1. Goal

Many types of adaptive learning environments (ALE) rely on sophisticated metadata models that help describe the parameters of learning objects (LO) presented by these environments. These parameters vary from merely descriptive terms facilitating cataloguing and discoverability of LO in content collections, to pedagogical properties of LO and semantic relations between them, which are used by adaptation components of the environment and define the logic of its adaptive intervention. This is why content and knowledge creation for ALE is a very complex procedure that requires considerable time investment and is especially demanding from the point of participating authors' expertise. It is also characterised by some other factors (see fig.1). This problem has long hindered the dissemination and adoption of adaptive and intelligent technologies in e-Learning.

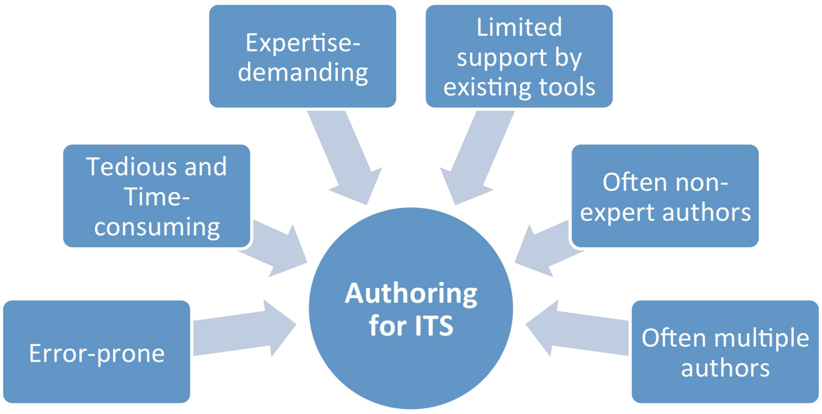

Fig. 1. Factors contributing to the problem of authoring semantic learning content for ALE/ITS

The ultimate goal of the ISASLC project is to advance the state of the art in the field of authoring technologies for different types of ALE, and especially, for intelligent tutoring systems (ITS). Content and knowledge creation for ITS is a very complex procedure that requires considerable time investment and is especially demanding from the point of participating authors' expertise. This problem has long hindered the dissemination and adoption of adaptive and intelligent technologies in e-Learning. Its solution has not been possible before. But now, with recent advancements in Artificial Intelligence and Human-Computer Interaction, and corresponding development of new Web technologies we might have just enough tools and resources to take the next step towards solving this problem.

2. Approach

The ISASLC project attempts to take this step by relying on the methods from such fields as Social Computing, Semantic Web, and Data Mining. From the practical perspective, ISASLC seek to widen the population of potential ITS authors by providing aid to inexperienced authors when it comes to error-prone and expertise-demanding authoring tasks, such as new content creation, metadata authoring, interactivity authoring, error detection and quality control.

The R&D activity within ISASLC is divided into five work-packages:

- Interactivity authoring support: the goal is to improve the current practice of authoring interactive exercises in ActiveMath through usage of collaborative authoring support and partial generation of solution graphs and metadata.
- Collaborative authoring support: the goal is to widely implement the effective authoring patterns developed on the Social Web and to find the benefits and limitation of collaborative authoring technology for semantic learning content creation.
- Metadata authoring support: the goal is develop a technology that will make authoring semantics for learning content available for a broader category of users;
- Gap detection: the goal is to develop a technology for effective detection of knowledge gaps in learning content;
- Open-corpus content discovery: the goal is to develop a technology for harvesting learning content from online sources and its provision with appropriate metadata.

3. Results

In all of the work-packages the project has achieved most of the expected outcomes by designing, implementing and evaluating technologies and tools for supporting semantic content authoring.

The main technical outcome of the project is the new authoring platform developed for the ActiveMath/Math-Bridge intelligent learning environment. This platform has been implemented as a Web-application and allows authors to create individual learning objects and assemble them into courses. Several dedicated tools have been developed to extend this platform with unique functionality. EXAMAT component supports authoring of various interactive exercises. Semantic Gap detection tool provides authors with an option to verify the correctness of metadata they have supplied (or did not) for the newly created learning objects. Exercise difficulty calibration tool allows post-hoc re-annotation of exercise difficulty. The entire authoring platform supports collaborative authoring of learning objects.

3.1 Metadata Schema

ActiveMath uses a fine-grained knowledge representation to encode LOs and a rich metadata schema to specify their properties, primarily, based on OMDoc and LOM representation standards. Each LO corresponds to about one paragraph of text and has a type specifying its primary function. ActiveMath's learning objects are divided into two main categories: concept items and satellite items. Concept items include general types of (mathematical) knowledge, such as definitions and theorems. Satellite items provide additional information about concepts in the form of examples, exercises and texts.

Metadata requirements are necessary to serve as a guideline to check if the metadata in the LOs is being used properly. In the proposed approach, these requirements are formalized as an OWL2 ontology. OWL2 allows to specify hierarchies of classes to represent the network of LOs and the metadata elements, and instances or individuals of such classes to represent the LOs and the values of those metadata elements. To describe how LOs relate to each other and to assign metadata to the LOs, OWL2 object properties are used. Object properties are relations, which link elements in the ontology, and can have restrictions that define the semantics and usage of those relations. A complete metadata ontology was created to explicitly formalize the metadata usage in ActiveMath. This ontology contains all descriptive, pedagogic and semantic metadata that is used in the LOs. Fig. 2 shows extracts from this ontology.

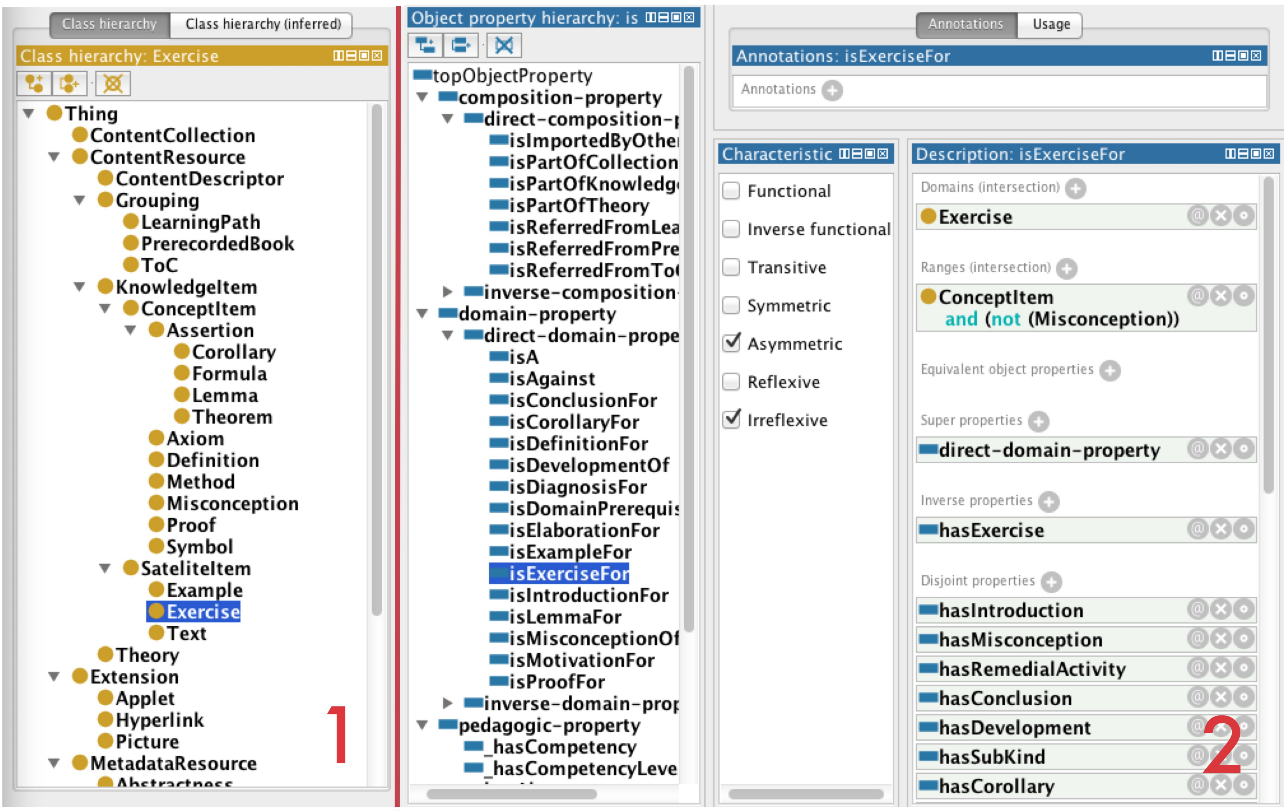

Fig. 2. ActiveMath metadata schema (1. hierarchy of LO types and metadata elements; 2. set of relations between LO and metadata elements, including the restrictions on applicability of these relations)

3.2 Semantic Metadata Gap Detection Tool

OWL2 axioms have different types, each representing a particular restriction or property in ontologies. In our case, these axioms allow specifying constraints on applicability of metadata elements, and rules enabling inference of new knowledge about the LOs. For example, the *DisjointObjectProperties* axiom indicates metadata relations that cannot be applied to the same pair of LOs; the *InverseObjectProperties* axiom helps to define pairs of inverse metadata relations, such that if LO1 is linked to LO2 with one of them, LO2 is linked to LO1 by another. Our algorithm analyses the axioms according to their types. Overall, eight types of gaps have been identified:

1. External LO does not exist (ELO): a LO metadata element references another LO from an external collection, but the external LO does not exist.
2. External collection does not exist (ECO): a LO metadata element references another LO from an external collection, but the external collection does not exist.
3. Undefined LO or metadata value (UND): a LO metadata element references another LO that does not exist, or a metadata element has an undefined value (e.g. *hasDifficulty* metadata element can have one of the five predefined values: *VeryEasy*, *Easy*, *Medium*, *Difficult*, *VeryDifficult*; using any other value would introduce an UND gap).
4. Domain of a metadata element is wrong (DOM): a LO is annotated with a metadata element, that it cannot be annotated with according to the domain restrictions in this element (e.g. *isIntroductionFor* metadata relation can be used to annotate *Satellite* LOs, such as *Exercise* or *Example*, but not *Concept* LOs, such as *Theorem* or *Axiom*).
5. Range of a metadata element is wrong (RNG): a LO a metadata element has a value of a type that is incorrect according to the range restriction of this element (e.g. *isIntroductionFor* metadata relation can refer only to *Concept* LOs, but not *Satellite* ones).
6. Metadata element is functional (FUN): a LO is annotated with the same metadata element more that once, but according to the metadata schema, it can be applied only one time (e.g. the *isProofFor* relation is functional, which means a certain *Proof* can be a proof for only one *Corollary*, *Lemma* or *Theorem*).
7. LO or metadata value type definition (TYP): a LO or a metadata value is defined more than once. This is a gap, as each LO or metadata value is unique and cannot have multiple definitions.
8. Metadata elements are disjoint (DJN): two LO are connected by metadata elements, which are disjoint (e.g. in the ActiveMath metadata schema, object properties *isIntroductionFor* and *isConclusionFor* are disjoint, which means it is impossible for a LO to be at the same time and introduction and a conclusion to another LO).

After the metadata schema has been created, the metadata validation process can be fully automated. This process consists of the four steps(see Fig. 3).

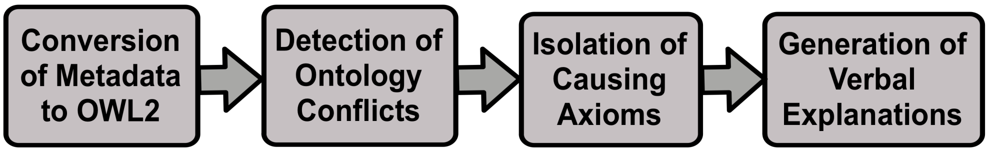

Fig.3. Phases of the semantic gape detection process.

The gap detection tool has been implemented as an extra functionality of the ActiveMath/Math-Bridge authoring platform (see right panel of Fig. 4).

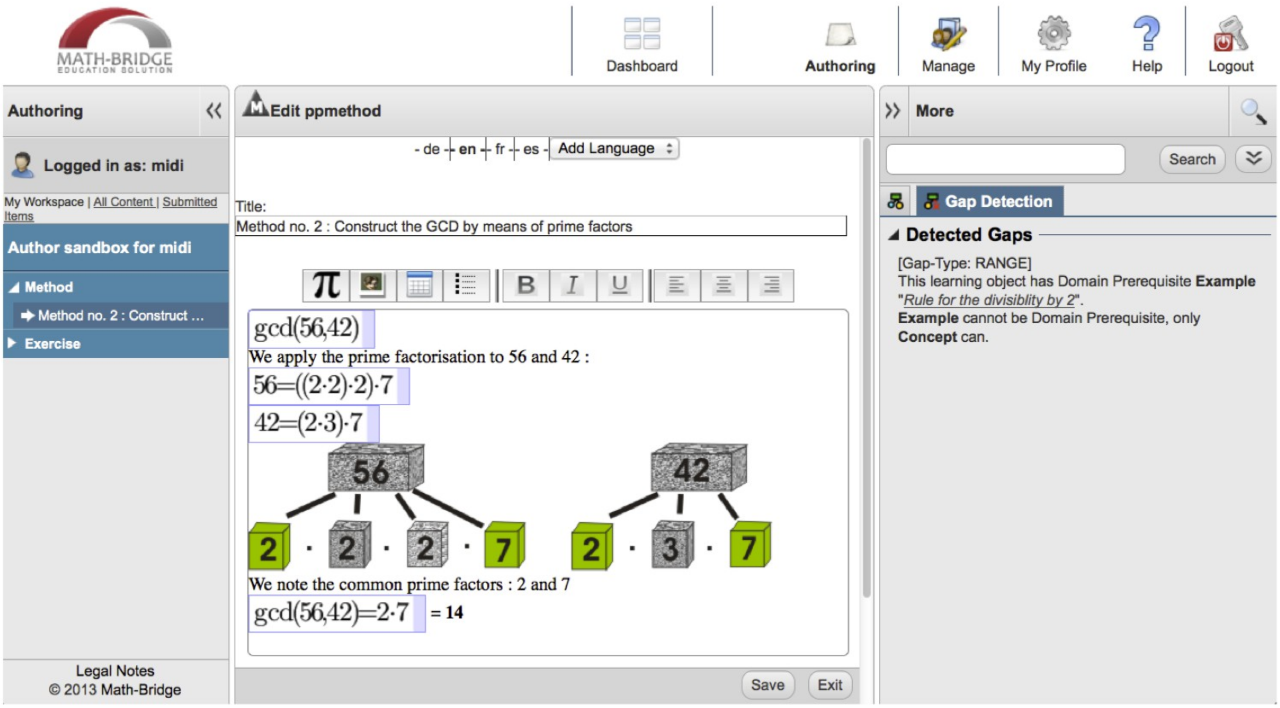

Fig. 4. Interface of the semantic gape detection tool.

3.3 Exercise Difficulty Calibration Tool

Inaccurate exercise difficulties will cause an Intelligent Tutoring System (ITS) to produce ineffective instruction:

- Presenting easy exercises that are incorrectly annotated as difficult, to a strong student can cause unnecessary drilling practice and boredom;

- Presenting difficult exercises that are incorrectly annotated as easy to an under-performing student might result in repeated failures and frustration.

Our approach aims at accurate calibration of exercise difficulty metadata in an ITS. First, we data-mine exercise activity logs to predict students' knowledge of underlying concepts (see Fig. 5); the students' knowledge predictions are used as estimates of their ability to solve the exercises. Then, they are combined with the observed outcomes of consequent exercise attempts to infer the difficulty of the exercises used in the system. The approach relies on two well-founded techniques widely used in the field of learning analytics for probabilistic estimation of student mastery and difficulty of assessment. Bayesian Knowledge Tracing (KT) is used to retrospectively compute student knowledge for the concepts involved in calibrated exercises. Item Response Theory (IRT) 2PL-model is applied to estimate the exercise difficulty based on its history of attempts.

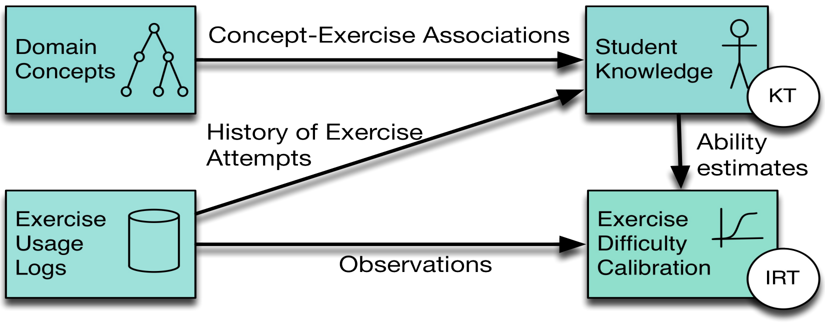

Fig. 5. Mane phases of the exercise difficulty calibration process.

3.4 EXAMAT – Authoring Tool for Interactive Exercises

A new authoring tool for interactive exercise has been developed (see Fig. 6). EXAMAT supports creation of various interactive exercises: one-step and multistep, testing and training, with and without graphics, MCQ, ordering, free-input, etc. Exercises are provided with exhaustive metadata describing their different pedagogical properties and linking them to concepts, theorems, axioms, etc.

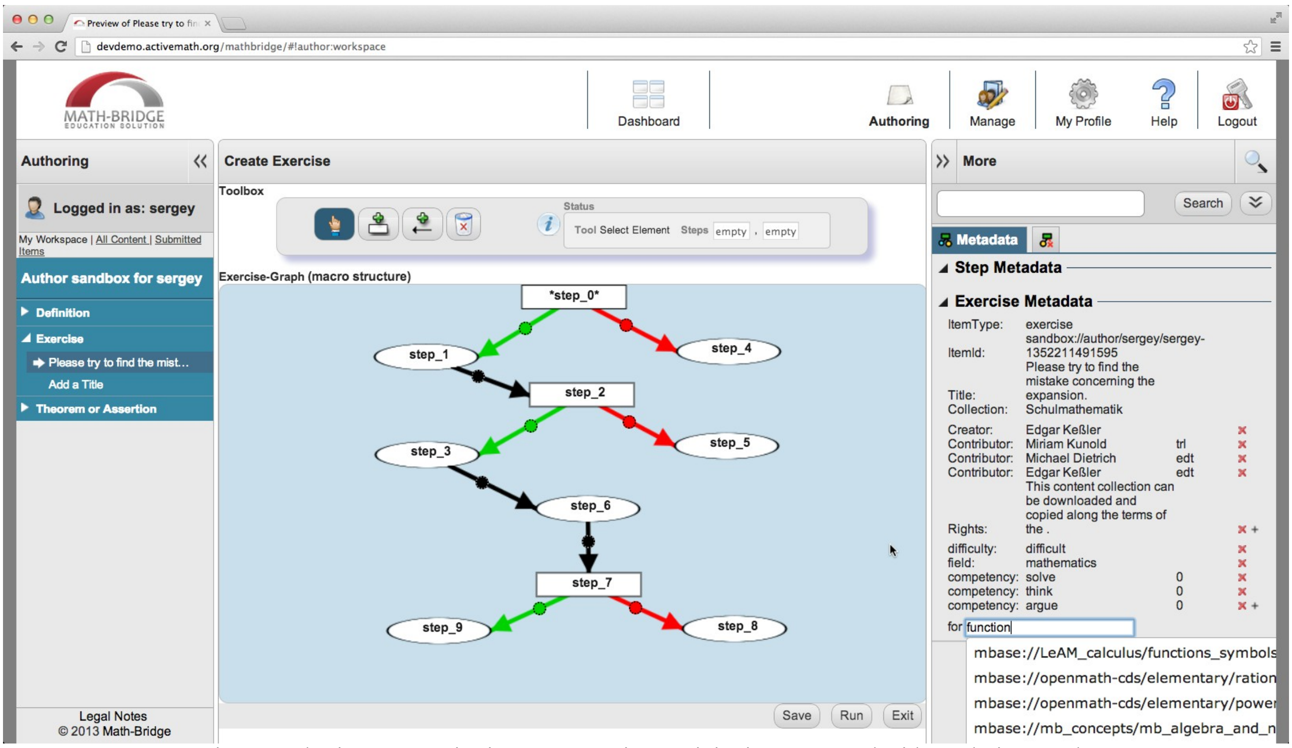

Fig. 6. Authoring an exercise in EXAMAT: interactivity is represented with a solution graph.