**.

A **LevelInstance** represents the contents of a particular **Level** within a particular **DimensionInstance**, and is defined by a set of **MemberInstances**.

A **MemberInstance** represents the appearance of a particular **Member** within a particular **LevelInstance**.

2 Operations

The ontology only represents metadata. By publishing an instance of the **DataCubeInstance** class, a data provider is stating that there exists presently a queryable data cube, conforming to the structure defined by the **DataCubeInstance**, whose content is made available at the **queryEndpointURL** of the **DataCubeInstance**.

At the **queryEndpointURL** of a particular **DataCubeInstance**, one can request a projection of the data cube content using a simple query language.