Memex Metadata (M^2) for Reflective Learning

Jane Greenberg*, Abe Crystal, Anuj Sharma
Metadata Research Center
School of Information and Library Science
University of North Carolina at Chapel Hill
Tel: +19199628066
janeg@ils.unc.edu (*chief contact), abe@unc.edu, nooj22@email.unc.edu

Eva Méndez
Universidad Carlos III de Madrid / SILS Metadata Research Center
Tel: +34916248620
emendez@bib.uc3m.es

John Oberlin, Michael Shoffner
ITS/Technology Assessment and Planning
University of North Carolina at Chapel Hill
Tel: +19199623441
john_oberlin@unc.edu
michael_shoffner@unc.edu

Abstract:

This paper reports on metadata research and development supporting memory and reflective learning that is being conducted as part of the Memex Metadata (M^2) for Student Portfolios project, University of North Carolina at Chapel Hill. The paper reviews learning, memory, and reflective learning strategies; introduces the M^2 project; presents a reflective learning scenario for a plant biology class; and reviews two metadata developments underlying the M^2 project: 1) A *context awareness framework* (CAF), and 2) An *extended metadata framework*. The paper concludes with a brief discussion of future research activities.

Keywords:

Metadata for learning; Higher education; Student's memories; e-Portfolios; Personal Information Management; PIM; Context awareness; Ontologies; Student-centered learning.

1. Introduction

Students learn best through engagement and active learning, and by using self-monitoring and reflection to guide their education (McCormick, 2003). This knowledge poses important challenges for developers of e-learning technologies, particularly the design of Personal Information Management (PIM) systems that students may use to store and organize their digital educational materials. In the context of undergraduate education, PIM systems can be used to store:

- Personal digital educational memories, captured via digital cameras, cell phones, email, IM (instant messaging), and other technologies.

- Student resources, generated individually or collaboratively, for a specific class (e.g., notes taking during class, paper drafts, project work).
- Standard course resources (e.g., professors' slides and lecture notes, whiteboard notes, readings, assignments, quizzes, exams, course evaluations).

PIM systems need to provide mechanisms for students to effectively manage the wide range of digital materials accumulated during the learning process (Johnson and DiBiase, 2004). PIM systems also need to store objects and events recorded via memory augmentation technology, such as Microsoft's SenseCam[1], if these data are to be incorporated into the educational process and used for innovative learning strategies.

Metadata is critical to these information management challenges. Metadata needs to be captured when digital memories are created, manipulated, and used; and it needs to be automatically harvested, extracted, and derived when memories are stored in PIM systems. PIM systems also need to support annotation so students can personally comment on their digital store of course material for future use.

The University of North Carolina is investigating these metadata challenges through the Memex Metadata (M^2) for Personal Educational Portfolios project, launched February 2006. M^2 is part of Microsoft Research's Digital Memories (Memex) project, which aims to investigate and develop the utility of SenseCam technology and MyLifeBits software.[2] Microsoft's initiative builds off of Vannevar Bush's notion of the Memex, as detailed in his landmark article, "As We May Think" (1945). Our research has focused initially on biology students engaged in specimen identification and learning plant taxonomy. This paper reports on the M^2 project and focuses on metadata developments associated with reflective learning, including the project's context awareness framework (CAF) and an extended metadata framework.

2. Learning and Memory

Learning occurs when recognizing something new and relating it to what is known. Students learn, for example, by solving problems or by transferring what is known into practice. Memory plays a central role in learning. Memorization and reciting facts are low-level activities, while understanding concepts and principles are high-level activities that build upon lower-level skills (Huber, 1993).

The traditional lecture/exam educational model emphasizes short-term memory. Students become successful test-takers, memorizing and reciting facts, although research indicates they don't necessarily develop an understanding of the discipline being studied (Norman and Schmidt, 2005). Bloom highlighted this limitation in the mid-1950's, when he found that most instruction methods were focusing on fact-transfer and information recall, rather than meaningful personal development (1956). Students caught in the memorization paradigm have difficulties solving complex problems within the discipline of study. Moreover, the emphasis on memorization can be an impediment to life-long learning.

---

[1] SenseCam technology is described below in Section 4, Memex Metadata (M^2) for Student Portfolios (see also Figure 1).

[2] Microsoft Research Digital Memories (Memex): http://research.microsoft.com/ur/us/fundingopps/RFPs/DigitalMemories_Memex_RFP_Awards.aspx.

3. Beyond Memory: Reflective Learning and New Technologies

Over the last two decades, a growing number of educators have been promoting educational reform involving inquiry-based learning and reflection (Gilmer & Alli, 1997; Roth, 1996). Research shows that students learn more effectively, and develop higher-level learning skills, when they are actively involved and engaged in the learning process. In education, or other endeavors, reflection promotes critical thinking, a higher order of cognition (Boud, Keogh and Walker, 1985).

We are now in a world of pervasive computing, and learning strategies need to account for all means by which students gather information and data. Additionally, memory-related technologies offer new opportunities for innovative learning, including reflective learning strategies. Among some of these technologies are digital cameras, videos, sound recording systems, PDAs, and cell phones equipped with many memory features. These technologies store memories, providing students with the freedom to reflect and the digital store to aid reflective learning. We need to explore the role new technologies can play in improving and accelerating education. And, we need to understand metadata requirements for successful implementation and use of new technologies.

The Memex Metadata (M^2) for Student Portfolios project, University of North Carolina at Chapel Hill, is exploring the potential of Memex (memory recording and storage) technologies for reflective learning; and, a key component of the project work focuses on a series of metadata questions.

4. Memex Metadata (M^2) for Student Portfolios

M^2's partnership includes the University of North Carolina's School of Information and Library Science, Metadata Research Center; Information Technology Services (ITS), and the Biology Department—all part of the Chapel Hill campus. We are an interdisciplinary team of researchers, technologists, and educators. Our project work extends three previous biology education and technology collaborations: the UNC Plant Information Center (PIC[3]), Project OpenKey[4], and Botnet[5].

We are exploring metadata questions and the usability of pervasive technologies for reflective learning. Our current technology includes Microsoft SenseCam technology and MyLifeBits (MLB) software (Figure 1). The SenseCam is a sensor-enhanced, wearable digital camera that automatically records images every 90 seconds, or when the sensors suggest the wearer has encountered a new environment or situation. MyLifeBits is a PIM system, built on a relational database that supports capture and retrieval of vast amounts of personal digital information.

---

[3] PIC (Plant Information Center): http://www.ibiblio.org/pic.

[4] OpenKey Project: http://www.ibiblio.org/openkey.

[5] North Carolina Botanical Information Network: http://www.ibiblio.org/botnet/flora/indexstart.html.

Figure 1: M^2 SenseCam (left) and MLB Calendar Display

It is important to note that we are exploring the use of technology and metadata to support students' memory and reflection. We are leaving aside, for now, the key issue of privacy, which introduces many complex legal and social issues—and require separate in-depth research.

5. M^2 Metadata Research Questions

Metadata is critical for effective use of pervasive computing in today's learning environment, and researchers need to explore metadata issues so innovative learning strategies, including reflective learning, can be successful. The M^2 research team has identified a number of significant metadata questions requiring investigation. Among some of the most pressing questions for our project are:

- What metadata is needed to support reflective learning scenarios using the SenseCam technology and other pervasive technologies?
- How can we extend the University's Context Awareness Framework (CAF) ontologies to support reflective learning and other *general* memory-related activities?
- What metadata is required for reflective learning for students *specifically engaged in* a Local Flora course?
- What is the most efficient and effective means generating metadata in this new learning environment?

The remainder of this paper discusses how we have been addressing these questions, and concludes with a brief discussion on future research

5.1 Metadata Needs for M^2 Reflective Learning

The M^2 team includes biology educators, who want to incorporate reflective exercises into their courses. These educators want their students take advantage of new technologies to better understand the discipline of plant biology and gain a deeper understanding of our natural world. One key challenge has been identifying reflective learning strategies for plant biology, and then identifying where metadata plays a critical role.

Plant biology has traditionally emphasized memorization of plant taxa and characteristics. To get a sense of how new technologies, including Microsoft SenseCam, can assist reflective learning, we have developed a series of scenarios. An example of a technology-enhanced reflective learning scenario for a Local Flora class is presented in Example 1.

Example 1: Technology-Enhanced Reflection: A Scenario

At the end of class on a Thursday, Kim's instructor reminds the students to complete their weekly review. Kim goes to a coffee shop to get a drink and complete her review. She opens her Memex, and sees a week's worth of SenseCam images, documents, emails, photos, and audio related to her Local Flora biology class. She browses through the collection, selecting artifacts that strike her as particularly memorable, useful or interesting. She is able to quickly combine these artifacts into a multimedia diary of her week's learning.

After creating the diary, Kim writes brief responses to the three review questions: What were the main themes this week? What did you not understand? What did you find most interesting? Her responses are stored in her Memex along with the multimedia diary entry.

Figure 1 and 2, Appendix A, includes two snapshots of Microsoft Research's MyLifeBits (MLB) digital memories shell. Figure 1 illustrates metadata annotation, and Figure 2 presents a contextual retrieval example with SenseCam images and a digital photograph of plant specimen.

Metadata requirements for reflective learning scenario, like the one presented above, generally fall into three metadata classes that Gilliland (2005) uses to describe the metadata landscape: *content*, *structure* and *context*. Examples of metadata in each of these classes, drawing upon our scenario, follow below:

- **Content metadata:** The subject of the SenseCam images, documents, emails, photos, and audio *objects* related to the Local Flora biology class.[6]
- **Context metadata:** When (date and time) and where (location and address) the objects are produced, manipulated, and used.
- **Structural metadata:** The format and architectural composition of the SenseCam images, documents, emails, photos, and audio objects (e.g., GIF, TIFF, MP3).

Gilliland's framework has helped us to identify M^2 metadata requirements for reflective learning, extend the CAF, and build a M^2 metadata framework.

5.2 Context Awareness Framework (CAF)

Prior to the initiation of the M^2 project, UNC's ITS was developing a Context Awareness Framework (CAF). The CAF has two key features:

1. An integrated framework with a series of ontologies applicable to a range of objects (people, classes, events, and so forth) active in the University's pervasive computing environment.

[6] Objects is used in a generic sense to include digital resources (e.g., a Word document, slides, an article), and happenings or experiences that might be recorded via the SenseCam or other technology.

2. Software agents that will be installed on students' computers to communicate with CAF ontologies and other rules sets specific to the university environment.

The goal of the CAF is to link software agents and ontologies to automate common educational needs, such as configuring a classroom for a presentation or sending needed class information to a student's laptop during the exact time of need. For example, when combined with GPS or RFID location information, a CAF agent can detect that a student had entered a classroom for a specific class, and load necessary software or Web pages for the class session.

The CAF integrates a series of ontologies constructed using the W3C's Web Ontology Language (OWL[7]). The Context Broker Architecture (CoBra[8]) serves as a base for the CAF. CoBra provides the Standard Ontology for Ubiquitous and Pervasive Applications (SOUPA[9],) ontology that encodes domain knowledge about location, time, schedule, and the like for the purposes of modeling meetings.

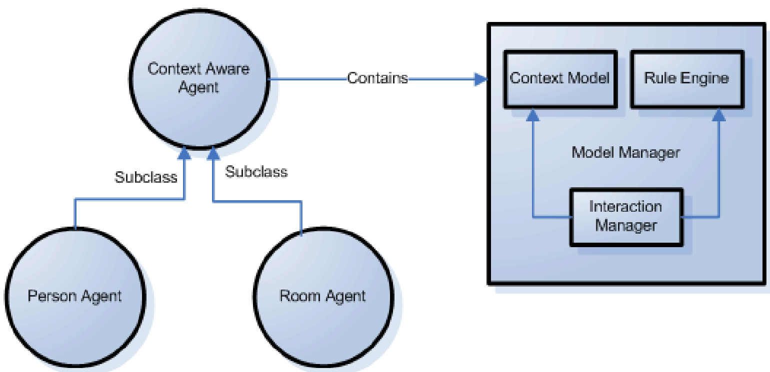

Figure 2: Context Awareness Framework Architecture

CAF software agents are being developed in Java using the Java Agent Development Environment (JADE) toolkit[10]. JADE is an implementation of the Foundation for Intelligent Physical Agents (FIPA)[11] intelligent agent standards. Each context-aware agent provides a model manager that contains its internal context model. The model manager uses ontologies and rule sets to understand context and make inferences about information that comes in from the environment. CAF agents depend on the CAF ontologies, specifically their metadata to recall digital memories and execute or recount a learning experience.

Integration between the CAF and MLB is currently being developed. This integration requires mapping context data, represented by the CAF's ontologies, into MLB's data structures.

5.3 M² Metadata Framework

[7] OWL: Web Ontology Language: http://www.w3.org/2004/OWL.
[8] CoBRA (Context Broker Architecture): http://cobra.umbc.edu.
[9] SOUPA: http://pervasive.semanticweb.org/soupa-2004-06.html.
[10] JADE: http://jade.tilab.com.
[11] FIPA: http://www.fipa.org.

The M^2 framework enhances the CAF to record object content, structure, and context by integrating and extending common metadata-supported functions (resource discovery, use, authentication) to facilitate context awareness (Greenberg, 2005). Our current framework supports effective contextual retrieval (Cutrell, et al, 2006) via our CAF.

Our initial framework has been designed to support biology students engaged in a Local Flora course that includes field study and laboratory work, although are framework and many of the components can be generalized to support students in other disciplines of study. There are three classes defining the metadata framework's architecture; they are: *main tasks*, *object types*, and *object formats* (Table 1).

Table 1: M^2 Metadata Framework

Main tasks ▪ Determine students' educational context/address, by answering the following questions: ≈ What is the student's location? (Physical location, from country to campus room number) ≈ What activity are they engaged in? (A meeting, study group, class, lecture, lab) ≈ Who they are with? (Students, faculty, staff, community member, professional) ≈ What are the student's cognitive state and personal behavioral characteristics? (Examples: Happy, Sad, Stressed, Busy, Early, Late)
Object types ▪ Activities: Lecture, Conversation, Project ▪ Documents: Slides, Handouts, Student notes Quizzes, Exams, Papers. ▪ People (Student, faculty, staff, community member, professional) ≈ Current emphasis is on biology students in field and lab biology classes, but the implementation can be extended to university students, faculty, and staff in other disciplines.
Object formats ▪ Audio, video, images, etc.

M^2 project metadata needs to be generated:

- Automatically when an object or happening is constructed.
- Harvested from existing educational resources (e.g., Teachers slides, or an assigned reading).
- Derived from system protocols when objects are stored in the PIM.
- Manually created by students to support future use of their digital store.

Although our work has been informed by a many developments, two key resources include: IEEE Learning Object Model (LOM) (IEEE, 2002) and Lagoze's work on event metadata (2000). Table 2 presents the base component metadata for documents and events. This foundation metadata module can be enhanced with other parts of our framework. Our goal is to generated as many of the elements via automatic means. The base component requires student input for only two metadata elements to initiate automatic metadata generation for other elements.

Key

Generation methods
A = Automatically generated
D = Derived
H = Harvested
M = Manually generated

Requirement
R = Required
O = Optional metadata
*The R/O designation only refers to manually generated metadata

Example 2: M^2 Base-level Metadata Component for Documents and Events

Element Name	Description	Example	Code
Object collector/ owner	Name of person who will store memory in MLB	Doe, Jane	D
Class Dept. Code	Four letter department code	BIOL, INLS	M/R
Class Number	Three digit class number	096, 156, 157	M/R
Class Section	Two digit code	01	A
Class Name	Course name	Local Flora	A
Professor	Last name and first name combination.	Smith, Paul	A
Scope Note	Automatic summary or keywords	Fieldtrip report from the arboretum tour (Automatically extracted from document text)	A,D, + H
Annotation	Field used by student at their discretion.	Focus of fieldtrip was gymnosperms	M/0
Starred Item	Designates an implied importance to the content	Binary value: Star or no star.	M/0
Location	Location where data type was created.	Derived using GPS of RFID	A

6. Conclusion and future work

This paper reviewed the M^2 project. Attention was given to learning, memory, and reflective learning strategies and M^2's metadata challenges. A reflective learning scenario for a plant biology class was presented, and the CAF and M^2 metadata framework were reviewed

Our current research focuses on metadata for objects supporting undergraduate biology students engaged in classroom activities and fieldwork. Although we are in the early stages of the M^2 project, we have been able to identify future research questions and directions. Future research goals include:

- Incorporating additional mobile devices (e.g., mobile phones and PDAs,) and sensors (e.g., environmental sensors) into our research design.
- Extending our work via the CAF and M^2 framework to other disciplines and communities.

- Identifying effective and efficient metadata generation patterns and sequences (e.g., description of event 1, followed by description of event 2, and so forth).

In relation to the last goal, we want to explore "When, during an activity, is the best time to automatically capture or provide an interface for manually creating metadata or annotation?"

In the larger world of e-learning and CAF developments, we are keeping abreast of how the CAF and other M^2 components might be used in the European Convergence process in Higher Education, where the ECTS credit system[12] is based on student workload. We are eager to connect our work to evaluation efforts measuring effective learning outcomes, particularly work that focuses on reflective learning and the use of pervasive technology.

In closing, the higher education community needs to embrace pervasive computing and learn how to connect new technologies and learning science knowledge to accelerate and improve education. The M^2 project is contributing to this goal by examining how metadata can support reflective learning and PIM for undergraduate students.

References

Bloom, B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

Bush, V. As We May Think (1945). The Atlantic Monthly, July 1945. Available at: http://www.theatlantic.com/doc/194507/bush.

Cutrell, E., Dumais, S. T. & Teevan, J. (2006). Searching to eliminate personal information management. *Communications of the ACM*, (49)1: 58-64.

Gilliland, A. J. (2005). Setting the Stage. In M. Baca (Ed.) Introduction to Metadata: Pathways to Digital Information, version 2.1. Los Angles, CA: Getty Information Institute: http://www.getty.edu/research/institute/standards/intrometadata/index.html.

Gilmer, P. J., Alli, P. (1997). Action experiments: are students learning physical science? Students as researchers: Creating classrooms that matter, In *Students as researchers: Creating classrooms that matter.* Steinberg, S.R. and Kinchloe, J. L. (Eds.). London: Falmer Press, pp. 199-211.

Greenberg, J. (2005). Understanding Metadata and Metadata Schemes. *Cataloging & Classification Quarterly*, 41(3/4): 1-27.

Huber, R.L. (1993). Memory is not only about storage. In D.D. Flannery (Ed.), Adult Development and Aging: A Life Span Perspective. New York: McGraw-Hill.

IEEE. (2002). IEEE Standard for Learning Object Metadata Standard. Institute of Electrical and Electronics Engineers, Inc.: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.

Johnson, G. & DiBiase, D. (2004). Keeping the horse before the cart: Penn State's e-portfolio initiative. *Educause Quarterly*, n. 4, pp. 18-26. Available: http://www.educause.edu/ir/library/pdf/eqm0443.pdf.

[12] European Credit Transfer System: http://europa.eu.int/comm/education/programms/socrates/ects_en.html.

Lagoze, C. (2000). Business Unusual: How "Event-Awareness" May Breathe Life Into the Catalog? *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web*, Library of Congress, Washington, D.C., November 15-17, 2000: http://www.cs.cornell.edu/lagoze/papers/lagozelc.pdf..

McCormick, C. B. (2003). Metacognition and Learning, in William M. Reynolds and Gloria E. Miller, eds., Handbook of Psychology, vol.7, Educational Psychology, pp. 79-102 New York: Wiley.

Norman, G. R. And Schmidt, H. D. (2005). The Psychological Basis Of Problem-Based Learning: A Review Of The Evidence: http://pblkurs.psi.uni-heidelberg.de/pbl_norman/pbl_norman.PDF.

Roth, W. M. (1996). Teacher questioning in an open-inquiry learning environment: interactions of context, content, and student responses, *Journal of Research in Science Teaching,* 33: 709-736.

Acknowledgements
We would like to acknowledge Microsoft Research and UNC's Information Technology Services for their support and sponsorship of this research.

APPENDIX A: SNAPSHOTS. MICROSOFT RESEARCH'S MYLIFEBITS (MLB) DIGITAL MEMORIES SHELL

Figure 1. Memex Annotation Metadata

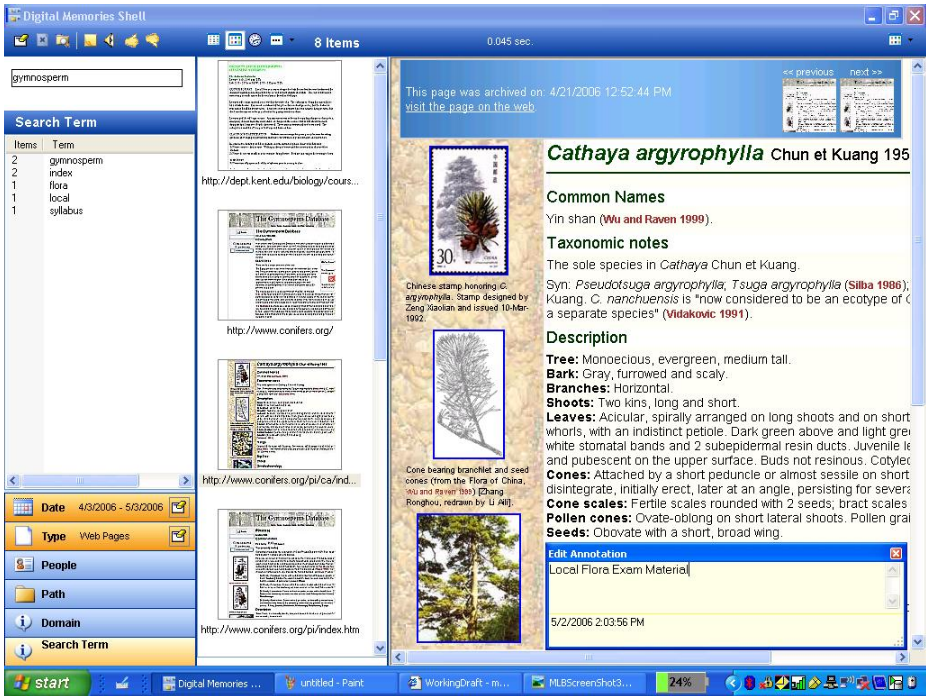

Figure 2. Memex Contextual Retrieval Example of SenseCam Images and Digital Photograph of Plant Specimen