ONTOLOGY OF TEST

Larisa Soldatova
Post doctoral researcher
Riichiro Mizoguchi
Professor
ISIR, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047
Japan
{larisa, miz}@ei.sanken.osaka-u.ac.jp

ABSTRACT

In the present paper design of test generation systems (TGS) based on test ontology and student's knowledge model is discussed. Main functions and regimes of TGS are described. Student's knowledge base is divided into two parts: domain independent- and domain-dependant knowledge. Suggested test ontology allows analyzing test characteristics, structure of test, and process of test composition. Some recommendations about selection of item form and scoring schema are given on the base on test ontology. Problem of generation of test questions is discussed and considered on the example "integration of function".

KEY WORDS

Ontology, knowledge model, test/problem generation

1. INTRODUCTION

For successful application of any IES (intelligent educational system) it is necessary to get information about a learner's knowledge. We can do it only by indirect measurement, and test is the most popular method. In order to design a system/subsystem of IES for measurement of learner's knowledge, to provide a feedback for the educational process, the designer has to know what functions are necessary, which architecture is appropriate, what test composition is, and how to represent the learner's knowledge by a model to measure it correctly.

The purpose of our research is analysis of test design process on the basis of ontology methodology, construction of ontology of test, and design of an information system for test composition. In the knowledge based community, ontology is defined as "a system of primitive vocabulary/concepts used for building artificial systems" [1]. "Ontology provides us with effective methodology and vocabulary for both analyzing and synthesizing knowledge-based systems" [2]. Test ontology can allow:

- to formalize a description of test and test tasks;
- to specify main concepts of test construction process and relationships between them;
- to standardize components of test design systems;
- to prove reusable components of test design system;
- to share vocabulary, what is especially important in interdisciplinary field as test design;
- to carry out the analysis of methods for test composition.

Unfortunately at the present, the process of constructing a test is still a kind of art. There are a number of rules and recommendations how to form test problems, but there are few technologies and information systems for the test construction. The suggested test ontology can help to understand how to create a good test.

2. TEST GENERATION SYSTEM

Users of information system for test generation TGS are: a professor or an instructor, who would like to design a test and/or analyze the test results, and a learner or an examinee whose knowledge the instructor would like to check. TGS can work in several regimes: *test composition* (either automatically or interactively with the instructor) and *test application* to check examinee's knowledge and to analyze the test result characteristics.

Functions of TGS:
Design: to determine the approach for the test composition, the test structure, and the method of scaling.
Input: the purpose of testing, the target group of examinees, requirements for the test characteristics.
Output: a recommendable test structure and a scoring scheme.
Adapt*:* to fit the test structure (in an interactive mode).

Input: instructor's preferences about test characteristics.
Output: a test structure and a scoring scheme.
Generate: to form test items.
Input: a type of item form, a level of difficulty
Output: test items
Select: to extract test items from data base of ready test items.
Input: the type of item form, the level of difficulty
Output: test items
Observe: to get information about the examinee's knowledge.
Input: test items
Output: answers on given questions
Evaluate: to analyze the examinee's answers to obtain quantitative characteristics.
Input: answers on given questions; quantitative or qualitative evaluation made by the instructor.
Output: quantitative and qualitative characteristics
Improve: to enhance quality of the test on the base of analysis of test result characteristics.
Input: test
Output: upgraded test.
Administrate: to provide security and access to the system recourses.
Input: commands.
Output: system information.

At the present time the process of test design and knowledge assessment can be automate only partially. The system can not evaluate essay or generate complicated test items. TGS needs to get some information from an instructor, such as his/her preference about test characteristics, test item reduction if necessary, and evaluation of open-questions items (see section 4).

3. STUDENT KNOWLEDGE BASE

TGS is based on a student knowledge model. We consider two parts of the knowledge base: domain independent- and domain-dependant knowledge [1, 3, 4]. The domain-independent part consists of cognitive ability as reasoning, capability, recognition; general knowledge as methodological, structural, classification knowledge; and mental skills.

The domain-dependant knowledge part consist of nodes, such as facts, rules, theorems, terms, and principles (see Fig.1, made in OE [5], where p/o is "part of" link with a slot, a/o is "attribute of" link, upper part of a slot corresponds to role of the concept, and right side of a slot represents a class), and different types of links. Example of the knowledge model is discussed in section 6.

4. TEST ONTOLOGY

TGS is mostly based on the test ontology. At the top level of test ontology is "world of tests" or set of tests which consists of particular tests (see Fig.2). On the other hand, test is a tool for measurement of students' knowledge. Any test has specification: purpose, type, and target group; structure as number of blocks of different types, each block has specific item form, instruction, scaling rule, examples, and set of items; test and result characteristics. The most important test characteristics are the following [6]:

- reliability, "degree to which individuals' deviation scores remain relatively consistent over repeated administration of the same test or alternate test forms";
- validity, as content validation about how test items represent problem domain, criterion-related validation of used criteria for making inference from the test result, and validation of test construction;
- precision of decision made on the basis of test results.

However, we can determine many characteristics, such as item reliability, validity, discrimination, correlation indexes, mean, and variance only after analysis of the test result.

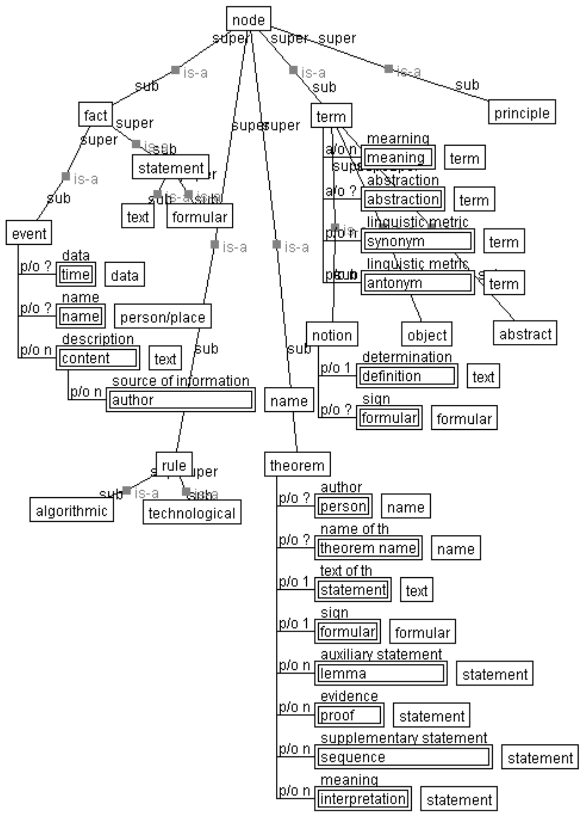

Fig. 1: Ontology of domain knowledge (fragment).

Purpose of test might be ranking of students, qualification or mastering, progress check, and diagnose of difficulties. Types of test are placement, admission, diagnostic, mastering, and examination (pre-exam, quizzes, mid-term exam, final exam, post exam). Test could be various kinds: computer based, paper, oral, listening, and demonstration.

To get good test characteristics it is necessary to select a correct approach for scaling. In classic test theory three broad approaches are considered [7]:

- subject-centered, which focuses on measurement of individual, his place in continuum of examinees group;
- response-centered, which focuses on measurement of individual correspondence to some criteria;
- stimulus-centered, which focuses primarily in locating the position of the items on the psychological continuum.

For each approach there are recommendable scoring schemes, such as nominal, ordinal, interval, and ratio [6].

The suggested ontology of test can help to an instructor who would like to create a test and designers of IES not to miss important components of test composition process, and to determine correctly the purpose and way of testing, to provide appropriate quality of the test. Many important solutions e.g. entering to company or university are made on the base of test results, but actually we can not confide to all tests. Quality of a test depends on many characteristics, and it is difficult to optimize all of them. We hope to complete the test ontology by knowledge about methods of optimization of different kinds of test.

4.1 Level of understanding

For composition of test we would like to use popular in pedagogy 6-leyers model of understanding, suggested by Bloom [8]. Compare with standard "flat" student's knowledge model 6-layers model gives much more information about student's stage and area of his/her difficulties. Each test item corresponds to definite level of understanding:

1st level, **knowledge** in narrow sense, is knowing facts and definitions of notions;

2nd level, **comprehension**, is understanding of meaning of notions, objects, abstracts, and knowing simple rules;

3rd level, **application**, corresponds to ability to apply known rules;

4th level, **analysis**, checks understanding of relationships between elements, ability to select and compound different rules;

5th level, **synthesis**, corresponds to ability to generalize knowledge;

and 6th level, **evaluation**, uses metaknowledge. The first levels correspond to the domain knowledge, and upper levels reflect mostly the domain independent knowledge. Thus, the system can generate test not only to get test score, but also to determine examinee's level of understanding.

1.

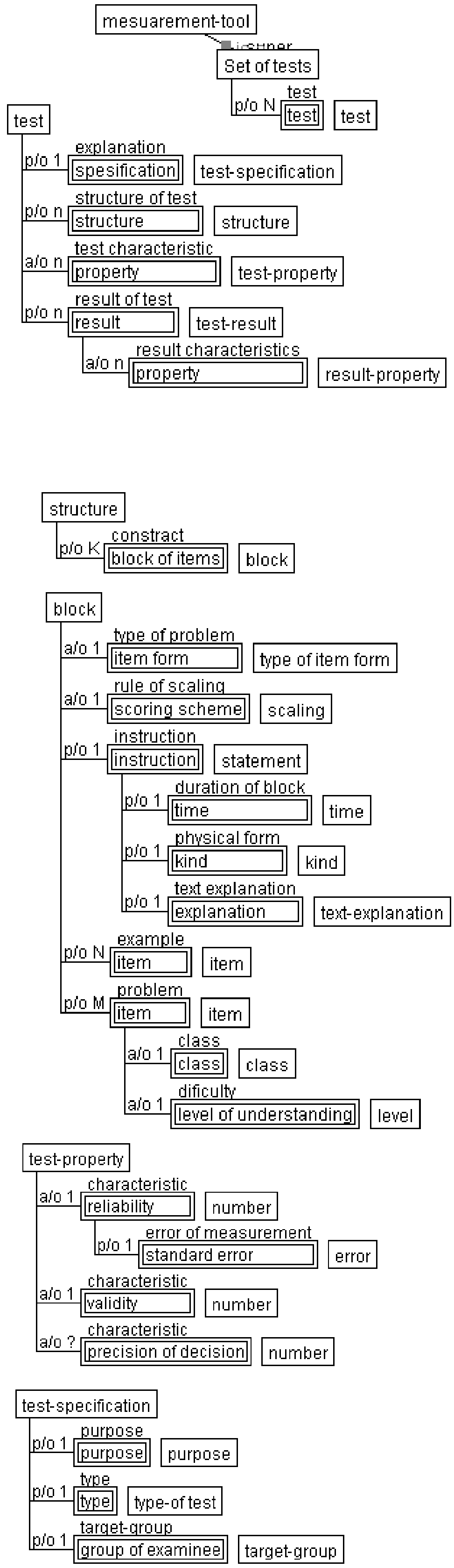

Fig. 2: Test ontology (fragment).

4.2 Item forms and structure of test

We can classify types of item forms as the following [9]:

1. **Open question**: fill-in-the-blank, give a short answer, and write essay.
2. The most popular form is **multiple choice** problems. We divide them according to amount of given answers (2, 3,…, etc.) and logic of answer (select one correct answer, select the most correct answer, or note all correct answers). The most famous examples of such item form are true-false (T/F) questions (is the statement true or false) and SAN – questions (select: the statement is correct sometimes, always, or never).
3. **Matching** problems with different amount of columns.

It is possible to give some recommendations about item forms for checking different layers of understanding (see table 1). For instance, fill-in-the-blank questions are good to check knowing facts, understanding of notion definition and meaning. High levels of understanding, synthesis and evaluation, we can check by essay or short answer with using nondichotomous scoring scheme.

Table 1: Recommendation for item form to check different elements of knowledge.

Level of understanding	Open question			Multiple choice			Matching
	Fill-in-the-blank	*Short answer*	*Essay*	*T/F*	*SAN*	*Other*	
I	***	**	--	**	*	*	*
II	***	***	--	***	**	**	***
III	*	***	--	**	**	***	***
IV	--	**	*	*	*	**	*
V	--	**	***	--	--	--	--
VI	--	*	***	--	--	--	--

Where sign – means not recommendable, * recommendable, ** very recommendable, and *** greatly recommendable.

From the classic test theory we can extract several practical rules for test construction. For example, property of placement test would be better if all problems are middle difficult. Cut score, the criteria for mastering, should not be 40-60%, otherwise quality of the solution, made on the base of test score, might be not good [6]. Such recommendations we summarize in table 2. Amount of test items have to be 100 and more to calculate well statistical test characteristics. So, if the purpose of a test is ranking and type of the test is placement, the system can generate the following structure of the test to get the best test characteristics (see table 3):

Table 3: Generated test structure for placement test (example).

Item form	**I**	**II**	**III**	**IV**	**V**	**VI**	**Total**
Fill-in-the-blank	3	7	4				14
Short answer	2	7	7	4	4		24
T/F:	2	7	5	2			16
SAN	1	4	5	2			12
Multiple choice	1	4	7	4			16
Matching	1	6	8	2			17
Essay					1		1
Total:	10	35	36	14	5		100

However, often professor/instructor prefers to minimize test time or amount of blocks, even if statistic characteristics of test could be worse. TGS can generate test according to instructor's preference with keeping as well test properties as possible (see table 4). Like that, T/F questions do not require much examinee's time, and can check the majority of knowledge elements. If an instructor ordered to check test results automatically, TGS will not include open-question items to the test.

Table 4: Optimization of test structure.

Instructor's criteria	**Reaction of TGS**
minimize time	Include more T/F items and less short-answer items
minimize amount of blocks	Use 2-4 the most recommendable item forms
preference of some item forms	Ex: cancel essay item
automate test generation	Exclude essay and matching items
automate checking of test results	Exclude open-question items

5. TEST ITEM GENERATION

In [10] we discussed how to generate calculation problems with required level of difficulty. Let's consider generation of simple test questions. Student knowledge

Table 2: Dependence of item forms on purpose of testing

Purpose of testing	**Type of test**	**Approach**	**Score scheme**	**Focus on level of understanding**	**Comments**	**Amount of items**
Ranking	Placement	Subject-centered	Interval	II – III	Middle difficulty problems	100
Qualification	Mastering	Criterion-referenced	Ratio Level of understanding	I - V	All problems	100
	Pre-exam			I - VI		25
	Quizze					10
	Mid-term					100
	Final exam					100
	Post-exam					25
Diagnose difficulties	Diagnostic	Subject-centered	Ordered Interval	I – II	The majority is easy problems	100
Progress check	Placement	Subject-centered	Interval	II – IV	Middle difficulty problems	100

base about problem domain consists of facts {F} (events {E} and statements {S}), rules {R}, theorems {Th}, terms {T} (notions {N}, objects {O}, and abstracts {A}), and principles {P} (see section 3). According to the structure of these elements TGS can generate some test items.

For instance, to check the 1st level of understanding the system has to form questions about notion definitions N_i(definition) and facts {F}, where N_i is particular notion. TGS can generate the following questions:

Short answer: Give definition of N_i.
T/F: N_i is N_i (definition).
Multiple choice: N_i(definition) is

a) N_j
b) N_i
c) N_l
d) N_k, where is $i \neq j, k, l$.

Fill-in-the-blank: E_i(description) was _______.
Where correct answer is E_i(data).
_______ E_i(description).
Where correct answer is E_i(person).
E_i(description)_______.
Where correct answer is E_i(place).
Short answer: Who/When/Where E_i(description)?
Where correct answer is E_i(person)/ E_i(data)/ E_i(place).
T/F: E_i(description) E_i(name).
E_k(description) E_i(name),
and so on, where is $i \neq k$.

For upper levels of understanding TGS uses data base of ready test items.

6. EXAMPLE OF PROBLEM DOMAIN: INTEGRATION

Knowledge model about function integration, as example of problem domain, was constructed [11] (see the fragment at Fig. 3). To check knowledge about function integration TGS can form the items like following:

1st level *Short answer:*
Give definition of definite integral.
Give definition of Riemann sum.
T/F: Calculation of area of region by integral was discovered by Leibniz.
Multiple choice:
Fundamental theorem of calculus was proved by

a) Newton
b) Fahrenheit
c) Leibniz
d) Lagrange

2nd level *Fill-in-the-blank:*
Meaning of definite integral is _______.
Short answer:
Explain the difference between differentiation and antiderivation.
T/F:
If the function is continuous on the closed interval, then the function is integrable on this interval.

3rd level *Short answer:*
Illustrate application of theorem about mean value.

4th level *Short answer:*
Analyze the difference between Riemann sum and area measure

5th level *Essay:*
Proof the theorem about max and min value of integral.

6th level *Essay:*
Proof the mean value theorem.

It is necessary to stress, that selection of test items by the system strongly depends not only on the purpose of testing, as we discussed it in the section 4.2, but also on a target group. For instance, if the target group is students of engineering specialties, the system has to generate test questions mostly for the 1st, 2nd, and 3rd level of understanding, because the main purpose of studying of function integration for them is application, not deep understanding of theory.

7. CONCLUSION

Suggested approach for design of information systems for test construction, based on the test ontology, allows systemizing, analyzing, and accumulating knowledge about test composition process. The test ontology describes structure of test, test tasks, test and result properties, types of test and items form, scoring schemas, and test composition approaches. For better diagnostics of student's knowledge Bloom 6-layers model of understanding is used. Also authors gave some practical recommendations about selection of item forms and scoring schema. Problem of test question generation is discussed and considered on the example.

Up to the moment the design of TGS is at the stage of knowledge acquisition and representation. TGS has several restrictions: such a system can generate only simple questions and calculation problems. To include to the test more complicated items the system has to use data base of ready test items. The system is not able to check *open-question* items. Meanwhile, TGS contains a lot of information about technology of test construction and can help to instructor to compose test and check the test results.

REFERENCES

[1] R. Mizoguchi & K. Sinitsa, Task Ontology Design for Intelligent Educational/Training Systems: Workshop on “Architectures and Methods for Designing Cost-Effective and Reusable ITSs", *Proc. ITS'96,* Montreal, 1996, 1-21.

[2] V. K. Marcke, Learner adaptivity in generic instructional strategies: *Proc. AIED*, 1995, 323-333.

[3] J.F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations* (Brooks/Cole, 2000).

[4] W. Chen, Y. Hayashi, L. Jin, M. Ikeda & R. Mizoguchi, An Ontology-based Intelligent Authoring Tool: *Proc. ICCE'98*, Beijing, China, 1998, 41-49.

[5] K. Kozaki, Y. Kitamura, M. Ikeda & R. Mizoguchi, Development of an Environment for Building Ontologies which is based on a Fundamental Consideration of "Relationship" and "Role": A workshop, *Proc. PKAW,* Sydney, Australia, 2000, 205-221.

[6] L. Crocker, J. Algina, *Introduction to Classical and Modern Test Theory* (Wadsworth Group, Thomson Learning, 1986).

[7] W.S. Torgerson, *Theory and methods of scaling* (NY: John Wiley, 1958).

[8] B.S. Bloom, et. al. *Taxonomy of Educational Objectives: The Classification of Educational Goals* (Handbook 1: Cognitive Domain. N-Y, David McKay Co., 1956).

[9] V.S. Avanesov, *Modern Methods of Learning and Control of Knowledge* (DVGTU, Vladivostok, 1999) (in Russian).

[10] L. Soldatova, Comparison of the Problems of Complicity: *Proc. ICCE’99*, IOS Press, V. 2, 169-171.

[11] L. Leithold, *The Calculus with analytic geometry* (Harper & Row, Publishers, 1976).

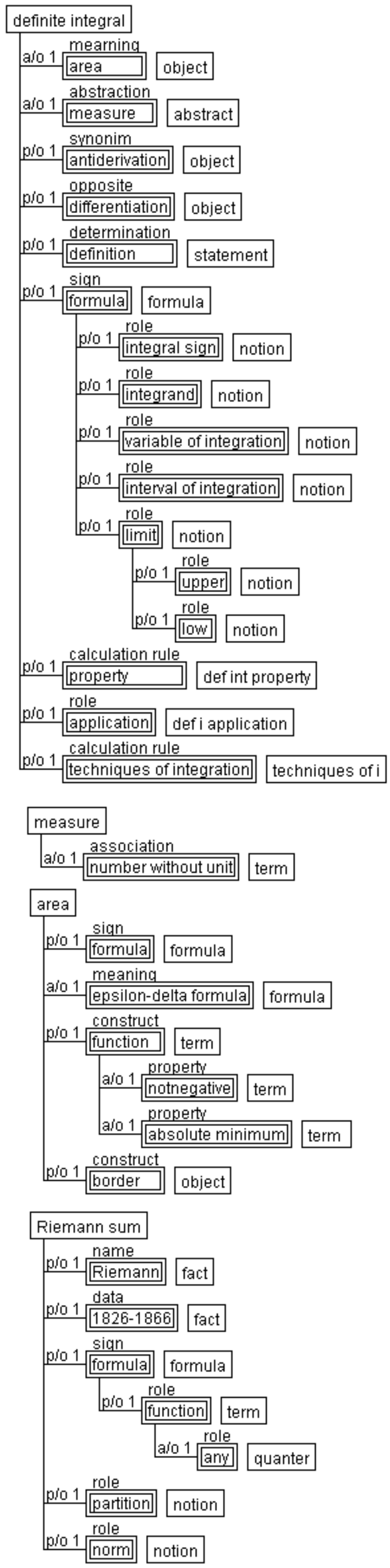

Fig. 3: Problem domain: integration (fragment).