USING REU PROJECTS AND CROWDSOURCING TO FACILITATE LEARNING ON DEMAND

Hong P. Liu and Jerry E. Klein
Address Embry-Riddle Aeronautical University, 600 S. Clyde Morris Blvd, Daytona Beach, FL 32174

ABSTRACT

With the increasing complexity of technology and large quantities of data in our digital age, learning and training has become a major cost of employers. Employee competence depends more and more on how quickly one can acquire new knowledge and solve problems to meet pressing deadlines. This paper presents a practical method to use REU (Research Experience for Undergraduates) projects and crowdsourcing to help students learn complex content as needed. The major question addressed in this work is to find ways to reduce the cost of learning for both the learners and their mentors. As a work-in-progress, we would like to share our preliminary design in using natural language processing (NLP) and data mining tools to retrieve and structure instructional materials semi-automatically.

KEYWORDS

Learn on demand, crowdsourcing, data mining, conceptual models, (Research Experience for Undergraduates) REU

1. INTRODUCTION

The digital age is characterized by big data, increasing complexity and rapid technology changes. Existing academic programs are inadequately structured to prepare students with the skills and knowledge needed for today's complex fast moving work. Learning and training has become a major cost to employers and many small businesses are showing a tendency to hiring graduates with some work experiences but average academic records rather than hiring new graduates with above average academic with no experience. This brings great challenges for universities to place their new graduates. Research Experiences for Undergraduates (REU) has been advocated in recent years as co-curriculum activities to provide students hand-on experiences to complement the traditional undergraduate programs. Hands-on research experiences bridge the gap between problems in textbooks and real-world problems and therefore serve as a good stepping stone for our students in obtaining their first job. Student success upon graduation will greatly depend on how quickly one can learn new knowledge as needed and solve emerging problems. The purpose of REU programs is to (a) motivate students to learn new domain knowledge as they need it in order to solve problems and to (b) cultivate students' problem solving ability.

However, REU projects are time consuming for both student team members and faculty mentors. The cost for the learners can be measured by the total time consumed from the starting point of identifying missing knowledge to the end point of being able to solve the problem. REU is intimidating to most undergraduate students because it requires them to spend significant time acquiring knowledge beyond the standard curriculum for their major. The lecturing type of traditional curricula can be called as just-in-case education. It is designed to prepare students with well-refined knowledge and fundamental skills associated with a particular discipline. As co-curriculum activity, REU can be called learning-on-demand education. Its learning environment is more like that in industry rather than an academic setting. Its purpose is to teach students to acquire new knowledge and emergent technologies quickly in order to solve the problems at hand. Consequently, the students in REU projects are learning as needed and learning by doing so that they can significantly reduce the destructive effects of forgetting. This type of learning also enables the learners to take advantage of the ubiquitous internet memory to extend their biological memory (Starmer, 2012).

The major cost for faculty mentors is the time involved in preparing and teaching the required domain knowledge needed to complete projects.

Because of the interdisciplinary nature of the REU projects, the mentor often does not have all the knowledge or the instructional materials on hand to solve the problems posed by the projects. In addition, mentoring REU projects is counted as services without teaching relief in most universities. Hence, REU is discouraging to faculty members. Our examples demonstrate a practical method to reduce the cost of learning for both the students and mentors.

A primary role of a faculty mentor supporting a project team is identifying a student's knowledge gap. The students should take the initiative to search and learn the knowledge from the internet, especially from the crowd sourced educational materials. While it is impossible or time consuming for the mentor to learn everything in advance and prepare the course material, the mentor may save time by focusing on understanding the relevant domain knowledge at the ontological level and building a conceptual model of the factors used to solve the assigned tasks in a logical and chronological order.

The next section of the paper illustrates our system engineering approach and modeling tool Opcat to scaffold REU. The third section introduces the challenges and strategy for mentoring on REU projects. The fourth section presents work-in-progress for identifying and organizing educational courseware using educational crowdsourcing technology (Eytan, 2012, and Karger 2012). The fifth section describes future work using GATE, an NLP tool and Weka, a data mining tool, to retrieve and extract instructional materials for mentoring students engaged in REU projects.

2. USE SYSTEM ENGINEERING APPROACH TO SCAFFOLD REU

Liu and Ludu (Liu 2012) identified three major challenges that hinder RE success:

1) Critical thinking: Many undergraduates just plug data into formulas to solve problems and show little ability to extend mathematics concepts beyond an algorithmic level.

2) Complexity: Undergraduates lack the experience and knowledge to divide complex problems into multiple small problems and the complexity of the research is beyond their current knowledge.

3) Applicability: Undergraduates have limited domain knowledge. Therefore, they cannot apply the knowledge to validate their mathematical models.

To address these shortcomings, Liu and Ludu established a system engineering-based REU program called ACE. Shown in Figure 1, ACE stands for Analysis, Computation, and Experimentation. Liu and Ludu have also established a nonlinear wave lab facility and Eco-Dolphin project to provide students a hands-on development environment for the validation and verification of objects they have designed and developed in their projects.
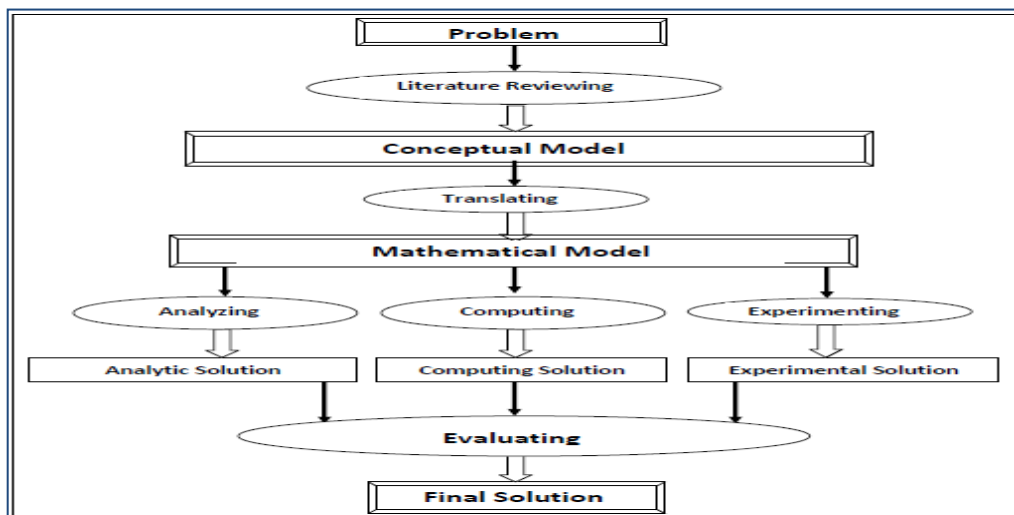

Figure 1. Opcat Diagram of the ACE Paradigm

We have selected Object-Processing methodology and its support tool Opcat for building conceptual models. Opcat is a user friendly tool with a gentle learning curve (Dov Dori, 2002).

Opcat offers dual representations: Object-Process Diagrams (OPD) and language (OPL), which enables an Opcat model to automatically include the graph legend. Opcat uses three entities as building blocks: objects, processes, and states. Objects are things that exist. Processes are things that transform objects by changing their states and creating or consuming objects. In OPD, objects, processes, and states are symbolized by rectangles, ellipses and rounded rectangles, respectively. The single integrated view of the Opcat models and the built-in syntax of Opcat based on OPM methodology help modelers identify missing components and breaks in logic links. The zooming in/out strategy for complexity management makes it easy to build and understand the conceptual models at almost any level in a hierarchy of a system and at different levels of granularity. These features make Opcat an ideal tool for quick prototyping and requirement gathering. The conceptual models help users to discover the critical paths at a project management level and identify the concepts, knowledge, and skills that a team needs in order to complete the project (Liu 2004).

Our system engineering approach shown in Figure 2 and 3 guides students to take small steps, analyze the problems incrementally and refine the results iteratively. The two Object-Process Diagrams have the same iterative elaborating loops for sense making (conceptual model) at the qualitative level and validating the mathematical model at the quantitative level. The data driven model validation is based on sound scientific principles. This approach can provide evidence based assessment for student learning. Students can evaluate their own solutions to problems without the teachers grading their work. Hence, our system engineering approach saves mentor's time by reducing the time needed to evaluate student work at the detailed level.
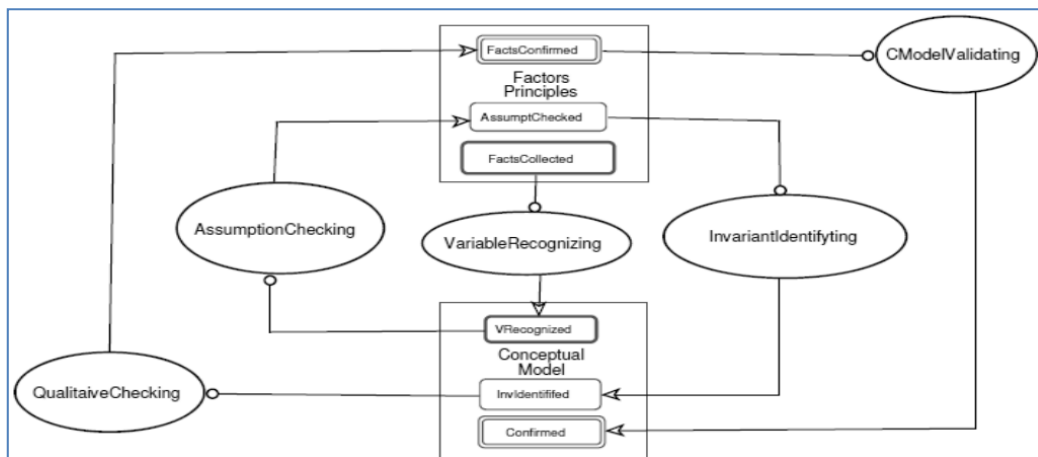

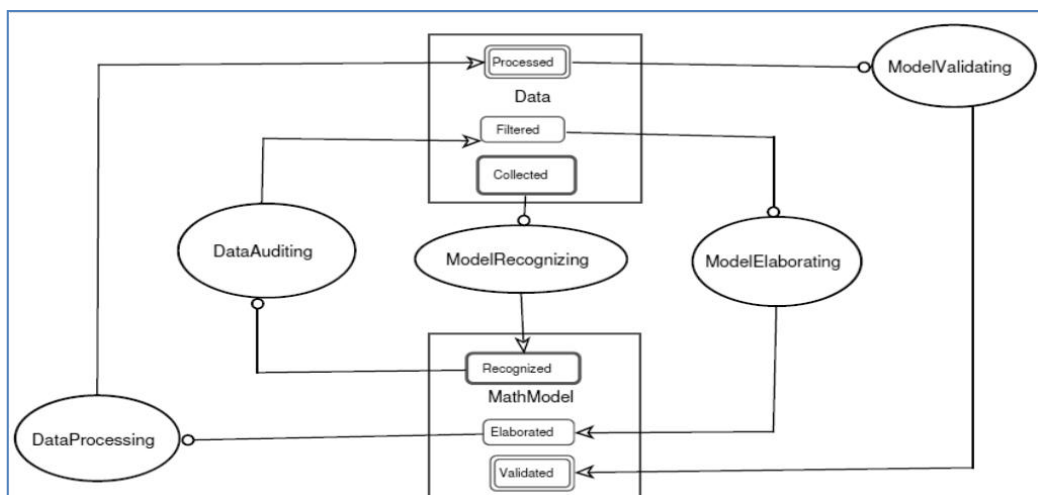Figure 2. Sense Making for Conceptual Model

Figure 3. Validating for Mathematical Models

Table 1. Computer Generated Sentences for OPL for OPDs in Figure 2 and Figure 3

Sense Making for Conceptual Models	Validating Mathematical Models
Conceptual Model can be Confirmed, InvIdentifed, or VRecognized.	MathModel can be Validated, Elaborated, or Recognized.
Confirmed is final.	Validated is final.
VRecognized is initial.	Recognized is initial.
Factors Principles can be FactsCollected, AssumptChecked, or FactsConfirmed.	Data can be Collected, Filtered, or Processed.
FactsCollected is initial.	Collected is initial.
FactsConfirmed is final	Processed is final.
	DataProcessing requires Elaborated MathModel.
	DataProcessing yields Processed Data.

3. LEARNING-ON-DEMAND STRATEGY FOR REU PROJECTS

The strategy to help undergraduate students employ learning-on-demand methods is to divide the REU projects into components (subsystems and tasks) at the proper level of granularity. The division of components not only needs to fit the nature of the problem, but also the academic requirements of team members. Figure 4 depicts the conceptual design of Eco-Dolphin project under the mentorship of Liu. Eco-Dolphin is the name of a fleet of adaptive and cooperative Automated Underwater Vehicles (AUVs) that a team of 6 SIAM Chapter students at ERAU have been working on since spring semester 2012. It will be designed to support future environmental science research and surveillance services in littoral water. Eco-Dolphin serves as a component of the ACE program (Liu 2012) that is sponsored by the department of mathematics and Honors Program at ERAU. The autonomous nature of robotics spans multiple disciplinary fields including mathematics, computer sciences, physics, mechanical engineering, electronic engineering, computer and software engineering. Consequently, the Eco-Dolphin project is a truly "trans-disciplinary" REU program. The 6 team members come from 6 different degree programs and are under the mentorship of three professors, whose expertise covers mathematics, physical sciences, computer sciences, software engineering, and Electronic Engineering. The team also includes a graduate student mentor in Mechanical Engineering.
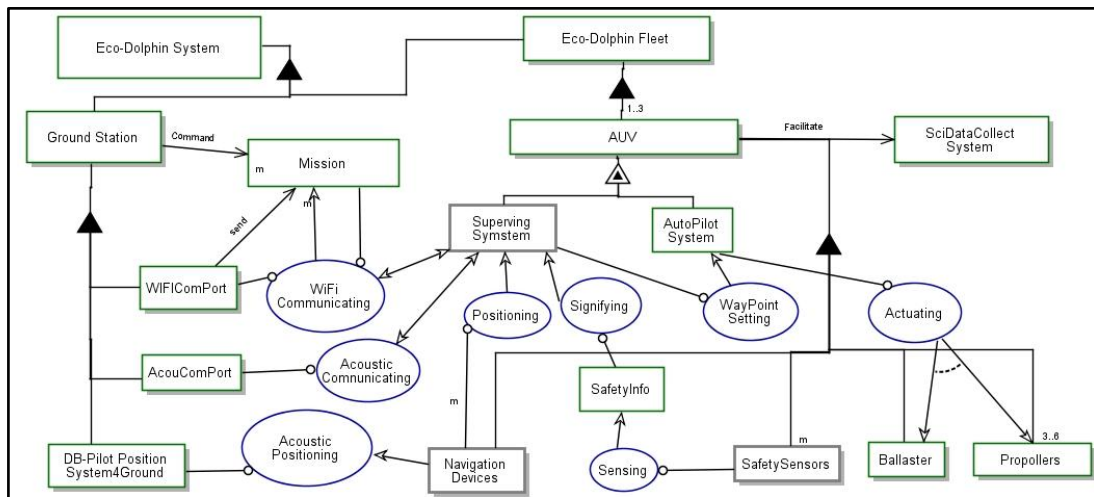

Figure 4. Top Level of System Conceptual Models of Eco-Dolphin Project

The interdisciplinary nature and hands-on REU project requires each member of the team to learn knowledge relevant to a wide range of assigned tasks. Since it is not feasible for a faculty mentor to teach each team member on a one-to-one base, the mentors and students have collected hundreds of relevant technical articles and organized them according to the subsystems named above. The files are shared by team members through Dropbox.

Our projects require mentors to learn more new knowledge than any student. For example, a robotics project typically starts as a mechatronics project and ends as a software engineering project. However, a faculty mentor may not have experience in either mechanical engineering or electronic engineering. To address this problem, we implemented a learning as-needed strategy by starting a literature review and identifying a few AUV projects as peer projects that the Eco-Dolphin team would emulate. The mentor tracked the references from the student reports and then traced the references on every subsystem. The other two faculty members served as advisors to the lead mentor of the project.

The OPL sentences in Table 2 help the modeler determine if the diagram in the OPD in Figure 4 confirms one's intention. The dual representation of the Opcat model facilitates communication among teammates and serves an idea tool for brain storming.

Table 2. Computer Generated Sentences Are a Sample OPL Associated with the OPD in Figure 4

System decomposition	Component dependent relationship
Eco-Dolphin System consists of Ground Station and Eco-Dolphin Fleet.	Actuating requires AutoPilot System.
	Actuating yields either Propollers or Ballaster.
Ground Station is physical.	Sensing requires SafetySensors.
Ground Station consists of DB-Pilot Position System4Ground, AcouComPort, and WIFIComPort.	Sensing yields SafetyInfo.
DB-Pilot Position System4Ground is physical.	…...
……	WayPoint Setting requires Superving Symstem.
	WayPoint Setting yields AutoPilot System.
AUV consists of Ballaster, 3 to 6 Propollerses, many SafetySensorses, and many Navigation Deviceses.	Positioning requires Navigation Devices.
	Positioning yields Superving Symstem.
……	……

After the mentors collected references, the student team leaders took the initiative to study the required knowledge associated with their assignments. The faculty mentor then prepared video lectures and online instructional materials for difficult components such as PID (process, integral and derivative) control and the Kalman filter needed for the navigation program and architecture design of software systems. In the first phase, the team successfully designed and built the hull. It includes the ballast, hydro pressure sensors and propeller, etc. But, the software component has not met the expected goal because of a lack of programmers. As the project progresses, the students and faculty mentor have to constantly learn new knowledge as needed. We have found that faculty learning time can be reduced by letting students lead the way. For example, students stay ahead of the faculty mentor in a new technology such as the Ardruino microcontroller programming and tutor each other.

The success of mentoring REU depends on assigning small tasks to students and frequently meeting with them to discuss issues in completing their tasks. While graduate student research advisors can assign tasks and meet them monthly, the mentors of REU assign tasks and meet students every week or every other week. However, the mentors of REU can save time by finding crowd-sourced materials and enabling students to learn on their own or from each other, instead of teaching traditional classes.

4. EXTRACTING MATERIAL FROM CROWDSOURCING FOR LEARNING ON DEMAND

Crowdsourcing is the process of assimilating many small contributions into resources of high-quality. Khan Academy, MIT Open Courseware, Stanford online Courses, Udacity, and Coursera have made a transformative impact on the "digital generation's" choice of college education and methods to gain knowledge. The scope of their collective impact and the influence on the infrastructure and subsequent impact on college education (e.g. virtual open universities) are too significant for any university to overlook.

Free courseware can also be exploited as supplementary education resources to provide students personalized education and support learning on demand. In addition, it can save a significant amount of time for mentors of REU projects to prepare instructional materials. In most subject fields, there are adequate learning materials from crowdsourcing. In order to leverage this free courseware, three major issues need to be addressed:

1. How to automate the process of identifying and extracting the online materials.
2. How to track the dynamics of web updating and trace breaking links.
3. How to evaluate the quality and the appropriateness of materials based on the needs of learners.

The authors are currently working with student research assistants on a project to build an Open Courseware Identification Subsystem (OCIS), which is expected to be completed in the summer 2014. The OCIS project has set two goals: the first goal is to collect calculus course materials that are compatible to our syllabi as complementary tutorial material; the second goal is to explore the information technology to automate the data retrieval and information extracting process. Collecting open courseware for a few courses manually is more efficient than developing a web application to automate the process. Consequently, the authors have manually collected computational science education materials online for a course in Mathematical Modeling and Simulation. However, the manually implemented process neither scales to large set of courses due to the prohibitive labor cost nor supports the dynamic content change of the online courseware.

Instead of writing our own natural language processing program, we chose to use a NLP tool called GATE (General Architecture for Textual Engineering) to fulfill the second goal mentioned above. GATE is to NLP as MATLAB is to numerical computation or Mathematica is to symbolic computation. Novice users can apply GATE's built-in resources to process common tasks such as name recognitions in the same way as novice MATLAB users call several MATLAB commands to perform matrix operations. Advanced GATE application developers may need to write GATE programs to process complicated tasks such as identifying and summarizing the particular newsletters of interest to their sponsors. A GATE application includes three components: language resources, process resources and an application pipeline made of built-in processing resources, third part plug-ins or user made plug-ins. Besides the corpus in data store, the GATE users typically need to create another language resource called an annotation schema. There are many options in GATE. It can be either a simple gazetteer with the keywords of a domain or an ontology driven gazetteer with logical relationships of the keywords. The user inputs one's domain knowledge into the system from the gazetteers. Figure 5 shows the flowchart of the processes for a web mining and KDD system (Wang 2012). Figure 6 gives a different view of such a flowchart for a particular project (Wang 2013). Figure 7 gives the detailed flowchart of the IE subsystem on the left side of Figure 6.
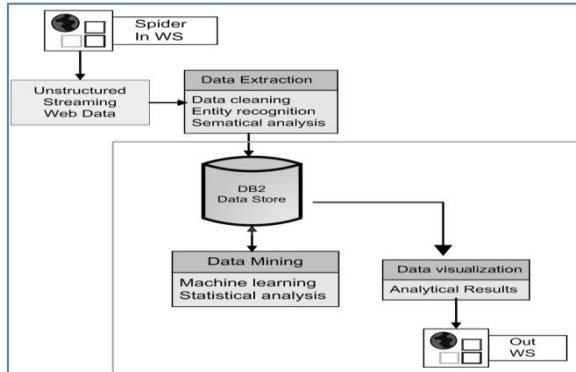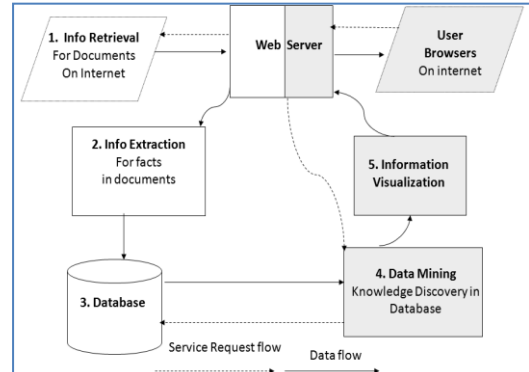

Figure 5. Flowchart for Web Mining and KDD,

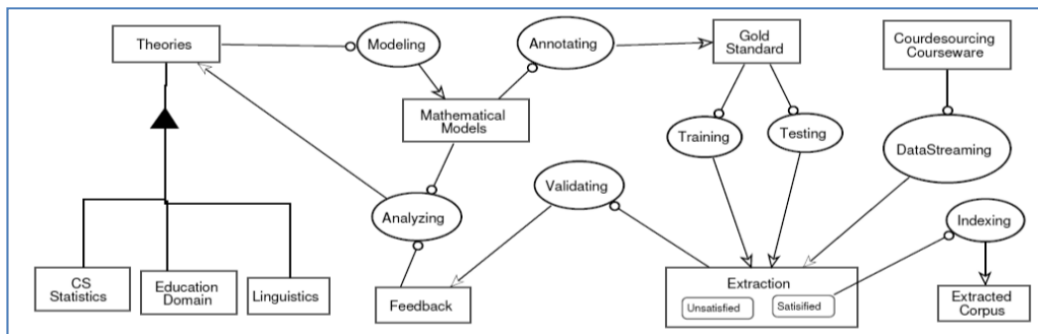Figure 6. Web Application to Graduate School Matching

Figure 7. Information Extraction (IE) Framework for Education Web Mining.

We will not completely automate the Information Extraction (IE) processes for OCIS because it is difficult to connect several tools from different vendors in a pipeline. Instead, we will manually launch each tool in the pipeline and exploit the strength of the tools one at a time. The IE in Figure 9 has three steps:

Step 1, we use xml-sitemaps - to download the website maps of courseware providers such as Khan Academy or the MIT Open Courseware. The free version allows us to crawl and download up to 500 pages each time. To download the sitemap of MIT open courseware, we input it intoxml-sitemaps.

Step 2, we use GATE to process the sitemap document and extract the links of interest. We first save the loaded files to the data store of GATE and then create a gazetteer with the keywords of the calculus courses. Next, we create a GATE application by pipelining the ANNIE processes. The output of this GATE application is the links of the courseware that we provide to our students. We can run the GATE application once a semester to check for new updates and break links autonomously.

Step 3, we will check the quality and appropriateness of the website. Google webmaster can be used to find the download rate and other properties of the websites. We will read the reviews for the small number of websites in the OCIS project. Finally, we randomly select sample content of the selected courseware and send it to colleagues and student assistants to review for quality and appropriateness. The links and property summary of websites as shown below will be saved to our database.

Address = http://www.youtube.com/watch?v=AUqeb9Z3y3k&feature=relmfu
Keywords_ in_link = {Mathematical, Modeling, Simulation}
FrequentlyUsedWords = {Matrix, Transformation, Scarce, toeplitz}
Download rate = 98234, number of reviewers = 57
Document_properties = {6368986 Bytes, Joe John, 2008/12/01, 2010/4/05}

As a relevant previous project, Figure 6 shows the flowchart for a web application for matching graduate schools for graduate seniors in colleges. The shaded half on the right hand is the Graduate Research Project (GRP) completed by Wang under the mentorship of the first author in 2012 (Wang 2013).

In addition to developing the web application, the major theoretical component of the GRP project is to use the Weka data mining platform a to analyze several data mining algorithms based on the performances of the graduate school matching problem. The GRP covers the data mining component and OCSI covers the IE component of figure 6. Combining the two projects, allows us to cover all the technology described in Figure 6. Ideally, we should have used the same application to cover the entire system in Figure 6. However, we made the choice for the projects based on the interests of the graduate student and the sponsor of OCIS. Many components are reusable for future work because the GRP project, OCIS, and the information extraction for REU projects are similar educational web mining applications (C. Romero, 2007).

5. FUTURE WORK

As the OCIS project continues, we would like to use data mining to match the open courseware stored in database and the syllabi of our corresponding courses. The task is to complete the shaded steps 4 and 5 on the right side of the Figure 6. If we replace the profiles of graduate schools by the content tables of the open courseware and the graduate applicant forms by the syllabus (including the topics for each lesson) of our corresponding course, the data mining application is almost identical to the GPR project mentioned above. The output will be the matching of content between the open courseware and the lessons of our corresponding courses. We are doing this matching manually for the OCIS project because the scope exceeds our one year time frame. Our next application is to semi-automate the process for finding open education materials associated with our particular REU projects. The major difference between finding courseware for the known courses and materials for REU is that the logic order of the former is well known in standard curricula and that of the latter is most likely unknown.

Another goal is to automate the process to obtain structured user review summaries from unstructured online user reviews. GATE can be used to obtain semantics level recognition (e.g. name recognition) easily. GATE developers usually need to write Java script to recognize hidden relationship of words based on contexts for summarizing documents. These patterns are unlikely to be recognizable at the semantic level. For example, the two OPLs shown in Table 1 for mathematical model validation and sense-making conceptual models have the exact same structure and logic relationship. However, shallow semantic level comparison cannot identify the two diagrams because there are no identical words used in the two diagrams.

Nevertheless, the OPL in regular language contains a clear pattern based on the few syntactical relationships of three entities (object, process and states). Hence, the ontological model in Opcat can be used by the GATE program to identify hidden relationships based on structure similarity between the training corpus and targeted corpus. GATE actually supports ontology driven unsupervised annotation.

This paper addressed the problem of how to motivate students to learn as needed by providing them hands-on research projects and giving them the opportunity to learn by doing. A system engineering approach called ACE is introduced to help students evaluate their own research solutions. Conceptual modeling and its support tool, Opcat, is advocated to scaffold learning of complex systems. We offered the first hand experiences about how to take advantages of crowdsourcing education materials to save time for preparing learning materials. While the traditional just-in-case education requires students to memorize formulas, definitions and methods to pass tests, just-in-time learning and learning by doing can significantly reduce the destructive effects of forgetting and let the learners take advantage of the ubiquitous internet memory to extend their biological memory. A major challenge for faculty mentors of REU projects is the time consuming tasks to prepare learning materials out of the main stream curricula. We shared the designs of our previous project, an ongoing project, and future work about how to use natural language processing tools and data mining tools to find quality and relevant open educational materials online efficiently.

ACKNOWLEDGEMENT

REFERENCES

Book

Dov Dori, 2002. *Object-Process Methodology:* Springer, Berlin Heideberg.

Journal

Hong Liu, D. P. Gluch, 2004. Conceptual Modeling with the Object-Process Methodology in Software Architecture: The Journal of Computing Science in Colleges, Vol. 19, Num. 3, 2004, P.10– 21.

Hong Liu, Jayathi Raghavan, 2009. A Mathematical Modeling Module with System Engineering Approach for Teaching Undergraduate Students to Conquer Complexity: Lecture Notes in Computer Sciences, Vol. 5545, Part II, pp 93-102.

C. Romero *, S. Ventura, 2007. Education Data Mining, A survey from 1995 to 2005, Expert Systems with Applications: Expert Systems with Applications, Vol. 33, Number 1.

Ming Wang, Hong Liu, Drew Hwang, 2012. Applying Data Analytics to Development of the Web-based Information Security Career System: International Journal of Information and Electronics Engineering: Vol. 2, No. 5, pp 806-809.

Conference paper or contributed volume

Hong P. Liu, Andrei Ludu, 2012, ACE - A Model Centered REU Program Standing on the Three Legs of CSE: Analysis, Computation and Experiment: Procedia Computer Science. Omaha, USA, 1773-1782.

C. Frank Starmer, 2012. Exploring Google + Internet Resources as a memory amplifier, spending more time thinking and less time navigating web sites and forgetting? Available: http://frank.itlab.us/IT/education.pdf .

Daniel S. Weld1 Eytan, Adar, et. al, 2012. Personalized Online Education—A Crowdsourcing Challenge, Available: http://homes.cs.washington.edu/~chrislin/papers/hcomp12c.pdf .

David Karger, 2012. Crowdsourced Personalized Online Education Resources, Available: http://frank.itlab.us/IT/education.pdf .

Wenyu Wang, 2013. A Prototype of Web Application for Applicants to Match Graduate Schools Based on Data Mining,https://dl.dropboxusercontent.com/u/73148290/Preprint/WenyuWangGRPPaperERAU.doc .