STI-DICO: a Web-Based System for Intelligent Tutoring of Dictionary Skills

Alexandra Luccioni  
Roger Nkambou  
Jean Massardi  
Université du Québec à Montréal  
201 av. du Président-Kennedy  
H2X 3Y7 Montréal, QC, Canada  
vorobyova.alexandra@uqam.ca  
nkambou.roger@uqam.ca  
massardi.jean@uqam.ca

Jacqueline Bourdeau  
Claude Coulombe  
LICEF, TELUQ  
5800, rue Saint-Denis  
H2S 3L5 Montréal QC, Canada  
jacqueline.bourdeau@teluq.ca  
claude.coulombe@gmail.com

ABSTRACT
In this paper, we describe an innovative project where Web technologies are exploited to develop an Intelligent Tutoring System (ITS) that uses a Learning Management System (LMS) as its learning interface. The resulting ITS has been instantiated into a specific system called STI-DICO which aims at helping future French primary school teachers to acquire the knowledge and skills needed to use the French dictionary. The learning process in the ITS takes place via a number of authentic learning scenarios that represent situations that the future teachers will face in the classroom. By using a LMS as the learning interface component of the system, we enable it to be directly deployable on the Web to a large population of students, all the while retaining the adaptive components of an ITS to deliver a personalized learning experience to its users.

Keywords
Intelligent tutoring system; learning management system; service-oriented architecture; Web-based learning

1. INTRODUCTION
In the last two decades, the Web has become an irreplaceable tool in our personal and our professional lives. It has also become a key component in the democratization of education, giving learners all over the world instant access to up-to-date information. The Web is especially useful in the field of Educational Technology (ET), since it helps address the issue of increasing classroom sizes by providing the opportunity of using Web-based tools in classrooms (Kalyanpur, 2005). One of the most widely-used applications of ET are Learning Management Systems (LMS), which are easy to use and can empower teachers to use the Web to share content and exercises with their students. While LMSs are not adaptive on their own, in the hands of a skilled post-secondary teacher, they can become powerful tools to add content from the Web as a complement to textbooks and classroom activities.

Another powerful tool of ET are Intelligent Tutoring Systems (ITS), machine tutors that apply artificial intelligence techniques to guide the user throughout the learning process (Woolf, 2010). ITS are increasingly using web-based technologies and adopting service-oriented architectures in order to reach more students on a global scale, lower development costs, and permit the integration of content from the Semantic Web (Hsiao et al., 2013; Weber and Brusilovsky, 2015; Nye 2015). Adaptivity in ET tools is important because it enables the personalization of the learning activities to the objectives, strategies, knowledge and even emotions of learners. This changes the relationship between the learner and the content to be learned, traditionally static, to make it interactive, engaging, collaborative, and pertinent to the user's context of learning.

Our study focuses on the integration of LMS with ITS, to exploit the scalability and ease of use of the first with the adaptive guidance and intelligence of the second. If this integration is carried out successfully, this could provide ITS with a springboard towards their usage on a larger scale in classrooms and by independent learners. This, in turn, would permit a more vast collection and analysis of the data obtained from interactions with the LMS-ITS hybrid systems and the usage of Educational Data Mining techniques to further improve their efficiency.

In the second section of this paper, we expose our hypothesis and methodology to develop STI-DICO, an adaptive ITS for learning dictionary-related skills, using an LMS as learning interface and deploying a Service-Oriented Architecture (SOA). The third section presents the architecture and components of STI-DICO. In the fourth section, we present a concrete learning situation in STI-DICO and the way in which it will function. The paper ends with some words on ongoing and future work, followed by a conclusion.

2. HYPOTHESES AND METHODOLOGY
In order to ensure that we maintain the essential link between theory and practice and to advance both components synchronously, we have chosen the Design Based Research (DBR) methodology (Reeves et al., 2005). DBR has the advantage to be situated in a concrete application context, which enhances partnership and collaboration between researchers and practitioners, while focusing on the concrete results of the innovation tested. DBR works via an iterative process, each iteration representing both a progression in the complexity of the system tested and of the proposed theoretical knowledge.
Our operational hypothesis is that a service-oriented architecture integrating an LMS interface with ITS back-end components is the best integration approach given the needs of our project. While there are only a few existing predecessors of LMS-ITS integration (Rice, 2011; Alev et al., 2015b) we believe that it is a promising development path for the next generation of ITSs, resulting in adaptive technology that can easily be used and modified by teachers according to their needs while providing a rich source of educational data that can be analyzed for the benefit of course authors, teachers and learners.

While the system architecture proposed is generic and can be applied to a variety of knowledge domains and skills, we have chosen to instantiate it within the domain of linguistics, and more specifically that of dictionary usage. This is due to the fact that in Quebec, a gap has been observed between the requirements that the Ministry of Education has of its students and the reality of teaching in the classroom. More specifically, French primary school teachers are expected to teach their students how to use dictionaries, but themselves have not received adequate training in lexicology to foster the learning of the needed skills by their students (Tremblay 2009, Anctil 2011). Having established the demand for dictionary-specific training in our joint work with Tremblay (Tremblay et al., 2013), our aim is to develop an ITS which can be offered in complement to existing courses in teacher training, and extended to other users and to other languages.

Since the knowledge domain of our ITS is considered to be an ill-defined domain (Fournier-Viger et al., 2010), our fundamental hypothesis is that a formal modeling of the knowledge and skills to be learned via the system is necessary for the learner to better understand and assimilate the knowledge targeted by our ITS. We therefore accompany our technical implementation with research in linguistics and learning sciences, and more specifically, dictionary usage (Nesi, 1999; Lew, 2013) in order to increase the effectiveness of our system.

We have already carried out several steps of our methodology, notably the literature review and needs analysis. Another step of the project is underway, consisting in the development of the content to be modeled in our ITS, both in terms of dictionary skills and lexicological knowledge as well as the learning activities. We have currently developed a referential of dictionary skills consisting of 125 skills in 4 categories (conceptual knowledge, linguistic competency, dictionary competency and practical skill) linked with an ontology of 25 linguistic concepts to ensure adequate theoretical grounding. The referential will be transformed into a series of databases used by STI-DICO and linked to the learning activities to model the skills acquired by the learner. The data collection of this first iteration is underway, consisting in an evaluation of the referential being performed by three experts in linguistics.

3. EXPLORING ITS AUTHORING TOOLS

The modular architecture chosen for STI-DICO is based on traditional 4-component ITS architectures (Anderson et al., 1985; Nkambou et al., 2010). However, due to the limited scale of our project, it was not possible to envisage developing the ITS interface from scratch. Likewise, we wanted to employ a Service-Oriented Architecture (SOA) to ensure that our ITS was coherent with current trends in the domain (Nye, 2015). As a result, we decided to utilize the interface of an LMS platform and integrate it with the reasoning components of an ITS.

In the current section, we consider the different implementation possibilities for our ITS, and more specifically, the interoperability of the various tools available to us. In this perspective, we will analyze Open edX, GIFT and CTAT.

Open EdX Software Architecture

Open edX is an open-source LMS platform designed using a stack of effective technologies such as Python programming language, Nginx, Django, MySQL and MongoDB (edX, 2015). Due to its Web service-oriented architecture, Open edX is able to provide services to and accept services from other software using standard Web protocols. As such, Open edX aims to provide its services to hundreds of thousands of students while adjusting to an increasing or decreasing demand for computing resources using cloud infrastructure. This type of architecture enables deployment from a laptop or a small cloud server to a multi-server infrastructure to serve tens of thousands of students.

Furthermore, the modular architecture underlying Open edX facilitates content composition and reuse, especially since it supports the LTI (Learning Tools Interoperability) standard, whose main objective is to establish standard means of integration of distance learning applications (providers) with hosted course platforms (consumers). LTI allows a wide range of integration scenarios (IMS, 2012) and its main purpose is efficient user authentication (See Fig. 1). Open edX also allows course authors to insert custom JavaScript problems and HTML5 widgets directly into courses, allowing these custom elements to be evaluated in the same manner as exercises created using Open edX templates. This gives the course author a higher degree of freedom in the creation and evaluation of their course.

Figure 1. Overview of LTI v1.1

Due to all of the characteristics mentioned above, we see Open edX as a powerful LMS interface component for our project. Furthermore, it is also the LMS platform used in our educational institution and therefore we have access to a local version of Open edX as well as adequate support and logistics.

Building the Core of Intelligent Tutoring Systems

For the back-end intelligent components of our ITS, two authoring tools were examined: GIFT (Generalized Intelligent Framework for Tutoring) and CTAT (Cognitive Tutor Authoring Tools). From an architectural point of view these two systems are different and therefore offer different means of integration with our LMS interface. In the current section, we will discuss the advantages and disadvantages of each and justify our final choice of authoring tool.

GIFT Software Architecture

GIFT is an open-source, modular, service-oriented architecture whose goal is to make automated authoring, instruction and effect analysis easy and cost-effective (Sottolare, Brawner, Goldberg, and Holden, 2012). It provides a series of
tools for an expert system developer, from course authoring tools to create activities, lessons and guidance, to survey authoring tools for questionnaires which appear during the learning process.

GIFT is a Java-based Message-Oriented Middleware (MOM) software with a messaging API compatible with JMS (Java Message Service) protocol. Although JMS can be more robust and reliable than Web services, JMS middleware must control all the software components that it communicates with. This means that GIFT cannot communicate with external servers on which it has no control, but may however communicate with various rich clients interfaces developed with the Java Swing library.

Another disadvantage of GIFT is that it currently only supports plug-ins with a small number of interface types, notably those created using Microsoft PowerPoint and Visual Basic, and is not up to par in terms of interoperability and authoring services such as knowledge tracing (Aleven, 2015a). That being said, in recent years, GIFT is being re-engineered in order to integrate Web services and support LTI. However, this re-engineering is not complete at the moment of writing this article (Personal communication from K. Brawner, Sept 2015). Since an integration of an ITS tool with Open edX would require it to have full Web capabilities and LTI capacity, it is impossible to consider GIFT as adequate to develop the intelligent back-end of our ITS.

**CTAT Software Architecture**

CTAT is a suite of authoring tools for ITS as well as a factored architecture for developing and delivering tutors. It supports multiple ways of authoring tutors as well as multiple options for developing the tutor front-end and back end (Flash, Java, HTML5) (Aleven et al., under review).

CTAT supports the creation of two types of tutors: example-tracing tutors and rule-based cognitive tutors, the key difference between them being that example-tracing tutors can be applied in problems that have a limited-branching solution space, and can be created without programming but using problem-specific authoring (Aleven, McLaren, Sewall, & Koedinger, 2009). On the other hand, rule-based cognitive tutors require AI programming to build a cognitive model of student but can handle problems even with larger solution spaces (Aleven 2010). CTAT is a Web application conceived as a factored architecture for tutoring, with components and interfaces (APIs) between its components (Aleven, McLaren, & Sewall, 2009; Aleven et al., 2015a, 2015b). Furthermore, CTAT is LTI-compliant, enabling it to use a standard method to establish a secure link that allows single sign-on (SSO) from most LMS platforms.

However, there are still issues regarding CTAT. For instance, it is not an open-source software (although it may be used freely for research purposes), which prevents modification or personalization of its code to better correspond to the needs of ITS developers. Furthermore, the stable versions of CTAT are only compatible with Flash and Java technologies to build the tutor interface. Using Flash raises a host of issues (security, mobile device support, compatibility with browser plug-in technologies), whereas using Java is complicated for non-programmers, which many ITS developers are. Fortunately, CTAT is now transitioning from Flash to HTML5 for the tutor front-end, which will simplify the process of integrating it with Open edX (Personal communication from V. Aleven, December 2015).

4. **PROPOSED STI-DICO ARCHITECTURE**

Our architectural analysis has led us to choose the integration of Open edX with CTAT for the architecture of our ITS. Not only is the integration with CTAT smoother than that with GIFT, but CTAT also implements the provider side of LTI, compatible with the LTI consumer capacity of Open edX. Furthermore, a recent project has demonstrated the technical feasibility of the integration of CTAT with Open edX (Aleven et al., 2015b) using the LTI standard. In the architecture proposed by Aleven and colleagues in their CTAT-edX experiment (Aleven et al., 2015b), the tutor’s inner loop functionality, meaning its personalized guidance within a problem, and its outer loop functionality, meaning the selection of problems based on the student model (van Lehn, 2006), were handled separately. Moreover, the inner loop of the tutor was moved into the student’s browser by re-implementing it in JavaScript, which had a positive effect on the scalability of this approach.

We would like to reproduce and extend the CTAT-edX experiment using different, more complex domain knowledge and a more complex ITS architecture. Our first prototype of STI-DICO will be built using CTAT, hosted on the CMU servers, which will be called upon by the course web pages hosted on local Open edX servers. The tutor will therefore be able to seamlessly run in Open edX course pages. A simplified version of the proposed tutor architecture is shown in Figure 2.

![Figure 2. Integration of a Tutor Tool in an LMS Platform](image-url)
acquired and link them with their corresponding activities. Since the learners we are targeting are sufficiently advanced in the domain and possess adequate contiguous knowledge, we believe that it is important for them to have access to their student model and follow their own learning path in real time.

5. STI-DICO LEARNING SCENARIO

STI-DICO is aimed at future French teachers in primary schools in Quebec, in order to guide them in the acquisition of skills and knowledge necessary to teach the use of French dictionaries (more specifically, electronic dictionaries) in the classroom. In order to ensure authentic, situated learning (Herrington and Oliver, 2000), we have created a series of authentic learning activities that resemble the tasks the teachers carry out in the classroom: correcting and improving students' texts, designing classroom activities, etc. The activities offered in STI-DICO put the users in these familiar situations in order to develop their existing skills as well as to teach them how to exploit the dictionary to improve their performance.

For instance, in order to help the user learn how to use a monolingual dictionary in order to find the synonyms of a word, we will place them in a situation where this knowledge is directly mobilized: a student text in which the same word is repeated several times (Figure 3).

When faced with such a text and a sufficiently vague guidelines (“Improve the quality of this text”), the users must not only recognize the presence of redundant words, and therefore the need to find synonyms, but they must also be able to find the corresponding dictionary section that presents a word's synonyms, and finally replace the redundant words with their synonyms.

STI-DICO not only evaluates the user's final answer, but also the entire sequence of the activity: the user must click the words in the text that are repeated, indicate how it is possible to replace the words with their synonyms, and also carry out the correct search in an instance of electronic dictionary accessible via the interface. Our ITS can therefore evaluate different skills and types of knowledge that the user has acquired or not, and use this knowledge to propose different activities to address missing concepts or skills.

To accompany the authentic learning scenarios, STI-DICO also proposes explanations of fundamental concepts that are needed for dictionary use. For instance, if a user is not capable of completing the activity described above, it is possible that he is not familiar with the concept of synonymy. If this is the case, STI-DICO diagnoses this knowledge gap and proposes an explanation of the concept of synonym (Figure 4), accompanied by examples and presents the learner with a series of more theoretical exercises (for instance, drag and drop exercises to match words with their synonyms) in order to explore this concept in more depth. STI-DICO will also highlight the links between concepts (e.g. synonymy and antonymy are both types of lexical relations) in order to help the learner build their concept hierarchy.

A third type of exercise proposed in STI-DICO are those that aim to develop learners' dictionary-specific skills, for instance the usage of wildcards to search in electronic dictionaries (e.g. table* for all words that start with table), or interpreting abbreviations within a dictionary entry (e.g. 'n.f.' for nom féminin, feminine noun in French). These skills, while extremely important for successful dictionary usage, do not depend on any fundamental notions and are often dictionary-specific. We chose the most popular electronic dictionary in Quebec, Antidote, and limit the activities to its usage, while ensuring that the students acquire an in-depth knowledge of its functioning after having completed STI-DICO's activities (Figure 5).

6. ONGOING AND FUTURE WORK

The work on STI-DICO is in full swing, with an ongoing collaboration with the CTAT team (Aleven et al., 2015) in order to seamlessly link the CTAT reasoning motor with OpenEdX and the HTML5 STI-DICO interface. We believe that this integration is important because it will permit for STI-DICO to be light and platform-independent. In parallel to this, we are working on integrating the GNT ontology (Polguère, 2010) with our referential of dictionary skills and knowledge and indexing these two components with the learning activities in STI-DICO. This will permit the system to carry out cognitive diagnosis of a learner's skills and knowledge based on their behavior in the system's activities.
Our plans for the near future include several components: 1) Completing the interface of STI-DICO in HTML5 and OpenEdX; 2) Converting the existing exercises and course modules in lexicology (Tremblay, 2009) to populate our learning object database; 3) Think-out-loud experimentation with experts in lexicology in order to validate the skills and activities chosen; 4) Evaluation of STI-DICO with future primary school French teachers completing their diploma at our university, 5) Analysis of the results of the evaluation and improvement of the system.

We are also planning a project in partnership with INRIA Sophia-Antipolis, to ensure that the learning objects that we develop are compliant with the EEE Standard for Learning Object Metadata (P1484.12). We will therefore be able to share our activities with knowledge object pools such as ARIADNE (http://www.ariadne-eu.org/) and GLOBE (http://www.globinfo.org/). We believe that this is an important step for the future development of intelligent tutoring systems, which need to follow recent trends in computing, such as service-oriented architectures, distributed and ubiquitous computing, and the pooling of resources to ease their entrance to mainstream educational contexts (Nye, 2015).

From an architectural point of view the main challenge will be the scalability of the solution. With very large numbers of users, a server-based tutor engine can be faced with severe server load. To tackle this issue in their CTAT-edX experiment, Aleven and colleagues developed a JavaScript version of the tutor's inner loop in order to distribute a large part of the computational workload to the student's computer. A long-term solution for the scalability problem is to distribute the workload of the CTAT/TutorShop server from one unique server to many identical servers behind a load balancer, which will allow horizontal scalability. While this solution would require some significant software re-engineering, it is feasible and stable. We have also developed a contingency plan involving the usage of the JS Input mechanism if ever LTI integration becomes an issue.

7. CONCLUSION

A work-in-progress research project of an intelligent tutoring system has been described, using a service-oriented architecture and integrating an LMS platform with activities developed in HTML5. This system, STI-DICO, enabling the learning of dictionary skills, will eventually be offered as an online course module for primary school French tutors studying at our university. Based on existing components, such as an ontology of linguistic concepts (Polguère, 2010), course modules in lexical didactics (Tremblay, 2009) and empirical research on dictionary use (Nesi and Haill, 2002; Lew, 2013), the resulting ITS will be a both fundamentally sound and fully functional in terms of learner cognitive diagnosis and learner interaction.

Difficulties and limits of our project include integrating the various components of the system and assuring their communication as well as the addition of functional dictionary excerpts to permit users to consult them during their learning process. Furthermore, while we will save all learning data collected via our experimentation, we have not yet explored the potential of the project in terms of learning analytics and educational data mining. We plan to integrate this dimension in further work on STI-DICO.

We believe in the success of our project due to the sound cognitive foundation of our system, the positive results regarding the impact of teaching of dictionary skills explicitly (Bishop 2001) as well as the usage of ITS in classrooms (Mark et Greer 1993, Van Lehne et al. 2005, Roscoe et al. 2013). We are also following recent trends of integrating ITS with LMS (Aleven et al., 2015) and using service-oriented architectures and web-based interfaces (Nye, 2015). Evaluation results are expected in mid-2016 and will enable us to improve our system and its functionalities and assure its pertinence for the target audience.

ACKNOWLEDGMENTS

Thanks to the GIFT and CTAT teams, especially Vincent Aleven, Octav Popescu and Jonathan Sewall from CMU and Keith Brawner from the GIFT team, as well as Johan Kayser for work on the initial STI-DICO prototype.

REFERENCES


927


