Improving the Efficiency of Automatic Knowledge Generation through Games and Simulations

Mark Floryan[1], Beverly Park Woolf[1]

[1] Department of Computer Science, University of Massachusetts, Amherst
140 Governors Dr. Amherst, MA USA
{mfloryan, bev} @ cs.umass.edu

Abstract. We have created a generalized algorithm for automatically constructing domain level knowledge bases from student input. This method has demonstrated greater efficiencies than when knowledge is hand crafted by subject matter experts (SMEs). This paper presents two related methods for improving automated knowledge acquisition by leveraging the properties of games and simulations. First, we discuss game mechanics that, when added to our intelligent tutor Rashi, lead to higher quantity and quality of student input. In a separate but related analysis, we present a novel game type called a knowledge refinement game (KRG) to improve the knowledge in an expert knowledge base. This game motivates SMEs to refine the generated knowledge base, especially for data in which the system has low confidence. Utilizing an anonymous agreement policy ensures the quality of SME responses and results show that small amounts of KRG activity leads to noticeable improvements in the quality of the knowledge base. We assert that these two results in unison provide evidence that gaming has a powerful potential role in improving artificial intelligence techniques for education.

Keywords: expert knowledge bases, serious games, game mechanics, ill-defined domains, increased student input.

1 Introduction

Research in educational software is often focused on development of expert knowledge bases that support intelligent algorithms [2], which in turn are intended to improve a student's experience with intelligent tutoring systems (ITS) by offering customized interactions. In contrast, other research communities apply gaming and simulation mechanics to increase the efficacy of educational software, often by increasing user engagement and motivation [1][7][9][22]. This paper describes research to incorporate lessons from game design to optimize a tutor's ability to automatically learn domain level knowledge. In particular, we have developed a set of algorithms that examine student actions while using an intelligent tutor and use this student input to construct an expert knowledge base (for details on this process see [6]). This approach produces relatively small but precise domain models that are useful for generating automatic feedback to students [3].

The focus of this paper is on applying game mechanics to improve this automated knowledge generation process. We have automatically created a domain knowledge base with greater than 70 percent precision [6] in about 300 student work hours. This paper describes how to improve these precision numbers by incorporating games and simulations into the knowledge generation process. We present two experiments: one focused on *reducing the time* required to automatically build an expert knowledge base and the second on *increasing the quality* of the resulting expert knowledge base.

Our first effort involves incorporating game mechanics into tutors to increase the quantity and quality of student work. If this is accomplished, then we predict that our automatic knowledge generation process can be achieved with greater efficiency because students are contributing more data per hour, and the automatic knowledge acquisition process is directly dependent on the quantity and quality of student data provided. Several researchers have explored the impact of games on digital tutors and many have determined that student engagement and motivation increases [7][9][22].

We also present techniques for *improving the quality* of generated domain models. To this end, we present a novel type of game that invites subject matter experts (SMEs) to correct and vet existing nodes and arcs in the expert knowledge base. We incorporated game mechanics from the nascent field of “games with a purpose” [12]. For example, the ESP Game [13] uses crowd sourcing to support developers’ need to collect large amounts of labels for images to improve computer vision algorithms.

This paper provides background information, methods, and results for two experiments. We describe our core tutor and its features in Section 2 and describe the methods and results of an experiment to *reduce the time* needed to develop expert models in Section 3. Specifically we observed that an increase in student motivation led to more student work. Section 4 describes the game to *improve the quality of the expert knowledge base*, while Section 5 describes the results of having SMEs use this game. Section 6 closes with an analysis and discussion of future work.

2 Reducing Knowledge Acquisition Time

This section describes design decisions for game mechanics added in part to improve our tutor’s ability to conduct automatic knowledge acquisition. We also examine the impact of these game features on the quantity of student work.

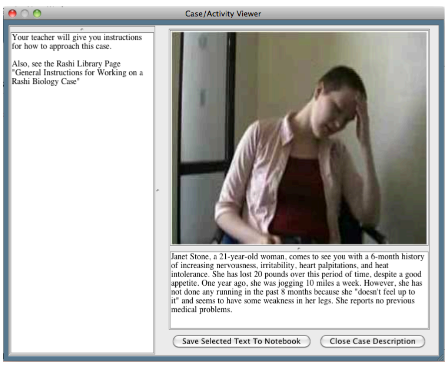

Figure 1: Case Description. Students question the virtual patient by using the Interview Environment.

This research was conducted within Rashi, an intelligent inquiry tutor used by thousands of students over the past five years to learn human anatomy. In the *Human Biology Tutor,* students evaluate virtual patients and generate hypotheses about their medical condition. The tutor provides case descriptions for students to investigate, along with information about how to approach each problem [4]. Cases are presented as open-ended environments for student exploration and to acquaint students with methods commonly used by professionals in the domain. The basic Rashi tutor is domain independent, but this paper focuses primarily on the *Human Biology* domain.

The system contains two major components: a *procedural core* that supplies data collection mechanisms (e.g., interactive images, interview interfaces, video and dynamic maps) and a *content knowledge base* (e.g., an expert system) with knowledge about individual cases.

We created a methodology for constructing domain-level expert knowledge bases for Rashi automatically through crowdsourcing [6]. This approach involved collecting and analyzing the work of numerous students working within Rashi and using an intelligent algorithm to coalesce data from those student efforts to construct the domain model. We compared the knowledge created in a human crafted expert knowledge base (HEKB) with that resulting from our automated construction of the expert system to judge its quality and found that our algorithm does well and that the evolving expert knowledge base (EEKB) models can be generated in significantly less time [6]. However, we still need to optimize the quality of this generated model (we measured between 70 and 80 percent precision on average) and to further decrease the necessary build time (currently measured around 300 student work hours).

Game Mechanism to Reduce Knowledge Acquisition Time

We added game mechanics to Rashi to optimize the fantasy, urgency, and sense of reward [10]. We added a patient status panel, located within the main Rashi window that provides students with an easy way to monitor, in real time, the condition of the patient, see Figure 2. The patient status panel consists of three parts:

> ***Patient Character:*** *This animated character shows a visual representation of the patient, including a few emotions that are loosely correlated with his or her health.*
>
> ***Health Bar:*** *The health bar displays a quantifiable view of the patient's health. The bar is color coded to display healthy (green), sick (yellow), or critical (red) conditions. This health bar is dynamically updated depending on the patient's current condition and any currently applied treatment.*
>
> ***Treatment:*** *This panel displays a text representation of the condition for which proper treatment is currently being administered. Thus, once students set the treatment for the patient, they can observe both the treatment and its effects in unison.*

Students are also provided with a new *treatment button*. Upon pressing this button, the patient status panel is immediately updated to reflect that the patient is being treated for a condition. The treatment selected directly affects the "health" of the patient, reflected by the patient status panel. The effects of the treatment are revealed slowly, and the result of treatment is not monotonic. While students are waiting for the results of their treatment to become apparent, they are encouraged to explore other potential diagnoses.

These features are game mechanics primarily because they provide a concrete goal (making the patient well), incorporate urgency (fear the patient may be lost despite the fact that this is impossible in Rashi) and fantasy (the feeling of achieving victory when the patient is treated correctly) [10]. Previous Rashi versions asked students to construct arguments for diagnosis, but treatment or even formal diagnosis was not a part of the system. We posit that this sense of responsibility encourages students to be motivated differently than when these features are absent.

Methods and Results to Reduce Knowledge Acquisition Time

Our goal was to identify whether game mechanics led students to produce higher quantity and quality data in Rashi. To study this possibility, we selected two data sets that differed only in the presence of the game mechanics described above. The selected data sets were both from a large rural university class in Biology 101 in 2011 (no game features) and 2012 (game features) and were equivalent in virtually every other respect. Students were taught by the same teachers, used similar pedagogies, and evidenced a similar caliber of introductory biology knowledge. We first aimed to show that the 2012 class produced more Rashi data than did the 2011 class. We compared work produced by students in each class (Table 1). Student work includes generating hypotheses (e.g., "Patient has hypothyroidism") and relations between hypotheses and evidence (e.g., "Elevated TSH supports patient has hypothyroidism.").

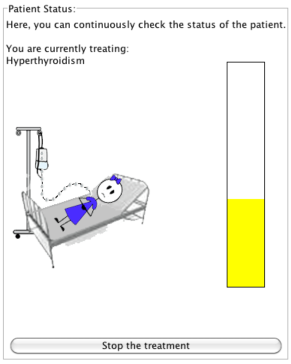

Figure 2: A cartoon patient is presented to Rashi users along with a health bar.

We see that students who used the game mechanics contributed more work per student than did the control group. Namely, we see a 59.5 percent increase in the raw amount of data contributed by students who used Rashi with additional game mechanics. Our previous work found that we needed 300 student work hours to create our EEKB, thus it appears that adding game mechanics to Rashi has the potential to decrease the number of student hours necessary to build our expert system automatically by up to 59 percent. This represents a vast improvement in efficiency.

Year	*Num Students*	*Num Data / Relations*	*Num Hypotheses*	*Total*	*Contributions / Student*
2011 (no game)	396	4342	2111	6453	**16.295**
2012 (game)	539	9328	4683	14011	**25.994**

Table 1. Amount of work contributed by students in similar university courses with and without game mechanics.

Additionally, we wish to confirm that the quality of this additional work is not less than that of its counterpart. Therefore, we developed an estimated argument quality metric for judging the strength of student created arguments. Because Rashi teaches in ill-defined domains, we cannot strictly judge the quality of work, but can make strong estimations based on several factors. The formula for estimated work quality is:

$$\text{Grade} = [\text{Correct(H)} / |H|]*W_1 + [\text{Correct(R)} / |R|]*W_2 + [|R| / |H|]*W_3 + [|H|]*W_4$$

Where:

H = the set of student hypotheses

R = the set of student relations

Correct: a function that returns the number of items in the input set that match to the expert knowledge base.

W = A weight ($0<=W_i<=1$) for each term of the grade. $W_1+W_2+W_3+W_4 = 1.0$

We estimated the quality of each student's argument using the metric above. Figure 3 provides a summary of the difference between each group. We see that the student's provided with Rashi game mechanics actually contributed higher quality work, in addition to the increased quantity reported above.

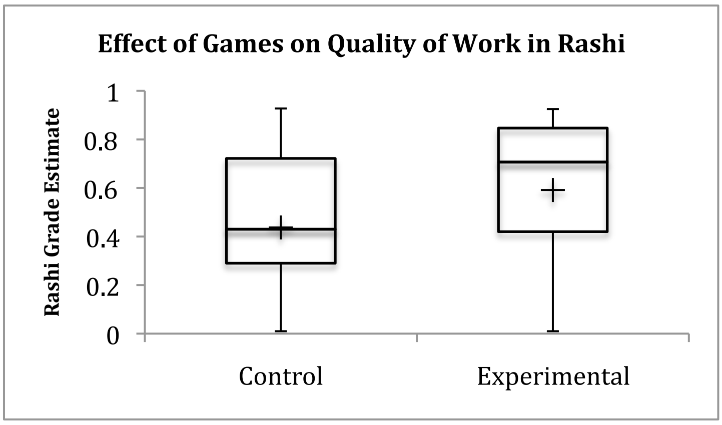

*Figure 3: Estimated student grades in Rashi across groups that contained and did not contain the new Rashi game mechanics *(p < 0.01).*

3 Improving Knowledge Acquisition Quality

Once an expert model is constructed automatically (as described in [6]), it is in our interest to examine its contents and modify the data in places where the knowledge base might have low confidence in the data. In this section, we present our design for a knowledge refinement game (KRG) we call "Dr. Doctor". We define a knowledge refinement game as any game that incorporates game mechanics and whose purpose is to alter the underlying structure of a data model.

The knowledge refinement game performs three major non-trivial tasks. First it analyzes the evolving knowledge base to identify areas of low confidence (improvement detection). Then, it represents expert system knowledge as questions (question generation). Thirdly, it updates the expert system with the knowledge provided in the SME's response (model update). Every EEKB entry (node or relation) has a confidence property with default value of 10% and this confidence rises linearly with every successive student who provides evidence of this entry.

The three step process is repeated for as long as the expert wishes to continue (Figure 4). SMEs are given a score and a level depending on the amount and quality of their contributions.

Figure 4: Screenshot of 'Dr. Doctor', a Knowledge-Refinement Game that accepts input from players and improves the quality of the Rashi evolving knowledge base.

The game creates varying types of questions in each phase as described below. The questions are completed for each phase before moving on to questions in the next phase because of dependent relationships that exist between question type and aspects of the knowledge that are not vetted beforehand.

> ***Phase 1. Verify Nodes:*** *Nodes in the expert knowledge base that have a relatively low confidence are retrieved so the SME can confirm whether or not they belong in the EEKB. An example questions of this form: "Is this hypothesis valid for this domain: <Patient is pregnant>?"*

> ***Phase 2. Verify Relationships:*** *Once the Nodes are verified, the same is done for the relationships between those nodes. The relationships are presented to the SME and the responses are reflected in updated confidence values. For example, "Reduced TSH" strongly supports "Patient has hypothyroidism"*

> ***Phase 3. Combine Similar Nodes:*** *The game uses string search techniques to estimate whether pairs of nodes reflect the same semantic data. The SMEs are presented with such pairs and asked if they are indeed the same topic or idea, and whether they should be combined in the EEKB.*

Dialogue between SME and Refinement Game

After asking the SME a particular question and receiving the response, the system dynamically adjusts the evolving knowledge base to reflect this information, on the assumption that an expert response is generally accurate. The game phrases its queries to require only yes / no responses, and so the system must only deal with the responses from this constrained interaction.

Table 2 summarizes the question types, responses, and reactions within the knowledge refinement game. We see that the expert responses directly impact the probabilistic confidence values of the EEKB data.

Phase	Question Type	SME Answer	Response of KRG System
1	Node Verify	*YES*	Increase node confidence
		NO	Decrease node confidence
2	Relation Clarify	*YES*	Increase relation confidence
		NO	Decrease relation confidence
3	Combine Nodes	*YES*	Combine nodes into single node
		NO	Don't ask about these nodes again

Table 2: A simplified summary of the Knowledge-Refinement Game's question types, potential expert answers, and responses made by the game.

In the next section, we discuss the game elements woven into 'Dr. Doctor,' and argue that these game mechanics are essential in promoting accurate SME feedback.

Game Mechanics in the KRG

Dr. Doctor incorporates game mechanics in order to motivate SMEs to work longer and to make the experience more enjoyable [10]. In particular, these mechanics are designed to accomplish two goals: motivate SMEs to continue contributing for extended periods of time and offer incentives for SMEs to provide accurate information [12].

Feedback Statistics: *Dr. Doctor displays dynamically changing statistics regarding experts' contributions, Figure 3, top left. This includes the number of all time contributions these SMEs have made to the expert system, the confidence of the EEKB, and the percent increases for which these players are responsible.*

Points: *Players are awarded points for answering questions, Figure 3, top right. This provides rewards to SMEs for their contributions and motivates them to continue*

Levels: *SMEs progress through increasing levels as they garner points, e.g., undergraduate, graduate, professor, and players are given a higher status at each increasing level.*

Agreement Bonus: *Players are rewarded with a score bonus when their answers are in agreement with other SMEs.*

An anonymous agreement policy is implemented through the agreement bonus described above. This policy helps ensure quality input by forcing a first order optimal strategy of agreeing with fellow SMEs. Since other player's identities are anonymous, the optimal strategy then becomes to input truthful responses. In addition, the application does not update the knowledge base permanently unless multiple SMEs have agreed on the validity of a change.

Methods and Results to Increase Knowledge Acquisition Quality

To evaluate the knowledge refinement game, we posted the application on the web, making it accessible on demand. We then recruited three teaching assistants from the biology department of a large rural university who agreed to play the game at their leisure over the course of one week. We asked that participants contribute at least 100 responses within the game. Two of the participants were upper-level undergraduates, while the other was a graduate student. Additionally, in order to contextualize the game features of 'Dr. Doctor', we offered a prize (gift card) to the participant who achieved the highest score within the game.

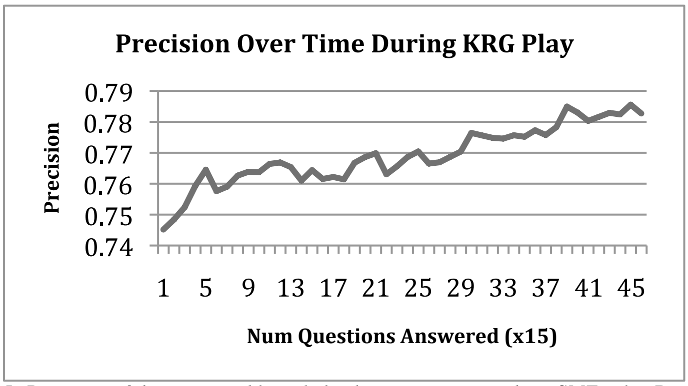

Figure 5: Precision of the generated knowledge base over time as three SMEs play Dr. Doctor

After one week of the game being available, two of the three participants contributed more responses than asked for, answering 250 and 235 questions respectively. The third participant answered just above the minimum, logging 105 questions answered. We analyzed how the quality of the evolving expert knowledge base changed over time by saving the state of the knowledge after every 15 inputs. For every snapshot of the knowledge base, we judged quality using two metrics. The first was *precision*, which is calculated as the percentage of the generated knowledge that is in agreement with knowledge created by a human expert. More information on precision can be found in [6].

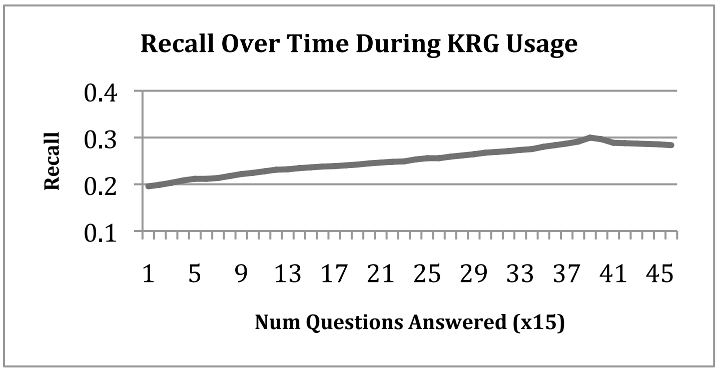

Figure 6: Recall (breadth of knowledge acquired) over time as SMEs played Dr. Doctor.

We found that as students played our game, the precision of the evolving knowledge rose. We observed a four percent increase in precision (Figure 5). We also measured knowledge *recall* or the breadth of knowledge acquired through our automated knowledge generation process, see [6] for more information. A display of the change in recall over time can be seen in Figure 6. We observed a nine percent increase in knowledge recall as our participants played Dr. Doctor.

4 Conclusions and Future Work

In conclusion, this paper presented two experiments that highlight opportunities for applying games to benefit automatic knowledge acquisition within intelligent tutoring systems. We first presented game mechanics that led to higher quantities and quality of student input within an inquiry tutor. Because students contribute more data when presented with game mechanics, automatic knowledge acquisition [6] can be accomplished more efficiently.

Games can also be applied to help optimize the quality of automatically generated knowledge. We presented a novel game called a knowledge refinement game (KRG) that motivates SMEs to judge the quality of generated knowledge. Although the game was designed for SME players, we tested the game with upper class undergraduate and graduate students. We observe that a small amount of KRG game play leads to a four percent increase in the quality of the knowledge base, and a nine percent increase in the breadth of acquired knowledge. We also observe that two of our three participants contributed more than twice the amount of data requested.

Although we present our results on a single example tutor, we believe that these approaches can generalize to an array of tutors that utilize various expert models. It is easy to see that incorporating game approaches can be beneficial for a variety of tutors. In addition to this, knowledge refinement games can be applied to many models. In particular, the KRG requires three essential steps. First the game must be able to identify and locate areas in an expert model that require updating. Then the game must be able to construct a question from this identification, and lastly be able to update the model appropriately once an answer is provided. Any tutor / model that can accommodate these three stages is capable of benefitting from this approach.

Our future work in this area will focus on extending the process of automatic knowledge acquisition and refinement, both to domains outside of Human Biology as well as to additional iterations of the knowledge building process within Rashi. We hope to provide further evidence that large amounts of student data can be obtained automatically from students within tutors, and that knowledge can be efficiently refined and improved by players of knowledge refinement games. We also hope to provide evidence that knowledge refinement games can be effective when played by non-subject matter experts, in an attempt to widen their potential usage and application. Lastly, we wish to explore more deeply the necessity and benefit of game mechanics on the successful usage of the KRG. We believe that the game mechanics had a direct effect on the success of our experiment, but wish to confirm these beliefs experimentally.

Acknowledgements

This research was funded by an award from the National Science Foundation, NSF 0632769, IIS CSE, Effective Collaborative Role-playing Environments, (PI) Beverly Woolf, with Merle Bruno and Daniel Suthers. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

1. Bader-Natal, Ari *"Incorporating Game Mechanics into a Network of Online Study Groups"*. In Proceedings of the Workshop on Intelligent Educational Games 2009. Brighton, England.
2. Dragon, T. *"The Impact of Integrated Coaching and Collaboration within an Inquiry Learning Environment"*. Doctoral Dissertation. University of Massachusetts, Amherst (2013).
3. Dragon, T., Floryan, M., Woolf, B.P., Murray, T. *"Recognizing Dialogue Content in Student Collaborative Conversation."* In Proceedings of the 10th International Conference for Intelligent Tutoring Systems, ITS 2010. Pittsburgh, PA (2010).
4. Dragon, T., Woolf, B. P., Marshall, D. and Murray, T. *"Coaching within a domain independent inquiry environment."* Proceedings of the Fifth International Conference on Intelligent Tutoring Systems. Jhongli, Taiwan, Springer 4053, 144-153 (2006)
5. Floryan, M.; Woolf. B.P. "*Rashi Game: Towards an Effective Educational 3D Gaming Experience*". Proceedings of the IEEE International Conference on Advanced Learning Technologies. Athens, GA (2011)
6. Floryan, M.; Woolf, B.P. *"Authoring Expert Knowledge Bases for Intelligent Tutors through Crowdsourcing"* International Conference on Artificial Intelligence in Education (2013)
7. Hallinen, Nicole; Walker, Erin; Wylie, Ruth; Ogan, Amy; Jones, Christopher. "*I Was Playing When I Learned: A Narrative Game for French Aspectual Distinctions"*. In Proceedings of the Workshop on Intelligent Educational Games 2009. Brighton, England.
8. Lynch, C. Ashley, K. D. Pinkwart, N., Aleven, V. "*Concepts, Structures, and Goals: Redefining Ill-Definedness"* In Aleven, V., Lynch, C. Pinkwart, N., Ashley, K. (Eds). *International Journal of AI in Education; Special Issue on Ill-Defined Domains*. Volume 19. Number 3. pp. 253-266, (2009)
9. McAlinden, Ryan; Gordon, Andrew S. Lane, H. Chad; Pynadath, David. "*UrbanSim: A Game-based Simulation for Counterinsurgency and Stability-focused Operations"*. In Proceedings of the Workshop on Intelligent Educational Games 2009. Brighton, England.
10. Prensky, Marc; *"Digital Game-Based Learning"*. McGraw-Hill, 2001.
11. Rowe, Jonathan P. Mott, Bradford W. McQuiggan, Scott W. et al... "*Crystal Island: A Narrative-Centered Learning Environment for Eighth Grade Microbiology"*. In Proceedings of the Workshop on Intelligent Educational Games 2009. Brighton, England.
12. Siorpaes, Katharina; Hepp, Martin. *Games with a purpose for the semantic web*. IEEE Intelligent Systems, 23(3):50–60, 2008.
13. Von Ahn, L. and L. Dabbish, "Labeling Images with a Computer Game," *Proc. ACM CHI*, 2004.