LIS S686 Web Archiving and Preservation

Department of Library and Information Science
Indiana University School of Informatics and Computing, IUPUI
Fall 2021 Syllabus (last update on Aug 1 2021)

Instructor: Ayoung Yoon, Ph.D.
Email: ayouon@iupui.edu
Office Phone: 317 278-2388
Office hours: By appointment
Course site: https://iu.instructure.com/courses/2012250
Pre-requisite: S500
Students are highly encouraged to take S581 Archives and Records Management before taking this course.

COURSE DESCRIPTION

The World Wide Web is the primary delivery mechanism for digital content, and as source of information and knowledge for archivists, the importance of the Web has been indisputable. This course aims to provide knowledge of the role and potential of the Web as a medium and as a source for archival collection development. Archivists working in the Internet world need to be familiar with the tools and appropriate techniques for preservation of information delivered through the “surface” Web (static Web pages, blogs, e-mail discussion lists, etc.) and information that is part of the “deep” Web (e.g., databases, streaming media, and authenticated resources). Once Web content is captured and brought into a preservation environment, preservation administrators are responsible for transforming that content into persistent formats and data structures.

Through lectures, presentations on specific topics, analysis of Web archives, and hands-on work, students will gain insight on relevant issues arising from the nature and characteristics of the Web and will learn how to make correct and fruitful use of the Web in their archival work. About half of this course will expose students to existing and emerging tools for capturing Web content, with an emphasis on hands on work and practice using the current generation of Web crawlers. Students will also learn about current preservation formats and how to migrate various types of Web-based content from their native formats to persistent formats.

COURSE LEARNING OUTCOMES with Revised Bloom’s taxonomy level

Upon completion of this course, students will

<table>
<thead>
<tr>
<th>RBT*</th>
<th>PGPL</th>
<th>ALA Comps.</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4, 5</td>
<td>1</td>
<td>2A, 2B, 2D</td>
<td>Discussion</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2A, 2D</td>
<td>Discussion, reaction paper</td>
</tr>
</tbody>
</table>
Upon completion of this course, students will

3. Analyze the challenges of acquiring, downloading, storing, and providing access to Web-based content.  4 1 2A, 2B, 2C, 2D  

4. Analyze legal and policy constraints on Web archiving.  4 1 2B, 2C  

5. Evaluate existing tools for Web archiving and preservation.  5 1 2B, 2C, 2D  

6. Create a focused archive of Web content from an understanding of existing standards and best practices for sustainability of archived Web content.  6 1, 2 2A, 2B, 2C, 2D  

*RBT: Revised Bloom’s taxonomy; PGPL: Principles of Graduate and Professional Learning

**MLS Program Outcomes**

**PG 3: Curate Collections for Designated Communities**

- Information professionals are responsible for curating various types of information and/or collections while collaborating with diverse designated communities throughout the information/data lifecycle. Curation is the selection, deselection, management, preservation, collection, and archiving of information, and it establishes, maintains, and adds values to information for present and future use. Curation is essential stewardship to enhance and maintain the long-term value of data, information, and collection.

**Core Competencies of Librarianship** as formulated by the American Library Association were approved and adopted as policy by the ALA Council, January 27, 2009. A person graduating from an ALA accredited master’s program in library and information science should know and, where appropriate, be able to employ the following:

2. Information Resources

2A. Concepts and issues related to the lifecycle of recorded knowledge and information, from creation through various stages of use to disposition.

2B. Concepts, issues, and methods related to the acquisition and disposition of resources, including evaluation, selection, purchasing, processing, storing, and deselection.

2C. Concepts, issues, and methods related to the management of various collections.
2D. Concepts, issues, and methods related to the maintenance of collections, including preservation and conservation.

Principles of Graduate and Professional Learning

This course addresses the following:

- Demonstrating mastery of the knowledge and skills expected for the degree and for professionalism and success in the fields.
- Thinking critically, applying good judgment in professional and personal situations.
- Communicating effectively to others in the field and to the general public.
- Behaving in an ethical way both professionally and personally.

How we will conduct “class”

While online courses afford great convenience, they also demand extra effort from both an instructor and students. Because there is no face time for lectures, discussions, group work, or other activities, all this must take place through the course site on Canvas. This involves extensive writing and creation of slides, videos, and other media we will use to communicate our ideas and questions. The syllabus, assignments, and many of the readings (unless available online) will be posted on Canvas. In each week I will provide slides or notes and brief “lectures” on important points (there are a few weeks of exception depending on the topic), but much of the work of the course will take place in the forums via discussion of the readings, videos, and slides, and your own work (discussion, assignments, quiz, and exercises).

Software requirements for online class

In your LIS courses you will need to create documents or presentations, manipulate spreadsheets, and more.

You have three sources for software: IUware, IUanyWare, Office 365

- **IUWare** ([https://iuware.iu.edu/](https://iuware.iu.edu/)) allows students, faculty, and staff to download software at no charge. See: [What is IUware?](https://kb.iu.edu/d/agze)
- **IUanyWare** ([https://uits.iu.edu/kuanyware](https://uits.iu.edu/kuanyware)) uses a web browser or mobile app to run certain IU-licensed software applications without your needing to install them on your device. See: [About IUanyWare](https://kb.iu.edu/d/bbbr) [Note: You will be asked to download and install Citrix Receiver the first time you use the full service.]
- **Office 365** ([https://uits.iu.edu/office365](https://uits.iu.edu/office365)) is a subscription-based service free to all IU currently enrolled students that provides multiple options for accessing the newest versions of Microsoft Office. See: [About Microsoft Office 365 at IU](https://kb.iu.edu/d/bexq)
- For more details, see [How to get university-licensed software at IU?](https://kb.iu.edu/d/acl0)
- Technical Assistance. For help with Canvas technical issues or if you are having difficulty with any of the technology used for the course, check the materials at [https://kb.iu.edu/d/bc11](https://kb.iu.edu/d/bc11)
Readings and Sources

Readings are on Canvas or links on the World Wide Web. It is expected that students will have read the materials every week, as we will be referring to them in lectures and in the exercises.

Required reading materials:

- This class has one required Textbook.
  - As we all are having different types of difficulties influenced by Covid-19, which may also include some financial concerns, **I would not require students to purchase the book this semester.** I will provide a copy of chapters from the previous edition as necessary.
  - As a substitute, I also assigned Julien Masanes (Ed)’s Web Archiving, Springer, 2006, as IUPUI library has eBook. It is a bit outdated, but I’ll complement the time gap with my lecture notes as well as a few chapters from Adrian Book.

- Some readings are available online and URLs are offered in the agendas.
- Required readings have been selected to encourage participation in class discussion forums and assignments. All required and recommended readings are available on the Canvas worksite for S686.

Useful Resources (journals, conference proceedings, and professional organizations):

- PoWR, the Preservation of Web Resources Handbook. The JISC-PoWR Team. (2008).
- Web Archiving. Alex Bal. UKOLN, University of Bath.
- International Internet Preservation Consortium (IIPC). [http://netpreserve.org/about/index.php](http://netpreserve.org/about/index.php) (Links to an external site.)
- Web Archiving Bibliography, Austrian On-Line Archive. [http://www.ifs.tuwien.ac.at/~aola/links/WebArchiving.html](http://www.ifs.tuwien.ac.at/~aola/links/WebArchiving.html) (Links to an external site.)

Canvas

Canvas will be an integral part of S686. If you have problems accessing Canvas, please contact the University Information Technology Services (UITS) Support Center at 317-274-HELP. All course announcements will be found in Canvas along with the class syllabus, course schedule, exercises for grade, practice exercises, and other course documents.
Assignments and Grading

Grades will be based on class participation (including assigned tasks), discussion, and exercise with a series of assignments. All assignments are due **every Tuesday midnight (11:59 pm)**. **When applicable, a response to other students are due after one day, Wednesday midnight (11:59 pm).**

Grades will be based on various types of assignments.

1. Discussions (2): 11 points
2. Review of Web archives: 10 points
3. Archiving tool analysis: 6 points
4. Selection/Scope exercise: 10 points
5. Metadata update: 8 points
6. Archive-It report review: 5 points
7. Preservation plan: 14 points
8. Reaction paper: 8 points
9. Visual presentation and peer evaluation: 13 points
10. Final web crawl report: 15 points

Evaluation

The final letter grade will be assigned based on the following range with a total of 100 points possible:

- 96-100 = A
- 84-86 = B
- 60-69 = D
- 90-95 = A-
- 80-84 = B-
- Below 60 = F
- 87-90 = B+
- 70-79 = C

The meaning of the letter grades follows the DLIS Grading Policy:

**A:** Outstanding achievement. Student performance demonstrates full command of the course materials and evinces a high level of originality and/or creativity that far surpasses course expectations. The grade of A+ is not granted in DLIS, except in very exceptional cases.

**A-:** Excellent achievement. Student performance demonstrates thorough knowledge of the course materials and exceeds course expectations by completing all requirements in a superior manner.

**B+:** Very good work. Student performance demonstrates above-average comprehension of the course materials and exceeds course expectations on all tasks defined in the course syllabus.

**B:** Good work. Student performance meets designated course expectations, demonstrates understanding of the course materials, and has performed at an acceptable level.
B-: Marginal work. Student performance demonstrates incomplete understanding of course materials.

C: Unsatisfactory work and inadequate understanding of course materials.

D: Unacceptable work; course work completed at this level will not count toward the MLS degree.

F: Failing. May result in an overall grade point average below 3.0 and possible removal from the program.

Note: The course instructor will issue the grade F if a student does not finish all required assignments.

A final grade of "I" or "Incomplete" will NOT be given except in extreme situations.

Inform the instructor if you are having difficulty completing the requirements of this course.

**Late Submissions**

In fairness to students who turn in assignments on time, late papers will be penalized. The earned grade will be lowered one grade level (e.g., from A- to B+) for each day that the assignment is late.

*Important note on plagiarism and academic ethics*

Unless otherwise specified in an assignment, all submitted work must be your own, original work. Any experts from the work of others must be clearly identified as a quotation, and a proper citation provided. Be aware of the IUPUI policy on plagiarism. All cases of plagiarism (unattributed quotation or paraphrasing) of anyone else's work, (e.g. from published materials) will be officially reported by academic misconduct. Indiana University and the Department of Library and Information Science policies on academic dishonesty will be followed.

See the source to learn more about

- Plagiarism: [https://wts.indiana.edu/writing-guides/plagiarism.html](https://wts.indiana.edu/writing-guides/plagiarism.html) (Links to an external site.)
- Code of student rights, responsibilities, & conduct: [http://studentcode.iu.edu/](http://studentcode.iu.edu/) (Links to an external site.)

Students found to be engaging in plagiarism, cheating, and other types of dishonesty will receive a failing grade for the course.

*AES and Special needs*

If you need an accommodation for a disability or have any other special need, please make an appointment to discuss this with me. I will be most able to address special circumstances if I
know about them early in the semester. My office hours and contact information are listed at the beginning of this syllabus.

Accommodations will be made for qualified students with disabilities registered with Adaptive Educational Services. The AES office is located in Taylor Hall, UC 100. The phone numbers are 317/274-3241 or 317/278-2050 TTD/TTY, by fax 317/274-2051, or by email aes@iupui.edu.

ELECTRONIC PORTFOLIO PARTICIPATION

Students admitted to the DLIS, program beginning fall, 2011 and forward are required to participate in the electronic portfolio or ePortfolio system. In this site, you will upload materials from courses that address each of the MLS program goals. Talk to your instructor for guidelines and suggestions for submission of course work to this system.

ONLINE WRITING CENTER

The Online Writing Consulting is an experimental service that will allow you to meet with a writing center tutor in cyberspace if you are taking one of the selected online courses. https://liberalarts.iupui.edu/uwc/pages/onlineconsulting-folder/index.php (Links to an external site.)

ADMINISTRATIVE WITHDRAWAL

A basic requirement of this course is that you will participate in class and conscientiously complete writing and reading assignments. Keep in touch with me if you are unable to attend class or complete an assignment on time. If you miss more than half our class assignments without contacting me, you will be administratively withdrawn from this section. Administrative withdrawal may have academic, financial, and financial aid implications. Administrative withdrawal will take place after the full refund period, and if you are administratively withdrawn from the course you will not be eligible for a tuition refund. If you have questions about the administrative withdrawal policy at any point during the semester, please contact me." Learn more at https://studentcentral.iupui.edu/register/withdraw-from-iupui/index.html (Links to an external site.)

Course Overview at a Glance

Module 1. Overview of the course; Introduction to the Web

- Introduction to the class
  - Course logistics
  - Introduction of the students and instructor
  - Review of syllabus and assignments
- Internet and the World Wide Web: Challenges and Potentials for Archives
- Introduction to the concept of web archiving

Module 2. Basic concepts in Web Archiving
• What is web archiving?
• Web curation model
• Surface Web and deep Web

Module 3. Archiving different types of Web materials and social memory

• Types of materials on the Web
• Who does archives Web? Why they do? What are the needs?
• Web and organizational memory
• Archiving personal web & social media

Module 4. Exploration of existing web archives

• Review of existing web archives
• What are the different approaches current web archives take?

Module 5. Web Archiving technology

• Dynamics of websites and different technology
• Introduction to different web archiving tools
  o DeepArc (National Library of France)
  o Heritrix
  o HTTrack
  o PageVault
  o SIARD
  o WebCite
  o Archive-It
  o Web Curator Tool
  o WebRecorder (Rhizome)

Module 6. Selection & Scoping

• Selection policy and criteria
• Different selection approaches: domain, topic or event, media type and genre based

Module 7. Acquisition / Metadata & Description

• Dynamics of websites and different technology
• What information should be provided? (Metadata!)
• How are users of web archives’ needs different from users of active websites?

Module 8. Archiving Deep Web Content

• Review from Week 2: surface web and deep web
• Challenges, strategies and issues of archiving deep web
• Database archiving, transactional archiving
• Tools for capturing surface and deep web content

Module 9. Quality Control and Post Capture Processing
• Why is post-collection processing necessary?
• Different methods of quality control (types of tests)

Module 10. Preserving Web Sites

• Challenges of preserving websites (technical, financial, and organizational)
• Strategies (passive vs. active; emulation vs. migration)
• Significant properties of web

Module 11. Access and Use

• Web archives access tools
• What information should be provided? (Metadata!)
• Who uses web archives? How are their needs different from users of active websites? How are web archives used?

Module 12. Legal and Ethical Issues

• Intellectual properties
• Privacy
• Content reliability
• Human rights

Module 13. Visual Presentations of Crawls

Module 14. Final project

CODE OF CONDUCT

All students should aspire to the highest standards of academic integrity. Using another student’s work on an assignment, cheating on a test, not quoting or citing references correctly, or any other form of dishonesty or plagiarism shall result in a grade of zero on the item and possibly an F in the course. Incidences of academic misconduct shall be referred to the Department Chair and repeated violations shall result in dismissal from the program.

All students are responsible for reading, understanding, and applying the Code of Student Rights, Responsibilities and Conduct and in particular the section on academic misconduct. Refer to The Code > Responsibilities > Academic Misconduct at http://www.indiana.edu/~code/ (Links to an external site.). All students must also successfully complete the Indiana University Department of Education “How to Recognize Plagiarism” Tutorial and Test. https://www.indiana.edu/~istd (Links to an external site.) You must document the difference between your writing and that of others. Use quotation marks in addition to a citation, page number, and reference whenever writing someone else’s words (e.g., following the Publication Manual of the American Psychological Association). To detect plagiarism instructors apply a range of methods, including Turnitin.com. http://www.ulib.iupui.edu/libinfo/turnitin (Links to an external site.)
Academic Misconduct:

1. **Cheating:** Cheating is an attempt to use or provide unauthorized assistance, materials, information, or study aids in any form and in any academic exercise or environment.
   1. A student must not use external assistance on any “in-class” or “take-home” examination, unless the instructor specifically has authorized external assistance. This prohibition includes, but is not limited to, the use of tutors, books, notes, calculators, computers, and wireless communication devices.
   2. A student must not use another person as a substitute in the taking of an examination or quiz, nor allow other persons to conduct research or to prepare work, without advanced authorization from the instructor to whom the work is being submitted.
   3. A student must not use materials from a commercial term paper company, files of papers prepared by other persons, or submit documents found on the Internet.
   4. A student must not collaborate with other persons on a particular project and submit a copy of a written report that is represented explicitly or implicitly as the student’s individual work.
   5. A student must not use any unauthorized assistance in a laboratory, at a computer terminal, or on fieldwork.
   6. A student must not steal examinations or other course materials, including but not limited to, physical copies and photographic or electronic images.
   7. A student must not submit substantial portions of the same academic work for credit or honors more than once without permission of the instructor or program to whom the work is being submitted.
   8. A student must not, without authorization, alter a grade or score in any way, nor alter answers on a returned exam or assignment for credit.

2. **Fabrication:** A student must not falsify or invent any information or data in an academic exercise including, but not limited to, records or reports, laboratory results, and citation to the sources of information.

3. **Plagiarism:** Plagiarism is defined as presenting someone else’s work, including the work of other students, as one’s own. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged, unless the information is common knowledge. What is considered “common knowledge” may differ from course to course.
   1. A student must not adopt or reproduce ideas, opinions, theories, formulas, graphics, or pictures of another person without acknowledgment.
   2. A student must give credit to the originality of others and acknowledge indebtedness whenever:
      1. directly quoting another person’s actual words, whether oral or written;
      2. using another person’s ideas, opinions, or theories;
      3. paraphrasing the words, ideas, opinions, or theories of others, whether oral or written;
      4. borrowing facts, statistics, or illustrative material; or
      5. offering materials assembled or collected by others in the form of projects or collections without acknowledgment

3. **Interference:** A student must not steal, change, destroy, or impede another student’s work, nor should the student unjustly attempt, through a bribe, a promise of favors or threats, to affect any student’s grade or the evaluation of academic performance. Impeding another student’s work includes, but is not limited to, the theft, defacement, or mutilation of resources to deprive others of the information they contain.

4. **Violation of Course Rules:** A student must not violate course rules established by a department, the course syllabus, verbal or written instructions, or the course materials
that are rationally related to the content of the course or to the enhancement of the learning process in the course.

5. **Facilitating Academic Dishonesty:** A student must not intentionally or knowingly help or attempt to help another student to commit an act of academic misconduct, nor allow another student to use his or her work or resources to commit an act of misconduct.

**OTHER CAMPUS RESOURCES**

1. **Counseling and Psychological Services (CAPS):** Students seeking counseling or other psychological services should contact the CAPS office at 274-2548 or capsindy@iupui.edu. For more information visit [https://studentaffairs.iupui.edu/health/counseling-psychological/index.html](https://studentaffairs.iupui.edu/health/counseling-psychological/index.html) (Links to an external site.).

2. **Disabilities policy:** All qualified students enrolled in this course are entitled to reasonable accommodations for a disability. Notify the instructor during the first week of class of accommodations needed. Students requiring accommodations register with Adaptive Educational Services (AES) and complete the appropriate AES-issued before receiving accommodations. The AES office is located at UC 100, Taylor Hall (Email: aes@iupui.edu, Tel. 317 274-3241). For more information visit [http://aes.iupui.edu](http://aes.iupui.edu) (Links to an external site.).

3. **Emergency preparedness:** Know what to do in an emergency so that you can protect yourself and others. For more information, visit the emergency management website at [http://protect.iu.edu/emergency](http://protect.iu.edu/emergency) (Links to an external site.).

4. **IUPUI course policies:** Several campus policies governing IUPUI courses may be found at the following link: [https://policies.iu.edu/categories/academic-faculty-students.html?](https://policies.iu.edu/categories/academic-faculty-students.html?) (Links to an external site.)

5. **Sexual misconduct:** IU does not tolerate sexual harassment or violence. For more information and resources, visit [http://stopsexualviolence.iu.edu](http://stopsexualviolence.iu.edu) (Links to an external site.).

6. **Student advocate:** The Student Advocate assists students with personal, financial, and academic issues. The Student Advocate is in the Campus Center, Suite 350, and may also be contacted at 317 274-4431 or studvoc@iupui.edu. For more information visit [https://studentaffairs.iupui.edu/advocacy-resources/index.html](https://studentaffairs.iupui.edu/advocacy-resources/index.html)