

BASIC SCHEMA ON THE GEOGRAPHIC INFORMATION DEVELOPED FOR HIGHER EDUCATION

Morishige Ota

Abstract

In order to clarify and solve geographic problems of our society, the standardized schemas to describe geographic information (GI) included in GI standards are quite useful to share knowledge and to exchange information. However, there are very few chances to learn these standards at higher education in Japan. One of reasons is that the university aims to deepen "theories" and not "specifications". Meanwhile, standards are well structured, consistent with others and they are constructed on knowledge of Geographic Information Technology (GIT). Thus, we can see GI standards as a systematized gateway to reach knowledge of GIT. This is the reason that we developed simplified geographic information schemas (SGIS) for higher education as a profile of GI Standards, and use them in the introduction course of GIT. We have been practicing experimental lectures at the University of Tokyo since 2008.

Keywords: higher education, geographic information schema, spatial data infrastructure

geographic information schema

spatial data infrastructure

Mr. Morishige Ota

Kokusai Kogyo Co., Ltd.

2 Rokuban-cho, Chiyoda-ku, Tokyo, 102-0085, Japan.

Tel: +81 3 6361 2456, E-mail: morishige_ota@kkc.co.jp

INTRODUCTION

Today, the construction and utilization of Spatial Data Infrastructure (SDI) are common from national to international level, and related technology is progressing rapidly. This trend started from US [1] and necessary GI Standards for SDI are provided by ISO/TC211, OGC and other related organizations. "Basic Act on the Advancement of Utilizing Geographic Information" was enacted in Japan at 2007 [2]. In the same year, EU published INSPIRE Directive for establishing an infrastructure for Spatial Information in the European Community [3].

GIS Association of Japan and the Center for Spatial Information Science at the University of Tokyo co-organize the seminar called "the Studio for Spatial Information Standards" for the people working in the related industries since 2005 and more than 300 people have already participated. However, engineers establishing maintain and improvement of SDI are still not sufficient in Japan.

Meanwhile in Japan, the initiative on a GIT BoK was undertaken by the research project team (Leader: Masatoshi Arikawa) as a part of "Development of geographic information science curricula and sustainable web library systems for serving their contents" (Chair: Atsuyuki Okabe) by Japan Society for the Promotion of Science from 2005 up to 2007 [4][5]. And the new research initiative "Research project on the education of Geographic Thinking" chaired by Yasushi Asami was started in 2009 and it is following and deepen the previous research work.

Thus, we think it will be possible to supply engineers who have basic skill concerning SDI in near future, if we will be able to build fundamental knowledge of SDI onto GIS courses in the higher education. However, what is the fundamental knowledge? At first, this paper will introduce standardized geographic schemas as the fundamental knowledge for SDI. Then we will introduce each schema. And we will report the experimental lectures held at the University of Tokyo, since 2008. Finally, we will conclude the discussion and make mention of our future work.

SIGNIFICANCE OF EDUCATION ON GEOGRAPHIC INFORMATION STANDARDS

We need policies, standards, mechanisms and skillful people to develop, maintain and improve SDI. According to Executive Order 12906: Coordinating Geographic Data Access published by the White House in 1992 [1], 'National Spatial Data Infrastructure' is defined as "the technology, policies, standards, and human resources necessary to acquire,

process, store, distribute, and improve utilization of geographic data".

Meanwhile, Article 17 of Japanese NSDI Act enacted in 2007 [2] prescribes 'Interoperability of Fundamental Geographic Data (FGD) in Geographic Services' as "The State and local governments shall make efforts to maximize the interoperable use of exiting FGD when they prepare maps in those regions where they carry out their work and projects in such administrative functions that require to use maps as city planning, maintenance of public facilities, management of agricultural areas including agricultural fields and forests, cadastral survey, real estate property registration, taxation and statistics."

And, INSPIRE Directive published in 2007 [3] defines 'infrastructure for spatial information' as "metadata, spatial data sets and spatial data services; network services and technologies; agreements on sharing, access and use; and coordination and monitoring mechanisms, processes and procedures, established, operated or made available in accordance with this Directive;"

The common aim of these order, act, and directive is to increase interoperability and usability of geographic data. To achieve this aim, we need policies, standards, mechanisms, and human resources. Actually policies were made and now we have rules in various regions, countries and organizations. ISO/TC211, OGC and other related organizations have been providing international standards in the past decade. Mechanisms such as internet, clearinghouse, and registries have been established and improved rapidly. However, do we have enough human resources who understand both Information Technology and Geographic Information Standards? In case of Japan, education is one of the most important issues to realize and improve SDI in state and local governments. This is the reason that we should teach geographic information standards at departments of Geo-related Science and Information Technology in Universities.

SCHEMAS FOR GEOGRAPHIC INFORMATION EXCHANGE AND UTILIZATION

Model is a representation as a simple description of the object in the real world. Schema is a formal description of a model. Schema commonly approved in the society is called "standard schema". Geographic Information Standard such as those provided by the international organization use formal language to describe models. For example, ISO/TC 211 uses UML as the conceptual schema language [6]. These schemas are comprehensive and versatile enough for various applications. However, they are too complicated to teach in the limited time during the university education. Thus, we developed Simplified Geographic Information Schemas (SGIS) and use them in the half-year course for the introduction of GIT. Schemas included in SGIS are shown in Table 1.

Table 1. SGIS for the introductory course of GIT

Schema packages	Outlines	References
Basic Data Types	It defines fundamental data types such as number, string, and Boolean	ISO/TS 19103:2005 Conceptual schema language
General Feature Model	It defines feature type as abstraction of real world phenomena, its properties and relationships (inheritance, association).	ISO 19109:2005 Rules for application schema
Spatial Schema	It defines spatial characteristics of features	ISO 19107:2003 Spatial schema
Temporal Schema	It defines temporal characteristics of features	ISO 19108:2002 Temporal schema
Coverage Schema	It defines coverage as a function to return values from its range for any position within its domain.	ISO 19123:2005 Schema for coverage geometry and functions
Reference System Schema	It defines coordinate, location and temporal reference system. Coordinate reference system consists of geodetic datum, coordinate system and projection system. Location reference system defines gazetteer to get position from location identifier and vice versa. Temporal reference system defines calendar system and clock.	ISO 19111:2003 Spatial referencing by coordinates ISO 19112:2003 Spatial referencing by geographic identifier ISO 19108:2002 Temporal schema
Instance Model	It specifies how to implement instances (geographic data).	ISO 19118:2005 Encoding
Portrayal schema	It defines schemas for styling symbols and labels on graphic devices	ISO 19117:2005 Portrayal

BASIC DATA TYPES

Basic data types are grouped into three categories in ISO/TS 19103:2005 [6]. They are,

- Primitive types: Fundamental types for representing values, examples are *CharacterString*, *Integer*, *Boolean*, *Date*, *Time*, etc.
- Implementation and collection types: Types for implementation and representation structures, examples are *Names* and *Records*, and types for representing multiple occurrences of other types, examples are *Set*, *Bag* and *Sequence*.
- Derived types: Measure types and units of measurement.

However we defined the least basic data types for the introductory course shown in Figure 1. Italic typeface is used in the class diagram to show that the type is “abstract”. Simultaneously, terms using italic typeface in the main body indicates the types (classes) defined in the schemas.

- Primitive types: *Number*, *Integer*, *String*, *Image*, *RGB*, *Multiplicity*, *Object*, *Record*, *RecordType*, *NameTypePair* and *Null*.
- Collection types: *Vector*, *RecordSet*.

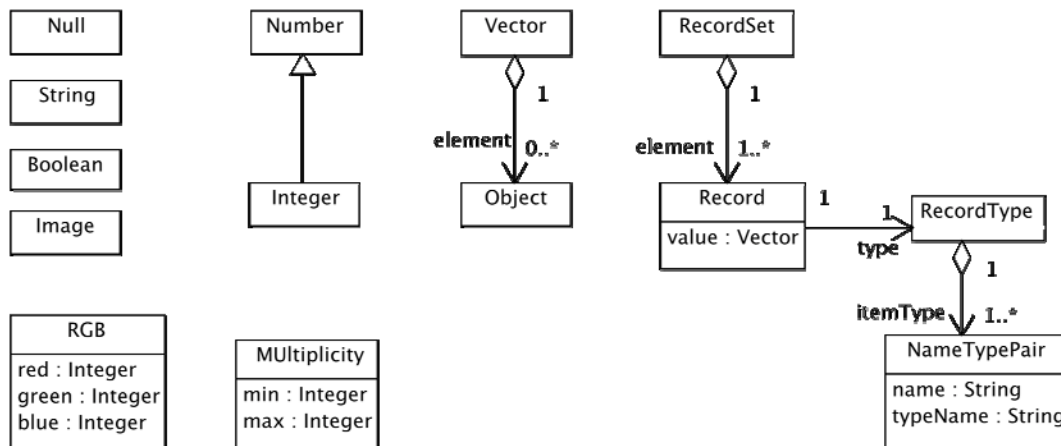


Figure 1. Simplified schema package for basic data types

In Figure 1, *Vector* is an array of objects. *Object* means any type. *Record* is a vector that associates with *Record type*. It defines name and data type of each object in a vector using *NameTypePair* (name: object name, typeName: object type). *Multiplicity* is used to define cardinality, in other words, minimum and maximal limits of multiple occurrences of data taking same type. *Number* is any type of number such as integer and real. *Integer* is used in the case to restrict the type is integer. *Null* means that value is not exist. *Boolean* takes a value of true or false. *Image* can be any type of image such as jpeg, gif, tiff, and png. *RGB* shows color defined by red, green and blue. The value of each component color is from 0 up to 255.

GENERAL FEATURE MODEL

General Feature Model (GFM) is the model of the concepts required to classify a view of the real world. It is expressed as UML class diagrams defined in ISO 19109 – Rules for application schema [7]. GFM is a meta-model that establishes a basis for the classification, structure and relationships of “features”. *Feature* is defined as abstraction of real world phenomena in ISO 19000 series. Its type is the representation of concept and its instance is the representation of each entity in the reality. UML notation rule is also a meta-model but GFM provides a basis for classification of any things in the real world. Schema designed for application in accordance with GFM is called application schema.

The relationship between feature types is expressed as association or inheritance. Feature type has properties that are attributes, operations and association roles. All these concepts are expressed as UML meta-classes. Figure 2 is a simplified GFM for the education in the introductory course.

GFM expresses that *FeatureType* is defined by its type name, definition and the flag indicating whether it is abstract type or not. Feature type may be the extension of single super type and links to other feature types through association. *AssociationType* includes feature types relating other types. Usually it includes a pair of feature types. However, if it includes only one feature type, it implies self-association. For example, in case a house has adjoining relation with other houses, its type has self-association. *AssociationType* is a subtype of *FeatureType*, because association also can be

thought as abstraction of real world phenomena. As association type is a feature, it inherits typeName, definition, and abstract. However typeName is optional, because there is a case that it is undefined.

Feature type has properties. *PropertyType* is classified into attribute, operation and association role. *AttributeType* may be spatial, temporal, location or thematic. Spatial means geometry and topology such as point, curve, surface, node, edge and face. Temporal means instant period, temporal node and edge. Location means geographic identifier such as address and postal code. Thematic means other attributes with basic data type such as character string, number, date, time, and image. These types are given as typeName in *AttributeType*.

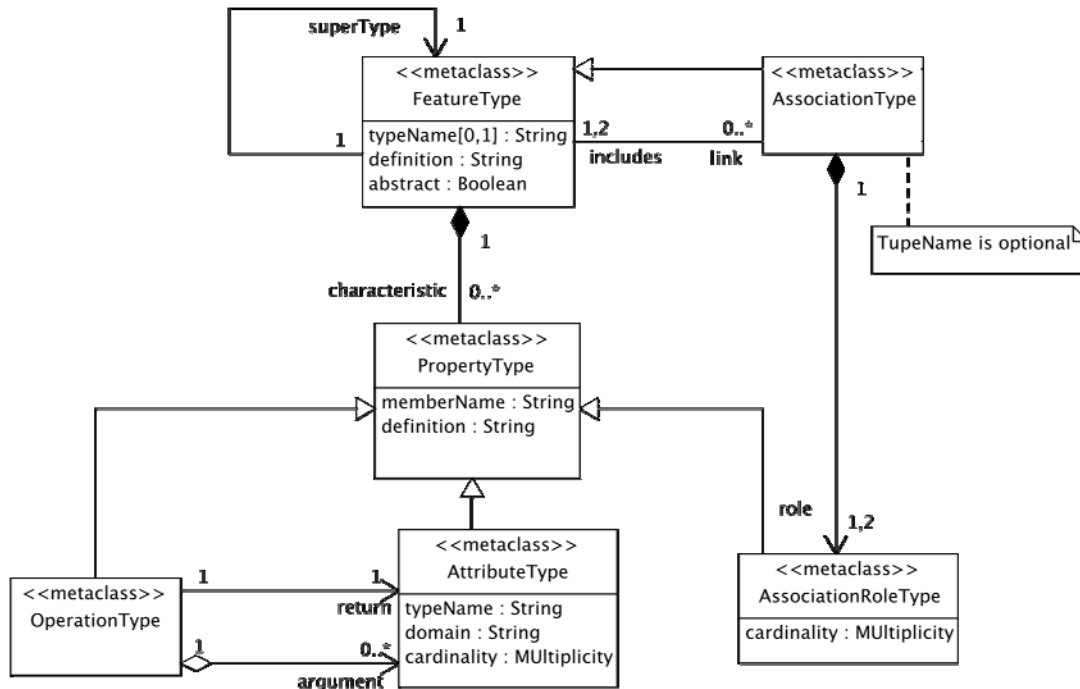


Figure 2. Simplified general feature model package

If a feature instance have more than one attribute value, cardinality (min and max occurrence) is specified.

OperationType is a type to define operation of a feature. For example, you can get a return value (e.g. reachable distance) by giving arguments (e.g. average speed and time) to the operation defined in the feature (e.g. car).

AssociationRoleType refers to a role of a feature for an associating feature. For example, a role of a building may be “school” for students and “department store” for consumers. Association role and “includes” (the role of *FeatureType*) have one to one relationship through association type.

SPATIAL SCHEMA

Spatial schema [8] includes data types for describing the spatial characteristics of features. It consists of two dimensional geometry and topology packages in SGIS.

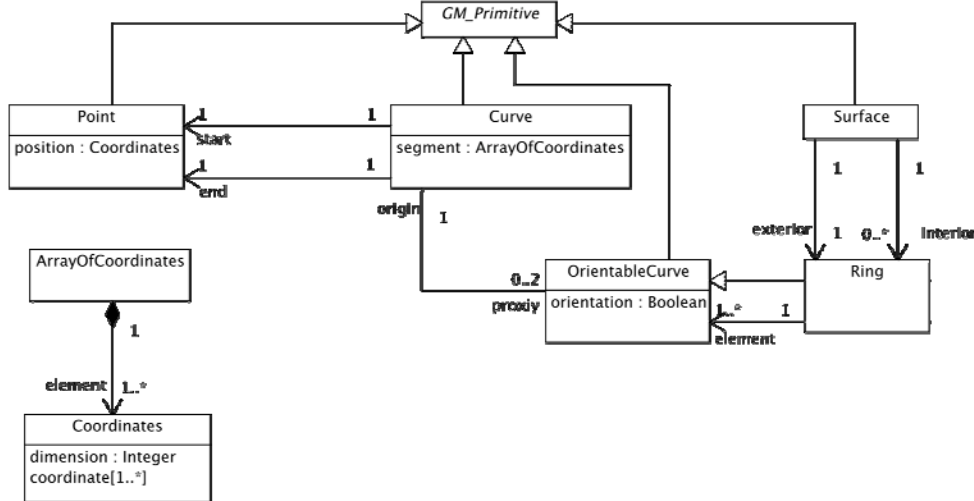


Figure 3. Geometry Package

Geometry package comprises geometric primitives and complexes. Geometric primitives are *Coordinates*, *ArrayOfCoordinates*, and *GM_Primitives*. *GM_Primitive* is an abstract root type of *Point*, *Curve*, *OrientableCurve*, *Ring*, and *Surface*. Figure 3 shows the geometry package. *Point* is a zero dimensional geometry represented by its position. *Curve* is an array of coordinates of which ends are associations to start and end points. *OrientableCurve* is a curve having orientation (forward: true, or backward: false). *Surface* is a closed area bounded by an exterior ring, and interior rings if there are holes in it. *Ring* is basically a closed orientable curve composed of a sequence of orientable curves. If it acts as an exterior boundary, the orientation is “forward”. However, if the same ring acts as interior boundary, the orientation is “backward”.

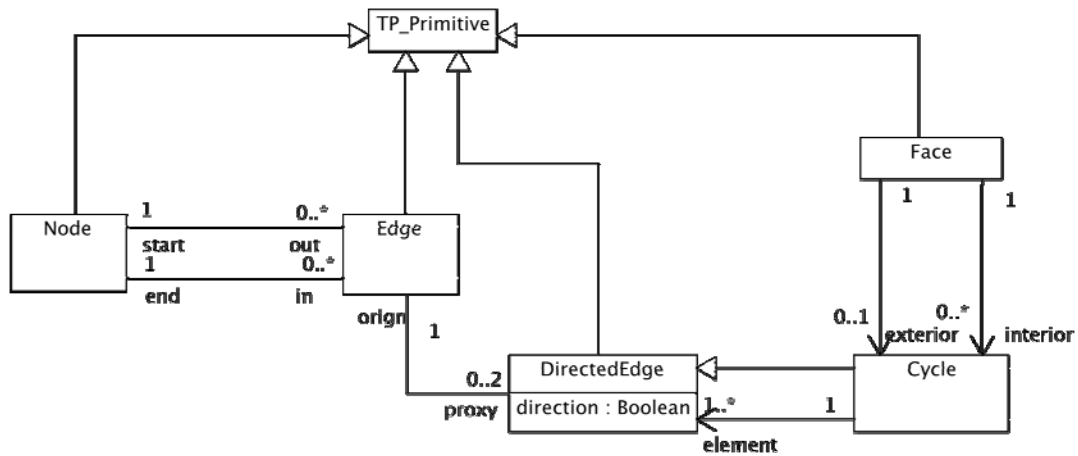


Figure 4. Topology package

Topology package is a representation of topological characteristics of geometries. It comprises *TP_Primitives* (*Node*, *Edge*, *Cycle*, and *Face*). *Node* represents a start and end of edges. *Edge* is associations of start and end nodes. *DirectedEdge* is an edge with direction, a component of *Cycle*. *Cycle* is a closed directed edge represented by a sequence of primitive directed edges. It has no boundary. *Face* is a two dimensional manifold bounded by exterior cycle, and interior cycles if there are holes in it.

TEMPORAL SCHEMA

Temporal schema [9] consists of a set of types for describing the temporal characteristics of features. It consists of geometry and topology primitives.

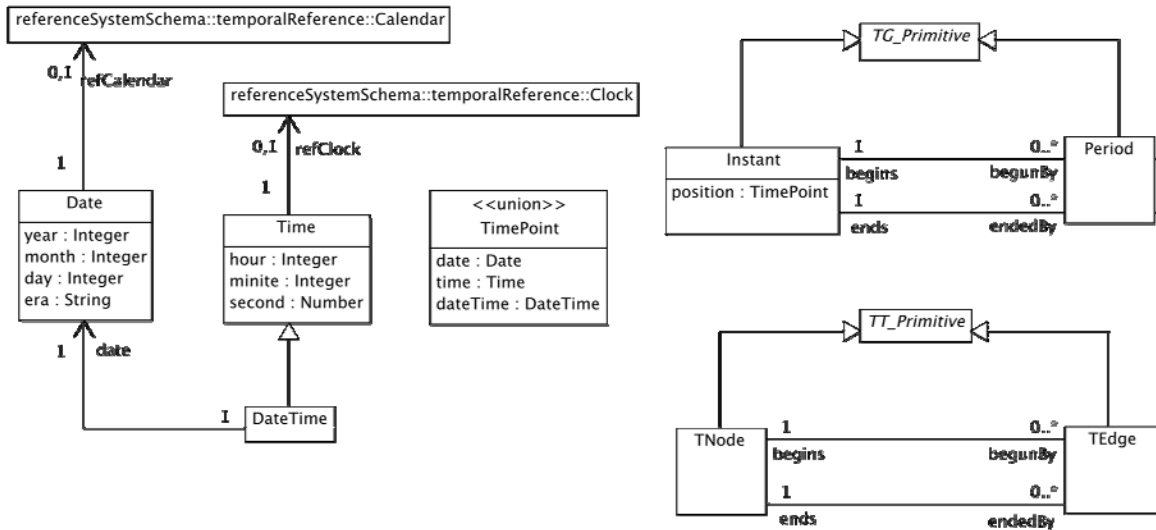


Figure 5. Temporal schema package

Temporal schema is shown in Figure 5. *Date* is a time point of which resolution is one day. It is defined by year, month, day and era (e.g. BC and AD in Gregorian calendar, and Meiji, Taisho, Showa, and Heisei in Modern Japanese Calendar). *Time* is a temporal distance from any origin point represented by hour, minute, and second. If the time is clock time, it associates with local clock system with offset from Universal Coordinated Time (UTC). *DateTime* is a time with date. It is used to identify the time point of which resolution is finer than one day. *TimePoint* is a position in time. User can choose *TimePoint* as date, time or *dateTime* as its stereotype is union.

TG_Primitive is an abstract type acting as a root type of temporal geometry such as *Instant* and *Period*. *Instant* is a temporal characteristic of a feature that exists shorter than temporal granularity. A time point abstracts its existence. *Period* refers to an interval from beginning to ending instant. *TT_Primitive* is an abstract type acting as a root type of temporal topological primitives such as *TNode* and *TEdge*. *TNode* and *TEdge* represents topological aspects of instant and period. *TEdge* has direction, and we cannot imagine cyclic sequence of temporal edges, because time has absolute directionality. The temporal network constructed by temporal nodes and edges shall be directed acyclic graph.

COVERAGE SCHEMA

Coverage is a feature that acts as a function to return values from its range for a position within its spatial, temporal, or spatiotemporal domain (e.g. a raster image, polygon overlay, or digital elevation matrix) [10]. However, the scope of coverage schema in SGIS is limited for implementation of spatial (point, line and surface) coverage only. *Coverage* is an abstract type that has geometry type, and attribute the coverage use. It also has an operation “evaluate”. It is used to get values corresponding to the given position. *DiscreteCoverage* returns the same values for every position within any single spatial object in its domain. It means that the evaluate operation returns the same value if the nearest spatial object from the given position is the same. *GeometryValuePair* is an element of a set that defines the relationships between geometry and value in coverage. For example, as a meteorological station has its position (geometry) and observation (value), a set of stations can be seen as discrete point coverage. You can get observation value of the nearest station to the position you give.

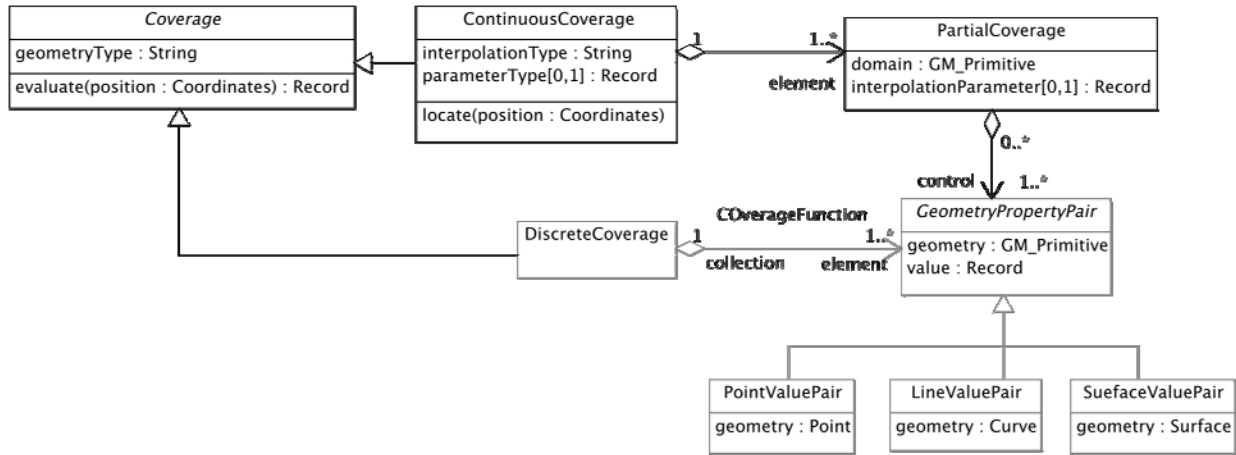


Figure 6. Coverage schema package

ContinuousCoverage returns different values for the same geometry at different positions by interpolation function. For example, you can get estimated value (e.g. by weighted mean) from a set of meteorological stations near the position you give. *PartialCoverage* is a set of *GeometryValuePair* (e.g. meteorological stations) near the given position. It is constructed by “locate” operation defined in *ContinuousCoverage*.

REFERENCE SYSTEM SCHEMA

Reference system schema consists of packages describing the spatial, location and temporal reference system for spatial coordinates, geographic identifier and time point.

Coordinate reference system (CRS) is a coordinate system related to an object by a datum [11]. There are many types of CRSs, for example, geodetic, vertical, image, and engineering. *GeodeticCRS* is defined by *CoordinateSystem* and *GeodeticDatum*. *CoordinateSystem* describes its type name (e.g. Cartesian, angular), dimension and association to the definition of each *Axis*. Parameters of *GeodeticDatum* (*PrimeMeridian* and *ReferenceEllipsoid*) describe the relationship of coordinate system to the Earth.

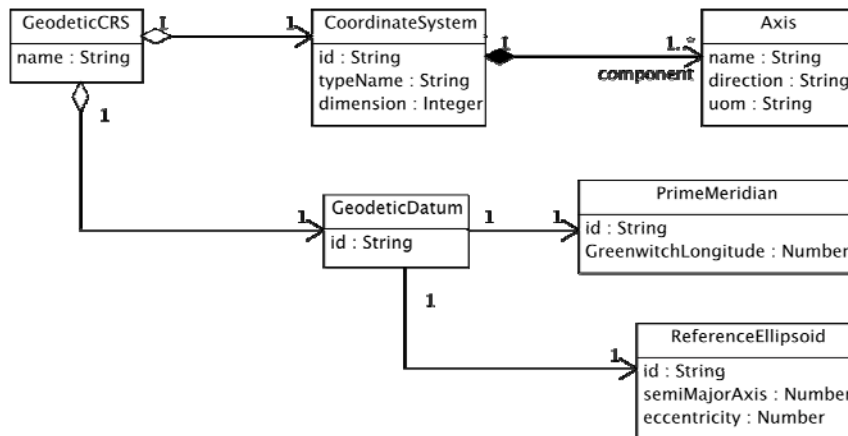


Figure 7. Geodetic coordinate reference system package

Location is a geographic identifier indirectly indicates the position of a feature [12]. *Gazetteer* is a directory of geographic identifier and it acts as the function to get a position from an identifier, and to get an identifier from a position using *LocationPositionPair*, a pair of identifier and position.

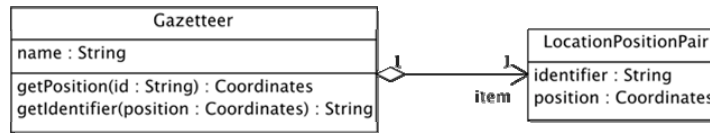


Figure 8. Location reference system package

Temporal reference system consists of *Calendar* and *Clock* [9]. In SGIS, we adopted *Calendar* as a date transferable to and from Julian date in order to transfer one type of date to other. *CalendarEra* is a division of a calendar composed of a sequence of periods of one type counted from a specified date. BC, AD for Gregorian calendar and Meiji, Taisho, Showa, and Heisei for Modern Japanese Calendar are examples of calendar eras. *Clock* refers to a local time system by its name and time difference from Universal Coordinated Time (UTC).

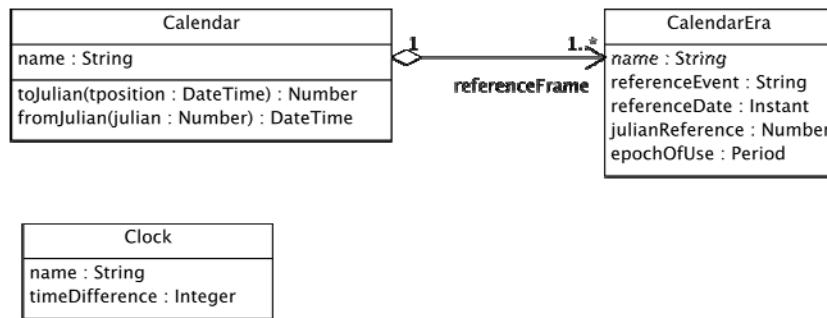


Figure 9. Temporal reference system package

PORTRAYAL SCHEMA

Portrayal means that presentation of information to humans [14]. In this paper, portrayal schema defines the specifications for styling the geographic data. It consists of symbol dictionary, label dictionary and portrayal dataset. Symbol dictionary (Figure 10) includes definitions of point, line and surface symbols. Label dictionary (Figure 11) defines how to represent labels such as geographic names and other annotations. Portrayal rule dictionary (Figure 12) defines which properties in a feature shall be visualized.

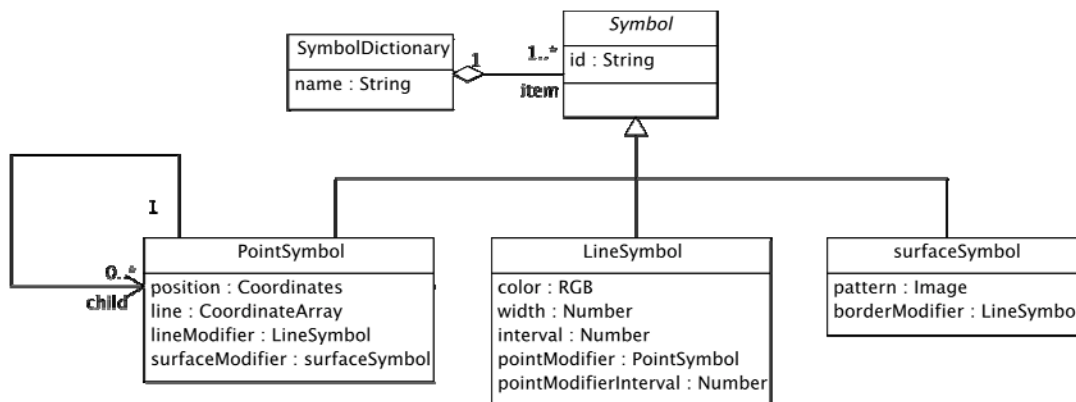


Figure 10. Symbol package

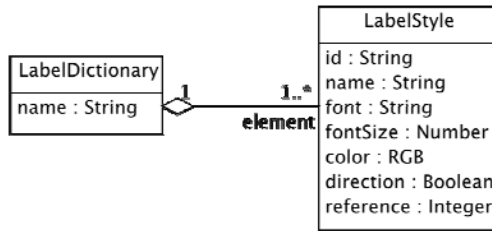


Figure 11. Label package

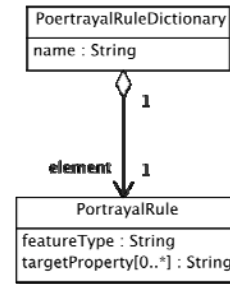


Figure 12. Portrayal rule package

Portrayal dataset (Figure 13) is a graphic dataset transferred from geographic data for the visualization on graphic devices such as graphic display and paper. The procedure of transformation is as follows.

1. Select a feature instance until the end of the feature instance file.
2. Lookup the portrayal rule that define which attributes are portrayed.
3. Search a symbol definition that is equivalent to a geometry attribute, if necessary.
4. Make figure as a combination of shape (geometric attribute) and symbol definition
5. Search a label definition that is equivalent to location or thematic attribute, if necessary.
6. Calculate a reference position on the graphic device using geometric attribute.
7. Make label as a combination of label string (thematic attribute), reference position, angle, etc.
8. Add figure and label in the portrayal data set. Then go back to 1.

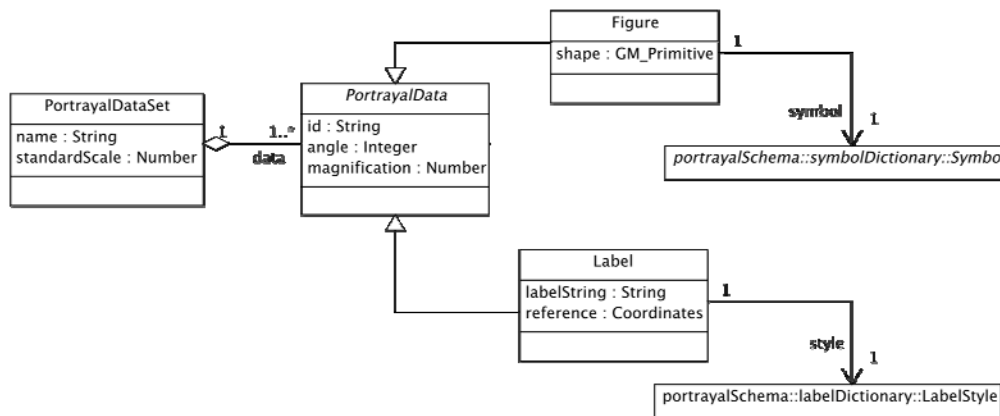


Figure 13. Portrayal dataset package

INSTANCE MODEL

Feature instance set is implemented in accordance with the instance model (Figure 14.) using meta-language such as XML. For example, if the TV Tower called “Tokyo Tower” is implemented in compliance with Instance model, the implementation in XML can be described as a document in Figure 15. A class *Landmark* is a feature type defined in compliance with GFM. “Tokyo Tower” located at the position pt_001 is an instance of *Landmark* class.

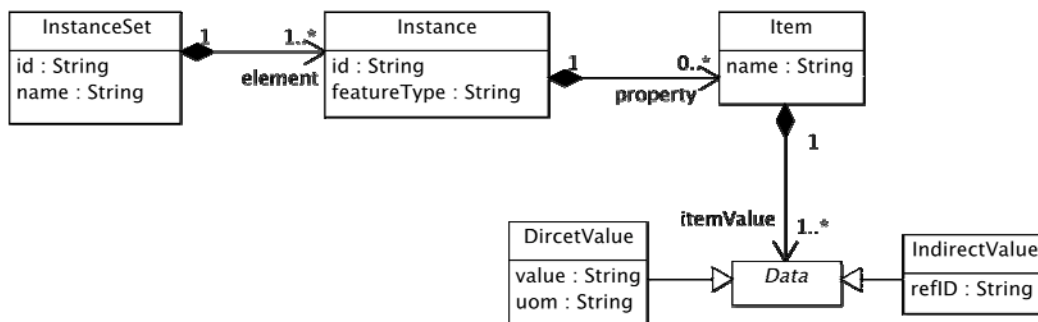


Figure 14. Instance model package

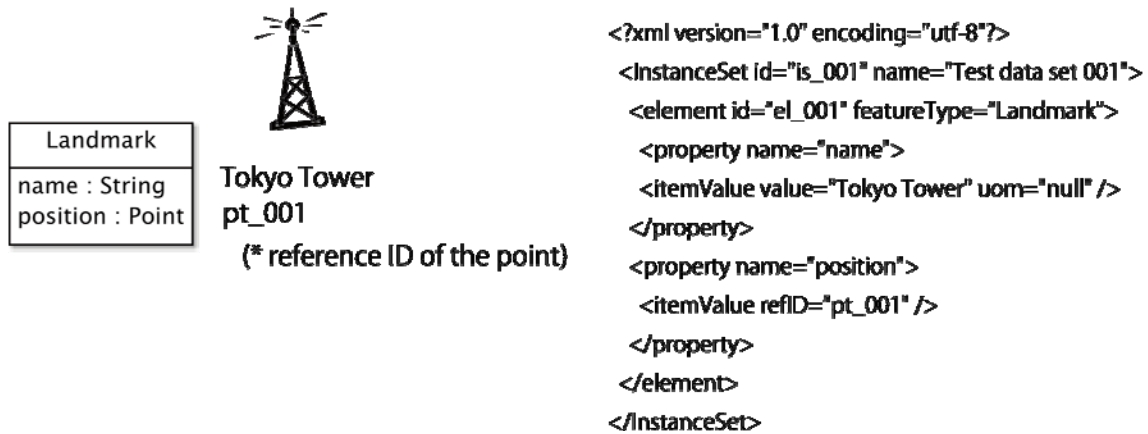


Figure 15. An example of feature instance implementation

EXPERIMENTAL LECTURES

The author lectured an Introduction of GIT at University of Tokyo for the second and third year undergraduate students in the department of urban engineering during the second semester since 2008 [5]. The number of students is about one hundred each year.

The purpose of the course is that students obtain the basic literacy of GIT. We prepared a curriculum for 14 lectures (seen under), texts, and four exercises.

1. Introduction (Geographic Information Technology and our society)
2. Spatial thinking and conceptual modeling
3. Model Formulation (Introduction to UML)
4. General Feature Model
5. Application schema
6. Spatial schema (1)
7. Spatial schema (2)
8. Temporal schema
9. Reference system
10. Coverage
11. Spatial Analysis
12. Portrayal
13. Geographic data acquisition
14. Geographic information processing

The students study 1) how to recognize the domain of discourse that a student confronts and select geographic features in the domain, 2) how to formalize the description of features and their relationships. At first, we focus on the model formalization of geographic phenomena and introduced the concept of object orientation, UML notation, General Feature Model (GFM), and application schema design in compliance with GFM. At second, we introduce spatial, temporal and location attributes, reference systems, and spatial analysis as feature operation. And at third, we introduce portrayal, data acquisition and geographic information processing.

CONCLUSION

We expressed that the education of GIT is important to realize and improve SDI, and the standard schemas for geographic information is useful for the education. Then, we introduced Geographic Information Schemas, which were designed for undergraduate students by simplification of schemas specified in ISO 19100 series. We believe that it is possible to implement and exchange fundamental geographic data in compliance with these schemas. And we have a plan to develop fundamental Geographic Information System by applying these schemas. In addition, we are presenting the texts using in the course on the Web for the people who has interested in the education of GIT.

REFERENCES

- [1] White house, *Executive Order 12906: Coordinating Geographic Data Access*, 1992.
- [2] Murakami, H., *New Legislation on NSDI in Japan: "Basic Act on the Advancement of Utilizing Geospatial Information*, Bulletin of the GSI Vol. 55, <http://www.gsi.go.jp/ENGLISH/Bulletin55.html>, Geographical Survey Institute ,2008.
- [3] European Union, *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007*, Official Journal of the European Union, <http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML> , 2007.
- [4] Ota, M.,Arikawa, M., Makino, H., Kubota, K., Okabe, A.*The Research Initiative on a Curriculum for Geographic Information Technology in Japan*, Proceedings of the 6th Asian Symposium on Geographic Information Systems from a Computer Science & Engineering Viewpoint [ASGIS2008], Niigata, Japan, pp.5-12 , 2008
- [5] Ota, M., Arikawa, M., *A Curriculum for Geospatial Information Technology and Its Experimental Practice*, Proceedings of 6th International Symposium on Digital Earth, Beijing, China, Sept., 9-12, 2009
- [6] ISO, ISO/TS 19103 Geographic Information – Conceptual Schema Language, 2005
- [7] ISO, ISO 19109 Geographic Information - Rules for Application Schema ,2005.
- [8] ISO, ISO 19107 Geographic Information - Spatial Schema ,2003.
- [9] ISO, ISO 19108 Geographic Information - Temporal Schema ,2002.
- [10] ISO, ISO 19123 Geographic Information – Schema for coverage geometry and functions,2005.
- [11] ISO, ISO 19111 Geographic Information - Spatial Referencing by Coordinates,2007.
- [12] ISO, ISO 19112 Geographic Information - Spatial Referencing by Geographic Identifier, 2003
- [13] ISO, ISO 19118 Geographic Information – Encoding, 2005
- [14] ISO, ISO 19117 Geographic Information – Portrayal, 2005

BIOGRAPHY



Mr. Morishige Ota is the Fellow of Kokusai Kogyo Co., Ltd. and an adviser of Infrastructure Innovation Institute, Inc. in Japan. He was a visiting professor of the Center for Spatial Information Sciences at the University of Tokyo from 2005 up to 2007, and he is a part-time lecturer for the Department of Technology at University of Tokyo since 2008. He is a member of Japanese delegation to ISO/TC211 since 1995. He was a project leader for Japan Industrial Standard (JIS) X 7108 - Temporal Schema and X 7115 - Metadata, which are identical to ISO standards. He was awarded as the International Standardization Contributor from the Ministry of Economics, Trade and Industry in 2007. Now, he is a project leader for JIS X 7136 – GML, and a member of OGC Global Advisory Council.