Internet Search Strategies

Identifying Web Sites

The World Wide Web is a system of Internet servers that supports hypertext and multimedia to access several Internet protocols on a single interface. The World Wide Web is often abbreviated as the WEB or www.

The World Wide Web was developed in 1989, and now consists of hundreds of millions of electronic documents called web pages residing on computers all over the world. Many pages have built-in links to other documents. A collection of related Web pages is a Web site, which is stored on a computer Web server. Most Web sites have a starting place, called a home page, which is like a table of contents for the site.

Each Web page has a unique address, much like a postal mailing address, called the URL, or Uniform Resource Locator. These URLs can be a simple combination of letters such as *http://www.startribune.com,* or a complex combination of letters, numbers and slashes such as *http://www.senate.gov/general/capcom.htm.* Reading the URL can give a general idea of where the resource originates and who is responsible for the site. This is the format of the URL: protocol://host/path/filename.

http://www.senate.gov/general/capcam.htm

The structure of this URL is:

Protocol: **http://**
Host computer name: **www**
Second-level domain name: **senate**
Top-level domain name: **gov**
Directory name: **general**
File name: **capcam.htm**
(Note how much information about the content of the file is present in the URL)

The top-level domains in the United States are:
.com   Commercial enterprise
.edu   educational institution
.gov   U.S. government entity
.mil   U.S. military entity
.net   network access provider
.org   usually nonprofit organization

Additional domain names you might see are .biz, .museum, .info, .pro (for professionals) .name (for individuals), .aero (for the aerospace industry), and .coop (for cooperatives)

In addition, dozens of domain names have been assigned to identify and locate files stored on servers in countries around the world. These are referred to as country codes, and have been standardized by the International Standards Organization.
Internet Search Tools

Subject Directories

Directories are very useful when you have a general idea of where you want to start. The first page usually gives you the most general or broadest categories, like *Computers and Internet.* You then click your way down the hierarchy to the right category, e.g. to *News and Media,* and then to *Computers and Technology Blogs*, and finally to select the Web site you find most relevant to your search.

Subject directories, unlike search engines, are created and maintained by human editors, not electronic "spiders" or "robots". The editors review and select sites for inclusion in their directories on the basis of previously determined selection criteria. The resources they list are usually annotated. Directories tend to be smaller than search engine databases, typically indexing only the home page or top level pages of a site.

Search Engines

Use a search engine when you have a unique or obscure topic or when you want to search the full-text of millions of pages. Search engines are "robots" (software) that read web pages and put parts of the text into a large database or index that you may access. Once the spiders get to a web site, they typically index most of the words on the publicly available pages at the site. None of the search engines cover the whole Internet, but some of them are quite large. Search engines should be the first choice when you know exactly what you are looking for.

Search engines check out the title field and scan the headers and text near the top of the document. Some of them assess popularity by the number of links that are pointing to sites; the more links, the greater the popularity, i.e., value of the page. On the down side, the sheer number of words indexed by search engines increases the likelihood that they will return hundreds of thousands of responses to simple search requests. Remember, they will return lengthy documents in which your keyword appears only once.

Today, the line between search engines and subject directories is blurring. Search engines no longer limit themselves to a search mechanism alone. Across the Web, they are partnering with subject directories, or creating their own directories, and returning results gathered from a variety of other guides and services as well.

Meta-Search Engines

In a meta-search engine, you submit keywords in its search box, and it transmits your search simultaneously to several individual search engines and their databases of web pages. Within a few seconds, you get back results from all the search engines queried. Meta-search engines do not own a database of Web pages; they send your search terms to the databases maintained by search engine companies. Few meta-searchers allow you to delve into the largest, most useful search engine databases. They tend to return results from smaller and/or free search engines and miscellaneous free directories, often small and highly commercial. It is recommended that you go directly to search engines to get the most precise results, and use meta-searchers if you want to explore more broadly.

Specialized Databases

These are databases that are, for example, maintained by academic institutions and government agencies. They are high quality informational sites, and sometimes the content is not indexed by general search tools. Many times the database has its own search box to search the contents of its site. Some databases are subscriptions and available to authorized users with a password like the Minnesota West Library databases.

Comparison of Search Tools

Subject Directories	Search Engines	Meta-search Engines	Specialized Databases
http://dir.yahoo.com/	*http://www.google.com*	*http://www.metacrawler.com*	*http://www.usa.gov*
http://www.ipl.org/	*http://www.ask.com/*	*http://www.dogpile.com*	*Minnesota West databases*
http://www.about.com/	*http://www.bing.com/*	*http://www.search.com*	
http://infomine.ucr.edu/	*http://www.altavista.com/*	*http://www.webcrawler.com/*	

SUBJECT DIRECTORY	SEARCH ENGINE	META-SEARCH ENGINE	SPECIALIZED DATABASES
Created and maintained by human editors who review and select sites based on certain criteria.	Created by computer robots or spiders, which travel the Web gathering and indexing information.	Sends your search query to several individual search engines at one time and compiles the results.	Databases are high quality information sites.
Arranged in hierarchical subject categories and sub-categories.	Collected information is searched with keywords or phrases.	Useful for simple searches and determining which search engine to use.	Database content is generally not indexed by general search tools.
Does not search text, but rather the title and descrip-tion of the site.	No browsing or subject categories.	Typically they only catch the top 10% of search results in any of the search engines they visit.	Some of the databases are password-protected.
Often indexes only the home page or top level pages of a site.	Cover a much larger part of the Web than the subject directories.	Searches are considered to be "quick and dirty"	
Best for browsing or searching using general terms.	Has features to narrow and refine a search.		

Constructing a Search

Identify and Develop a Topic. Make sure it isn't too broad or doesn't contain too many issues to be addressed effectively in one paper.

➢ From the library's home page, click on **Research Quickstart** to select a broad subject area you're interested in.

➢ Choose library-subscribed journal, magazine or newspaper indexes to determine if there is enough information on your chosen subject. You will also find the link to the library's online catalog of books that are both print and electronic ebooks form.

➢ Research Quickstart also has list of high-quality, reliable Internet sites to you help you in your research, as well as style manuals and writing guides.

Develop a Search Strategy. A search strategy begins with selecting keywords for your topic sentence.

Topic Sentence: College students engage in risky behavior

1. Identify synonyms to help you obtain more results. Merriam Webster's Thesaurus Online http://www.m-w.com/ offers these synonyms for risky:

   Synonyms: GRAVE, GRIEVOUS, HAZARDOUS, JEOPARDIZING, MENACING, PARLOUS, PERILOUS, RISKY, SERIOUS, THREATENING, UNHEALTHY, UNSAFE, VENTURESOME

2. Choose examples of specific risky behaviors:
   a. High-risk drinking   b. substance abuse   c. sexual risk-taking
3. Obtaining a variety of terms allows you to develop a more sophisticated search strategy.

Search Strategy: "college students" and "risky behavior"

Use quotes around exact phrases. Use AND to narrow your search to include BOTH keywords (hazardous and risky), use OR to broaden your search to include ANY of the keywords (serious or unsafe), and NOT to exclude one meaning of a word (Drugs not Prescription). In a web search you can also use + or – in front of a word saying you MUST or MUST NOT include a word. There is no space between the sign and the keyword.

Use Research QuickStart or Choose a Search Tool for the Internet

➢ When you begin a search, start with the Minnesota West Periodicals/Databases or through Research Quickstart.
➢ If you are browsing and trying to determine what is available in your subject area, start out with an Internet subject directory like *yahoo* or *About.com*.
➢ If you are looking for specific information, go to a search engine, such as *Google* or *Alta Vista.* Search engines will search for information buried in the text of web pages.
➢ If you are in a hurry, use a meta-search engine such as *metacrawler* because they are best at providing a quick overview on a subject or a unique term.

Evaluating and Refining a Search

The number of resources on the Internet is growing rapidly. The number of Web pages is too many to count, but easily in the tens of millions. There is no accurate source that can count all the pages online, it's very important to evaluate everything available on the Internet.

- Anyone can publish anything on the Web
- Web sources are often not verified by editors as are print publications
- Information on the Internet is not regulated for quality or accuracy
- Web pages are fluid. Unlike a book, the page you cite today may be altered or revised tomorrow, or it might disappear completely.

Therefore, it is necessary to learn how to judge the validity of the information you find on the Internet. Librarians and other information professionals have developed a set of criteria, the **Five Points of Reliability**, which you can use to evaluate whether a particular website has value for research.

To demonstrate these Five Points, you will be using the Website, http://www.oncolink.upenn.edu

Authority

To evaluate the authority of a Web page, you must look at both the authority of the page itself and the authority of the site on which the page resides. Before looking at the authorship of the page, it is helpful to first return to the site's home page to analyze the authority of the site.

Click on Oncolink, and then come back to read more about *authority.*

Notice how the Oncolink and University of Pennsylvania Cancer Center symbols appear together prominently at the top of the page. They also serve as links back to the Oncolink home page that, if followed, reveals that the University of Pennsylvania Cancer Center is the organization directly responsible for the contents of Oncolonk. In addition,

- The page's information is copyrighted by the Trustees of the University of Pennsylvania.
- An email address is provided for the site.
- The design consistency not only helps users recognize that they are still within the Oncolink site, but also serves as an additional aid to verifying the authority of the page.

Accuracy

Accuracy is the extent to which information is reliable and free from errors.

Go back to the Oncolink home page and notice that there are many links to news items and articles.

- By clicking on one of these you will see the article is **free from grammatical or spelling errors.** This indicates that care was taken in producing the page.
- Not only are the **sources of the factual information** named. But links are also provided to many of the original sources.
- In addition, it is clear that the Editorial Board has ultimate responsibility for the accuracy of the information provided.

Currency

This is the extent to which material can be identified as up to date.

Go to Oncolink and note the current date and last revision dates of the page.

- We can readily determine the currency of the Oncolink page by noting when the page's contents were last revised
- In addition, the current date is displayed at the top of the page.
- Each of the news items linked to the page is dated and the source is given.

Objectivity

Objectivity is the extent to which material expresses information or facts without distortion by personal feelings or other biases.

Because of the possibility of influence by an advertiser or sponsor on the objectivity of the information, it is important to first look at any advertising or sponsorship present.

- An analysis of the Oncolink home page clearly indicates the site has corporate sponsors and provides a link to the home page of each of them. With this information, a user can learn more about the nature of what type of sponsorship is being provided.
- Click of the "About Oncolink" link. This page retrieves the mission statement of the organization. The mission statement conveys the rationale behind the page and gives stated objectives.

Coverage

Coverage is the range of topics included in a work and the depth to which those topics are addressed. The intended audience is the group of people for whom the material was created.

- The disclaimer located at the bottom of the page offers some insight into the topics covered and the intended audience.
- The disclaimer states "Onclolink is designed for educational purposes only and in not engaged in rendering medical advice or professional services."
- Click on the disclaimer at the bottom of the page. The topics listed on the home page help the user determine the types of information that will be found at the site.

The next page has websites for you to practice the Five Points of Reliability on as listed above.
Practice Sites

Most Websites can be categorized into one of these six types:

- Organization/Advocacy
- Educational/Entertainment
- Informational
- News
- Personal Web page/Blog

It is important to analyze a Web site page-by-page rather than assuming that all pages at a given site will be of one type. For example, it is common to find both information and advocacy pages at the same site, and also common to find sites that have business pages combined with entertainment pages.

Personal Web pages often include entertainment items about a favorite musician, provide information about a research topic, advocate a favorite cause, and try to sell a used bicycle all at the same time.

Organization/Advocacy

An organization/advocacy site that is promoting their ideas or product, has the primary purpose of influencing public opinion. It may influence people's ideas or it may encourage activism. A single individual or a group of people may be responsible for the page.

Examples include the Democratic and Republican Parties, The National Right to Life Committee, and the National Abortion Rights Action League.

The URL frequently ends in .org (organization) if the page is sponsored by a nonprofit organization.

Go to the homepage of the Environmental Defense Fund page (http://www.edf.org/)

Apply the Five Points of Reliability:

- The URL is .org which indicates that the page may be from a nonprofit organization.
- The goals of the organization are to promote environmental solutions to today's problems.
- There is a link with membership information, encouraging you to join the cause.
- The Web site is clearly trying to influence the user's opinion.

Educational/Entertainment

An educational/entertainment Web site has the primary purpose of providing enjoyment while at the same time educating the user in some aspect.

Examples are the National Gallery of Art (.gov), Ellis Island.org, and Enature.com.

The URL may have a variety of endings depending on what type of site the page is coming from.

Go to the Smithsonian Institution, http://www.si.edu

Now apply the Five Points of Reliability:

- The URL is .edu, indicating an educational institution
- The site is clearly sponsored by the Smithsonian, which has an outstanding reputation
- While it provides enjoyment in the form of virtual tours of the collections, it also educates with a direct link to teaching resources and ideas.
- Information is factual with no advertising
- Events and exhibits are current
- Accurate, no grammatical or spelling errors.

Informational

An informational Web page is one with the primary purpose of providing factual information.

> Examples include government research reports, census data, and factual information such as that found in encyclopedias.

> The URL address may have any one of a variety of endings because many Web page types provide information.

Go to the Environmental Protection Agency site, http://www.epa.gov

Now apply the Five Points of Reliability:

- The purpose of the site is clear.
- It is clear who is responsible for the page – The United States Government
- Sources of information are listed or linked to/from the home page.
- Information is accurate with sources listed.
- It has been recently revised.
- Many topics to choose from as they relate to the US Environmental Protection Agency.
- Links to other reliable complementary resources.

News

A news site's main purpose is to provide current information on local, regional, national, or international events. Numerous news sites are devoted to one particular topic such as business news, technology news or legal news.

> Examples include newspapers, television and radio stations, and Web-based news organizations without a print counterpart.

> The URL address frequently ends in .com (commercial)

Go to the *Washington Post*, http://www.washingtonpost.com

Now apply the Five Points of Reliability:

- The date and revision time reflects current news

- Note at the bottom half of the page how the editorial and opinion material is clearly labeled.
- In the upper right corner, you can click on the print version for access to the actual print version of the Washington Post.
- Minimal advertising
- It is clearly a reliable news site with factual information.
- Author of articles is clearly identified and includes author's credentials.

Personal/Blog

A personal Web page is created by an individual who may or may not be affiliated with a larger organization.

Anyone can and does publish anything, including favorite hobbies.

Often showcases a talent or artistic ability of the site's creator.

May have links to advocacy pages, informational pages, entertainment or even business pages.

The URL may have a variety of endings depending on what type of site the page is coming from

Go to a personal Web page/Blog, http://www.momsminivan.com/blog/

Now apply the Five Points of Reliability:

- Owner of the page is a mom who likes to take her kids on adventures in a minivan
- Links may provide useful information
- Cannot always determine when written or if the information is factual.

Other Sites to Evaluate

Site that you may need additional software to view it (like Shockwave, or Flash)
http://www.nabiscoworld.com

Site where a fee is required to access all/part of the site:
http://www.consumerreports.org/main/home.jsp

Additional sites to evaluate for accuracy:
http://descy.50megs.com/mankato/mankato.html
http://www.dhmo.org/

Citing Your Sources

The Minnesota West Libraries have copies of all style manuals in the reference section. These print sources include accurate updated examples for all types of sources you may encounter. There is also a link to MLA and APA examples on the Minnesota West Library page.

Publication Manual of the American Psychological Association. 6th ed. Washington, DC: American Psychological Association, 2010. Print. (Call Number: REF BF76.7 .P83 2010)
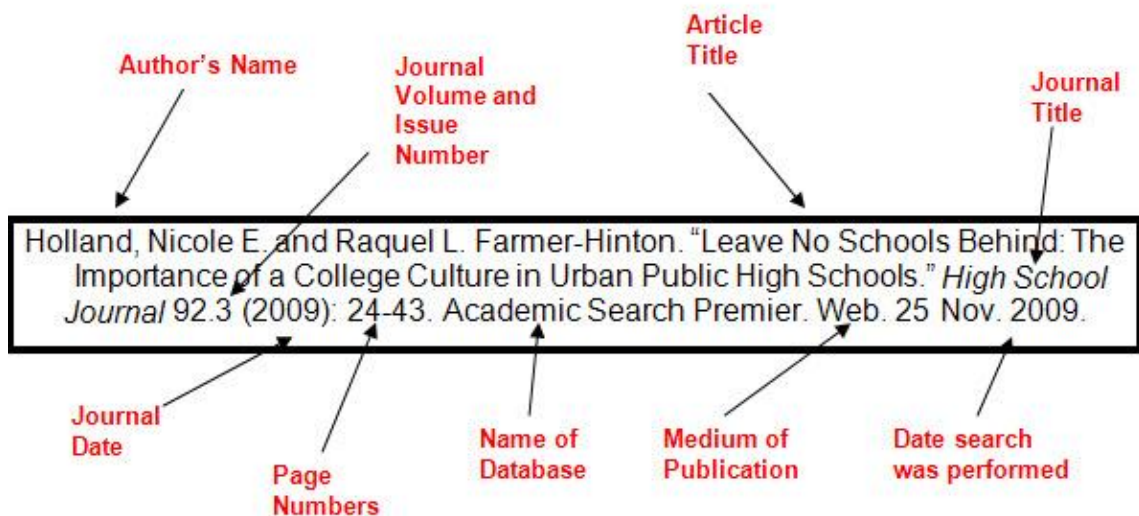
Gibaldi, Joseph. *MLA Handbook for Writers of Research Papers.* 7th ed. New York: Modern Language Association of America, 2009. Print. (Call Number: LB2369 .G53 2009)

The Chicago Manual of Style. 16th ed. Chicago: The University of Chicago Press, 2010. Print. (Call Number: Z253 .U69 2010)
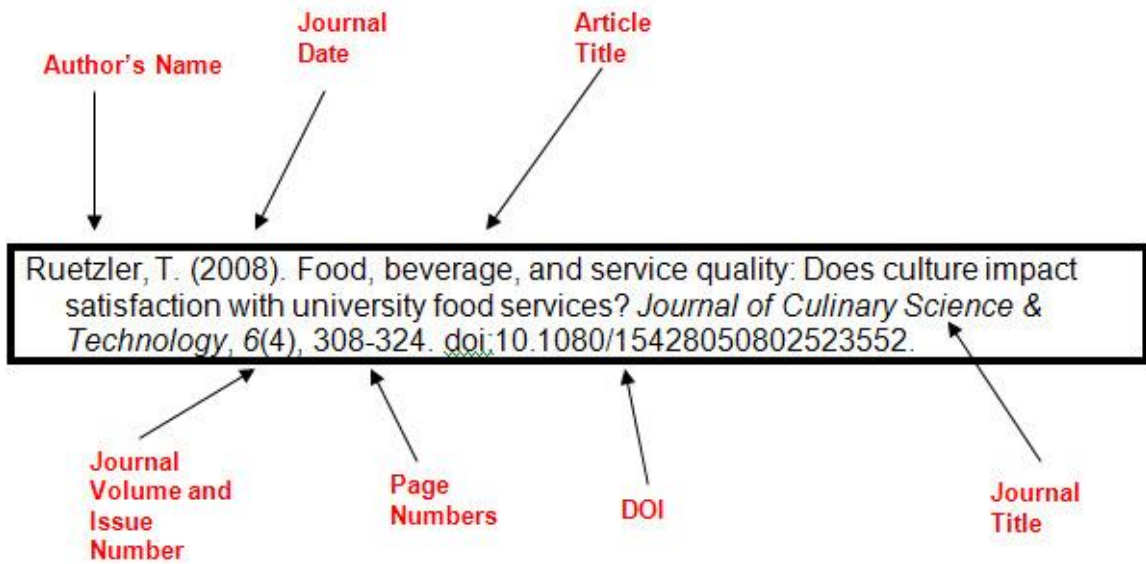
For information on preparing an annotated bibliography, go to:

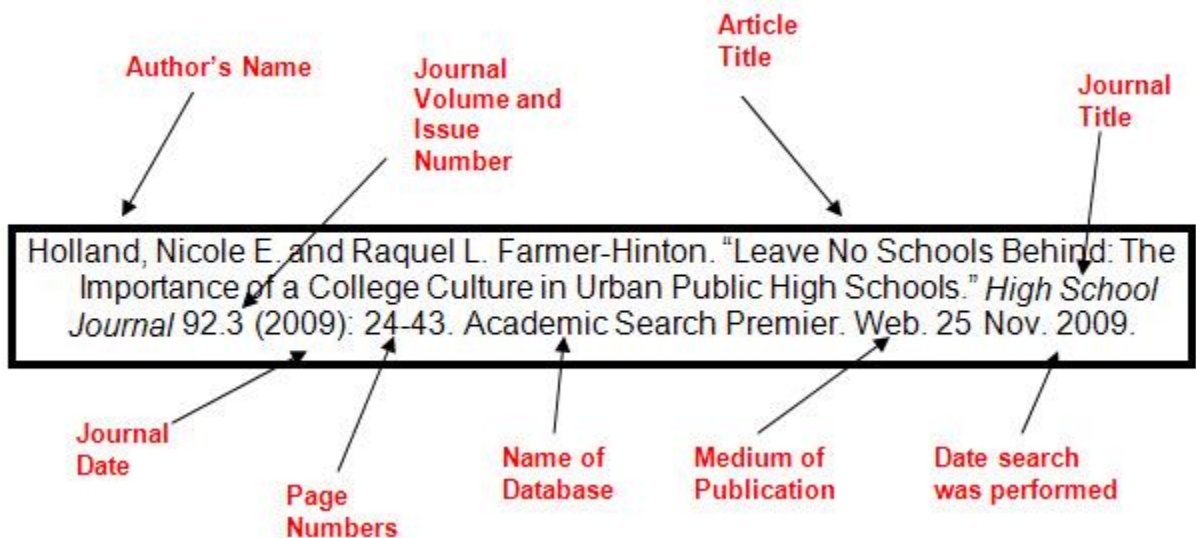http://owl.english.purdue.edu/owl/resource/614/01/

An example of an APA reference without the DOI (Digital Object Identifier):

An example of an APA reference with the DOI (Digital Object Identifier):

Author's Name   Journal Date   Article Title

Ruetzler, T. (2008). Food, beverage, and service quality: Does culture impact satisfaction with university food services? *Journal of Culinary Science & Technology*, 6(4), 308-324. doi:10.1080/15428050802523552.

Journal Volume and Issue Number   Page Numbers   DOI   Journal Title

An example of an MLA style works cited:

Author's Name   Journal Volume and Issue Number   Article Title   Journal Title

Holland, Nicole E. and Raquel L. Farmer-Hinton. "Leave No Schools Behind: The Importance of a College Culture in Urban Public High Schools." *High School Journal* 92.3 (2009): 24-43. Academic Search Premier. Web. 25 Nov. 2009.

Journal Date   Page Numbers   Name of Database   Medium of Publication   Date search was performed

An example of a Chicago reference with the DOI:

Author's Name Article Title   Journal Title

Hodge, Adam R. "Pestilence and Power: The Smallpox Epidemic of 1780–1782 and Intertribal Relations on the Northern Great Plains." *Historian 72, no. 3* (Fall 2010): 543-56. doi:10.1111/j.1540-6563.2010.00270.x

Journal Date Page Numbers   DOI   Journal Volume and Issue Number
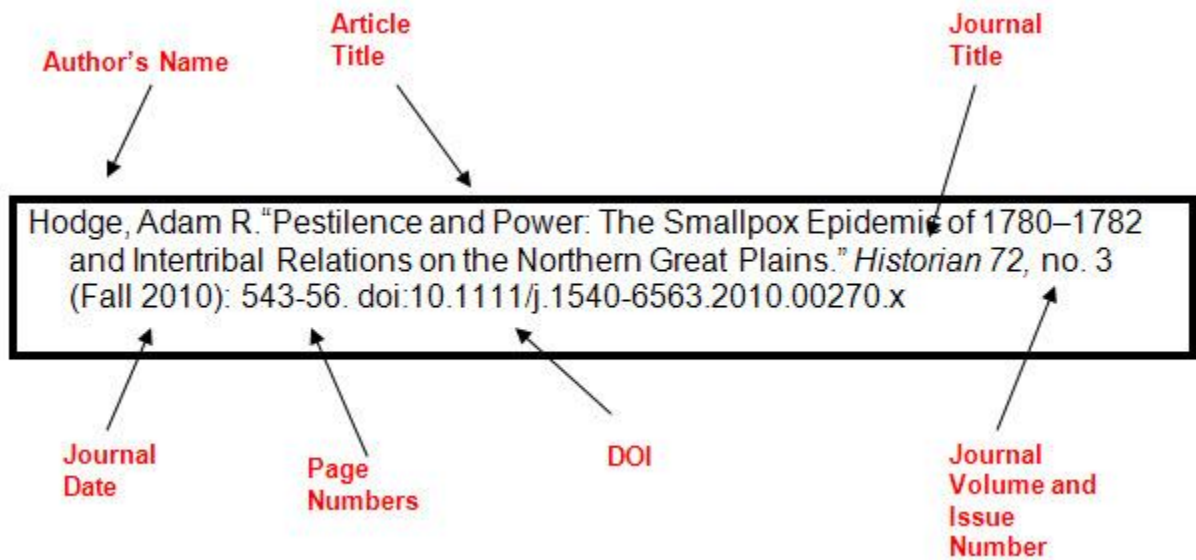
Find more examples of citations using APA, MLA, and Chicago styles, or

APA, MLA, and Chicago model papers through the Purdue Online Writing Lab.