Using Schema Analysis for Feedback in Authoring Tools for Learning Environments

Harrie PASSIER & Johan JEURING
Faculty of Computer Science, Open University of the Netherlands
Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands
Email: harrie.passier@ou.nl & johan.jeuring@ou.nl

Abstract. Course material for electronic learning environments is often structured using ontology and schema languages. During the specification and development of course material, many mistakes and errors can be made. In this paper we introduce schema-analysis as a technique to analyse structured documents, and to point out (possible) mistakes introduced by an author during authoring. With this technique we are able to produce valuable feedback. We show the technique at work using six categories of mistakes and two types of schemata.

Introduction

Electronic learning environments (LE’s), comprising eLearning systems, intelligent tutoring systems etc., are often complex tools. Since instances of such systems, for example for a particular course, are often written by non computer science experts, authoring tools have been developed to support the development of such courses. More open-ended authoring tools for LE’s allow for more flexibility in both the form of the content and the order of the steps to design an LE [11]. This flexibility implies a higher probability of mistakes such as inconsistencies and inaccuracies. To improve the quality of LE’s, an authoring tool should include mechanisms for checking the authored information on for example accuracy and consistency. Murray [11] mentions several such mechanisms. In this article we introduce schema-analysis, with which we are able to detect a number of the possible mistakes that can be introduced by an author during authoring a course.

The results presented in this paper are part of a project in which we will investigate general feedback mechanisms to learners as well as to authors [12]. To be able to produce semantically rich feedback we imagine an environment that contains several types of knowledge, like domain, task, education and feedback knowledge, which are represented by ontologies. These ontologies are the arguments of a general feedback engine, which observes student and author activities and matches these against the argument ontologies. This framework, a general feedback engine and the use of ontologies as arguments, supports the constant requirement for flexibility, adaptability and reusability of knowledge structures in LE’s [2]. In this article we focus on feedback to support authoring LE’s.

During authoring different aspects of a course, for example content, structure, and the ontologies, will be authored. In this article we focus on course structure and domain ontology, which we represent by IMS Learning Design (IMS LD) [8] and RDF. Using IMS LD an author defines the structure of a course in a flexible way. IMS LD supports a wide range of pedagogies in online learning. Rather than attempting to capture the specifics of many different pedagogies, it does this by providing a generic and flexible language. With such a flexible language, an author can easily make mistakes. These mistakes can be partly prevented by using templates. Some drawbacks of templates are: loss of flexibility, because an author must follow the steps prescribed by a template, and problems with maintainability: it is hard to maintain documents produced by means of templates. With schema-analysis we maintain flexibility, are able to produce feedback when an author makes a mistake, and leave the author, as a didactic professional, free to accept or not accept the feedback information [3]. The freedom to accept or not accept feedback is important. When a (possible) mistake is signalled, it is the author’s decision to reject or accept the warning. Sometimes, it could be the author’s intention to deviate from rules. What the system signals as a possible mistake may be correct from the author’s point of view.

To determine the quality of a course, we want to detect whether or not the following properties hold for a course: If such a property holds, this may signal (the absence of) a potential mistake: (1) Completeness: Are all concepts that are used in the course defined somewhere? (2) Timely: Are all concepts used in a course defined on time? (3) Recursive concepts: Are there concepts defined in terms of itself? (4) Correctness: Does the definition of a concept used in the course correspond to the definition of the concept in the ontology? (5) Synonyms: Are there concepts with different names but exactly the same definition? (6) Homonyms: Are there concepts with multiple, different definitions?

Since a course and course related material are represented by means of schema languages such as IMS LD and RDF, we can use schema analysis techniques to answer the above questions, and to produce feedback about possible mistakes for authors. We have implemented the mentioned analyses as six distinct schema-analyses, which we show at work in a simple course structure and domain ontology. We have yet to develop examples in the context of real courses.

Schema analysis techniques are based, amongst others, on mathematical results about fixed points. Since these results are not widely known, we will explicitly show how to use them in the context of schema analyses. Schema analyses will be expressed in the functional, declarative, programming language Haskell, since this allows us to stay close to the mathematical results we use. We briefly explain Haskell, for more information see [7].

This article is structured as follows: Section 1 briefly explains what we mean with schemata and introduces the languages we use to represent them. Furthermore, we extend IMS LD and RDF to be able to define more structure. Section 2 introduces schemata and Haskell constructs we use are introduced in section 2.1. Sections 2.2 and 2.3 presents the necessary data structures and algorithms. The last two sections discuss related work (section 3) and conclude (section 4).

1. Schemata and representations

An ontology specifies the objects in a domain of interest together with their characteristics in terms of attributes, roles and relations. Using an ontology many aspects of a certain domain can be represented, such as categories (taxonomic hierarchy), time, events and composition [13]. A composite object contains objects related to other objects using ‘has part’ or ‘uses’ relations. Any object that consists of parts is called a composite object. A composite object has structure: the parts and their relations. Such a structure description is called a script or a schema. In this article we focus on schemata.

Domain ontology - To represent a domain ontology we use RDF, which can be used to represent meta-data as well as the semantics of information in a machine accessible way. RDF is a universal language that describes resources. The basic building block of RDF is a triple: (resource, property, value), which defined concepts and related concepts. For example the concept ‘cycle_wheel’ consists of (has parts) the concepts ‘rim’ and ‘spoke’ (‘<cycle_wheel, has_part, rim>’ and ‘<cycle_wheel, has_part, spoke>’).
Course structure - XML is a language for structuring documents. A data type definition (DTD) describes the type of a set of XML documents. IMS LD [8] is a DTD developed to represent structures of e-courses. The content of a course is presented in a structured way, and activities in an activity-structure. For the examples in this paper we focus on the Activity-model, which consists of several elements: Metadata, Objectives, Prerequisites, Environment and an Activity-description. An activity-description consists of nine elements. One of them is the What-element, which contains the instruction for the activity to be performed. Possible instructions are grouped together by the parameter entity Extra-p. To be able to add more specific annotations to content and structure we introduce two new elements in the Extra-p element: Definition and Example. Furthermore, we introduce a new attribute Educational-strategy of the element Activity with two possible values: Inductive and Deductive. Introducing such elements will make it possible to structurally analyse educational material. These elements serve as examples to illustrate the analysis techniques at work. In practice many elements can be added, depending on the desired analyses. Listing 1 shows only the relevant elements and attributes related to the activity-model together with the newly defined elements example and definition. The new elements and attribute are marked in bold.

Listing 1. Parts of the activity-model in IMS LD definition

The definitions of the new elements Definition and Example are presented in listing 2.

Listing 2. Definition of the new elements

2. Schema analysis to detect authoring problems

The schemas given in Section 1 represent structural aspects, which can be analysed. In this section we give some examples of schema-analyses that determine whether or not certain properties hold. The results of these analyses form the basis of feedback to the author. The analyses take the schemata as input. In this paper we perform two types of analyses: 1) the analysis of structural properties of a schema, for example the recursive property, and 2) the comparison of a schema with one or more other schemata, for example to test the correctness of a definition.

2.1 Haskell preliminaries

The tuple data type \( (t_1, t_2, \ldots t_n) \) is constructed from component types. It consists of values \( (v_1, v_2, \ldots, v_n) \), in which \( v_i: t_i \) etc where \( :: \) means ‘is of type’. Function \( fst \) selects the first element of a pair, \( (x, y) \rightarrow x \), and function \( snd \) selects the second element. We use the data type list extensively. The empty list is denoted by \([\ ]\), and the concatenation of two lists \( x \) and \( y \) is denoted by \( x ++ y \). Prepending an element \( x \) to a list \( x \) is denoted by \( x : x \). In a list comprehension \( [x \mid x \leftarrow xs, \text{test} x] \), a new list is generated from the list \( xs \). Each element \( x \) of \( xs \) is tested, and, if the test succeeds, added to the new list. Function \( map \ f \) takes a list and applies function \( f \) to all elements in the list, \( \text{so } \text{map } f \ xs = [f \mid x \leftarrow xs] \). Anonymous functions can be constructed using lambda notation \( \lambda \), so function \( \lambda (x, y, z) \rightarrow (x, y) \) selects the first two components of a triple. Function \( null \) tests if a list is empty: \( \text{null } [\ ] = \text{ True} \). To check if an element \( x \) is an element of list \( xs \), we use the expression \( \text{elem } x \text{ } xs \). Function \( zip \) takes two lists and returns a list of corresponding pairs: \( \text{zip } [1,2] [3,4,5] = [(1,3), (2,4)] \), where extra elements in the longer list are discarded. Functions \( head \) and \( tail \) extract the first and the remaining elements of a nonempty list, respectively. Function \( inits \) returns the list of initial segments of its argument list: \( \text{inits } "abc" \rightarrow ["", "a", "ab", "abc"] \), and function \( tails \) returns the list of all tail segments of its argument list: \( \text{tails } "abc" \rightarrow ["abc", "bc", "c"] \). Function composition composes two functions: the output of the second function \( g \) becomes the input of the first function \( f; \ f \ g \ x = f \ (g \ x). \) The type of a function \( f; \ f \ g \) can be read as: function \( f \) takes two arguments of types \( t_1 \) and \( t_2 \) and returns a value of type \( t_3 \). Not all arguments have to be mentioned in a function definition, so \( \text{completeCourse} c = \text{null } \) \( (\text{undefinedConcepts } c) \), with type \( \text{completeCourse } : \text{Course} \rightarrow \text{Bool} \), equals \( \text{null } \text{. undefinedConcepts} \). Functions can be passed as parameters. For example, in \( \text{map isEven } [1,2,3,4] \) the type of \( \text{map } = \lambda (\text{Int} \rightarrow \text{Bool}) \rightarrow \text{Int} \rightarrow \text{Bool} \). Choice between conditions is represented by a vertical bar \( \mid \). For example \( \text{max } x y \mid y > x \mid \text{otherwise } = y \)
m

means: if the guard \( y > x \) is true then \( x \), otherwise return \( y \).
Completeness — We distinguish three kinds of (in)completeness: (1) within a course, (2) within an domain ontology and (3) between a course and an domain ontology. If a concept is used in a course, for example in a definition or an example, it has to be defined elsewhere in the course. The undefined concepts in a course are calculated in three steps: (1) determine the set of concept id’s that appear in the right- and left-hand sides of concepts within examples and all concept id’s that appear in the right hand side of concepts within definitions (used concepts), (2) determine the concept id’s that appear in the left-hand side of concepts in definitions (defined concepts) and (3) check that each of the used concepts appears in the set of defined concepts. A course is complete if all concepts used appear in the set of defined concepts. Completeness can also be applied to an (domain) ontology, and between a course and an ontology. The first one checks whether all used concepts in the ontology are defined in the same ontology, the second one if all used concepts in a course are defined in the ontology. The same three steps are performed in both functions.

Timely — A concept can be used before it is defined. This might not be an error if the author uses an inductive instead of a deductive strategy to teaching, but issuing a warning is probably helpful. Furthermore, there may be a large distance (measured for example in number of pages, characters or concepts) between the definition and the use of the concept, which is probably an error. We define the function timely to determine whether or not concepts in a course are defined in time and a function outOfOrderConcepts to list the concepts that appear to be out of order.

outOfOrderConcepts = function outOfOrderConcepts, function extractActivities returns for every activity in the course the tuple (Exercise, [Extra p]) and puts these tuples in a list activities. Then, using functions init and tails every [Extra p] list is split as follows: for every element e in the list [Extra p] the list is subdivided into a left part (e1), which contains all elements to the left of element e, and a right part (e2), which contains all elements to the right of e. For example, in the list [[1], [2], [3], [4], [5]], where e = 4 is example and d is definition. Finally, function intime tests the timely constraints for all tuples (ex, epl, epr): if the first element of epr is a definition and the educational strategy is deductive, then: 1) a related example appears after the definition, and 2) no related example appears before the definition (tested by elemByEqConcept e in the code below). In case of an inductive activity, a related example appears before the definition and function intime is always true if epr is empty or the first element of epr is an example.

2.3 Solving authoring problems with schema analysis
In this section we describe six algorithms (four briefly and two in more detail), which can be used to signal the (possible) mistakes listed in the introduction of this paper. The complete code can be obtained from:
recursiveOntology :: Ontology -> Bool
recursiveOntology = not . null . listRecursiveConceptsOntology
listRecursiveConceptsOntology :: Ontology -> [Id]
listRecursiveConceptsOntology = recursiveConcepts . extractUIConceptsOntology

Function recursiveConcepts calculates for every concept all reachable concepts, as explained in section 2.2. Every concept in reachable is checked for recursiveness: a concept is recursive if the concept’s Id is a member of the set of reachable concepts. The recursive concepts are collected in a list.

recursiveConcepts :: ([Id], RelatedConcepts) -> [Id]
recursiveConcepts allConcepts = listNonTerminalConcepts = filter (not . null) . nElements . allConcepts
reachable = reachable nonTerminalConcepts allConcepts
...