

# Using Graph Edit Distance to Diagnose Student's Science Process Skill in Physics

Ming-Xiang Fan<sup>1</sup>, Maiga Chang<sup>2</sup>, Rita Kuo<sup>3</sup>, and Jia-Sheng Heh<sup>1</sup>

<sup>1</sup> Dept. of Information and Computer Engineering, Chung-Yuan Christian Univ., Taiwan

<sup>2</sup> School of Computing and Information Systems, Athabasca University, Canada

<sup>3</sup> Dept. Of Digital Design, Mingdao Univ., Taiwan

fs601254@yahoo.com.tw, maiga@ms2.hinet.net,  
ritakuo924@gmail.com, jsheh@ice.cycu.edu.tw

**Abstract.** Science process skill is important to students when they are learning sciences. We have built a story-based virtual experiment environment in Physics to train student's science process skills. However, the teacher currently doesn't know how the students' science process skills are while they are using the virtual experiment environment. For this reason, this research tries to apply the graph edit distances to design the graph-based diagnosis methodology and uses the methodology to compare the student's science process skill graph and the expected graph in order to give both the teacher and the student feedback regarding the student's science process skills. This research also proposes the experiment design and plans to do the experiment in the end of this May (May, 2010) to prove the effectiveness of the proposed graph-based diagnosis methodology in analyzing students' science process skills in Physics.

**Keywords:** Science Process Skill, Virtual Experiment Environment, Story-based Virtual Experiment, Physics, Problem Solving, Graph edit distance, Diagnosis.

## 1 Introduction

There are many researches use the virtual experiment environment to teach Science topics [4][5]. We have built a story-based virtual experiment environment in Physics to train students' science process skills. However, the most important concern the teacher may have is how students' science process skills change after they used the system.

For this reason, this research wants to diagnose a student's science process skills after s/he completed the virtual experiment and to give the teacher feedback with descriptive explanations of the student's skills and quantitative score as the measurement of the skills.

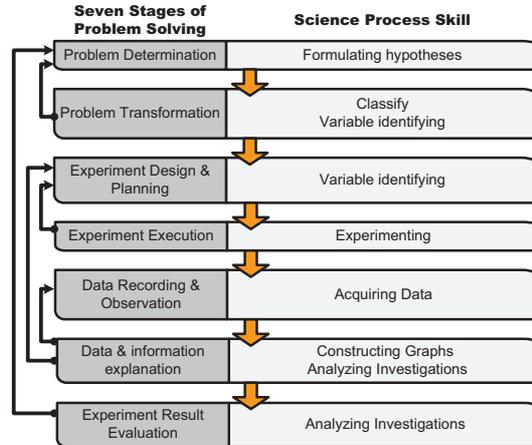
The virtual experiment environment we have built combines the science process skills and seven stages of problem solving [17]. Each problem solving stage has been mapped to different stages of doing virtual experiment. The student's manipulation behavior at each stage then can be considered as his/her performance in specific science process skill. This research transforms the student's manipulation behavior at

each stage to a graph, and then uses the revised graph edit distance methodology to generate diagnosis result to the student and the teacher.

Section 2 describes relevant research which we use to develop the virtual experiment environment and to design the graph-based diagnosis methodology, e.g. science process skills, seven stage of problem solving, and graph edit distance. Section 3 introduces the virtual experiment environment we have built. Section 4 explains the methodology of using graph edit distance to diagnose the student’s science process skill. In Section 5, we talks about the experiment design we are going to do in this Summer. At the end, Section 6 discusses possible future works.

## 2 Research Background

This research designs a graph-based diagnosis methodology to diagnose students’ science process skills when they do Physics experiment in the virtual experiment environment. The virtual experiment environment [7][16] allows students doing experiments by using virtual equipments [4]. Experiment is a process of solving problem. Researchers, e.g. Polya (1957) and Mitchell and Kowalik (1989), have proposed different problem solving strategies [8][11]. Polya’s problem solving strategy involves four stages: understanding the problem, devising a plan, carrying out the plan, and looking back. Mitchell and Kowalik’s problem solving strategy has six stages: mess finding, data finding, problem finding, idea finding, solution finding, and acceptance finding.



**Fig. 1.** Revised relations between science process skills and the seven stages of problem solving based on Kuo et al. (2000)

The seven stages of problem solving can lead student to solve science problem while doing experiment [1]. Kuo and her colleagues (2000) have analyzed the relations between the seven stages of problem solving and the science process skills as Fig. 1 shows [5]. Students who follow the stages one by one to do Physics experiments can practice different science process skills at different stage. If they think they

have mistakes at one stage, they could go back to the previous stages. Each stage associates with different science process skills listed at the right hand side of Fig. 1.

The story-based virtual experiment environment we have created can automatically generate different virtual experiment from a knowledge structure. Many researchers discuss how to construct the knowledge structure according to human's mental model. For examples, Concept Map [10] is widely used by psychologists and educators; Knowledge Map [6] is also widely used in the e-learning field to store the learning contents. This research uses the Context-Aware Knowledge Structure which is proposed by Wu et al. (2008) to store the knowledge for generating virtual experiment. The context-aware knowledge structure [15] has three layers:

1. **Domain layer:** denotes the subjects and the topics of the learning environment.
2. **Characteristic layer:** stores the characteristics of the subjects and the topics.
3. **Object layer:** represents the real learning objects.

As long as this research wants to design a graph-based diagnosis methodology, we need to have some definitions and symbols to represent graphs first. Graph is a useful data structure which can be used to represent various objects and concepts. Using graph can transform the behavior diagnosis problem to a graph matching problem. Edit distance is a method which can measure the difference between two data structures suchlike strings [14], trees [13] and graph [12].

Three operations in graph edit distance are insertion, deletion, and relabeling for both nodes and edges. The error-correcting graph matching [2] between two graph  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  has six graph edit operations for changing  $G_1$  to  $G_2$ : substitute a node in  $G_1$ , delete a node from  $G_1$ , insert a node into  $G_1$ , delete an edge from  $G_1$ , insert an edge into  $G_1$ , and substitute an edge in  $G_1$ .

This section summarizes how a virtual experiment environment can be used as a platform to train students' science process skills; how we make computer generate virtual experiment automatically; and, how two graphs can be compared. Before we design the graph-based science process skill diagnosis methodology, Section 3 first reveals the virtual experiment environment and introduces how the student's science process skills are recorded for further diagnosis.

### 3 Virtual Experiment Environment

We have designed a Story-based Context-aware Knowledge Structure and a revised seven stages of problem solving to generate the virtual experiment automatically. The virtual experiment environment is a web-based application. Both students and teachers use web browser such as Microsoft Internet Explorer and Firefox to access the application. Teachers can build their own knowledge structure for specific science course and access their students' learning progress. Students will be asked to do virtual experiment in the virtual experiment environment.

In the rest of this section, we are going to use a real scenario to explain how the virtual experiment environment works and how the environment trains the student's science process skills. Ms. Cool is a Physics teacher and wants to train her students' science process skills in Dynamics. She leads the students to the computer lab and

asks them to do Dynamics relevant virtual experiments in the virtual experiment environment.

Alex is a student in Ms. Cool's Physics class. When he gets into the virtual experiment environment, the system first shows him a story book and asks him to choose the story described in the book he interests. Alex chooses the *cliff* story that he thinks it should be fun. After Alex chose the story, the system asks him to think about the cause and the effect of the hypothesis for the story-based problem as Fig. 2 shows, i.e. formulating hypotheses skill, one of science process skill in Fig. 1. He chooses "to make people's mass increase will let the stone's height increase" as the cause and the effect of the hypothesis.

The screenshot shows a web-based interface titled "STORY SCENE QUESTION". The main text asks "how can we solve the problem in the story scene?". Below this, there are two rows of dropdown menus. The first row is labeled "make" and has "people" and "mass" selected. The second row is labeled "increase". The third row is labeled "will let" and has "stone" and "height" selected. The fourth row is labeled "increase". Below these dropdowns, the text "to solve this problem" is displayed. At the bottom of the interface is a button labeled "make hypothesis".

**Fig. 2.** The cause and the effect of the hypothesis for the story-based problem

With the cause and the effect of the hypothesis Alex have chosen, he then needs to choose the relevant animation to the story-based problem, i.e. classify skill. When Alex makes his choice at this stage, he will be asked to watch different animation clips, and he finally picks the *bullet shoot to block* animation up.

After he made choice and picked-up an animation he thought that it may be relevant to the story-based problem, the system asks him to do object mapping between the problem and the animation, i.e. classify skill. Alex maps the people to the bullet and the stone to the block, under such circumstance, the stone and the people are the story objects and the bullet and the block are the animation objects. After Alex figured out the object mappings, he needs to pick-up the Physics quantity that he should observe in the following experiment in order to solve the story-based problem for each object, i.e. variable identifying skill. Alex thinks the bullet's mass and velocity and the block's mass, velocity, and height are important Physics quantities and worth to observe for solving the problem.

After Alex chose all the Physics quantities for corresponding object, the system asks him to categorize these quantities into the manipulated variable, the responding variable and the control variable(s), i.e. variable identifying skill. Alex identifies the bullet's mass as the manipulated variable, the block's height as the responding variable, and he thinks that other object quantities belong to the control variables. Alex then watches the animation with controller, i.e. play, rewind, and pause, to observe and to record these Physics quantity values. The system will generate relevant charts with these records and Alex can take a look at the charts and think about the relations

among different objects' Physics quantities, i.e. constructing graphs skill and analyzing investigations skill.

At last, Alex needs to evaluate his hypothesis to see if the hypothesis and the chosen animation can solve the story-based problem for him, i.e. analyzing investigations skill. He can repeat these stages to solve other story-based problems.

This section describes the virtual experiment environment and uses a practical scenario to show how the virtual experiment environment works in training students' science process skills. Section 4 talks the design of the graph-based diagnosis methodology, this research uses the methodology to compare the student's virtual experiment manipulation behavior with the expected one and to give both the teacher and the student the diagnosis results as feedback.

### 4 Diagnosis Method

In this research, we use the graph edit distance to analyze students' manipulation behaviors in the virtual experiment environment and diagnose students' science process skills. For this purpose, a student's virtual experiment manipulation behavior and the correct one are needed to transform to graphs first. The graph transformation method treats the options at every stage of a virtual experiment as the graph's nodes, and considers the student's manipulations and choices as the graph's edges.

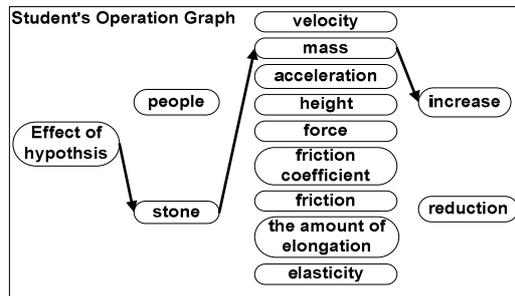


Fig. 3. Student's manipulation graph of choosing the hypothesis' effect

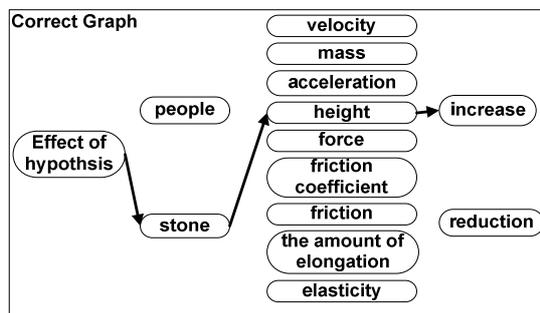


Fig. 4. The graph to represent the correct hypothesis' effect choice

For example, Fig. 2 is a virtual experiment stage which asks the student to choose the hypothesis for specific Physics problem. In this example, the student needs to choose the cause and the effect of the hypothesis. The cause and the effect options are physical objects, physical quantity, and the state that this research uses it to present whether the quantity value is going up or down. These options will be transformed to the nodes by the graph transformation method. If the effect a student chose is that the stone's mass will increase, the graph nodes, i.e. stone (physical object), velocity (physical quantity), and increase (the state), will be connected as Fig. 3 shows. Fig. 4 shows the correct options of the hypothesis.

In Fig. 4, the most left node is the root node, is using to distinguish which stage the graph represents. After the student's graph and the correct graph are transformed, the distance of the two graphs can be used to measure the difference between these two graphs. This research uses the graph edit distance to compute the distance. Different from abovementioned graph edit distance research, this research doesn't take node costs into consideration due to the two graphs in this research have identical node sets. Also, the cost of substituting an edge does not take into consideration due to all edges are undirected and there is no different meaning attached on the edges, i.e. all edges are same. This research considers only if two nodes are connected.

This research defines node level from right to left, which means the rightmost nodes are at level 1. Symbols that this research uses to describe the graph and to design the diagnosis methodology are: (1) uses  $node_{(k,i)}$  to represent a node in the graph, where  $k$  is the level the node belongs to and  $i$  indicates the node is the  $i$ -th node; and, (2) uses  $e_{\{node_{(k,j)}, node_{((k-1),j)}\}}$  to represent the edge between the  $node_{(k,i)}$  and  $node_{(k-1,i)}$ .

The two costs that this research considers are

$cost_{edge\_delete}(e_{\{node_{(k,j)}, node_{((k-1),j)}\}})$ : the cost of deleting an edge between two nodes from adjacent two levels.

$cost_{edge\_insert}(e_{\{node_{(k,j)}, node_{((k-1),j)}\}})$ : the cost of inserting an edge between two nodes from adjacent two levels.  $\square$

For instance, the cost of changing Fig. 3 to Fig. 4 will be

$$cost_{all} = cost_{edge\_delete}(e_{\{stone, mass\}}) + cost_{edge\_insert}(e_{\{stone, height\}}) + cost_{edge\_delete}(e_{\{mass, increase\}}) + cost_{edge\_insert}(e_{\{height, increase\}}),$$

in order to delete the edge between *stone* and *mass* and the edge between *mass* and *increase*, and to insert the edge between *stone* and *height* and the edge between *height* and *increase*.

In addition to the cost, every node in the graph in this research has level attributes. The left nodes have higher level values, because the left-hand-side nodes and its edges represent the precedent actions and/or choices that the student has made in doing the virtual experiment. So the right-hand-side nodes can't be connected without its parent nodes, i.e. the left-hand-side nodes, being connected first. For science process skill diagnosis, the higher level a node has, the more important the node is.

This research thus defines the level weight of a node as  $weight_{level}(node_{(k,j)})$ . The level weight of each node depends on whether the node has connected with other nodes or not in the correct graph. First, all nodes except the root node have same level weight, which is 1 in this research. Second, the children nodes' weights will be added to its parent node. Taking Fig.5 as example, the level weight of node C is 1 plus the summation of its children nodes' weights, i.e.

$$\begin{aligned}
 &weight_{level}(node_C) \\
 &= weight_{level}(node_{(2,1)}) \\
 &= 1 + weight_{level}(node_F) + weight_{level}(node_G) + weight_{level}(node_H) \\
 &= 4.
 \end{aligned}$$

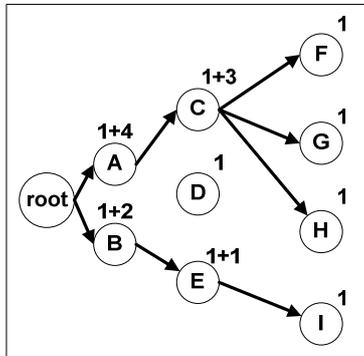


Fig. 5. Level weights of nodes in the graph

Beside the level weight can be used to distinguish the importance of nodes at different levels, the nodes at the same level may have different importance due to its conceptual meanings. For example, in a Dynamics experiment, let's give the student a hypothesis and three physical quantity concepts: velocity, mass, and modulus of elasticity, and asks the student to choose the effect of the hypothesis. Would it be an issue if the student chooses stone's modulus of elasticity instead of choosing either stone's mass or stone's velocity? The modulus of elasticity of an object actually belongs to another Physics topic rather than Dynamics. If the student chooses another topic's physical quantity, it might imply that the student may have big misconception and his/her specific science process skill may need to improve significantly.

This research defines the conceptual weight for each node  $weight_{conceptual}(node_{(k,j)})$  in order to tell the difference between a student's choice and the correct one when diagnosing the student's science process skills. The conceptual weights of two nodes at the same level represent how different conceptual meanings the two nodes have. The different concept weights may influence a student's science process skill diagnosis result. But in our virtual experiment environment, all conceptual weights are considering same due to the system only uses Dynamics relevant concepts which stored in the knowledge structure to generate the virtual experiment.

The level weight and the conceptual weight will be used as the coefficient of the graph edit cost:

$$(weight_{level}(node_{(k-1,j)}) * weight_{conceptual}(node_{(k-1,j)})) * (cost_{op}(e_{\{node_{(k,j)}, node_{(k-1,j)}\}}))$$

The  $cost_{op}(e_{node_{(k,i)}, node_{(k+1,j)}})$  has three operations:

1. insert an edge,  $cost_{edge\_insert}(e_{node_{(k,i)}, node_{(k+1,j)}})$ ;
2. delete an edge,  $cost_{edge\_delete}(e_{node_{(k,i)}, node_{(k+1,j)}})$ ;
3. and, no change,  $cost_{n/a}(e_{node_{(k,i)}, node_{(k+1,j)}})$ .

The  $cost_{n/a}$  happens when the connectivity of two nodes is exactly same in both graphs, which means, there is no necessary to edit the graph.

Taking Fig. 5 as example, we suppose the conceptual weights of all nodes are 1. If we want to delete the edge between node A and C, we need:

$$(weight_{level}(node_C) * weight_{conceptual}(node_C)) * cost_{op}(e_{\{A,C\}}) \\ = (4 * 1) * cost_{edge\_delete}(e_{\{A,C\}}). \quad \square$$

The total cost (the difference) of two graphs can be summarized as:

$$\sum_{k=1}^{(l-1)} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} \{ weight_{level}(node_{\{k,i\}}) * weight_{conceptual}(node_{\{k,i\}}) * cost_{op}(e_{node_{(k,i)}, node_{(k+1,j)}}) \}$$

where  $l$  indicates how many levels the graph has,  $n_k$  represents how many nodes in level  $k$  and  $n_{k+1}$  represents how many nodes in next level.

The total cost can represent how different the student's graph is from the correct one. Furthermore, if the teacher wants to know the student's specific science process skill level, the maximal cost has to be found. Using the maximal cost, we can understand how the student performed via

$$(1 - \frac{cost_{total}(the\_student)}{Max(cost_{total}(possible\_graph_j(the\_graph)))}) * 100\% . \quad \square$$

where  $cost_{total}(possible\_graph_j(the\_graph))$  means total cost of changing a graph,  $possible\_graph_j(the\_graph)$ , which has different connectivity combinations, to the correct graph,  $the\_graph$ . The percentage computed by the diagnosis methodology represents the degree that the student has completed in specific stage, i.e. specific science process skill, therefore the percentage could be represented as the student's completion degree of specific science process skill. The quantitative approach's results can then be used to generate the feedbacks to both of the teacher and the student.

## 5 Experiment Design

This research plans to have an experiment to assess the graph-based diagnosis methodology in finding students' science process skills in this Summer. The participants of

the incoming experiment will come from either middle school or high school, because our virtual experiment environment currently cover the Physics topics from seventh grade to twelve grade and have hundred Physics experiment animations for generating different Physics virtual experiment automatically. We plan to focus on Dynamics relevant virtual experiments due to these virtual experiments have been approved by Physics teachers and professors.

The students who are going to be participating in the incoming experiment are called as the experiment group students. We will collect their genders and past academic achievements in Physics information for further experiment evaluation. The experiment group students will use the virtual experiment environment that we have built and the graph-based science process skill diagnosis methodology proposed in this paper. The incoming experiment requests the teacher to lead the experiment group students to computer lab for using the virtual experiment environment and the science process skill diagnosis system.

The incoming experiment will collect the following data for doing quantitative and qualitative data analysis: student's manipulation behavior logs, computer attitude questionnaire, computer literacy questionnaire, science process skills and its mastery levels (via computerized diagnosis system), genders, past academic achievements in Physics, pre-test results, post-test results, two revised tests of integrated science process skills [3][9] (one will be done before the teacher leads the students to computer lab and another will be done after the students used the virtual experiment environment), and interviews.

For assessing if the graph-based diagnosis methodology is useful and understanding the accuracy of the diagnosis results, two ways will be used. First, the researchers will compare the consistency and the agreement degrees among the student's post-test results, the two revised tests of integrated science process skills, and the science process skill computerized diagnosis results. Second, the Physics teacher will be getting involved in analyzing the student's manipulation behaviors in virtual experiment environment and grading the student's corresponding science process skills. The researchers will then compare the agreement degree between the computerized diagnosis results and the human grading results.

## 6 Future Work

This research proposes a graph-based diagnosis methodology to compute the student's completion degree of specific science process skill. In the following research, we plan to define the student's achievement level according to the percentage and the virtual experiment difficulty. Similar completion percentages of specific science process skill in two different virtual experiments may have different meanings. For example, Alex had 70% and 50% completion degree of classify skill in both virtual experiment A and B, which doesn't mean Alex's classify skill in the two experiments are different if virtual experiment A is an easy one and virtual experiment B is a difficult one. So our future work is to find the method to identify the achievement levels for different virtual experiments and to give both the teacher and the student a more objective feedback about the student's science process skills.

## References

1. Assessment of Performance Unit (APU): Science in school. Age 15. Report No. 4 DES, England (1986)
2. Bunke, H.: On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters* 18, 689–694 (1997)
3. Dillashaw, F.G., Okey, J.R.: Test of integrated process skills for secondary science students. *Science Education* 64(5), 601–608 (1980)
4. Heh, J.-S., Li, S.-C., Chang, J.-C., Chang, M.: Providing Students Hints and Detecting Mistakes Made by Students in Virtual Experiment Environment. *IEEE Transactions on Education* 51(1), 61–68 (2008)
5. Kuo, L.-P., Dong, D.-X., Hsu, C.-K., Heh, J.-S.: Design an Enhanced Virtual Experiment Environment Using Science Process Skills on WWW. In: *The Proceedings of the 12th AACE World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-Media 2000)*, Montreal, Canada, June 25-July 1, p. 1785 (2000)
6. Kuo, R., Chang, M., Dong, D.-X., Heh, J.-S.: Applying Knowledge Map to Intelligent Agents in Problem Solving Systems. In: *The Proceedings of the 14th AACE World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-Media 2002)*, Denver, Colorado, USA, June 24-29, pp. 1053–1054 (2002)
7. Martin-Villalba, C., Urquia, A., Dormido, S.: Object-Oriented Modelling of Virtual-Laboratories for Control Education. *Intelligent Systems, Control and Automation: Science and Engineering* 38, 103–125 (2009)
8. Mitchell, W.E., Kowalik, T.F.: *Creative Problem Solving* (1989), <http://www.qub.ac.uk/directorates/sgc/learning/Resources/Managingstress/Filetoupload,119297,en.pdf> (retrieved on January 24, 2010)
9. Myers, B.E., Dyer, J.E.: Effects of investigative laboratory instruction on content knowledge and science process skill achievement across learning styles. *Journal of Agricultural Education* 47(4), 52–63 (2006)
10. Novak, J.D.: Applying learning psychology and philosophy of science to biology teaching. *The American Biology Teacher* 73, 12–20 (1981)
11. Polya, G.: *How to Solve It*. Doubleday, Garden City (1957)
12. Sanfeliu, A., Fu, K.-S.: A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on System, Man, and Cybernetics* 13, 353–362 (1983)
13. Tai, K.-C.: The tree-to-tree correction problem. *Journal of the Association for Computing Machinery* 26(3), 422–433 (2003)
14. Wagner, R.A., Fischer, M.J.: The String-to-String Correction Problem. *Journal of the Association for Computing Machinery* 21, 168–173 (2001)
15. Wu, S., Chang, A., Chang, M., Liu, T.-C., Heh, J.-S.: Identifying Personalized Context-aware Knowledge Structure for Individual User in Ubiquitous Learning Environment. In: *The Proceedings of the 5th International Conference on Wireless, Mobile and Ubiquitous Technologies in Education*, Beijing, China, March 23-26, pp. 95–99 (2008)
16. Yan, Y., Liang, Y., Du, X., Saliah-Hassane, H., Ghorbani, A.: Putting labs online with Web services. *IEEE Computing* 8(2), 27–34 (2006)
17. Fan, M.-X., Kuo, R., Chang, M., Heh, J.-S.: Using Story-based Virtual Experiment to Help Students Building Their Science Process Skills. In: *Proceedings of the AACE 22th World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2010)*, Toronto, Canada, June 29-July 2 (2010)