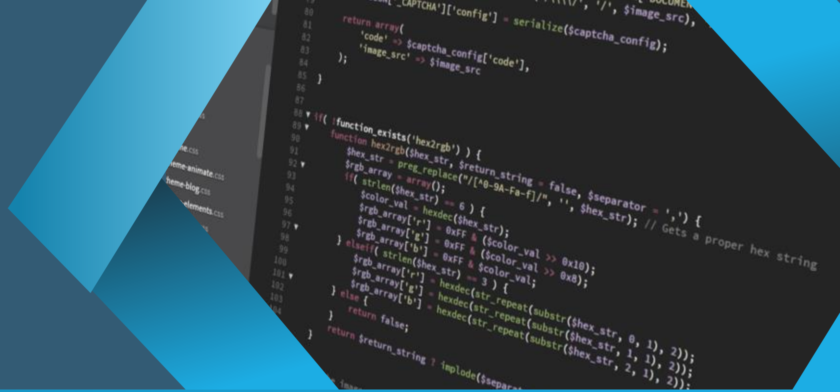
Clean CaDET

Clean Code and Design Educational Tool



A CONCEPTUAL MODEL OF THE KNOWLEDGE-BASED RECOMMENDER SYSTEM

Executive Summary

Via its Smart Tutor module, Clean CaDET aims to ensure efficient recommendations of learning materials for understanding and addressing code smells. Smart Tutor is envisioned as an adaptive e-learning system consisting of content, learner, and instructional model, with the goal of integration of a plethora of educational techniques and AI-based mechanisms that will enable teaching of a specific subject matter and development of cross-cutting skills at scale. In this document, we first give an overview of each model of the adaptive e-learning system. As the instructional model of Smart Tutor is envisioned as a Hybrid Recommender System, we use this document to present an Ontology-Based Knowledge-Based Recommender System, which will serve as a basis of the latter system.

Key Takeaways

1. *Ontology-based recommenders are Knowledge-Based Recommender Systems that use ontology for knowledge representation. The ontology describes the domain concepts and the relationships between those concepts.*
2. *When organizing the content model, the collection of learning objects should be divided into subcollections, where each should teach a particular skill or knowledge type. With this organization, each of the subcollections will contain all instructional components necessary to teach that skill.*
3. *An adaptive e-learning system takes information available in the learner model to tailor learning materials and teaching methods to each learner. Therefore, it's important to accurately assess a learner's current knowledge and identify its personal traits.*
4. *Professional instructors and clean code researchers will define rules which will represent the basis of our instructional model. With this, we're avoiding the knowledge acquisition bottleneck, which is a common problem for Knowledge-Based Recommender Systems.*

Contact

Jelena Slivka

slivkaje@uns.ac.rs

https://clean-cadet.github.io/

1 Introduction

Recent years have seen the rise of specialized e-learning software, centered around the use of intelligent tutors. Various studies showed that these tutors helped students outperform students from the conventional classes in the majority of the observed cases. As an educational tool, Clean CaDET aims to ensure efficient learning with personal advice and educational materials. As a part of our project, we aim to develop an AI-based educational system for delivering personalized multimedia to the learner, detailing the nature and impact of the smell, as well as the recommended refactoring solution. A part of the Clean CaDET that represents an AI-based educational system that personalizes the learning using a Recommender System (RS) is Smart Tutor.

The Smart Tutor is envisioned as an adaptive e-learning system that selects the most suitable content from the repository of educational material to facilitate learner's learning. Its instructional model represents Hybrid RS which will combine Knowledge-Based (KB) and Collaborative Filtering (CF) approaches. The main benefit of hybridization is to take advantage of the strength of each particular technique while overcoming the limitations of individual techniques [1]. Hence, our hybrid approach will alleviate problems such as cold-start and rating sparsity, which limit the performance of recommendation. As the KB approach tends to work better than others at the beginning of their use, and the CF requires collecting user ratings of the recommended material, the KB component will be developed first.

KB RS recommends items based on specific domain knowledge about how certain item features meet users' needs and preferences and, ultimately, how the item is useful for the user. Similarity score in KB RS can be directly interpreted as the utility of the recommendation for the user [1]. Three major types of KB RS are Case-based, Constraint-based, and Ontology-based. When it comes to e-learning, there is no single model for the learner profile or the structured content, which makes the need for ontology even more essential [2]. Ontology-based recommenders are KB RS that use ontology for knowledge representation. The ontology describes the domain concepts and the relationships between those concepts. [3]

The KB component of the instructional model of Smart Tutor is envisioned as the Ontology-based RS, where the ontology will be used for educational content representation. This will be combined with user preferences and behavior representation to define the knowledge, represented via rules, that will be used to recommend learning material to the learners. In this document, we're providing an overview of models of our Smart Tutor and an overview of our KB RS.

2 Conceptual model

Our Smart Tutor consists of three aggregates that support its intelligent tutoring, as illustrated in Figure 1.

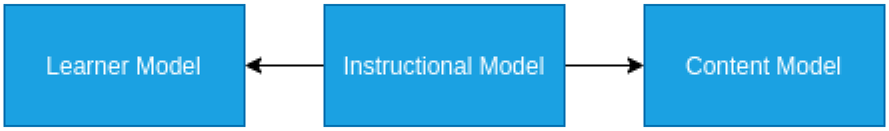


Figure 1. Conceputal model of Smart Tutor

The Content Model is charged with structuring and storing the learning objects that describe concepts from the quality code engineering domain. The Learner Model hosts information related to the user. This includes an estimate of the user’s current knowledge and personality traits. The Instructional Model connects Learner Model and Content Model via rules for constructing knowledge nodes of learning objects to produce recommendations that best suit the user. A detailed overview of the three modules is given in the following sections.

3 Content model

The requirements for any content model can be grouped into two categories [4]:

1. Requirements of the delivery system, which has to be content independent, robust, flexible, and scalable;
2. Requirements of the learning content that is to be delivered, which prescribe that content must be composed in a way that the delivery system can adapt it to the needs of the particular learner, meaning that the content will have to be built to the same specification.

Organizing content via learning objects (LOs) fulfills the requirement of building to the same specification. LOs are small and reusable components such as video demonstrations, tutorials, stories, assessments, etc [4]. Another research [5] defines a learning object as the knowledge of:

- *Facts (Know what)* - statements that explain the relationship between two objects or events.
- *Concepts (Know that)* - classes of items, words, or ideas that are known by a common name, share some common features, and include multiple specific examples.
- *Principles (Know why)* - types of relational rules that can be used as “if-then” or “cause-effect” relationships.
- *Mental Model* - a network constructed from facts, concepts, and principles.

LOs may be selectively applied, either alone or in combination to meet the learner’s needs for learning. When organizing the content model, LOs collection should be divided into subcollections, where each should teach a particular skill or knowledge type. With this organization, each of the subcollections will contain all instructional components necessary to teach that skill [4].

When defining the content model, knowledge structure has to be a part of it, to allow dependency relations to be established in an e-learning system. This represents the basis for [4]:

- Assessment - identifying the current status of a particular LO
- Cognitive diagnosis - if there’s any problem, identify the source
- Instruction - which LOs need to be taught next to fix a problem or to present a new topic.

Elements of knowledge structures are knowledge nodes. Each node has an associated collection of LOs, and can be classified in terms of types of knowledge as [10]:

- *Basic Knowledge (BK)* - What part of the content (definitions, examples, formulas, etc.)
- *Procedural Knowledge (PK)* - How part of the content (step-by-step information, relations among steps, etc.)
- *Conceptual Knowledge (CK)* - Why part of the content (relational information among concepts and the explicit connections with BK and PK).

Restricting a node to a single knowledge type limits the scope of what can be in any single node, thus ensures that the educational content is broken to an appropriate size [4].

With our content model, we'll first define LOs of different types and roles, that will facilitate the learning of domain concepts. After that, we'll group LOs in the knowledge nodes:

- *BK* - terminology and specific facts, containing definitions, formulas, and illustrative examples
- *CK* - interrelationships of basic elements within a larger structure, containing mind maps, complex examples, analogies, and consequences
- *PK* - how to do something with the facts and concepts, containing algorithms, techniques, methods, and heuristics for when to use them.

As the content model contains domain-related knowledge, ontology can be used to create associated structures and dependencies. Ontology is recognized as a term referring to the shared understanding of some domains of interest, which is often conceived as a set of classes (concepts), relations, functions, axioms, and instances. It is a formal description of concepts in a domain of concepts (class), properties of each concept describing various features and attributes of the concept (slot), and restrictions on the slot (facets) [5]. Using ontology, an adaptive e-learning system can first decide what needs to be taught and then decide what to teach.

While researching legacy ontologies, we weren't able to find one that we can completely reuse. Therefore, we shall develop our ontology. When developing an ontology, we aim to follow the purpose-oriented model defined by Lee et. al [5], as it proved to be a widely adopted practical process for ontology development. The steps that we shall take are:

1. Definition of explicit domain and purpose of our ontology
2. Survey of all important terms in our explicit definition of the domain
3. Definition of concepts (classes) in our domain and their hierarchy strongly depending on the expert's view and the purpose
4. Definition of features and attributes (slots) of classes that strongly depend on the purpose of our ontology
5. Creation of instances of classes while taking care that the instances that are unconcerned with the ontology purpose should not be created.

4 Learner model

The learner model incorporates various components to construct a learner profile tailored to the learning domain. It can be defined as a model that contains the learners' current knowledge and personal traits [6]. This definition serves as the basis in our implementation of the learner model.

As the learner's current knowledge is an essential part of the learner model, it is important to successfully assess the learner. Shute and Towle [4] name two aspects of knowledge assessment:

1. *Domain-dependent information* - use of pretest and performance data to initialize a learner model concerning LOs. With this, it's possible to eliminate those that are already known, and focus on the improvement of weak areas;
2. *Domain-independent information* - learner profile data. With this, it's possible to pick and serve optimal LOs sequences and formats.

In our implementation, we plan to use pretesting and completion tracking to identify learner's current knowledge.

A personal trait is an underlying user characteristic that defines a learner as an individual and can be their learning style, cognitive style, and emotion. Research shows that the learners who are aware of their preferences based on their personal traits can [6]:

- Improve their understanding and confidence during the learning process
- Increase their motivation
- Retrieve suitable learning materials in learning situations that constantly change.

The selection of appropriate personal traits to integrate into an adaptive e-learning system is challenging. Authors of the recent systematic study [6] classified personal traits in adaptive e-learning systems into four categories: cognition, affective, behavior or psychomotor, and mixed. Their study shows that the most frequently used traits are from the cognition learning domain category. Cognition is related to a pattern of information processing that uses rational thought to create and gain knowledge during the learning process. As cognition includes personal traits such as learning style, cognitive style, working memory capacity, personality type, etc., we plan on researching detailed differences between them to decide which ones we support.

The same study [6] shows that computer-based detection techniques are most frequently used to identify personal traits. Computer-based detection techniques usually use a combination of two or more Machine Learning (ML) or non-ML techniques to analyze implicit user input to automatically determine personal traits. Research shows that this approach is more accurate than others because it responds quickly to changes in the learner's personal traits, which enables frequent updates to the learner information in the learner model, resulting in a better learner model. As some computer-based detection techniques require large amounts of training data to achieve accurate identification [7], and information of new learners is initially insufficient to build an appropriate learner's profile [8], we'll explore a hybrid technique for the identification of personal traits, that will include pretesting.

5 Instructional model

The instructional model manages the representation of educational material and ensures learner mastery by monitoring the learner's model concerning the content model, addressing discrepancies, and prescribing an optimal learning path for a learner [6]. Gagne's [9] events of instruction represent guidelines for designing good e-learning environments, as they provide necessary conditions for learning and serve as the basis for designing instruction and selecting appropriate media. Research shows that this nine-step model is a good way to facilitate learners' successful acquisition of knowledge and skills in the adaptive e-learning system [6]. We plan on following these guidelines when developing our instructional model.

Professional instructors and clean code researchers will define rules which will represent the basis of our instructional model and KB RS of Smart Tutor itself. With this, we're avoiding the knowledge acquisition bottleneck, which is a common problem for KB RS [1], by having knowledge of domain experts explicitly encoded into a formal and executable representation.

To determine the quality of rules, and the recommendation itself, we plan on two evaluations of our KB RS. Firstly, we'll conduct the offline evaluation at the design time, which will help us verify the prediction power of developed rules. We'll assume user behavior and collect enough data, so we can make reliable decisions based on the simulation. Secondly, we'll conduct the user-centric evaluation with a focused user study after the system has been launched. This will allow us to test user behavior when interacting with KB RS, and the influence of the recommendations on user behavior. It will also allow us to collect qualitative data that is, as research shows [1], often crucial for interpreting the quantitative results.

6 References

[1] Ricci, F., Rokach, L. and Shapira B., 2015. Recommender System Handbook. Springer US.

[2] George, G. and Lal, A.M., 2019. Review of ontology-based recommender systems in e-learning. Computers & Education, vol. 142.

[3] Harrathi, M., Touzani, N. and Braham. R., 2017. A Hybrid Knowledge-Based Approach for Recommending Massive Learning Activities. 14th International Conference on Computer Systems and Applications (AICCSA), pp. 49-54.

[4] Shute, V. and Towle, B., 2003. Adaptive E-Learning. Educational Psychologist, vol. 38(2), pp. 105-114.

[5] Lee, M., Ding, Y.Y., and Tzone, W., 2005. Java learning object ontology. Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), pp. 538-542.

[6] Afini Normadhi, N.B., Shuib, L., Md Nasir, H.N., Bimba, A., Idris, N. and Balakrishnan, V., 2018. Identification of personal traits in adaptive learning environment: Systematic literature review. Computers & Education, vol. 130, pp. 168-190.

[7] Lukasenko, R. and Grundspenkis, J., 2010. Adaptation of Intelligent Knowledge Assessment System Based on Learner's Model. In Proceedings on the 16th International Conference on Information and Software Technologies, Kaunas, Lithuania, pp. 1-8.

[8] Baldiris, S., Graf, S. and Fabregat, R., 2011. Dynamic User Modeling and Adaptation Based on Learning Styles for Supporting Semi-Automatic Generation of IMS Learning Design. In 2011 IEEE 1th International Conference on Advanced Learning Technologies, pp. 218-220.

[9] Gagne, R.M., Briggs, L.J. and Wager, W.W., 1992. Principles of instructional design (4th ed.). Fort Worth, TX: Harcourt Brace.

[10] Krathwohl, D.R., 2002. A revision of Bloom's taxonomy: An overview. Theory into practice, vol. 41(4), pp. 212-218.

The Project is funded by the Science Fund of the Republic of Serbia, Grant No. 6521051, AI - Clean CaDET

Science Fund of the Republic of Serbia