

## HIST 390 MIDTERM EXAM (KEY)

	PART ONE (20pts)	PART TWO (30pts)	MIDTERM (50pts)
A+	19.50	29.25	48.75
A	18.50	27.75	46.25
A-	18.00	27.00	45.00
B+	17.50	26.25	43.75
B	16.50	24.75	41.25
B-	16.00	24.00	40.00
C+	15.50	23.25	38.75
C	14.50	21.75	36.25
C-	14.00	21.00	35.00
D+	13.50	20.25	33.75
D	12.50	18.75	31.25
D-	12.00	18.00	30.00

The whole exam and each part of the exam will have both a letter grade and a number grade. The numeric grade must be *at least* the number in the cell in order to earn that letter grade. A 44.99 score on the exam, for instance, would earn a B+.

**PART ONE (20 points):** Pick 20 out of the following 30 questions and give a definition and/or example for each one. If the question includes an acronym, please also say what the acronym stands for (no points off if you can't, but it will help your score on the question if your definition is iffy). Note: if you answer more than 20 questions, we will grade only the first 20.

- 1) What is an archive?

An archive is a collection of historical materials that have been preserved for research. One example is the National Archive. Source: classroom discussion.

- 2) What is Creative Commons?

Creative Commons is a project whose purpose is to enable people to put licenses on their creative work so that others know how (and whether) they can re-use it without asking permission. Creative Commons licenses are human-readable, lawyer-readable, and machine-readable. An example of a Creative Commons license is a CC-BY license, which means that the item can be re-used as long as it is properly attributed to its creator. Source: classroom discussion, *Digital History* "Owning the Past - Protecting Your Intellectual Property."

- 3) What is a client?

A client is "the computer requesting a file from the server." "Client" can also mean the software on the client computer. An example of a client is my own laptop (or the web browser on my own laptop, such as Firefox) requesting a web page such as <http://google.com> from Google's server. Source: *Digital History*

“Getting Started - The Web, Websites, and Web Pages.”

- 4) What is crowdsourcing?  
Crowdsourcing is using the Internet to ask a large number of people to do work for free. Wikipedia is an example of a crowdsourced project. Source: classroom discussion.
- 5) What is a database?  
A database is a structured set of data that can be queried. According to *Digital History*, client-based database programs include MS Access and Filemaker, while server-based database programs include Oracle and PostgreSQL. Source: *Digital History* “Getting Started - Databases and XML” and Appendix (not assigned).
- 6) What is the DNS?  
The Domain Name System is the technology that translates numeric IP addresses such as 128.0.0.14 to alphanumeric domain names / URLs such as google.com. Source: *Digital History* “Getting Started - Naming Your Site.”
- 7) What is a domain name?  
A domain name is the main part of the alphanumeric “web address” or URL that you type in to get to a web page. An example of a domain name is google.com. Strictly speaking (according to *Digital History*), a domain name consists of the “second-level domain” (google) plus the “top-level domain” (.com). In the URL <http://groups.google.com>, “groups” is a “subdomain” and is not part of the domain name “google.com.” Source: *Digital History* “Getting Started - Naming Your Site.”
- 8) What is “dpi”?  
“DPI” is “dots per inch.” It is a way of measuring the resolution of a digital image. For example, a high-resolution digital image may be scanned at 600 or 1200 dpi. Source: *Digital History* “Becoming Digital - Digital Images.”
- 9) What is fair use?  
Fair use is the part of copyright law that allows people to copy materials without permission. The four factors that determine whether a use is fair concern the nature of the work, the purpose of the use, the amount used, and whether the use adversely affects the market for the original. Source: class lecture, *Digital History* “Owning the Past? - Fair Use.”
- 10) What is the “history” of a Wikipedia article?  
On a Wikipedia article, clicking the “History” or “View History” tab (depending on the version of Wikipedia) will let you view all the different versions of that article with the changes highlighted. Source: class lecture / exercise.
- 11) What is an HTML tag?  
HTML stands for Hypertext Markup Language. A tag in HTML is a piece of

code. An example of an HTML tag is <p>. Source: class lecture.

12) What is ICANN?

ICANN is the Internet Corporation for Assigned Names and Numbers. It is the organization that manages domain assignments. Source: class lecture and *Digital History* "Getting Started - Naming Your Site."

13) What is an IP address?

An IP address is an Internet Protocol address such as 160.111.76.139. It is the numeric "address" for a computer on the network. Source: class lecture and *Digital History* "Getting Started - Naming Your Site."

14) What is the memex?

The memex is the associative information retrieval device imagined by Vannevar Bush in his 1945 article "As We May Think." He describes it thusly: "A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility." It is often cited as a kind of precursor to the Internet. Source: "As We May Think."

15) What is metadata?

Metadata is data about data. It is descriptive information about an object or file, such as the title and author of a book (or digital book). Source: class lecture (it's in *Digital History* but not in assigned parts).

16) What is OCR?

OCR is Optical Character Recognition. It is the technology that enables digital images of text to be transformed into machine-readable digital text. Source: *Digital History* "Becoming Digital - How to Make Text Digital."

17) What is a pixel?

A pixel is one unit (dot) of a digital image or digital screen. Images and screens are measured in pixel height by pixel depth, as in an image measuring 900 pixels by 1500 pixels. Source: *Digital History*, "Becoming Digital - Digital Images."

18) What is the public domain?

The public domain is "the realm free of copyright restrictions." When an item is in the public domain, it can be shared, reused, adapted, and sold without restriction. Source: *Digital History*, "Owning the Past - Sharing the Property of Others."

19) What is RSS?

RSS is Really Simple Syndication. It is a technology that enables people to "subscribe" to a frequently updated website or blog with an RSS reader. Source: "RSS in Plain English."

20) What is a “salted” password?

A “salted” password is a password with characters added to the beginning of it before it has been run through a hash (encoding technique). This means that people can’t “reverse engineer” the password by looking it up in hash tables that list what common passwords look like when encoded by common techniques. Source: “Passwords Under Assault” p. 3 (quote: “Salting appends several unique characters to each account password before running it through a cryptographic function, a process that blunts the value of rainbow tables and other types of precomputed attacks”).

21) What is a server?

A server is a computer that “serves” files (including web pages) to the Internet. Source: *Digital History*, “Getting Started - The Web, Websites, and Web Pages.”

22) What is the Sonny Bono Copyright Term Extension Act?

The Sonny Bono Copyright Term Extension Act is an act passed in 1998 that extended the time in which an item published in the U.S. remains under copyright by 20 years, from 75 years after the death of the author to 95 years after the death of the author. Source: *Digital History*, “Owning the Past? - A Brief History of Copyright.”

23) What is TCP/IP?

TCP/IP is the transmission protocol (i.e., standard) that allows information packets to travel between local networks. Source: “History of the Internet.” (It stands for Transmission Control Protocol / Internet Protocol, but that’s not mentioned anywhere in the readings.)

24) What is TEI?

TEI is the Text Encoding Initiative. It is set of guidelines for marking up texts, first in SGML (Standard Generalized Markup Language) and then in XML (eXtensible Markup Language). Source: *Digital History*, “Becoming Digital - To Mark Up, or Not to Mark Up.”

25) What is a TIFF?

TIFF is a digital image format. Uncompressed TIFF files are very large, and they tend to be the “industry standard” for high-quality (especially archival) images. It stands for Tagged Image Format. Source: *Digital History*, “Becoming Digital - Digital Images.”

26) What is a web host?

A web host is the organization (or server) where a web site is stored and delivered to the web. An example of a commercial web host is GoDaddy, which will sell server space to individuals and organizations who want to build web sites. ISPs and universities can also host websites. Source: *Digital History*, “Getting Started - Serving Your Website” and class lecture.

27) What is a web page?

A web page is a file written in HTML. When published, the page will have a single URL such as <http://example.com/webpage.html>. Source: *Digital History*, “Getting Started - The Web, Websites, and Web Pages.” (Quote: “Although you can produce a website in many ways, beginners should recognize that a website is basically a collection of web pages, and a web page is simply a file produced or stored on one computer in a particular format that is sent to another computer that has requested it.”)

28) What is a web site?

A web site is a collection of web pages. See “web page.”

29) What is a WYSIWYG interface?

A WYSIWYG interface is a “What You See Is What You Get” interface. The term is usually used for software interfaces that help you write web pages without writing HTML. Source: *Digital History*, “Getting Started - Text and Images.”

30) What is XML?

XML is eXtensible Markup Language. It is a markup language rather like HTML that allows you to wrap textual elements in metadata within one text file. Source: *Digital History* “Getting Started - Databases and XML” and Appendix (not assigned). (Quote: “XML is much like an HTML document: pure text with tags surrounding words or passages, in this case, representing the definitions such as author or date.”)

**PART TWO (30 points):** Pick one of the following three questions and respond in an essay.

Essays are graded on two criteria: 1) How well the essay makes a coherent overall point and supports it with evidence, while still acknowledging drawbacks and opposing points of view (in other words how well-written and lucid it is); and 2) how many of the specific issues listed below the essay addresses. If other issues are raised besides those listed below, they should be a) relevant to the question, and, ideally b) derived from the course reading.

- 1) You are an archivist at the National Archives and have been assigned to digitize a collection of items related to a fairly obscure nineteenth-century American Civil War colonel. The collection includes 10 boxes of records (hand-written receipts, legal contracts, and similar items), 1 box of maps, 1,000 handwritten letters, 100 photographs, and 500 nineteenth-century books from the colonel's library. Your budget will only allow you to digitize some of these items, and although the Archives has web developers and programmers on staff, they are working on other projects and can't give you more than a few hours per week of help. Your goal is to help military historians and members of the public understand this colonel's historical importance. Write an essay describing how you would go about fulfilling this assignment. What factors will help you decide what to digitize? How will you go about digitizing as much as you can with as little money and programming help as you can? Where will people be able to find these digital items on the Internet? What information about the items will you make available?
  - a. Factors: expense of digitizing -- it will be more expensive to do full-text scanning of books than image scanning, and even more expensive to do full-text scanning with OCR and then get the OCR corrected. It will also be expensive to transcribe the letters.
  - b. Factors: difficulty of digitizing -- the books and images will be the easiest to scan, while the records and letters will be the hardest. (More on maps later in the semester.)
  - c. Factors: existence of other online resources -- many of the books may already be online, especially because they are in the public domain, since they were published before 1923.
  - d. Factors: historical importance of the items -- the question doesn't tell you which items have the most important or interesting content, but the essay should probably mention this as an issue.
  - e. Strategies: crowdsourcing -- the Archives already has a Citizen Archivist crowdsourcing program that could help get volunteers to transcribe the letters from images for free.
  - f. Finding items on the Internet -- digital images of the photographs (or even images of the documents) from this collection could be placed on Flickr Commons or Wikimedia Commons as well as on the National Archives's own website. The collection itself should be findable in Archive Finder.

- g. Information about the items -- most items should have title, author/creator, date, names of people mentioned or portrayed, names of relevant places.
  - h. Sources: the essay might mention *Digital History*, especially the chapter on “Becoming Digital.”
- 2) You are writing a historical article about women’s role in the civil protests in East Germany that led to the fall of the Berlin Wall in 1989. You are particularly interested in a particular East German protestor named Marta Brüler. The Wikipedia article about Marta Brüler displays a photograph of her carrying stones or bricks to a pile, and states that she was one of the first people to begin dismantling the wall physically. A newspaper article in the *New York Times* published on November 9, 1990 (the one-year anniversary of the Fall) shows the same photograph, and implies that Brüler was not dismantling the wall, but helping to construct a new crossing from East to West Germany. What do you do and why?
- a. Check the Wikipedia “History” page to find out when the photo about Brüler and its accompanying claim was added and who added it.
  - b. Raise the issue on the Wikipedia “Talk” page
  - c. If possible, contact the person who added the photo to the “Talk” page
  - d. Edit the Wikipedia article to take out the claim
  - e. Add the 1990 *New York Times* article to the References section of the Wikipedia page
  - f. Try to track down the original photograph
  - g. Consult photography experts to see if what’s happening in the photo can be determined from objective evidence
  - h. Go to Berlin and see if you can find the exact place in the photograph and see if that tells you anything
  - i. Sources the essay might mention: Jon Udell, “Heavy Metal Umlaut,” Errol Morris, “Which Came First?”
- 3) You are building a website about the history of North Dakota, and you want to include photographs of North Dakota from all periods from 1889 to the present. How do you go about finding photographs you can use on your website, and how do you determine whether you are allowed to use them?
- a. Source: Creative Commons Search
  - b. Source: Flickr Commons (and Flickr)
  - c. Source: Wikimedia Commons
  - d. Source: Internet Archive
  - e. Source: Archive Finder
  - f. Source: Google Images Advanced Search
  - g. Public domain - anything published in the U.S. before 1923 can be used
  - h. Fair use - essay should discuss all four factors (listed above) to determine whether the use is fair
  - i. Creative Commons license: Photos marked with a CC license can be used under the stipulated conditions

- j. Asking the creator for permission (and possibly paying for it)
- k. Taking the photographs yourself
- l. Sources: the essay could mention *Digital History*, “Owning the Past”; “Copyright Criminals,” and “The Amen Break.” Also class discussion and/or the image-finding resources listed on the course blog.